Crime Survey
for England & Wales

Research on Transforming the Crime Survey for England and Wales

Work package A: Developing and testing an online version of the CSEW questionnaire

KANTAR PUBLIC

Office for National Statistics

Contents

1. Summary, conclusions and recommendations

1.1 Context and objectives

The Crime Survey for England and Wales (CSEW), conducted by Kantar Public on behalf of the Office for National Statistics (ONS), is a large-scale face-to-face victimisation survey which asks people about their experiences of crime in the 12 months prior to the interview. The survey has been running since 1981 and, at its core, the method by which crimes are measured and counted has remained largely unchanged.

The core part of the survey includes crime screeners questions and victim modules that allow the measurement of both the prevalence and incidence of crimes, with incidents being coded to offence categories to ensure accurate classification of crimes which align with police-recorded definitions.

In recent years there has been a shift in the UK and internationally towards moving social and government-based surveys to multi-modal or online data collection, and this work forms part of the ONS transformation programme which has been investigating the feasibility of moving ONS household surveys to online and multi-mode data collection.

In 2020, following suspension of face-to-face fieldwork during the pandemic, a telephone version of the survey (TCSEW) was successful in producing crime estimates that were broadly comparable with the historical time series. However, the fieldwork challenges posed by the pandemic further highlighted the need to move forward with multi-modal and online data collection to help futureproof the survey.

Against this background, the core purpose of 'Transformation Work Package A' was to build on previous work conducted by Kantar Public to develop the CSEW as an online instrument and to provide a more robust assessment of the validity of the online version of the questionnaire using a live trial[1].

Throughout this report we use the following abbreviations:

- CSEW refers to the face-to-face version of the crime survey
- TCSEW refers to the telephone version of the crime survey
- WCSEW refers to the online (web) version of the crime survey
- RCT refers to the randomised control live trial which compared online (RCT Online) and telephone (RCT Telephone)

1.2 Background to the previous work to develop an online crime survey

The previous work conducted in 2017-18 by Kantar Public comprised an extensive scoping stage and around 100 qualitative interviews which involved iterative cognitive and usability testing of a draft online questionnaire.

The initial scoping work concluded that a move to online data collection would bring both opportunities and risks. The key opportunity is the chance to the make the questionnaire more

[1] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/redesignofcrimesurveyforenglandandwalescsewcorequestionsforonlinecollection/2018-07-19

streamlined, tailored and user-centred: without interviewers to encourage participation, guide respondents through the questionnaire, and maintain engagement this is essential. However, considerable challenges were noted, the primary one being that the method of counting and classifying crimes in the survey is extremely complex and not intuitive for respondents. This makes it difficult to replicate the survey instrument as a self-completion survey.

The initial development work in 2017-18 made significant progress towards developing an online self-completion instrument which worked for most respondents who either experience no crime or who have a 'simple' experience of crime: for example, a single incident or two clearly demarcated incidents. However, without an interviewer to resolve discrepancies and maintain accuracy, the online survey was not successful when respondents had more complex crime profiles: for example, if they had experienced multiple or repeat victimisation, or crimes which involved multiple features and were therefore susceptible to double counting. Accurately capturing and counting fraud crimes alongside non-fraud crimes was also found to be problematic in the context of an online self-completion.

1.3 Scope and limitations of this research

The Transformation A research programme built on the previous work by further developing the questionnaire and testing this more robustly in the context of a large-scale field trial, referred to here as a 'live trial'.

As in the previous work, the development and testing work did not cover the whole CSEW questionnaire. Instead, the review was confined to the sections of the questionnaire (screener questions and victim modules) which collect the data required to allow detailed offence coding and hence estimate victimisation prevalence and incidence rates. Investigating how an online crime survey could be used to collect data on highly sensitive topics such as domestic abuse or experience of drug taking is the subject of other work being undertaken by ONS.

It is also important to note that the timescale for developing the online questionnaire for the live trial was very restricted, and as a result it was not possible to conduct as thorough a redevelopment as would have been ideal, or to take forward all the recommendations from the first stage of the online development.

1.4 Summary of methodology

The project involved a series of sequential stages as shown in the table overleaf.

Table 1.1 Stages of the development work	
Stage	***Summary***
1	Scoping review which consisted of a rapid evidence review of the international literature on conducting online surveys for estimating crime
2	Review of the previous WCSEW questionnaire to build in as many as improvements as possible within the timescale, as well as to ensure the survey was fit for purpose for the live trial
3	Redevelopment of the existing prototype WCSEW script to take on board findings from Stages 1 and 2; further amendments were also made after Stage 4.
4	Qualitative cogability pre-testing[2] of the re-developed WCSEW with 15 people who had 'complex' crime profiles - that is experiencing at last three crimes in the last 12 months
5	A live trial based on a randomised control trial (RCT) to assess the reliability of an online version of the CSEW questionnaire on a larger scale. This was conducted on Kantar's Public Voice Panel[3] and involved: - A split sample experiment where participants were allocated either to a telephone or online version of the live trial questionnaire - A secondary stage where a subset of victims to the online live trial were followed-up by telephone and asked to complete it again in this mode The aim of the live trial (RCT) was to determine the impact of an online-based data collection mode (WCSEW) on measurement, prevalence and incidence of crimes, and how this differs between the online mode (RCT Online) and telephone mode (RCT Telephone). Given differences in methodology, no attempt was made to compare the live trial findings with the main face-to-face CSEW or TCSEW.
6	10 post-hoc qualitative depth interviews targeted on respondents with especially complex crime profiles to assess their experience of completing the survey.

1.5 Conclusions and recommendations

Conclusions and recommendations are drawn from across all stages of the work outlined above and are presented within broader themes below.

1.5.1 Complexity of crime measurement and how this translates to an online survey

The scoping review comprised a rapid evidence assessment (REA) of the existing literature on international crime surveys, focussing where possible on surveys that have undertaken any

[2] Refers to a hybrid between cognitive and usability testing of questionnaires

[3] Kantar Public's random sample panel

development or transformation work to transition from an interviewer-administered mode to a self-completion online mode. In total, 15 studies were included in the final review.

Complexity of crime measurement and the difficulty in translating this complexity into an online data collection tool was a common theme that emerged as part of the literature review.

The main findings were as follows:

Based on the international evidence review, there are no other examples of crime surveys which have replicated the full complexity of the CSEW in any mode. While the US-based National Crime Victimization Survey (NCVS)[4] and the New Zealand Crime and Victims Survey (NZCVS) are the surveys most similar to CSEW in terms of what they attempt to measure, none of the surveys reviewed were as ambitious as the CSEW in terms of covering the full complexity of the CSEW design: high number of screener questions, including sensitive crimes and fraud; measuring incidence (counts) as well as prevalence; differentiating between incidents that are similar (part of a series) and different (separate) where there are multiple incidents within crimes type; and collecting information about each incident that is comprehensive enough and of high enough quality for offence coding. *(Section 3.4)*

The evidence review of 15 international crime surveys found that only four5 had used a 'push to web' approach where respondents were prompted to log into an online survey platform to complete the survey. Most other surveys use online surveys in a more limited way, either alongside interviewer-administered data collection as part of a multi-mode approach or as a self-completion (CASI) module to collect data on more sensitive topics within an interviewer-administered design. Some crime surveys, all much simpler in content and structure than CSEW, also use a paper questionnaire as an alternative to online self-completion to avoid digital exclusion - an option that would not be possible for CSEW in its current form due to its complexity. *(Section 3.4)*

Importantly three of the four examples of push to web crime surveys benefit from using population registers with named individuals as sampling frames which makes the process of targeting the survey to one individual in a household much easier. As the CSEW is reliant on an address-based sampling frame, either a two-stage selection would be required to randomly select one individual, or all adults in the household could be asked to complete the survey. Both these approaches are challenging to administer online without an initial interviewer-led stage first. *(Section 3.3)*

When other crime surveys have moved online, researchers have accepted that this must involve compromise and a break in the time series. Examples of changes associated with online transition include: radical questionnaire redesign/re-structure to remove reliance on an interviewer and better suit online completion (Finland, Canada, Sweden); the need to compromise and simplify measurement aims (Finland, Sweden); an acknowledgement that results cannot be compared with the previous time series (Canada) or some kind of adjustment to make previous results comparable with the previous time series (Netherlands, Sweden). Mode effects resulting in changes to crime victimisation rates are also mentioned (Sweden, Canada) - see section 1.5.5 below for more discussion on this. While the US-based NCVS has not yet moved online, the extensive redevelopment work being conducted by the Bureau of Justice Statistics has highlighted similar transitional issues to those surveys that have already moved online. *(Section 3.4)*

[4] Developing and testing an online (CAWI) self-completion version of the NCVS is part of the ongoing research programme although development so far has been based on CATI and CAPI modes

[5] Finland, Netherlands, Sweden and Canada

Recommendations

Recommendations here will depend on a final proposed design but in summary we recommend the following:

- Given the complexity of the CSEW, an online-only data collection approach is unlikely to be feasible and therefore considerable thought needs to be given to a multi-mode survey design and how this would work. While the aim should be to produce a single data collection instrument that works for all respondents, the pros and cons of using a different data collection approach with victims who have 'complex' crime histories (which would need to be defined) should be considered: for example, a self-completion online approach as the default but with some interviewer-administered element for those flagged as having a complex crime history.
- The focus of any re-designed data collection instrument should be to reduce the cognitive burden on respondents by making the survey simpler and more respondent-centric. In the context of a multi-mode approach, future re-design work should focus primarily on developing a survey instrument that works online and only then consider how to adapt it for other modes (see section 1.5.5 for further discussion on mode).
- As this development work only focused on the parts of the CSEW that measure victimisation, in exploring a respondent-centric design it will be important to consider the questionnaire more holistically, also including the other parts of the CSEW questionnaire not tested here. Otherwise, participants who have not experienced crime in the last 12 months may feel less engaged (this point is covered further at 1.5.3). The need to better engage non-victims has been considered by the US-based NCVS team in their redevelopment work and it would be useful to draw on their work further as they move into their online-testing phase.
- It should be accepted that a change in design and mode from the current survey will require compromises to be made in terms of the complexity of the data collection instrument and this will almost certainly lead to a disruption of the longer-term time series. Greater clarity on what simplifications or compromises to the current survey might be acceptable by data users would help provide a more structured framework for future re-design work.
- Sufficient time needs to be allowed to fully develop and iteratively test the content and design of a new multi-mode survey design. Due to time and budget constraints, this package of work was unable to explore some of the more complex recommendations from the initial scoping work, as well as additional scoping work for alternative designs developed since. By way of comparison the US Bureau of Justice Statistics (BJS) has a long-term package of work lasting more than eight years to redevelop and test the NCVS questionnaire and new mixed-mode approach.

1.5.2 Double counting or duplication of crime incidents

'Double counting' refers to when a respondent selects more than one screener question to cover the same incident. For example, if a break-in also involves a bicycle theft and criminal damage the respondent might record this single incident at three separate screener questions. Unless this duplication is detected, this could feed into the crime estimates as three separate crimes, when it should be counted as only one offence based on established prioritisation rules. Apart from leading to inaccurate crime count measurement, double counting is problematic because under the current CSEW design it results in respondents being asked to record details of the same incident in more than one victim module, which adds confusion, burden, repetition and length.

The WCSEW script (and the comparison telephone version) included a complex series of check questions which attempted to detect if crimes mentioned across two or more screeners were interlinked; where a link was established check questions aimed to prompt the respondent to correct any incidents of double counting by 'deselecting' duplicate incidents. This was based on an approach developed in the previous redesign work. Although it was known from the previous qualitative testing that these check questions were problematic for respondents, unfortunately there was insufficient time to redevelop or test an alternative approach before the live trial, and therefore the same approach was used again. As such it was predicted in advance of the live trial that this approach would remain problematic. While this was indeed found to be case, the live trial data together with an analysis of participants' verbatim comments and qualitative testing, helped to quantify and flesh out the nature and scale of these problems.

We refer to findings from the live trial before these adjustments as 'pre-review' and after the live trial as 'post review'. The main findings were as follows:

Online respondents were more likely than telephone respondents to report being a victim of crime, as defined by reporting an incident at one or more screener question; they were also more likely to report multiple incidents. In the live trial, before any adjustments were made because of the double counting checks (pre-review), 46% of online respondents reported at least one incident compared with 42% of telephone respondents, while 23% of online respondents selected two or more screeners compared with 19% of telephone respondents. In particular, the rate of fraud screeners (25% vs. 21%) and vehicle screeners (9% vs. 7%) was higher for online compared with telephone respondents. *(Section 7.1.1)*

Post-review, the online and telephone victimisation rates dropped slightly. After double counting checks had been made, victimisation rates for online respondents fell from 46% to 44%, and for telephone respondents from 42% to 40% (this is net change). *(Section 7.1.2)*

A significant proportion of victims[6] who selected at least one screener changed their responses at the double counting checks. Among this subgroup of live trial respondents, 33% of online respondents and 21% of telephone respondents changed their responses due to these checks (this is gross change). The lower level of change among telephone respondents may be linked to interviewers helping (even if inadvertently) to pre-empt incidences of double counting meaning that related crimes were less likely to be mentioned more than once. The high rate of change has both positive and negative implications: positive in that it shows the approach does seem to reduce the rate of incidents which hopefully reduces double counting to some extent; but also negative in that this is a marker of complexity for respondents. After these in-survey corrections (post-review) the higher rate of victimisation among online respondents (net change) remained. *(Section 7.1.2)*

Attempted crimes and confidence fraud incidents are particularly likely to be associated with double counting, these screeners being the ones which were most likely to be interlinked with other crimes and 'deselected' because of double counting corrections. The most deselected screeners[7] were attempted theft from outside a dwelling (52%); attempted assault (49%); attempted theft from a vehicle (48%); and attempted theft away from home (47%). In addition, the rate of deselection was relatively high for both confidence fraud (39%) and attempted confidence fraud (39%). Further analysis revealed that attempt incidents

[6] In this context we have used 'victim' to refer to anyone who answers 'yes' to at least one screener, regardless of the outcome of offence coding.

[7] These figures are each based on all respondents who initially selected this screener and one or more other screeners.

were particularly likely to be recorded as part of a linked incident that encompassed multiple incident types (for example, attempted assaults were commonly featured in incidents that also involved threats and harassment).In addition, analysis revealed that some respondents may have selected both the actual crime and the attempted crime at a paired screener, when they were actually referring to the same incident. One hypothesis is that respondents were thinking of the events in sequence; in other words, an attempt ultimately leads to an actual crime. *(Section 7.1.3)*

Qualitative interviews further reinforced the challenges associated with double counting and found that 'deselecting' crimes that had been double counted does not always fit with participants' mental models. Understandably, participants taking part in the cogability interviews wanted to answer about their experiences at the point in the interview where these felt relevant, and it could feel unnatural to discount duplicated incidents even when participants understood the rationale for this. The double counting checks were shown to work more effectively when the circumstances were relatively simple (for example, comparing one crime against another). However, once more numerous incidents were involved - both across and within crime type - the checks could become extremely complicated and from the participants' perspective stopped being helpful and instead added confusion. In the WCSEW, each time a subsequent screener was selected the new incident(s) was checked against all incidents already reported, which could feel illogical to the participant. For example, it could feel insensitive and irrelevant to compare a sexual assault against vehicle damage and ask if they were part of the same incident. *(Section 5.3.2)*

Qualitative testing indicated that double counting correction often led to inaccurate data. The qualitative stages confirmed that participants with complex crimes profiles were the ones who struggled the most with the WCSEW survey design. The difficulties they experienced when attempting to resolve double counting meant that significant errors were picked up in the observed interviews that would not necessarily be detected in the live trial analysis. *(Section 5.3.2)*

Qualitative testing also found evidence of merging of crimes within victim modules, which might be considered the opposite of double counting. Some qualitative online participants who had experienced multiple victimisation overlapped incidents within victim modules, finding it difficult to separate them out. The victim module includes a series of 'check' questions to classify different features of the incident – so for example a participant may be answering about a theft and when asked if the theft incident involved damage, they might at this point draw in a different incident where damage had occurred. This meant there was sometimes a disconnect between crimes recorded at the screener questions and the victim modules. *(Section 5.3.12)*

Recommendations

- More generally, we should revisit the approach to crime measurement, drawing more on mental models of how respondents view an incident of crime) but also within the context of user measurement needs (see section 1.5.7 below).
- One more specific recommendation which has been carried through from the previous development work (there was no time to develop this for the live trial) was to restrict the double counting checks to only cross-reference an incident against already mentioned incidents that took place in the same month rather than against *all* previously reported incidents. This is the approach that has been adopted by the US-based NCVS. While this represents a less comprehensive approach to checking for potential double counting it would make it simpler for respondents and reduce confusion. It also seems more

intuitive to ask if two (or more) incidents that happened at a similar time were connected in any way.

- The method by which attempted crimes are recorded at the screener questions needs further consideration. The screener questions were set up so that each crime and the equivalent attempted crime were on the same screen so that respondents saw the two questions together – the original development testing found this worked well and appeared to help reduce double counting. However, the live trial analysis suggests that these paired screeners might have led to more double counting. One option worth considering in the future is to ask about actual and attempted crime in a single screener question rather than try to isolate these at the outset. Sorting out whether something was an actual crime, or an attempted crime, could then be resolved later in the victim module.
- As a related point, it may be beneficial to consider the value of using more 'composite' screeners which mean that incidents that tend to be linked (as they are similar) would be considered in the same screener (this approach was used for the violence screener – see section 1.5.6 below). For example, it was found that damage to the home was often linked to damage to vehicles and other personal items. One possibility would be to wrap these into one single question about damage or vandalism (covering homes, vehicles and personal property) with a list that can be multi-coded. The count would then relate to all incidents of 'damage' and the victim module could be used to establish the more specific attributes of the incident. Similarly, amalgamated versions of screeners could also be considered for thefts both inside the home and outside the home (for example, all incidents relating to a respondent's property) and to group together fraud-related crimes (since testing found that fraud incidents are more difficult to pin down to a specific screener, given that respondents often don't know how the fraud event occurred). The descriptor 'tag' (see section 1.5.9) could also be tailored further depending on which item(s) in the list were selected. However, this approach needs to be weighed up against wider evidence that shows that individual forced choice questions lead to higher rates of selection than the same items appearing in a muti-coded list[8].
- The check questions incorporated into the WCSEW were an attempt to identify and correct any double counting of incidents before the start of the victim modules and so prevent respondents being asked to record details of the same incident in more than one victim module. However, even if a simplified set of check questions could be developed as part of further development work, it is unlikely that they would successfully detect every incident of double counting. It is therefore worth investigating further how respondents can record duplicate or overlapping incidents at the start of each victim module. This would allow respondents to bypass a victim module if they had already answered questions about the same incident in a previous victim module. However, this would need to be carefully tested to ensure that this bypass was not used as a means of satisficing (short-cutting) in the survey.
- In relation to overlapping incidents with victim modules, without an interviewer on hand to help separate out incidents, the only way to address this issue within the current design would be to add further checking screens to confirm the participant is thinking about the correct incident. However, we would hesitate to recommend further check screens due to the already complex nature of the instrument. Instead, we recommend exploring simpler ways of ensuring that respondent with complex profiles can focus on specific incidents.

[8] Pew Research Center, May 2019, "When Online Survey Respondents Only Select Some That Apply"

1.5.3 Victim module eligibility

After the completion of live trial fieldwork, all victimisation modules were reviewed by the same team of coders who work on the main CSEW to determine whether what had been reported represented a crime or not and, if so, what offence code should be assigned to the crime[9].

In the published CSEW statistics, a 'victim' of crime is not based on reporting an incident at a screener question or completing a victim module, but instead on having a valid offence code. It was therefore useful to examine how the victimisation rates change based on different definitions: reporting an incident at a screener question vs. reporting an incident which is assigned an eligible offence code.

Since the proportion of ineligible victim modules can be seen as a measure of data quality it is also useful to examine any differences in the ineligibility rate between online and telephone respondents.

The main findings were as follows.

There was a higher rate of ineligibility in the online survey compared with the telephone survey. In total, a third (33%) of all victim modules generated at the screener stage by online respondents were classed as ineligible compared with 21% of victim modules generated at the screener stage by telephone respondents. *(Section 7.2.2)*

There are three reasons why a victim module does not generate an eligible offence code: the victim module is skipped by the respondent; the incident is outside of the 12 months date reference period; or no valid offence code can be assigned. *(Section 7.2.2)*

The reasons contributing to ineligibility are discussed separately below.

- **Online respondents were more likely than telephone respondents to opt to skip a victim module when given the opportunity.** People who reported an incident of violence, threats or harassment were given the option to skip the relevant victim module due to sensitivity and/or privacy concerns10. Overall, 22% of live trial respondents reporting an incident of this type chose to skip at least one victim module: 25% of online respondents and 14% of telephone respondents. *(Section 7.2.2)*
- **Online respondents were more likely than telephone respondents to report an incident that was date-ineligible.** In the live trial, an incident was classed as out of scope if it happened more than 12 months before the interview. Additionally, while some incidents were within the last 12 months, and so in scope, the respondent was not able to recall the exact month in which the incident happened. Online respondents were less likely than telephone respondents to report an incident that was in-scope and with an exact month according to date (89% compared with 96%). This was driven by online respondents being more likely to not know the exact month of the incident (7% compared with 3% of telephone respondents) and to record an incident which happened more than 12 months ago (4% and 1% respectively). Difficulty reporting dates was also uncovered in the qualitative research, where it was found that participants reported difficulty if the incident was minor, where it was not associated with a specific event, or where different incidents were linked or overlapping. *(Section 7.2.2)*
- **Online victim modules were also more likely than telephone victim modules to be given an invalid offence code.** The reasons for an invalid offence code include:

[9] Unlike the main CSEW, no additional checking was undertaken to quality assure the codes derived by the coders and supervisors

[10] This skip function is also used in the CSEW survey where incidents of domestic violence or sexual assault can be skipped, either at the request of the respondent or by the interviewer if they judge it is not appropriate to ask the questions

incidents outside the survey's coverage11; incidents where there is not enough information to be sure it reaches the threshold of an offence12; and incidents which could not be coded due to inadequate or incorrect information13. Overall, 24% of online victim modules which were coded were given an invalid offence code compared with 16% of telephone victim modules. The main reason for this difference was the higher proportion of incidents reported online outside the survey's coverage: 16% of all online victim modules and 9% of all telephone modules. This difference was particularly notable for fraud cases: 11% of online victim modules compared with 6% of telephone victim modules were fraud cases outside the survey's coverage. *(Section 7.2.2)*

Fraud, violence, threats and harassment were associated with the highest levels of victim form ineligibility among online respondents. The main reason for high ineligibility rates for violence, threats and harassment was skipped modules (see above). When skipped modules are excluded, the highest rates of ineligible victim forms for online respondents are fraud (40%), home-based burglary/theft/damage (27%) and vehicle incidents (26%). *(Section 7.2.2)*

Once ineligible victim forms are removed from the data, the victimisation rate drops significantly, and telephone respondents reported more eligible incidents. Based on all eligible victim modules, the victimisation rate fell from 44% to 30% for online respondents, and from 40% to 33% for telephone respondents. This means that although respondents initially report more incidents in an online self-completion survey, this will not necessarily translate into higher victimisation rates once cases are coded. *(Section 7.2.3)*

There was a high degree of association between screener questions and offence codes for non-fraud crimes, with a higher match rate for the telephone compared with the online survey. Within broad non-fraud crime categories, the match rates for online ranged from 77% to 89% and for telephone the match rates were in the range 81% to 93%. This suggests that the telephone survey was better at picking up incidents at the 'right' screeners. *(Section 7.2.4)*

Recommendations

- One of the benefits often cited for online self-completion surveys is that respondents may feel more comfortable providing details about sensitive crimes compared with an interviewer administered survey. However, the much higher 'skip' rates in the online vs. telephone survey suggest that respondents are not necessarily using the skip function for the safeguarding or sensitivity purposes it was intended for. It instead suggests that respondents are using this as a form of short-cutting (or satisficing). If this is true, then an online crime survey may significantly under report sensitive crimes such as physical or sexual violence compared with an interviewer administered survey if a 'skip' option is offered. Balancing this loss of information against the need to ensure privacy and safety concerns will be an important issue for any future online crime survey. We therefore recommend reviewing this feature to try to ensure it is only used for the purposes it is intended.
- Linked to the above, one interesting experiment would be to include the skip function for *all* crimes, not just sensitive ones, to see if there were also higher rates of skipping for

[11] For example, personal crimes that happened to someone else other than the respondent, or incidents of fraud where the respondent was not the specific intended victim

[12] For example, incidents that might be accidental damage rather than criminal damage, possible lost property rather than theft, or disputes over goods or services that could be trading standards issues rather than fraud

[13] For example, insufficient information provided, duplicate incidents, or incidents recorded under the wrong victim module (i.e. a fraud crime recorded under a non-fraud victim module or vice versa)

more minor crimes. If this is the case, it would reinforce the hypothesis that the skip function is being used in a way which is not as intended.

- There is clear evidence that online respondents are less likely to provide valid date information which leads to a high rate of date-ineligible incidents. In an interviewer-led survey, attempts are made by interviewers to help respondents recall the exact month an incident happened. However, no such assistance is available for respondents in a self-completion survey. This means that we may need to provide more prompts to respondents to help them recall an exact date and/or stress the important of this. Incidents which are out of scope because they occurred longer than 12 months ago should be filtered out before reaching a victim module which will help reduce the ineligibility rate.
- The high rate of ineligible victim forms for the RCT survey (and especially for the online survey) leads us to conclude that we are placing additional burden on respondents, as respondents spend time competing victim modules which are ultimately not used as part of published statistics. This suggests that the survey instrument requires further refinement to better steer respondents to only reporting relevant incidents.

1.5.4 Cognitive burden for multiple and repeat victims

Beyond the double counting complexities, the study provided wider evidence of excessive burden among respondents with more complex crime profiles: this refers to respondents who have experienced multiple and/or repeat victimisation and where two or more crime types might be interlinked (with the potential for double counting).

Excessive burden is an important issue as a survey which places too many cognitive demands on participants risks a higher rate of survey drop-out, poorer data quality and inaccurate offence classification. While such issues are applicable to all data collection modes, they are likely to be particularly acute in an online self-completion mode, where there is no interviewer available to assist.

The main findings were as follows.

The length of the online survey for respondents who had experienced multiple or repeat victimisation was long. In the online live trial, the demographics and screener questions took an average of c. 6.5 minutes for non-victims increasing slightly to c. 8 minutes for victims. However, each victim module added around 4-5 minutes to the total length meaning that the average total interview length14 for the online survey varied from an average of c. 8 minutes (no victim modules) to c.13 minutes (one victim module) to c. 34 minutes (six victim modules). Given that only part of the crime survey questionnaire was tested, once added to other parts of the CSEW, this indicates that online interview lengths will be excessive for people with complex crime profiles unless the script is simplified for victims of multiple crimes. *(Section 8.3)*

However, the average length of the online survey was shorter than the telephone version of the same script. Telephone interview lengths were considerably longer than online interview lengths: for non-victims an average of c. 13 minutes compared to c.8 minutes; and for victims and average of c. 28 minutes compared with 18 minutes. This is not surprising given the need for telephone interviewers to read out the questions and answer lists. This finding suggests that an online approach will be a less burdensome alternative to interviewer-administered modes. However, this may also raise concerns about data quality as it could

[14] Also includes extra time at the end to cover usability questions

suggest that online respondents are completing the survey more quickly due to increased attention deficit and satisficing (see below). *(Section 8.3)*

Ratings of difficulty in the live trial survey completion provided another measure of cognitive burden. Based on a score of 0-10 in terms of how easy or difficult people found the survey, ratings were categorised into 'easy' (score of 0-3), 'moderately difficult' (score of 4-6) and 'very difficult' (score 7-10). *(Section 8.1)*

Online respondents were more likely than telephone respondents to rate the survey as 'difficult' A third (34%) of online respondents rated the survey as moderately (6%) or very (28%) difficult, compared to 15% of telephone respondents (moderately difficult 7%, difficult, 8%). (Section 8.1)

Respondents completing multiple victim modules on the online survey were considerably more likely to rate the live trial survey as 'difficult'. 25% of non-victims rated the WCSEW as moderately or very difficult (compared with 9% of telephone non-victims). However, this rate increased by number of victim modules. The proportion of victims finding the WCSEW difficult (score of 4+) ranged from 40% of those completing one victim module to 67% of those completing four to six victim modules. Among the latter group, almost half (46%) rated the survey as 'very difficult'. *(Section 8.1)*

There was also evidence that vehicle and household crimes were associated with higher difficulty scores. This could be because these are more likely to be proxy crimes affecting other household members which respondents may find more difficult to answer about. Conversely personal theft or damage crimes were associated with the lowest difficulty scores. Younger people also found the survey more challenging than older people, possibly because they are less likely to be householders and therefore find it difficult answering questions about household crimes. *(Section 8.1)*

Although online respondents were more likely than telephone respondents to find the survey difficult, they encountered fewer problems of a more specific nature. In the live trial, telephone respondents were more likely than online respondents to find the survey repetitive (33% vs. 18%), that the questions didn't fit their experiences (32% vs 13%), that the survey was too long (10% vs 3%) and that the order of the survey felt unnatural (10% vs 1%). This provides some indication that the online survey is more user-focussed than the telephone survey. *(Section 8.1)*

Victims were more likely than non-victims to find the online survey repetitive and confusing, while non-victims were more likely to feel that the survey was not relevant to them. In the online live trial, victims were more likely than non-victims to find the survey repetitive (31% vs. 7%), confusing (9% vs. 1%), and too long (7% vs. 1%). Conversely, 89% of victims said that the survey questions were relevant to their circumstances compared with only 50% of non-victims. *(Section 8.1)*

Qualitative testing reinforced the finding that victims with complex crime profiles found the survey more confusing and burdensome. The intricate nature of asking for a specific count, the detailed questions on whether incidents were part of a series or were separate, as well as the month in which the incident(s) occurred, caused confusion for participants who had experienced complex crime. The requirement to type in a verbatim description of the incident added extra burden. *(Chapter 5)*

Qualitative testing uncovered clear evidence of respondents using satisficing tactics to reduce burden. When a questionnaire appears burdensome, respondents can look to short-cut (or satisfice) when completing the survey. Qualitative testing also provided some evidence that participants could 'learn' the structure of the questionnaire; that is by answering 'Yes' to a screener this leads to a loop of follow-up questions. This presents a risk that in a live context

respondents might self-edit their responses to reduce length and repetition (see section 1.5.3 for a notable example of this). 'Read-only' instructions screens were also problematic with evidence that some respondents skim-read or skip over questions where they didn't have to 'do' anything. *(Sections 5.3.1, 5.3.4)*

Recommendations

- Although the redevelopment work did not cover the whole CSEW questionnaire, it is a positive finding that the online version of the screener and victim modules is shorter than the interviewer-led mode equivalent, and it appears that the interview length is more manageable for both non-victims and victims with simple crime profiles. However, a shorter online interview length compared to telephone could also be linked to higher attention deficit and short-cutting when using the online version. The much longer interview length for victims with more complex crime profiles further reinforces need to simplify the questionnaire for complex victims (see section 1.5.1)
- Although online interview lengths were shorter for non-victims and those with a simpler crime profile, the fact that 25% of non-victims still found the online WCSEW survey 'difficult' means that the survey is perceived as challenging even for those with simpler crime profiles which further reinforces recommendations around simplification.
- As also discussed in section 1.5.1, the fact that non-victims find the survey isn't as relevant to them is linked to the fact that the development work only covered the victimisation parts of the questionnaire. In a full survey, we recommend a wider review focussed on the experiences of both victims and non-victims to ensure that that the whole survey (when developed) covers topics which feel relevant for everyone. This should be combined with clear messaging around the objectives of the survey to emphasise the wider objectives of the survey, not just victimisation.
- Simplifying the questionnaire and reducing cognitive burden associated with double count checks (see section 1.5.2) should make the survey more user-focussed. More specific recommendations associated with user-focussed design are to i) consider the loop-structure of the questionnaire to ensure that online respondents with multiple crimes are deterred from self-editing their responses based on 'learning' the pattern following a 'yes' response to a screener and ii) to reduce the text at 'read only' screens which convey important information to increase engagement and reduce the level of skim-reading.

1.5.5 Mode-related differences in victimisation rates

When combining survey data collected through multiple modes, or transitioning from one mode to another, it is important to consider how this may affect analyses. Mode effects are generally taken to mean differences in observed responses to survey items which are due solely to the mode of data collection, rather than to different sample profiles.

Despite attempts to make questions comparable across modes (often termed a 'unimodal' approach), mode effects are usually unavoidable as the two approaches can never be truly identical. Some examples of why measurement may still vary between interview-led and self-completion modes include the ability of interviewers to provide motivation or clarification when required, and the reduced inclination of people willing to disclose sensitive personal information or socially undesirable opinions/behaviours to an interviewer compared with online.

The live trial and other international crime surveys provide evidence of mode effects in the context of crime measurement.

The main findings are:

Mode effects are recognised in every crime survey included in the literature review where there has been a change in survey delivery and administration. Where change has been accompanied by robust experiments (for example, in Finland and Sweden) the results across these other surveys have found that online self-completion tends to produce higher victimisation rates, especially for sensitive crimes such as violence and harassment, where higher rates are thought to be attributable to reduced rates of social desirability bias. *(Section 3.4)*

Consistent with the wider literature, the WCSEW online live trial respondents were more likely than telephone respondents to report being a victim of crime during the screeners. After applying double counting adjustments within the data, online respondents were more likely than telephone respondents to say 'yes' at any screeners (44% vs. 40%). However, it is worth noting that the actual rate of victimisation, once offence coding has been conducted, is much lower than this (see section 1.5.3). *(Section 7.1.2)*

Focussing on crime type, online respondents reported more fraud and vehicle incidents than telephone respondents. This was apparent both before and after applying double counting checks (pre-review and post-review). Post-review, 23% of online respondents selected a fraud screener compared with 20% of telephone respondents, and 9% selected a vehicle screener compared with 7% of telephone respondents. The higher online prevalence rate of fraud crimes might be linked to social-desirability bias in the telephone version, as respondents might be reluctant to admit to an interviewer that they have fallen victim to a scam. If this is the case, this would point to the online data providing more accurate measurement of these types of crimes, once other aspects of survey design are accounted for.

The higher level of vehicle incidents online (which will include the relatively common incidents of damage to vehicle) might be related to greater uncertainty about whether these more minor incidents constitute a crime as opposed to accidental damage. Participants might be more reluctant to mention these to an interviewer thinking they seem trivial, or an interviewer might steer respondents away from mentioning incidents that might not immediately appear to be crime-related. *(Section 7.1.2)*

Recommendations

— When moving to a multi-mode design it is worth bearing in mind that it is likely to be impossible to eliminate mode effects. And in any case wider evidence[15] now points to the recommendation to develop an 'optimode' design where design is optimised for each mode rather than unified across all modes. This further reinforces the recommendation to prioritise making the online experience as user-centred as possible rather than focussing too closely on ensuring uniform presentation across modes.

1.5.6 Approach to asking questions about sensitive topics

There were some specific findings and recommendations in relation to the approach for asking about more sensitive topics such as physical and sexual assault, threats and harassment.

The redevelopment work highlighted the need to differentiate between assaults and threats/harassment more accurately. Other CSEW- and TCSEW-related work has indicated that legally defined common assaults such as spitting and shoving are often captured under threats/harassment rather than assault as respondents don't necessarily regard these incidents as an 'assault'. If the assault screener is too narrowly worded on crimes that involve violence this can create confusion. In order to more clearly differentiate between assaults and harassment/threats and to reduce confusion and double counting, we decided to change the

[15] https://analysisfunction.civilservice.gov.uk/policy-store/a-user-centred-design-approach-to-surveys/

approach in the live trial questionnaire to ask about assaults as a multi-choice rather than a single yes/no screener, where the multi-choice included the full range of assaults from spitting/shoving etc. to physical violence and sexual assaults – it was hoped that by including all of these in the list this would allow more accurate classification of crimes at the screener and reduce double counting against other screeners. *(Section 4.4.3)*

The victim module tested was not found to be a very good fit for threats and harassment. The WCSEW included new screener questions on threats and harassment and a minimal number of new questions were covered within the victim module to capture details of these incidents for offence coding. However, the new victim module questions were limited, and the qualitative interviews with victims of threats/harassment suggested the sequence of questions did not capture the key features of the incidents which were most relevant for victims. *(Section 4.7.5)*

The development work also highlighted the need to be fully transparent about the questionnaire coverage from the start. Respondents should know at the outset that the survey includes more sensitive topics such as physical assaults, sexual assaults, threats and violence (including domestic violence) to prepare participants and to allow them to find a private space to answer these questions if they wish. *(Sections 4.4.3, 5.3.1)*

Reassuringly only 2% of live trial online respondents said they found the survey too personal or sensitive when asked to evaluate the live trial survey. However, this increased to 10% among victims of assault, and 7% among victims of threats or harassment. *(Section 8.1)*

Among those experiencing an assault, threat or harassment, online live trial respondents (25%) were more likely than telephone live trial respondents (14%) to use the 'skip' function to avoid answering a full victim module. This feature was included as a 'safety net' in the live trial to allow respondents who had experienced more sensitive crimes to skip the victim module. This issue is discussed in more detail in section 1.5.3. *(Sections 8.1, 7.2.2)*

Recommendations

- A further review should consider how best to group and order the more sensitive questions so that respondents can safely answer all the questions at once rather than coming back to the incident(s) later, when they may no longer be in a private safe location.
- In the future it would be better to create a separate 'Threats and harassment' block within the victim module so that the question sequence feels more relevant for people who have experienced these types of incidents. In addition, it would be helpful to expand on victims' experiences of these incidents to allow them to provide further details that will feel relevant to them, drawing on the wider harassment work developed as part of the CSEW.
- Problems associated with the safety net 'skip' function have been detailed in section 1.5.3.

1.5.7 Mental models of crime victimisation

The previous development work was based around a redesign of the CSEW, but one which was still restricted to fit the same basic 'model' of the CSEW: that is screener questions followed by victim modules, and the requirement to develop a solution which led to both prevalence and incidence (counts). Furthermore, most (if not all) international crime surveys included in the review adopt a similar approach for screeners, including the handful that used an online self-completion mode.

Up until this point, there has not been the scope or time to move away from the central design of the CSEW screener/victim module approach. However, in recent years, survey development has followed a more respondent-centred design[16]. In light of this, the qualitative testing explored how participants describe their experience of crime in their own words ('mental models').

Mental models focussed on incidents of crime rather than the constituent features of the incident which doesn't always align with the existing screener/victim module approach. Within the qualitative interviews, a common mental model strategy was to take the most serious incident (or series of incidents) and talk about this in detail before moving onto the next one. Each incident may have involved one or more screeners (for example theft of vehicle, theft from vehicle and vehicle damage). Within these mental models, some respondents also included incidents that had happened to friends and family or in the general neighbourhood, rather than to themselves directly. In some cases, participants mentioned salient incidents that happened longer ago than the 12-month time frame. *(Section 4.4.3)*

Qualitative interviews indicated that the WCSEW tended to prompt participants to recall other more minor and less salient incidents that were not mentioned at the mental model stage. For the depth interview participants, it was possible to compare their mental model accounts with what they had reported in the live trial and regularly the two accounts did not align. However, this was mainly a positive finding as the screener approach was found to help jog participants' memories of more minor incidents and emphasise that these types of incidents still 'count'. *(Sections 5.2, 9.3.2)*

Recommendations

- The findings reinforce the need to consider redevelopment in the light of mental models, where people think about crimes as holistic incidents rather than in terms of the different features of incidents. However, purely focussing on mental models may conflict with user-needs so it is likely that an approach which balances respondent needs against user needs will be required.

1.5.8 Series crimes

In the CSEW and WCSEW, where a respondent has experienced more than one incident associated with the same screener, they are asked if these incidents were 'similar' in nature. And if all crimes are similar, they are treated as a 'series' of crimes which means that the respondent is only asked about the most recent incident, and the same offence code is assumed for all earlier crimes in the series.

The definition of 'series' crimes is complex when a respondent has experienced multiple incidents of the same crime type because the line between what is 'similar' and what is 'different' can become blurred, and categorising individual incidents into a 'similar' bucket or a 'different' bucket can be cognitively challenging. For the live trial, a pragmatic decision was made to treat all cases where there was a mix of both series and separate incidents as a series. However, this was a temporary 'workaround' solution and in any main stage, a better solution needs to be developed which is less confusing for respondents. It is notable that how the differentiation between separate and series incidents is one area which the NCVS treats very differently from the current CSEW. *(Section 4.7.3)*

[16] For example, see: Wilson, L. and Dickinson, E., 2021. Respondent Centred Surveys: Stop, Listen and Then Design. SAGE.

Recommendations

— One possible solution to simplify this would to be to avoid asking about 'similar' and 'separate' crimes and to instead apply a cut-off, for example to always treat 1-2 crimes as separate, and to always treat 3+ incidents as a series. The NCVS asks respondents whether they know enough details to distinguish between different incidents, which is an alternative way of designating incidents into a 'series'.

1.5.9 More specific questionnaire issues

The discussion above has covered some of the more general issues affecting user experience of the questionnaire. Overall, there were relatively few question wording issues encountered in qualitative testing outside of the more structural complexities such as double counting checks (covered in section 1.5.2 above). This was largely because the questionnaire built on questions which had been iteratively tested in the earlier work. However, there were a small number of more specific issues that arose as part of this later qualitative development work.

Definition and boundaries of the 'home' was not always clear. For example, some respondents queried whether 'your home' included driveways, annexes, garages etc. and 'outside your home' was in one case interpreted very literally as a café in the town centre. This could lead to crimes being captured at the 'wrong' screener. Although it does not matter from a scripting point of view where respondents record the incident, the crime description 'tag' used to define the incident in follow-up questions might be incorrect and so cause confusion for the respondent (see below). In response to this issue, the definitions at these screeners were clarified and more specific examples of what counts as 'home' and 'outside the home' were added. *(Sections 4.4.2, 5.3.9)*

Short-hand crime 'tags' did not always provide an accurate representation of the incident. This refers to tags which were used as text substitutions to indicate which incident type is being referred to in follow-up questions on counts and dates, and later in the victim module. However, the tag always relates to the first screener where an incident is recorded, which may not always be the most appropriate screener, especially if an incident involves composite features. For example, if a crime involved theft of tools from the shed and an attempted bicycle theft then the victim module would refer to the attempted bicycle theft rather than the actual theft from the shed, as this screener is asked first. Given this, the reference was changed to incidents from '*…the […] incident'* to *'…the incident that involved […]*' in an attempt to better describe composite incidents. *(Section 4.4.4)*

The cogability testing indicated some further comprehension problems associated with these tags. Some of the tags were felt to be worded too narrowly, which could cause problems later in the victim module as respondents felt that the description of the crime in the tag did not accurately reflect what happened to them. As a result, changes were made to some tags to define theme less narrowly. More generally, some of the tags were found to be a bit long and confusing. This was a particular issue in the fraud section where the tags are longer and less clearly distinct (given that by their nature fraud crimes can be vague if it is unclear how the fraud took place). *(Section 4.3.10)*

The crime description 'tags' could also bring about confusion in cases of high frequency crime as it was not always possible to differentiate between incidents in a meaningful way, particularly when they had taken place within the same month.

Recommendations

— The findings point to a need to ensure that crime tags reflect the incident more accurately, taking into account that in the current design the crimes are picked up at the first rather than the most relevant screener. Although more complex to develop, it would

be worth reviewing the possibility of creating crime tags that take into account composite features of an incident, which would also help people to differentiate between incidents that took place in the same month. However, this would need to be balanced against a competing recommendation that crime tags should be short and easily comprehensible. Given how important these tags are in helping respondents to focus in on the correct incident, the development of suitable crime tags could become a separate development project in its own right.

— The qualitative findings should be reviewed in detail to see if there are any more specific questionnaire wording issues that could be addressed in the next iteration of the WCSEW.

2. Introduction

2.1 Context

The Crime Survey for England and Wales (CSEW) is a face-to-face victimisation survey which asks people about their experiences of a range of crimes in the 12 months prior to the interview, as well as their attitudes towards different crime-related issues and perceptions of crime. The survey is asked of around 34,500 people aged 16 and over resident in households in England and Wales every year. It provides a rich source of information on a range of offences, as well as the nature and circumstances of crime incidents. The survey includes screeners, victim modules, self-completion modules for sensitive topics, and other crime-related topics, as well as a separate questionnaire for children. In contrast to police recorded crime, the survey captures incidents that are not reported to the police, and the estimates produced are unaffected by changes in police recording practices.

The CSEW produces both prevalence and incidence rates of crime. Prevalence relates to the proportion of the population who have experienced any crime in the last 12 months and incidence is the estimate of the number of crimes experienced in the last 12 months (see section 2.2). Offence coding is used after data collection to ensure accurate classification of crimes which align with police-recorded definitions.

The content of the CSEW has remained broadly consistent since its introduction in 1982, to enable comparability of data with previous years.

Like other government surveys, the CSEW has traditionally been conducted by interviewers asking people questions in-home. However, over recent years, there has been a policy shift towards making government services 'digital by default'. As part of this strategy, the Office for National Statistics (ONS) intends to move its household surveys to multi-mode data collection, with a priority focus on online self-completion. The overall aim of this transformation is to reduce costs, increase flexibility for participants, and minimise the time and burden associated with responding to government surveys.

Before the Covid-19 pandemic, initial exploratory work to assess the feasibility of transitioning the CSEW questionnaire from an in-home interviewer-administered instrument to a mixed-mode instrument which can be self-completed online was undertaken by Kantar Public[17]. Although considerable progress was made in developing a new survey instrument, the complexities of estimating crime through online self-completion made it clear that further work was required to move this forward. Maintaining the core measures of prevalence and incidence is central to the design of the CSEW and the complexity of these measurements presents one of the main challenges to the development of a multi-modal survey.

The disruption caused by the Covid-19 pandemic led to the suspension of CSEW face-to-face interviewing on 17 March 2020. Before resumption of face-to-face interviewing in October 2021, a telephone-operated crime survey for England and Wales (TCSEW) was successful in producing crime estimates that were comparable with the historical time series of estimates from CSEW. Although face-to-face fieldwork within the industry has now resumed, the pandemic has highlighted the need to move away from a sole focus on face-to-face fieldwork

[17] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/redesignofcrimesurveyforenglandandwalescsewcorequestionsforonlinecollection/2018-07-19

and the increased importance of moving forward with multi-modal data collection with a focus on online self-completion.

More widely there has been a shift both in the UK and internationally towards moving social and government-based surveys to multi-modal or online data collection. In the UK, the ONS now provides an online self-completion mode to respondents for government surveys, such as the Opinions and Lifestyle Survey and the Labour Market Survey[18].

The core purpose of 'Transformation Work Package A' was to build on previous development work to develop the CSEW as an online instrument and to provide a more robust assessment of the validity of the online version of the questionnaire.

2.2 Glossary and terminology used in this report

Throughout this report we use the following abbreviations:

- **CSEW:** refers to the face-to-face version of the crime survey
- **TCSEW:** refers to the telephone version of the crime survey
- **WCSEW:** refers to the online (web) version of the crime survey
- **RCT Online:** refers to the live trial group that completed the WCSEW online
- **RCT Telephone:** refers to the live trial group that completed the same questionnaire as the WCSEW but it was completed over the telephone
- **CAWI:** Computer-assisted web interviews
- **CAPI:** Computer-assisted personal interviews
- **CATI:** Computer-assisted telephone interviews
- **CAVI:** Computer-assisted video interviews
- **PAPI:** Paper and pen interviews

Key terms are defined below.

- **Screeners:** a series of 'Yes/No' questions that respondents are asked towards the start of the survey, to determine whether they have experienced any incidents that might be classified as a crime. There are 29 screener questions in total, each asking about a different type of incident.
- **Incident:** any single occasion on which a respondent has experienced something described in the screeners. A single incident could involve more than one type of crime; for example, an incident of mugging would likely also involve a theft and an assault.
- **Double counting:** an incident is 'double counted' when a respondent selects more than one screener to cover the same incident. Unless detected and corrected, this will result in them being asked to record details of the same incident in more than one victim module.
- **Victim:** any respondent who answers 'Yes' to at least one screener indicating they have been a victim of crime[19].

[18] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourmarketsurveytechnicalreport

[19] This is the definition adopted for the current project, although in CSEW victim classification is based on offence coding rather than screeners

- **Victim module:** for every screener to which the respondent answers 'Yes', they are asked to complete a victim module. This is a series of questions intended to capture more details about the incident in question.
- **Series of incidents:** a series is defined as a sequence of similar incidents, that happened in the same way, under the same circumstances, and possibly done by the same person/ people. The questionnaire aims to identify whether multiple incidents constitute distinct, separate incidents or a series of similar incidents. Where a series of incidents is identified, the respondent is asked to complete a single victim module in relation to the most recent incident. This helps to eliminate the repetition/ duplication that would entail from a respondent entering almost identical details into multiple victim modules.
- **Offence coding:** once the data from the victim modules has been processed, it is passed to a team of specialist coders who review the incident details and decide, using the Home Office Counting Rules, whether it constitutes a criminal offence and, if so, what type of offence.

Measurement of crime in the CSEW

- **Prevalence rate:** the proportion of the population who are victims of an offence once or more
- **Incidence rate:** the number of crimes experienced per household or per adult
- **Multiple victimisation:** defined as being the victim of more than one crime (either the same or different crime types)
- **Repeat victimisation:** a subset of multiple victimisation - defined as being a victim of the same type of crime two or more times (classified as either a 'series' of similar incidents or as separate incidents)

2.2.1 Risks and challenges of moving the crime survey online

It is important to consider this development work in the context of the unique challenges associated with measuring crime online.

The CSEW, formerly known as the British Crime Survey, was developed in 1981 as a relatively simple paper document. It was based on a central design which included a set of crime victimisation screening questions followed by a 'victimisation module' for each crime experienced in the last 12 months (up to a maximum of six). The classification and counting of crimes by the survey was designed to mirror police-recording of crimes. Over time, the survey has evolved to incorporate changing data collection technologies and policy priorities. However, at its core, the method by which crimes are measured and counted has remained largely unchanged. This is both a strength and weakness of the CSEW. Continuity in measurement has allowed robust tracking of trends in crime over time. However, on the flip side, there has been little scope to improve or update tracking questions and integrating new questions with existing time series questions has added length, complexity and repetition.

A movement to online surveying therefore brings both opportunities and risks. The key opportunity is the chance to the make the questionnaire more streamlined, tailored and user-focused. Without interviewers to encourage participation and maintain engagement, this will be essential.

However, the method of counting and classifying crimes in the CSEW is extremely complex and difficult to replicate in a self-completion survey. The CSEW provides several measures of

crime based on a 12-month recall period as detailed in section 2.2 above, including measures of repeat and multiple victimisation.

In deriving these measures, it is important to ensure that crimes are not ***double counted***. For example, if a burglary also involves a bicycle theft and criminal damage this should be counted as one incident, not three. Crimes are counted according to a prioritisation order applied during the classification process, designed to mirror as closely as possible the Home Office Counting Rules (HOCR).

The previous work on developing an online crime survey conducted by Kantar Public[20] has indicated that while prevalence is relatively straightforward to measure in any mode, the measurement of incidence, repeat and multiple victimisation is much more complex and does not easily translate into a user-focussed self-completion survey.

2.3 Background to the previous work to develop an online crime survey

The previous work conducted by Kantar Public in 2017-18[21] comprised an extensive scoping and development stage followed by around 100 qualitative interviews focussing on cognitive and usability testing of the draft questionnaire. The questionnaire was developed iteratively focussing on the risks and challenges associated with online measurement of crime as discussed above.

The initial development concluded that considerable progress had been made towards developing an online self-completion instrument which could work in the field for a large proportion of respondents. For example, in the most common victim scenario where a respondent experienced a simple, singular crime (such as theft of a car or an assault) the online questionnaire was quick and easy to complete.

However, it was concluded that the online survey did not work when respondents had more complex crime profiles, for example if they had experienced multiple or repeat victimisation, or single crimes which involved multiple features and were therefore susceptible to double counting. To disentangle the exact circumstances of crimes experienced, these respondents were faced with a complex series of check questions and validation screens which many found cognitively challenging. Accurately capturing and counting fraud crimes alongside non-fraud crimes was also problematic due to the complexities in the crime classification rules which is different for the two types of crime. It was concluded that these questionnaire complexities introduced clear risks in the context of an online survey, for example higher break-off rates and reduced respondent engagement. Although only a small proportion of respondents experience complex crime profiles of this nature, these respondents represent those most affected by crime and it is important the survey accurately represents the experiences of these victims.

2.4 Aims and objectives of Transformation Work Package A

In November 2021, ONS commissioned Kantar Public to develop an online self-completion questionnaire that can be used to estimate the prevalence and incidence of crime to an extent that is broadly comparable with the current interviewer-administered version of the CSEW (either by in-home or telephone interview), building on the earlier development work described in section 2.4.

This previous work was an important first step in systematically highlighting some of the challenges and risks that moving from an interviewer-administered to a self-completion survey

[20] Hamlyn, R., McGee, A., Willis, D. Re-Design of Crime Survey for England and Wales (CSEW) Core Questions for Online Collection: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/redesignofcrimesurveyforenglandandwalescsewcorequestionsforonlinecollection/2018-07-19

[21] ibid.

faced. At the conclusion of the previous work, recommendations were made about the focus of future development work and how some of the unresolved risks and challenges might be addressed.

Within this broad research objective, there were several more specific research questions that ONS was looking to address. These were:

i. What are the current examples of measuring crime online across international markets? What are the benefits and limitations of conducting crime surveys online?

ii. Do the online screeners identify victims of crime reliably? Are any changes needed to the screener questions?

iii. Would an online victim module, which collects data on the experience of individual crimes, measure all the necessary details for crime estimates reliably?

iv. Would an online questionnaire with victim modules be able to provide reliable prevalence and incidence rates for all crime types? When would a case become too complex to do this reliably?

v. What is the maximum length an online survey can be to ensure answers are reliable? Are there any methods that can be incorporated to ensure reliability of an online questionnaire?

2.5 Scope and limitations of this research

The research programme for the Transformation A project (detailed in section 2.8 below) continued the development of on online survey instrument by further development of the previous questionnaire developed by Kantar Public (see section 2.4) and testing this in the context of a live trial to provide a more robust assessment of the validity of an online version of the crime survey.

The focus of the Transformation A development work was the production of prevalence and incidence rates and, as a result, the development and testing work did not cover the whole CSEW questionnaire. Instead, the review was confined to the sections of the questionnaire which collect the data required to allow detailed offence coding which allows estimates of victimisation prevalence and incidence rates.

It is also important to note that the timescale for developing the online questionnaire for the live trial was very restricted, and as a result it was not possible to conduct as thorough a redevelopment as would have been ideal, as it was not possible to take forward all the recommendations from the first stage of the online development.

At the end of the previous WCSEW development project, it was concluded that some parts of the WCSEW questionnaire were not working well (in particular, the part of the script which attempted to manage double counting). However, to meet the constraints of the timetable, a decision was made to omit some of the more complex changes that would ideally have been put in place, but which were not possible within the parameters of the current project. This will now be picked up as part of future development work.

2.6 Ethical considerations

Prior to starting this work, the proposed programme of work was reviewed using the UK Statistics Authority's Data Ethics Self-assessment Tool together with the help and guidance of the Data Ethics team. This ensured that the work fully complied with the UK Statistics Authority's ethical principles. As part of its data collection transformation programme ONS had already considered the ethics of online data collection when sensitive topics are being asked

about and the ethical challenges posed by this[22]. This provided a good basis for considering the ethical and legal implications of this new work and how these might be addressed.

In the event the Data Ethics Team were happy with what was proposed and only raised two specific points both of which were addressed to their satisfaction.

One point was how any questions relating to experience of physical abuse, sexual assault or other potentially sensitive crimes would be presented to respondents and what support would be available for respondent who may be distressed or upset by the questions. This was addressed as follows:

- The fact that the survey included questions on sensitive topics such as physical or sexual assault was signposted to respondents at the start of the survey. Respondents were advised to find a private place to complete the survey and were also told they could skip any questions which they did not wish to answer.
- This information was repeated immediately before the sensitive questions relating to physical and sexual assault, threats or harassment.
- A decision was taken that anyone reporting sexual assault would not be followed up with a victim module which asked for more details about the incident.
- At the end of the interview, details of helplines such as Victim Support, National Domestic Violence Helpline, Rape Crisis, Mankind Initiative and the Samaritans were provided on request. In the case of the online survey, these were included at the end of the survey but only shown if the respondent indicated they wanted to see this information, while in the telephone survey the interviewer offered to provide details over the phone, by email, or through the post.

The second point raised by the NSDEC was around the issue of informed consent and at what point participants could withdraw from the research. As is the case with all research, participation is entirely voluntary and participants can withdraw at any point. This was addressed as follows:

- In the qualitative research (cognitive interviews and depth follow up interviews) both these points were included in the topic guide used by moderators and explained to participants before the start of the interview.
- In the case of the online and telephone survey it was felt that there was no need to explicitly say that participants could withdraw at any time because the research was conducted with members of Kantar Public's (random sample) Public Voice panel who would be aware of this fact. However, as noted above, the point about being able to skip any questions they did not want to answer was made clear to all participants. A screen at the beginning of the survey reminded respondents of this fact, and most questions included a 'Prefer not to say' option where relevant.

2.7 Summary of methodology

Answering all the research questions outlined in section 2.4 required several distinct stages of work, which are summarised in Table 2.1 below. The structure of this report mirrors these different stages.

[22] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/crimesurveyforenglandandwalestransformationresearchontheethicsofonlinedatacollectionrelatingtosensitivetopics/october2020

Table 2.1 – Outline of approach		
Stage	***Summary***	***Objective***
1	Scoping review consisting of a rapid evidence review of the international literature on conducting online crime surveys.	(i)
2	Review of the existing WCSEW online questionnaire developed as part of previous work to build in as many as improvements as possible within the timescale for the purpose of the live test.	(ii), (iii)
3	Redevelopment of the existing prototype WCSEW script to take on board findings from Stages 1 and 2; further amendments were also made after Stage 4.	(ii), (iii)
4	Qualitative pre-testing of the re-developed online questionnaire with 15 people who had 'complex' crime profiles (that is at least 3 crimes in the last 12 months) to explore whether the online screener questions and victim module were working as intended.	(ii), (iii)
5	A live trial (RCT) to assess the reliability of a wholly online version of the CSEW questionnaire (the WCSEW). This was conducted on Kantar's Public Voice Panel and involved: - A split sample experiment where participants were allocated either to a telephone (RCT telephone) or online (RCT online) version of the live trial survey - Panellists with a telephone number were randomly allocated to either the RCT telephone or the RCT online survey. - Panellists with no telephone number were automatically allocated to the RCT online survey - A sub-set of victims to the RCT online survey were followed-up by telephone and asked to complete the survey again. While the trial did not include a test of any within-household sampling method, as part of this stage of the work, Kantar Public's Methods Team developed a protocol that can be tested should an online CSEW seem feasible, drawing on evidence from recent trials that Kantar have undertaken and the evidence from Stage 1. The aim of the live trial (RCT) was to determine the impact of an online-based data collection mode (WCSEW) on measurement, prevalence and incidence of crimes, and how this differs between the online mode (RCT Online) and telephone mode (RCT Telephone). Given differences in methodology, no attempt was made to compare the live RCT findings with the main face-to-face CSEW or TCSEW.	(iii) - (v)
6	10 post-hoc cognitive depth interviews targeted on respondents to the live trial with especially complex crime profiles to assess their experiences of completing the survey in more depth.	(ii) - (v)

3. Stage 1: Evidence review

This chapter provides an overview of the rapid evidence review that was carried out as part of the project and outlines the key themes which emerged from this work.

3.1 Introduction

The first stage of work was to conduct a Rapid Evidence Assessment (REA) of the existing literature on international crime surveys. The primary aim was to identify and evaluate other surveys which currently measure crime using an online methodology or which have undertaken any development or transformation work to transition from an interviewer-administered mode to an online self-completion mode.

The first step involved holding a scoping workshop with the ONS Crime Statistics team to agree the review parameters and to pool our collective knowledge about examples of crime surveys and transformation work in other countries so that the review built on this knowledge. Following this scoping exercise, an evidence framework was set up to enable information to be compiled systematically. This contained summary fields for the overall methodology alongside key questions the REA sought to address. These fields are summarised in Table 3.1.

Table 3.1 – Summary fields and key questions for REA

Area	***Detail***
Basic background information	Country, survey name, frequency, when most recent survey was conducted and link to a source document
Sampling population and sample source	Population for the survey, sample design, and the nature of the sample frame (for example, is a population register available?)
Mode(s) of data collection	Which international crime surveys use online-self completion modes of data collection and how are sample members contacted? Is online self-completion used for the key parts of the questionnaire that measure prevalence and incidence? Is online self-completion used as a single mode or is it used in conjunction with other modes and, if so, how? Which international crime surveys have transitioned to online self-completion from previously interviewer-administered modes? Where this has happened what compromises, if any, were made and how were issues of comparability addressed? For surveys which have not yet taken this step, what feasibility work has been undertaken to explore the option of online self-completion and what are the learnings to date?

Table 3.1 – Summary fields and key questions for REA	
Changes to overall survey methodology over time	Apart from mode shifts, have other aspects of the methodology of international crime surveys changed over time? How did surveys manage fieldwork during the covid pandemic and were online or video methods used to enable continued data collection?
Approach to measuring crime	Which international crime surveys measure incidence (counts) as well as prevalence and how do they attempt to measure this in an online context (for example, whether all incidents are separate or part of a series)? Which international crime surveys use the 'screener' approach also adopted by CSEW and if so, what does this consist of? Which surveys, if any, conduct offence coding like the CSEW to formally classify crimes?
Survey content	What are the range of other topics covered by international crime surveys, including the extent to which sensitive crimes such as assault, sexual assault, threats and harassment captured?

3.1.1 Surveys included in the REA

A total of 18 international crime surveys were initially identified. Three were subsequently excluded from the review as the research team were unable to find sufficient information about the survey in English language. These were crime surveys in Denmark, Italy and Spain, meaning a total of 15 were included in the final review. The remainder of this chapter synthesises the overall review, structured around the questions presented in the table above.

The full evidence framework can be found in Appendix 1.

3.2 Basic background information

Table 3.2 summarises the surveys included in the REA and their key characteristics.

Table 3.2 – Surveys reviewed and key characteristics											
Country/ countries	***Survey name***	***ID***	***Link to main source***	***Named sample source***	***Uses CAWI***	***CAWI context***	***Exploring CAWI***	***Measures incidence (counts)***	***Collects series/ separate***	***Covers sensitive crimes***	***Conducts offence coding***
UK surveys											
England and Wales	Crime Survey for England and Wales (CSEW)	ID1	CSEW				✓	✓	✓	✓	✓
Northern Ireland	Northern Ireland Safe Community Survey (NISCS)	ID2	NISCS					✓	✓	✓	✓
Scotland	Scottish Crime and Justice Survey (SCJS)	ID3	SCJS		✓	Self-completion module only		✓	✓	✓	✓
EU surveys											
EU	Crime, Safety and Victims' Rights Fundamental Rights Survey	ID4	EU	✓ - some countries	✓	Strategies differed across the 10 countries using CAWI				✓	
Finland	Finnish National Crime Victim Survey (FNCVS)	ID5	FNCVS	✓	✓	Push to web		✓		✓	
France	Victimisation Survey - Living environment and Security	ID6	LES				✓				
Germany	German Victimisation Survey (DVS)	ID7	DVS	✓ sub group				✓			
Ireland	Crime and Victimisation, Quarterly National Household Survey (QNHS)	ID8	QNHS								
Netherlands	Safety Monitor	ID9	NSM	✓	✓	Push to web				✓	
Sweden	Swedish Crime Survey (SCS)	ID10	SCS	✓	✓	Push to web		✓		✓	

Table 3.2 – Surveys reviewed and key characteristics											
Country/ countries	***Survey name***	***ID***	***Link to main source***	***Named sample source***	***Uses CAWI***	***CAWI context***	***Exploring CAWI***	***Measures incidence (counts)***	***Collects series/ separate***	***Covers sensitive crimes***	***Conducts offence coding***
Rest of world											
Australia	Crime Victimisation Survey (CVS) (topic on the Multi-Purpose Household Survey)	ID11	ACVS							✓	
Canada	General Social Survey on Canadians' Safety (GSS)	ID12	GSS		✓	Push to web		✓		✓	✓
Israel	Crime Victimization Survey	ID13	ISCVS	✓						✓	
Japan	National Crime Victimization Survey (NCVS)	ID14	JNCVS		✓	Self-completion module only		✓		✓	
New Zealand	New Zealand Crime and Victims Survey (NZCVS)	ID15	NZCVS					✓	✓	✓	✓
United States	National Crime Victimization Survey (NCVS)	ID16	USNCVS				✓	✓	✓	✓	

3.3 Populations and sample source

Population for the survey, sample design, and the nature of the sample frame (for example, is a population register available?)

The sample frame available for any crime survey is an important consideration both in terms of the survey design and data collection mode, as well as influencing the likely level of response. Many crime surveys follow the model of the CSEW and seek to interview only one eligible adult in a household rather than seek to interview all eligible adults in a household.

Where a sample frame of individuals is available, for example in the form of a population register, this is a straightforward process and means it is possible to target selected individuals using any data collection mode. As well as names and addresses, some countries' population registers may also contain important contact details such as telephone numbers. In countries where only a household or address level sample frame exists (or where there is no sample frame) a two-stage selection process is generally employed: a sample of addresses is drawn initially and then at each address (or household) an eligible individual is selected for interview. While this approach works well with interviewer-administered surveys where the interview can carry out the second stage, it is less well tested on self-completion surveys where the number of occupants at an address is unknown and any selection process is reliant on the occupants at the sampled address following the correct protocol.

As with the CSEW, many of the surveys reviewed used address or household level sample frames without access to individual names. However, six surveys benefitted from sample frames that included individuals' names. These were:

- **Finland**: The Finnish National Crime Victim Survey (FNCVS, ID5) was first conducted in 1980 and has been conducted annually since 2012. The sample of people aged 15-74 is randomly selected from the Finnish Population Register.
- **Germany**: The German Victimisation Survey (DVS, ID7), last carried out in 2017, selected households via randomly generated landline and mobile telephone numbers for the base sample but included an additional separately sourced sample group of people with Turkish origin (3.3% of respondents). For this, a different method was used where individuals were sampled using a name-based classification procedure from entries in current telephone directories.
- **Netherlands**: The long-running Safety Monitor (ID9) draws a stratified sample of all non-institutionalised Dutch residents aged 15 years or older who are registered with their local municipality.
- **Sweden**: The Swedish Crime Survey (SCS, ID10), established in 2006, draws its sample from a population register of named individuals and can link to other databases to match to telephone numbers where possible.
- **EU**: The Crime, Safety and Victims' Rights Fundamental Rights Survey (ID4), carried out in 2019, was the first EU-wide survey on crime victimisation experiences. This survey differs to the others included in this review in that it covered 29 countries and the focus was on exploring experiences across a broad range of countries within the EU rather than on the detail required for National Statistics for a specific country. Sampling methods differed across the 29 countries. Available sampling frames were assessed and those that offered close to 100% coverage of the population nationwide were selected and a random probability sample drawn. These sample

frames could include sources such as population registers or registers of addresses. Where such sample frames did not exist, or were unavailable, the sample was selected in a multi-stage selection procedure, as described above.

- **Israel**: The Crime Victimization Survey (ID13), run annually since 2014, uses the Population Register as its sampling frame, this updated to April 2018. This enabled a representative sample of the Israeli population to be drawn, divided according to sector, sex, population group, age group, education level, economic situation as well as other characteristics.

It is noticeable that three of the crime surveys mentioned above that have been early adopters of an online self-completion approach (the Finnish FNCVS (ID5), the Netherlands Safety Monitor (ID9) and the Swedish Crime Survey (ID10)) all have access to a population register as a sample frame. Being able to sample from a population register is clearly an advantage when using an online data collection approach if only one eligible person in a household is being interviewed.

3.4 Mode(s) of data collection

Which international crime surveys use online-self completion modes of data collection and how are sample members contacted?

Most of the international crime surveys reviewed currently use interviewer-administered data collection modes: Computer-Assisted Personal Interviews (CAPI), Computer-Assisted Telephone Interviews (CATI), or a combination of the two. However, seven of the surveys reviewed make some use of online self-completion methods (CAWI) to collect information about victimisation. However, it is important to stress that these surveys differ quite widely in terms of the context in which online data collection is used and several of them use online data collection in conjunction with another mode (generally telephone).

Four of the seven surveys adopt(ed) the 'Push to Web' approach, that is named individuals are sent a letter which includes login details to the online survey and asked to go online to complete it. Three of these - the Finnish FNCVS (ID5), the Netherlands Safety Monitor (ID9) and the Swedish Crime Survey (ID10) - all benefit from using population registers as sampling frames while Canada's GSS (ID12) uses Statistics Canada dwelling frame as its primary source and where possible attaches telephone numbers to addresses which are taken from the Census and various administrative sources.

Many crime surveys which are primarily interviewer-administered face-to-face do include a self-completion module, typically for sensitive questions and topic areas. This is completed by the respondent on the interviewer's device (Computer-Assisted Self-Interview (CASI)) usually at the end of the main survey. For some international crime surveys, this is the segment that is conducted via, or has transitioned to, online self-completion, a simpler initial shift than the entire survey. The Scottish Crime and Justice Survey (SCJS, ID3) during the 2021/22 survey year and Japan's NCVS (ID14) used CAWI for the self-completion part of the questionnaire only, covering more sensitive types of crime. Interviewer-administered modes were used for the core parts of the survey and the online modules were conducted as a follow-up to the main part of the survey. The rate of attrition to Japan's NCVS was extremely low. Of the 6,000 adults sampled, 61.8% participated in the main survey (3,709) and 58.3% (3,500) completed the follow up self-administered questionnaire.

The USA National Crime Victimization Survey (NCVS, ID16) plans to implement the new online instrument in 2025 as part of a multi- or mixed-mode design alongside interviewer-administered modes. The redeveloped questionnaire has currently been tested via

interviewer-administered methods (face-to-face and telephone) with extensive work exploring its feasibility in an online context planned next, including a split sample test in 2024. For some surveys, it is clear from the outset that CAWI is not currently a feasible option. A notable example is Australia (ID11), where there are likely to be concerns about internet access in hard-to-access regions, where currently only telephone interviews are carried out.

Since 2019, the French Ministerial Statistical Department for Internal Security (SSMSI) has been carrying out work to redesign their victimisation survey, the Living environment and security (CVS, ID6) survey, which has been running since 2007. The intention is to implement a two-phase survey from 2022: a large screening phase (c.200,000) to measure prevalence and incidence of crime and identify victims of rare events followed by a smaller second phase to understand the nature of victimisation in detail. While it is understood that some part of this redesign will involve online data collection there is currently nothing published on the testing work undertaken by SSMSI prior to launch of the survey and attempts to gain information on this work proved unsuccessful.

Data collection and sampling methodology differed across the 29 countries included in the EU Survey (ID4) and so it has not been possible to explore the different approaches used across all 10 countries which used an online approach.

The following section explores the international crime surveys that use online-self completion methods in more detail.

Is online self-completion used for the key parts of the questionnaire that measure prevalence and incidence? Is online self-completion used as a single mode or is it used in conjunction with other modes and, if so, how?

Below are details of the surveys that use online-self completion to measure victimisation.

- **Finland**: The FNCVS (ID5) was conducted via CATI between 1970 and 2009, but this mode was found to be extremely costly. Since 2012, data has been collected via online and paper self-completion modes, the content of which are the same to allow comparability. Participants are sent a letter containing the paper questionnaire and can either fill this in and send it back or choose to complete the survey online. An experiment carried out in 2014, compared three modes of data collection (CAPI, CATI and CAWI), with the modes being as harmonised as possible to ensure comparability. The experiment found that the CAWI mode produced the highest prevalence estimates for most fear-related questions and property crimes and that the CATI mode produced lower prevalence estimates for violence compared with self-administered interviews (CASI or CAWI). The paper concluded that there was a social desirability effect for the interviewer-administered modes which led to a degree of under reporting, especially of sensitive crimes. It suggested that modes which offered a greater degree of confidentiality were likely to produce higher victimisation rates. Crucially the paper recommended that mixed-mode strategies should be explored, as none of the three 'single-mode strategies' were considered obviously better: for example, it suggested initially inviting respondents to take part in a CAWI interview but also offering non-responders an alternative mode after a certain period[23].

[23] https://academic.oup.com/jssam/article-abstract/2/4/459/2937101

- **Netherlands**: The annual Safety Monitor (ID9) began using CAWI in 2008 as part of a mixed mode data collection approach involving face-to-face, telephone, online and paper data collection. In 2011, an extensive experiment was conducted by researchers at the Central Bureau of Statistics (CBS) to look at total mode effects on various Dutch surveys, including the Safety Monitor. This found that respondents who completed the survey online tended to report more victimisation, feel less safe, and were more likely to report being bothered by anti-social behaviour (nuisance) compared with respondents who completed the survey by interviewer mode. The researchers concluded that these differences were primarily due to measurement effects rather than coverage or non-response differences. Following this work the Safety Monitor moved to self-completion mode only from 2012 using a standard push to web approach, with the sample being drawn from administrative records. Letters are mailed to the sample asking them to complete the survey online, with non-responders receiving two reminders, which also include a paper questionnaire[24].
- **Sweden**: The SCS (ID10) transitioned from mainly telephone interviews to online and postal questionnaires in 2017 using a population register as the sample frame. A driving factor was cost: the cost per CATI interview was increasing due to the extra effort required to achieve completed interviews and the desire to increase the sample size meant that CATI was becoming less of a viable methodology in the longer term. The switch to a self-completion approach in 2017 allowed a larger sample size at significantly lower cost. The SCS includes 14 screeners, these administered along with general attitudinal questions via either an online or paper self-completion questionnaire. There is also a separate telephone interviewer-administered (CATI) follow up of victims identified at the screener stage: this contains questions equivalent to the Victim Module of the CSEW, with victims being asked about up to three incidents. If duplicates are identified these are skipped, but no adjustment is made to the prevalence estimates.
- **EU**: The Crime, Safety and Victims' Rights Fundamental Rights Survey (ID4) adopted different modes of administration across the 29 countries. A feasibility assessment in 2015–2016 identified suitable sample frames for a representative online survey in each country and then pilot surveys were conducted to confirm which countries would be suited to online data collection. Of the 29 countries, 10 adopted an online approach, these were: Austria, Denmark, Estonia, Finland, France, Germany, Luxembourg, the Netherlands, Sweden and the United Kingdom. The remaining 19 countries used a CAPI approach with an interviewer contacting the respondents in person. In France and Germany, the survey was carried out with quota samples of people participating in online survey panels because access to a suitable sample frame could not be obtained.
- **Canada**: The GSS (ID12) runs every five years and up until 2019 used a mixed-mode (CATI and CAPI) approach. In 2019 an online survey was introduced as part of a mixed mode strategy, with sampled households being sent a letter asking them to complete the survey online and non-responders being followed up by telephone or face-to-face. In 2019, approximately 60% of all completed interviews were conducted online, with the remainder being conducted by CATI or CAPI. No experimental work was done to investigate the impact of introducing an online mode

[24] https://www.researchgate.net/publication/283733119_Disentangling_mode-specific_selection_and_measurement_bias_in_social_surveys

to the survey and so it is impossible to attribute whether any changes in estimates are due to the mode change or not. However, they do note the general evidence on social desirability bias and how mode might affect estimates and caution against comparing the 2019 results with previous iterations of the survey.

Surveys that use online-self completion as a supplementary mode for sensitive topics

- **Scotland**: The Scottish Crime and Justice Survey (SCJS, ID3) used a mixed-mode approach in the 2021/22 survey year. A 'Knock to Nudge' approach was used to offer respondents an interview by phone or video call, with the self-completion section (which covers more sensitive questions) completed by the respondent online or on paper.
- **Japan**: While the main part of the Japanese NCVS (ID14) is conducted face-to-face, a separate survey on stalking, domestic violence, child abuse and sexual incidents is conducted via self-completed questionnaires which could be submitted online.

Mode effects – summary

- Mode effects are recognised in every case where there has been a change in survey delivery and administration. Where change has been accompanied by robust experiments the results have been as anticipated: online self-completion tends to produce higher victimisation rates, especially for sensitive crimes such as violence. However, caution is needed as it is difficult to separate out coverage, non-response (sampling), and measurement (mode) effects. Considering this finding, it seems clear that more consideration needs to be given in the next stage of the redevelopment work for the CSEW about how mode effects should be addressed in any future mixed mode design. Also, since any change of mode seems likely to have an impact on the time series data, thought needs to be given as to how this is addressed and presented
- For the few surveys that have already adopted online self-completion modes this has been undertaken as part of a mixed mode strategy: either a combination of self-completion and interviewer-administered or self-completion but with an online or paper option. The use of paper self-completion is especially critical as it is generally accepted that CSEW (in its current form) is unsuitable for paper administration meaning that consideration does need to be given to avoiding digital exclusion in redesigning the survey.
- There is a trend towards dividing surveys into different components which are conducted essentially as separate surveys and by different modes due either to the complexity of the data being collected or the sensitivity of the subject matter: for example, the Swedish/French model of CAWI/PAPI followed by CATI (complexity of collecting victimisation details on a CAWI/PAPI survey) or the Scottish/Japanese model of CATI with CAWI/PAPI follow up (collecting data considered too sensitive for CATI) . Since this has both pros and cons attached to it the value of such an approach on the CSEW would need further thought.

Which international crime surveys have transitioned to online self-completion from previously interviewer-administered modes? Where this has happened, what compromises, if any, were made and how were issues of comparability addressed?

A key focus for the CSEW is finding ways to transition to CAWI while managing the balance between the need to simplify and streamline certain aspects of the current survey

to be suitable for a self-completion mode while also trying to retain and accurately capture the complexities of victims' experiences currently in the CAPI instrument.

Of the surveys which have already moved towards a self-completion mode either entirely or as part of a mixed-mode approach, there are two important findings to emerge which are of relevance to the CSEW:

- No other survey is as complex as the CSEW in terms of what it is trying to measure. All the surveys that have transitioned to online self-completion to date only measure prevalence and make no attempt to measure incidence rates (i.e. counts). Furthermore, in most cases the range of crimes that are measured are generally far less than the range of crimes measured in the CSEW. Details about each of these surveys is covered in the section below.
- The fact that surveys which have transitioned to online self-completion are generally simpler than the CSEW in content and structure means that it has been feasible to retain a paper self-completion option as part of the overall mode mix. It is universally accepted that the complexity of the current CSEW makes a paper questionnaire option unfeasible, unless it is a cut down version of the whole survey (e.g. just the screener questions).

Information about the learnings and compromises made from four surveys that have transitioned to online self-completion to some degree are discussed below:

- **Finland**: The FNCVS (ID5) first conducted in 1980, underwent significant redesign prior to 2012 to better suit self-administration through making the questionnaire easier and more accessible and to minimise dropout. The updated version focuses solely on crime victimisation rather than including additional sections covering topics such as accidents (for example, traffic and workplace accidents). The new 'slimmed down' version also includes fewer detailed follow up questions than the original version: for example, questions about the specifics of any injuries were excluded. A report on the development of the new survey was published in 2011 but is only available in Finnish[25].
- **Netherlands**: The Safety Monitor (ID9) has been conducted annually since 2012 using CAWI as the primary mode, with postal questionnaires included in reminder communications. The data are comparable from 2012 and adjustments for differences in methodology and the questionnaire have been applied to specific indicators to allow comparability with data collected between 2005 and 2012. The 2021 report notes that *"internal consistency and completeness is checked"* and that this takes place within the online questionnaire which suggests there is a checking process for inconsistencies (for example, possible double counting of incidents). Paper questionnaires are automatically checked post-completion and inconsistent or incomplete answers are edited using bespoke control and correction processes. It would be useful to understand the specific details of the changes made to the questionnaire to suit online administration and the checking processes used within it, but unfortunately no specific details have been sourced in English.
- **Canada**: As noted above, online self-completion was offered to survey respondents of the GSS (ID12) for the first time in 2019. As part of this move, changes were made to the questionnaire with question wording and answer categories being

[25] http://hdl.handle.net/10138/152504

modified to better suit online self-completion. Examples of changes included adding interviewer instructions to clarify which parts of the question should be read aloud and removal of pronouns to better suit self-completion. The 'refuse' option was removed and 'don't know' was retained only for questions where it was anticipated to be used frequently. It was also made possible to skip questions without giving any answer.

The potential impact of mode effects that this transition may have on the estimates is acknowledged and it is also noted that some questions around sensitive crimes are likely to be more prone to social desirability bias than simple 'factual' questions. While considerable effort was made at all processing, verification and dissemination stages to ensure the data were precise and of good quality, the 2019 quality report concludes that *"because of these changes, it is not appropriate to compare results from the 2019 GSS with previous iterations, or at the very least such comparisons must be accompanied with a warning"*.

— **Sweden**: The SCS (ID10) has a time trend dating back many years and the primary purpose of the survey is to measure change over time and compare different groups in the population rather than estimate exact levels of crime at any time point. Because of this it was important to develop a method to enable results collected before (2007-2016) and after (2017 onwards) the transition to self-completion to be compared. In transitioning the screener section of the Swedish Crime Survey to a self-completion approach the changes included: switching the order of questions to start with attitudinal questions relevant for all respondents, before moving onto the victimisation screeners; making explanatory text explicit rather than relying on interviewer discretion on a case-by-case basis; reducing the overall number of questions (for example, by merging questions where possible); and amending wording to suit online administration. Since these changes, prevalence of crimes has generally increased, especially for crimes of a sensitive nature including harassment, which may be attributable to reduced levels of social desirability bias in the online self-completion version. The Swedish Crime Survey team are considering shifting the follow up survey of victims currently conducted by telephone to be part of the main self-completion survey. This is primarily due to the level of attrition that occurs between the screeners and follow up victim survey as well as the cost of conducting the telephone interviews. However, this is still under consideration due to data quality concerns, including the fact that the current questions may not suit online administration.

For surveys which have not yet taken this step, what feasibility work has been undertaken to explore the option of online self-completion and are the learnings to date?

Other than the surveys already discussed, and the previous exploratory work undertaken on CSEW to investigate the feasibility of transitioning to a mixed-mode design involving an element of online self-completion, the other main survey to examine these issues in detail is the **National Crime Victimization Survey (NCVS)** conducted by the US Bureau of Justice Statistics (ID16). This work has several different objectives but primarily involves the complete redesign of the survey instrument to modernise the questionnaire, generate better and more comprehensive measures of crime, and engage non victims by adding questions on police topics such as performance and community safety.

An extensive multiyear package of research and testing is underway ahead of the goal of having a full redesigned instrument in place by January 2025[26]. Developing and testing an online (CAWI) self-completion version of the questionnaire is part of the research programme although no decision has yet been made about whether the NCVS will introduce an online component. The US Bureau of Justice Statistics (BJS) plan to conduct more field tests before making any changes to the current survey design and data collection mode, which is based on a panel design with an initial face-to-face (CAPI) interview at the first wave followed by telephone (CATI) interviews at subsequent waves.

While shifting the data collection mode of the NCVS to a mixed-mode approach involving some element of online self-completion is not the primary purpose of this redesign work the work is extremely useful as the NCVS is the crime survey most like the CSEW in terms of what it is trying to measure, and the measurement challenges it faces. The survey aims to measure both prevalence and incidence of a wide range of crimes and uses a screener and victim module structure which is like the CSEW questionnaire. It faces the same challenges as the CSEW in terms of, for example, avoiding the double counting of the same incidents and how to define separate incidents from a series of incidents.

Proposed changes to the questionnaire focus on the structure of the instrument rather than the content of questions and are broadly like the changes made by the Swedish Crime Survey (SCS, ID10) summarised above. Screener probes have been separated into shorter examples with Yes/No responses to each screener series; ‘behaviourally specific’ language has been improved; screener probes for rape or sexual assault have been expanded and vandalism added to the list of screeners. Terminology and information in the Crime Incident Report (CIR) (equivalent to the victim module on the CSEW) has also been improved to increase the amount of information collected. At the same time an online self-completion version of the redesigned instrument was developed for future testing.

Critically, the redesigned incident developed two different approaches for testing: an interleaved approach where follow up questions were included after each screener question to help correctly classify crimes at an earlier stage of the interview; and a non-interleaved approach where follow up questions were asked at the start of the CIR to help correctly classify crimes. In 2019-20 a large-scale field test was conducted which tested the two versions of the redesigned instrument against the existing NCVS questionnaire. This test involved only interviewer-administered data collection (CAPI and CATI) and so did not test online self-completion data collection.

[26] https://bjs.ojp.gov/content/pub/pdf/uncvsir_sum.pdf

The findings of this test found that the redesigned questionnaire produced higher victimisation rates for some types of crime (violence and property crimes) compared with the current screener; more information about incidents was collected in the redesigned CIR; and the redesigned instrument was better at correctly classifying incidents compared with the current survey. The test suggested that the non-interleaved version of the questionnaire performed better than the interleaved version in terms of correctly classifying incidents as crimes and was also generally easier for respondents to understand.[27]

The next stage of the work involves a split-sample design on the live survey from 2024 when respondents will be assigned either the current instrument or the new redesigned instrument based on the non-interleaved approach. This will allow comparison of the old and new designs as well as insights into the redesign's impact on victimisation rates and help inform whether statistical adjustments are needed to maintain trend data.

3.5 Changes to overall survey methodology over time

Apart from mode shifts, have other aspects of the methodology of international crime surveys changed over time?

International crime surveys included in this evidence review predominantly use interviewer-administered modes, either with only in-home or telephone interviews or a mixed CAPI and CATI approach. As noted, some surveys also include a self-completion module (CASI) as part of the interview, mainly for sensitive topic areas.

With interviewer-administered surveys, paper questionnaires have been relatively rare and, where they have been used, they tend to be used for sensitive topics either as an alternative to CASI or for subjects considered unsuitable for a telephone interview. However, the shift to an online mode has generally been accompanied by paper self-completion questionnaires as an alternative mode to avoid digital exclusion. Thus, paper questionnaires are used in Sweden (ID10), the Netherlands (ID9), and Finland (ID5) as part of the shift to a self-completion approach.

Other than the transition to online self-completion, the other key shifts in mode relate to sudden mitigating changes enforced by the covid pandemic, covered in the following section.

How did surveys manage fieldwork during the covid pandemic and were online or video methods used to enable continued data collection?

Among the surveys reviewed, there were three main strategies for managing in-person (CAPI) data collection during the covid pandemic:

- **Switch to remote interviewer-administered interviewing (CATI or CAVI).** This was the approach taken by CSEW and the Northern Ireland NISCS (ID2). The Scottish Crime and Justice Survey (ID3) moved from mainly CAPI to a blend of CATI and CAVI while the self-completion module shifted to CAWI. After a short suspension of fieldwork, the United States' NCVS (ID16) moved to a CATI design and extended the number of waves.
- **Switch to online self-administered interviewing (CAWI).** As mentioned above, the Scottish Crime and Justice Survey (ID3) began using online questionnaires for the self-completion element of the survey. The aim was to retain the more sensitive parts of the survey as a self-completion survey but using CAWI instead of CASI

[27] https://www.ojp.gov/pdffiles1/bjs/grants/303980.pdf

online, while maintaining interviewer-administration for the main part of the survey. As noted, the Swedish Crime Survey (SCS, ID10) and the General Social Survey on Canadians' Safety (GSS, ID12) had already shifted to an online/paper self-completion or interviewer-administered telephone approach in 2017 and 2019 respectively, which meant that Covid restrictions had negligible impact on their administration.

— **Temporarily suspend in-person fieldwork.** New Zealand Crime and Victims Survey (NZCVS, ID15) suspended fieldwork in 2020 during which time it was replaced by a telephone-based COVID-19 Justice Sector Survey. The United States' NCVS (ID16) also suspended fieldwork in March 2020 before switching to CATI as noted above.

3.6 Approach to measuring crime

Which international crime surveys measure incidence (counts) as well as prevalence and how do they attempt to measure this in an online context (for example, whether all incidents are separate or part of a series)?

The CSEW estimates both the prevalence and incidence of crime, the latter acknowledged to be a particular challenge in the context of an online self-completion interview. Of the international crime surveys reviewed, nine measure(d) incidence as well as prevalence, that is they collect a count for each type of crime. These were: Northern Ireland (ID2), Scotland (ID3), Finland (ID5), Germany (ID7), Sweden (ID10), Canada (ID12), Japan (ID14), New Zealand (ID15) and the United States (ID16). Of these, four attempted to untangle whether multiple incidents are part of a series (that is similar in nature) or should be separately counted (that is different in nature). These are: Northern Ireland (ID2), Scotland (ID3), New Zealand (ID15) and the United States (ID16).

Of the international crime surveys that measure incidence (that is they include count questions), three have developed an online self-completion instrument:

— **Finland**: The FNCVS (ID5) collects simple counts for all types of crime included in the questionnaire although the approach differs for property crimes and violent crimes (explained in more detail in the following section). The survey does not differentiate between incidents that are part of a series and those that stand out as being different.

— **Sweden**: The SCS (ID10) collects crime counts but does not try to untangle separate incidents from those that are part of a series. As noted, the key difference between this survey and CSEW is that they separate out prevalence from the information collected in the victim module and don't try to match the two.

— **United States**: The NCVS (ID16) measures incidence as well as prevalence and is the only survey included in this review, other than CSEW, that is exploring the feasibility of an online instrument that attempts to establish whether multiple incidents are part of a series or distinct (separate) from each other. However, this is done is a very different way from the current CSEW approach. On the NCVS if four or more incidents are entered at the count question, the participant is asked '*Do you recall enough details about each incident to be able to distinguish them from each other?*' If they can distinguish between them then the incidents are followed up separately at the Crime Incident Report stage. If they cannot distinguish between incidents, then they are considered a series and only the most recent is followed up in detail. This means that incidents can only be classified as all separate or all part of a series: a combination of both separate and similar incidents is not possible.

Which international crime surveys use the 'screener' approach adopted by CSEW and if so, what does this consist of?

The CSEW adopts a 'screener' approach to measuring crime where the participant is asked a series of 'Yes/No' questions, each covering a different type of incident. The rationale for this approach is that it encourages recall of incidents that are less salient and memorable and thus less likely to have been reported to police. The research team were unable to locate questionnaire wording for all surveys included in this review but from those that could be accessed, it was found that this approach is commonly used across international crime surveys including: Northern Ireland (ID2), Scotland (ID3), Finland (ID5), the Netherlands (ID9), Sweden (ID10), Canada (ID12), Israel (ID13), New Zealand (ID15) and the United States (ID16).

Of these, five surveys have a self-completion version of the questionnaire.

Finland: The FNCVS (ID5) uses a screener approach for all crimes covered although the approach differs for property crimes and violence crimes. Each of the 16 initial 'property' screeners are presented as 'Yes/No' questions and a follow up block contains a count question for each. The 16 property crime screeners are: Theft of car, Theft from inside car or of car parts, Car damage, Theft of boat, Theft of boat parts, Theft of motorcycle, Theft of bicycle, Theft of bicycle parts, Home break-in/attempted home break-in, Holiday home break in/attempted holiday home break-in, Storage space break-in/attempted storage space break-in, Damage to personal property, Theft of personal property, Non receipt of goods or services paid for, Use of credit card or money taken from bank account and Theft of personal identification.

In contrast, the violence and assault screeners are presented as a grid of 13 types of crime[28], which is broken down based on who the perpetrator is, rather than asked as simple 'Yes/No' questions. The four categories for the perpetrator are: Nobody, Former or present partner, Someone else you know closely and Someone you don't know or only know remotely. The respondent is then asked to code whether **any** of these assaults took place in these locations: At your workplace, In a restaurant or café, In a public place, In your own home, In someone else's home and Somewhere else. The subsequent count questions for violence and assault are for each **place** in which the assault(s) occurred rather than for each type of assault.

Netherlands: The Safety Monitor (ID9) 2021 questionnaire uses a screener approach (a total of 23) but differs from CSEW in that it uses a loop design to ask detailed questions (like those asked in the CSEW victim module) about the most recent incident after each screener. The questionnaire collects counts for each type of incident, capped at five times or more, for all types of non-fraud crime and only some types of fraud. The questionnaire attempts to detect double counting by asking whether subsequent incidents happened at the same time as earlier ones: for example, 'Did this [theft/incident] happen at the same time as a crime mentioned earlier?'. The questions about online crime are more extensive in comparison with CSEW.

The 12 non-fraud crime types are: Home break-in, Theft of personal items from car, Theft of car parts, Theft of car, Theft of bike, Theft of other motor vehicle, Personal theft, Other theft, Damage not including theft, Sexual assault, Threats and Physical assault. The 11 online crime types are: Purchase fraud, Sales fraud, Hacking device, Hacking account,

[28] Threatened to hurt you physically (not in person), Threatened to hurt you physically (in person), Obstructed your movement or grabbed you, Pushed or shoved you, Slapped you, Pulled your hair, Hit you with a fist, Hit you with a hard object, Kicked or strangled you, Used a weapon, Sexual assault, Attempted sexual assault, and Other kind of physical violence.

Phishing, Payment fraud, Identity fraud, Bullying, Stalking, Shame-sexting (revenge porn) and Other online crime.

Sweden: The SCS (ID10) includes 14 screeners covering the following areas: Bike theft, Car theft, Theft from vehicle, Home break-in, Identity theft, Being deceived out of money, Personal theft, Robbery or attempted robbery via threats or violence, Sexual assault, Physical assault, Threats, Harassment, Online bullying and Other crime. Participants are asked to record the number of times each incident took place in the last year. The self-completion questionnaire does not try to control or prevent double counting of the same incident at different screener questions and if duplicate incidents are identified these are skipped, but no adjustment is made to the prevalence statistics. Although this may mean that the prevalence estimates are inflated due to double counting, as the key interest is in tracking over time, this is not considered to be of huge concern based on the assumption that any measurement error remains constant over time.

Canada: The GSS (ID12) asks a series of 15 screeners, following each up with a count question, which is capped at 95. The 15 screeners are: Damage to property, Robbery/attempted robbery via threats or violence, Home break-in/attempted home break-in, Theft from outside home, Other/attempted other personal theft, Theft while on holiday in Canada, Theft/attempted theft of vehicle or vehicle parts, Vehicle damage, Other theft, Assault, Threatened assault, Sexual assault/attempted sexual assault, Unwanted touching in a sexual way, Other sexual assault and Other crime.

Where an incident has been recorded, subsequent screeners include the wording *'Excluding the incident(s) already reported…*' in an attempt to discourage double counting. Following the screeners, participants are given a summary of the total number of incidents they reported within each crime type and asked for confirmation or correction. If participants have not experienced any crime, this is also verified and corrected if something has happened to them in the last year which was not picked up in the screeners.

There is a further summary of incidents reported later in the questionnaire, just before the set of Crime Incident Reports (CIRs). There is an option at the beginning of each CIR to remove duplicate incidents ('Details already reported under another incident'). It is not clear from the published material whether there is a cap on how many incidents are followed up on.

United States: As summarised above, a lot of work is planned to redesign and test the NCVS (ID16) questionnaire before its introduction in 2024 with the aim being to increase the quality of information collected and the efficiency of the instrument flow. The redesigned questionnaire will maintain its two-stage measurement approach in screening and classifying victimisation.

A total of seven screeners are included covering: Vehicle theft/attempted vehicle theft, Vehicle part/attempted vehicle part theft, Theft/attempted theft, Home break-in/attempted home break-in, Vandalism (including the killing or injuring of an animal or livestock), Attack/attempted attack, and Unwanted sexual contact.

It is noticeable that the number of screeners is considerably lower than the CSEW (and many other crime surveys). This is partly because each screener covers several different combinations that are broken down into separate screener questions in other surveys: for example, all the screeners combine both actual and attempted crimes, while the theft screener covers all thefts and attempted thefts (except vehicle theft), whether from the home, outside the home, the person, or from vehicles. The NCVS also covers fraud and

cybercrime as a separate module from the main survey. This limitation on screeners has a practical advantage in terms of making the double counting check easier.

While different versions of the questionnaire have been tested (interleaved and non-interleaved), both approaches collect some key information after each screener:

- **A count** of the number of incidents within each crime type. If the respondent does not know the number of incidents, it is automatically set to a count of one.
- **Whether multiple incidents are similar or separate**: If four or more incidents are recorded, the participant is asked '*Do you recall enough details about each incident to be able to distinguish them from each other*?' If they can distinguish between them then the incidents are treated as separate but if the respondent cannot distinguish between then they are all considered a series (count of one) and only the most recent is followed up in detail. If two or three incidents are recorded then they are automatically treated as separate incidents and followed up individually.

Next, for each incident, up to a maximum of four, questions are asked about:

- **Date**: the month and year of the incident(s)
- **Double counting check**: Each incident is checked against all previously recorded incidents if the dates fall within the same month and year (or where the date is not given). Participants are asked '*Was this incident part of any other incident you have already mentioned?*' and, if so, they are shown a list of incidents already recorded and asked to link it to the correct one ('*Which incident was this part of?*'). The limit on the total number of screeners included in the survey (seven) and the cap on the number of incidents followed up at each screener (four) means the check is just about manageable from a respondent perspective but with a greater number of screeners and a higher incident cap it is easy to see how this approach could easily become unworkable. For context, the WCSEW includes 29 screeners (across both non-fraud and fraud crimes) and follows up a maximum of two incidents per screener. More detailed information on the redesign of the NCVS can be found on their dedicated website[29].

Which surveys, if any, conduct offence coding similar to the CSEW to formally classify crimes?

The CSEW classifies each incident recorded in a victim module using Home Office Counting Rules[30]. The open description and the victim module data is reviewed by specially trained coders who determine whether what has been reported constitutes a crime and, if so, what offence coding should be assigned to the crime. These data can then be compared with crimes reported to police to give an overall picture of crime in England and Wales.

Offence coding requires data to be of an extremely high quality to ensure enough information is available to classify the incident. Where offence classification varies across the sample population, for example across country or state, such as the EU's Crime, Safety and Victims' Rights Fundamental Rights Survey (ID4) and the US NCVS (ID16), offence coding is not practical.

[29] https://bjs.ojp.gov/programs/ncvs/instrument-redesign

[30] https://www.gov.uk/government/publications/counting-rules-for-recorded-crime

Of the international crime surveys reviewed, four others conducted offence coding to formally classify crimes. These were: the NISCS (ID2), the SCJS (ID3), the GSS(ID12) and the NZCVS (ID15).

3.7 Survey content

What are the range of other topics covered by international crime surveys, including the extent to which sensitive crimes such as assault, sexual assault, threats and harassment are captured?

The topics included across the international crime surveys selected covered a very wide range. In addition to measuring victimisation, topics covered included: perceptions and opinions of crime and criminal justice systems, experiences with and trust in the police, crime prevention and security, feelings of safety in the area you live in, nuisance and anti-social behaviour. More detail on topic areas is included in the full framework.

The CSEW includes almost all types of victimisation, including very sensitive crimes such as assault, sexual assault, threats and harassment. The collection of this type of data means there is a lot to consider when exploring the feasibility of transitioning to a new self-completion mode without an interviewer present to provide reassurance, clarification or to signpost support services. As noted earlier in this report, this package of work focused on the sections of the CSEW that measure victimisation (the screeners and victim modules), but the wider CSEW covers a much broader range of topics including further questions on domestic abuse and sexual victimisation, fear of crime, confidence in the police and Criminal Justice System, security measures, anti-Social Behaviour and risky behaviours. While development of these areas of the CSEW were out of scope for this piece of work, they would need to be included when considering overall development of the survey in any context.

Almost all the international crime surveys reviewed covered these very sensitive crimes regardless of mode of data collection. Some surveys covered the most sensitive aspects within a self-completion module so that the participant could complete it privately: either using a paper questionnaire or as part of a CASI module.

Two exceptions were surveys in Germany and Ireland.

- **Germany**: The DVS (ID7) does not include sensitive crimes although does include more detail about fraud and online crime than the CSEW, for example payment card fraud, phishing and malware damage.
- **Ireland**: The QNHS (ID8) does not include questions about sexual assault or domestic violence, as they were considered too sensitive and personal for inclusion in a general household survey.

3.8 Summary and conclusions

In summary, this evidence review explored the overall approach and design of 15 international crime surveys, focusing specifically on those which administered the key victimisation measures online via a self-completion questionnaire. Four of the surveys reviewed fell into this category, these adopting a ‘Push to Web’ approach. These were: Finland’s FNCVS (ID5), the Netherlands’ Safety Monitor (ID9), Sweden’s SCS (ID10) and Canada’s GSS (ID12). Except for Canada, all of these use a population register as a sampling frame which makes contact and potentially participation easier.

Other than CSEW, the other crime survey to carry out extensive exploratory work in investigating the feasibility of transitioning to a multi- or mixed-mode design which includes

online self-completion is the National Crime Victimization Survey (NCVS) conducted by the US Bureau of Justice Statistics.

A key finding of this literature review is that while the surveys mentioned above include one or a combination of the more complex components of victimisation measurement, none include the full list of complex aspects that the CSEW includes in its design: measuring incidence (counts) as well as prevalence, differentiating between incidents that are similar (part of a series) and different (separate) where there are multiple incidents within crime type and the collection of data of a high enough quality required for offence coding.

As such, a key consideration should be where they might be scope to exclude some of the content of the current CSEW to achieve a simpler and more respondent-centric design to suit a mixed-mode context.

4. Stages 2 and 3: Redevelopment of the online questionnaire for the live trial

This chapter provides an account of the steps involved in developing an online questionnaire that could be tested at scale in a live field trial. This covers the tasks involved in the following stages:

Stage 2: Review of the prototype online questionnaire developed in 2018 (see section 2.4) to assess its strengths and weaknesses, consider any further updates, and ensure it is fit for purpose for the live test.

Stage 3: Review of the prototype online questionnaire to take on board findings from Stage 1 and Stage 2.

The chapter covers the following sections:

- Introduction to the victimisation screeners and victim modules (section 4.1)
- Recommendations at the conclusion of the 2017-18 development work (section 4.2)
- Scoping stage (section 4.3)
- Redevelopment stage (section 4.4) – covering changes to introduction and demographics, victimisation screeners and victim modules
- Linking the screeners and victim modules and development of a prioritisation algorithm (section 4.5)
- Usability and closing questions (section 4.6)
- Limitations and recommendations for future development (section 4.7).

The final questionnaire is included in Appendix 2.

4.1 Introduction to the victimisation screeners and victim modules

The CSEW records participants' experience of crime via a series of crime screener questions to capture crimes experienced in the last 12 months, and then follow-up victimisation modules for each crime experienced to capture the nature of crimes and to provide information required for offence coding. The key features of these core modules of the face-to-face survey and how they are used to estimate crime are summarised below:

Victimisation screeners

- A set of Yes/No screener questions capture incidents experienced in the previous 12-month period by the household (such as vehicle or property crimes) or by the individual (such as fraud, robbery, or assaults).
- For each incident type the respondent is asked how many times this has happened in the last 12 months and when it happened (month/year).
- The questionnaire captures whether multiple occurrences of the same incident are part of a 'series' (defined as 'similar incidents where the same thing was done under the same circumstances and probably by the same people'). In this situation, only the most recent crime in a series is followed up in a victimisation module.

— Questions are worded to avoid double counting of incidents (i.e. reporting the same incident more than once at different screener questions) as far as possible. At the end of the screeners, the respondent is asked to verify that all incidents are distinct and not part of the same incident. The interviewer has the option to review and amend the respondent's recorded answers at this point.

Victimisation module and offence coding

— All those identified through the screener questions as possible victims of crime are then asked detailed questions about each incident, or series of incidents, in a 'victim module' which provides the detail needed for offence coding.

— There are two versions of the victim module, one for non-fraud crimes (all crimes excluding fraud and computer misuse) and one for fraud and computer misuse crimes.

— To combat respondent fatigue, a maximum of six victim modules are completed. If more than six separate incidents have been experienced in the reference period[31], then the CAPI program selects which incidents should be followed up according to an algorithm which prioritises more serious crimes over less serious ones.

— To further minimise respondent burden, if there are more than three separate incidents to be followed up only the first three priority offences are covered in detail (the 'long form'). For additional incidents up to the maximum of six, only limited details are collected, primarily those required for offence coding (this is known as the 'short form').

— Based on information collected and processed from the victimisation modules, outside of the interview a group of specially trained coders determine whether what has been reported constitutes a crime and, if so, what offence code should be assigned to it. This offence coding uses both answers to closed questions and an open-ended description of the incident to arrive at an outcome. This process has been developed to mirror the way incidents are coded as crimes by the police and has remained broadly unchanged since the survey began in 1982.

— If one incident involves several different offences (for example, burglary, car theft and criminal damage) then the crime is recorded with a single offence code according to prioritisation rules which are similar to the police-recording system – in this example the incident would be recorded as a burglary.

4.2 Recommendations at the conclusion of the 2017-18 development work

Based on the initial development and testing work, our key recommendations are set out in the table below.

There were limitations as to which of these could be taken forward as part of this development work due to timescale requirements. Section 4.6 below details the nature of these limitations and what we recommend carrying forward to future development work. The table below summarises the status of each of these recommendations and which ones we were able to take forward as part of either the previous study or this one.

[31] Only a very small proportion of victims (1% in 2016-17) experience 6 or more crimes.

Table 4.1 –Status of recommendations from earlier development work	
Recommendation at end of previous study	***Status summary***
Screeners	
Extend the screeners to cover attempted crimes more explicitly (these are not always picked up in the face-to-face survey). At the same time reduce the length and repetitiveness of the current questions by consolidating screeners and placing actual and attempted crimes on the same screen.	Incorporated as part of the original development work
Simplify, shorten and (where necessary) update question wording to improve respondent comprehension and engagement.	Incorporated as part of the original development work, although further improvements were made as part of the follow-up development work
Re-order the screeners so that household crimes are asked before vehicle-based crimes to help reduce double counting problems associated with the original order.	Incorporated as part of the original development work
Simplify the fraud screeners to remove duplication and reduce respondent confusion.	Incorporated as part of the original development work, although acknowledged that further work is still required to address this issue.
Consider trialling a re-structure of the questionnaire so that the non-fraud screeners/victimisation module and fraud screeners/victimisation module are asked in two separate blocks. This might help to improve flow and comprehension, although there are risks associated with respondents "learning" that saying yes to a screener leads to more questions – this would need to be carefully tested.	We did not have the scope to address any major re-structures as this as part of this programme of work
Move the count, date, and series definition questions to immediately follow the screener; this was shown to help improve flow and comprehension.	Incorporated as part of the original development work
Where respondents are unable to provide an exact number of incidents, allow them to provide a banded estimate (midpoints can then be used to estimate the count); this should reduce the volume of missing data from "don't know" responses.	Incorporated as part of the original development work
Re-word the "series" definition applied to multiple crimes of the same type to ensure more accurate classification and improved respondent comprehension.	Incorporated as part of the original development work

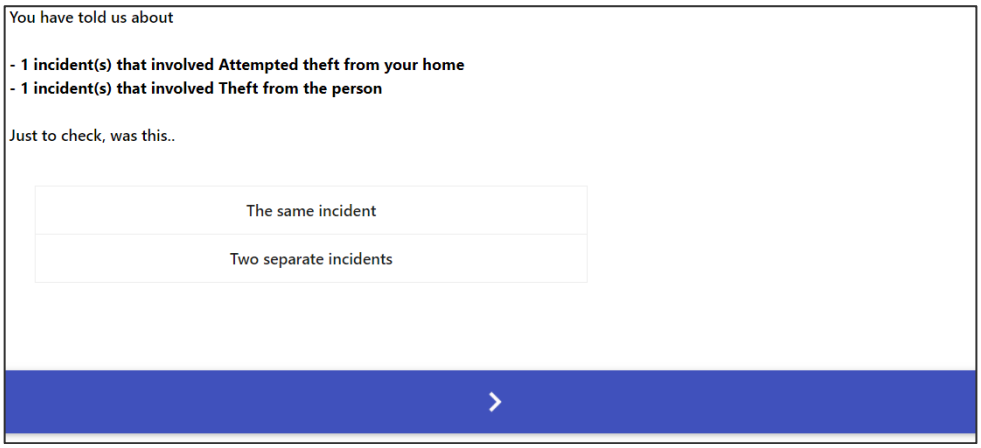
Table 4.1 –Status of recommendations from earlier development work	
Incorporate checks and verification screens to detect and correct instances of double counting. This includes providing clear upfront instructions, checking whether subsequent incidents are related to earlier ones, and giving the respondent the opportunity to review and correct incidents they have entered. This requires a complex series of scripted questions and checks which are not included in the face-to-face instrument (as the interviewer can handle this interactively).	Incorporated as part of the original development work (see section 4.2.1 below for more details on this) although this needs to be re-considered further as part of any future redevelopment work
Develop and trial a short animation/video to explain the key concepts to respondents at the start of the screener section	There was insufficient time to consider this part of this development work
Investigate whether it is possible to make the double counting checks, i.e. the checks that detect whether two or more incidents are related, more targeted (for example, only checking overlap between incidents which occurred in the same month).	There was insufficient time to consider this part of this development work
Victim modules	
Vary the order of the questions in the non-fraud victimisation module to be dependent on the screener the module is linked to. For example, if the module is triggered by an assault screener then the respondent should be asked questions about the assault first, before being asked if the incident also involved other features such as theft and criminal damage.	Incorporated as part of the original development work
Move questions about location of the incident to the end of the module so that the respondent is asked about the nature of the incident first, which will be more relevant to them.	Incorporated as part of the original development work
Reduce length and repetition by trimming back non-essential questions, combining and consolidating questions, removing duplication, simplifying wording and reducing the length of response lists.	Incorporated as part of the original development work, although further improvements were made as part of the follow-up development work

4.2.1 Addressing the issue of double counting

The ZRELATE approach developed in the 2018 development work was based on a direct approach to avoid double counting and involved the following stages[32]:

- A respondent answers 'yes' to an initial screener (Screener 1) and records the number of incidents in the past 12 months.
- As soon as a second incident is recorded (Screener 2), a check screen appears (**ZRELATE**) which asks the respondent if the second incident is related to the first one. **ZRELATE** is then repeated for any subsequent screeners recorded as 'yes'. The wording of **ZRELATE** was adapted depending on the number of incidents counted at the two crime types being compared. There are two versions of ZRELATE depending on the number of incidents counted at the two crime types being compared: a "simple" version when one incident is compared with another single incident, and a "complex" version when multiple incidents are compared.
- In the 'simple' version, participants can say that the second incident is related to the first. Where this occurs, the script automatically deselects the second incident, and it is not included in the final list of incidents.

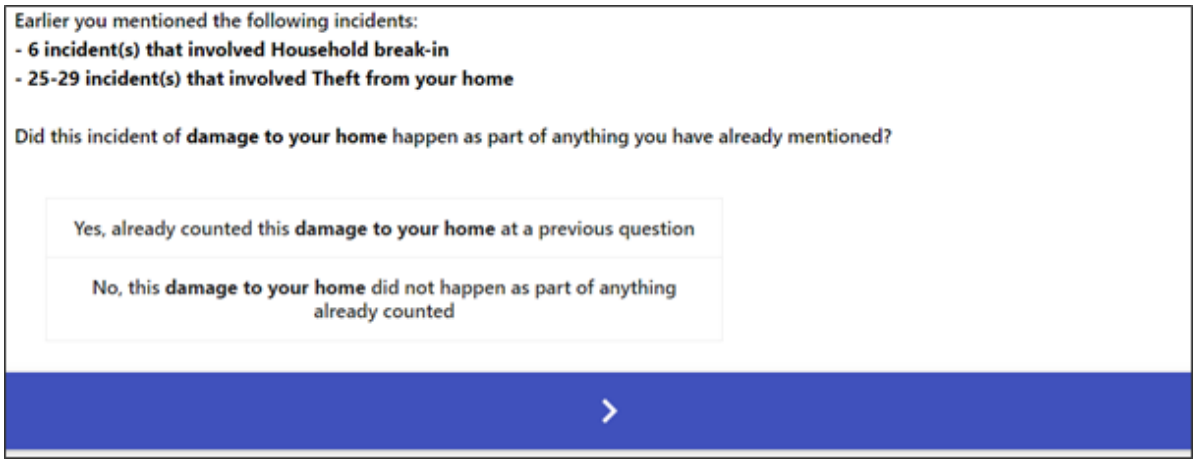
ZRELATE – simple version



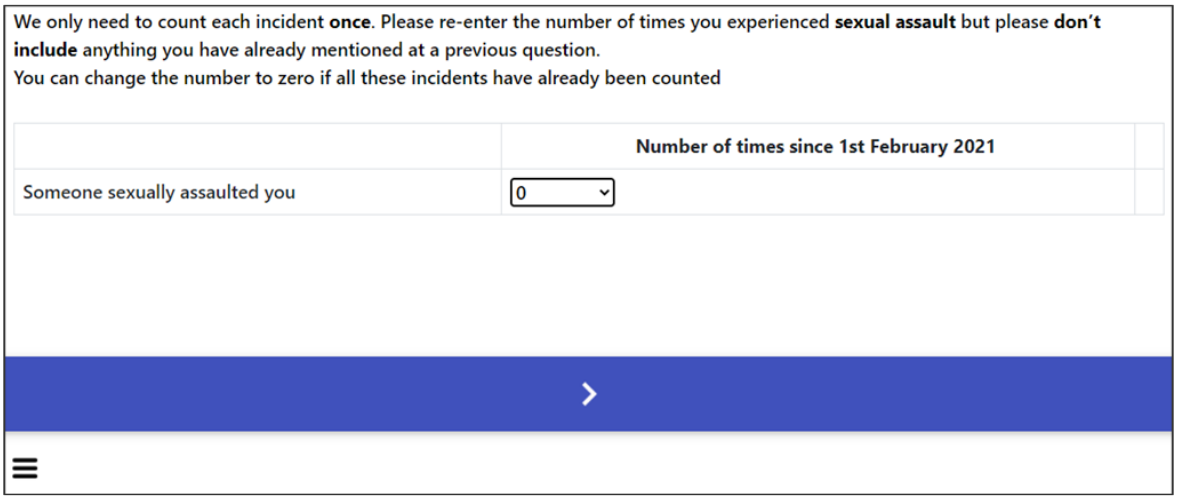
- In the "complex" version when the second/subsequent crime is counted as a multiple crime, and the respondent says it is related to an earlier crime, then we cannot assume/impute the number of crimes which can be discounted. In this situation the respondent is routed to a further screen (**ZCOUNTCHECK**) which asks them to re-enter the correct number of incidents.

[32] A more detailed explanation of this can be found in section 5.8.4 of the previous development report https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/redesignofcrimesurveyforenglandandwalescsewcorequestionsforonlinecollection/2018-07-19

ZRELATE – complex version



ZCOUNTCHECK



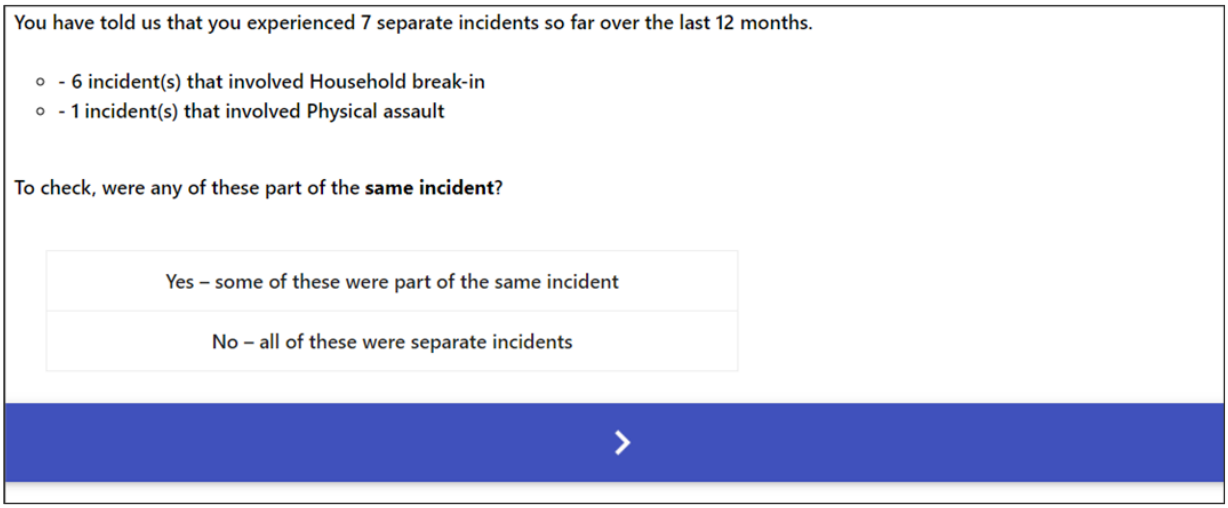
The main conclusion from the earlier work was that these checks were too cognitively challenging and confusing for respondents, and that a more radical approach to addressing double counting would be required going forward. However, given timescales, this was beyond the scope of this programme of work and therefore we continued to test the existing version of the WCSEW even though we knew this would need further development in the longer-term. Further details of this can be found in the previous development report (section 5.8.4 of that report).

In updating the previous questionnaire, a review of incidents (**ZREVIEW** sequence) was included at the end of each block of screeners (i.e. the non-fraud and the fraud screener blocks). The purpose of the review was to mimic the interviewer version of this screen. However, in the online version we wanted to use this to use this review to detect instances of double counting more explicitly.

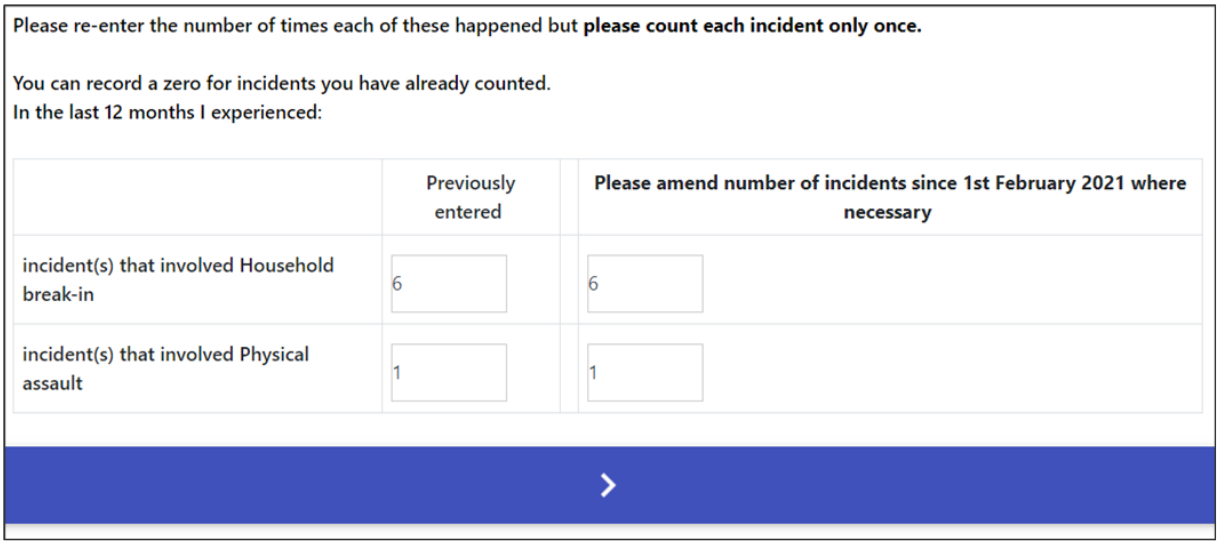
Therefore, at **ZREVIEW1**, in all cases where two or more incidents of different crimes were recorded a list was shown to the respondent. The respondent was then provided with a further opportunity to state whether any incidents were related. If any incidents were related, the respondent was asked to provide a recount at **ZREVIEW 2**. Finally, **ZREVIEW3** asked the respondent to confirm the correct number of crimes. This sequence of screens would only appear for the minority of cases when two or more separate crimes

were recorded. If a second/subsequent incident had been discounted because of **ZRELATE/ZCOUNTCHECK** and this resulted in only one crime type being recorded, then the series of **ZREVIEW** screens was not triggered. These screens are illustrated by way of example below.

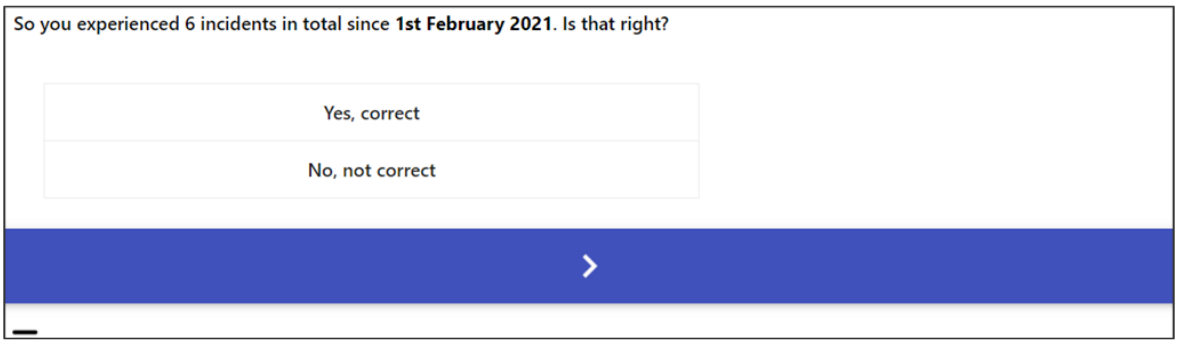
ZREVIEW1



ZREVIEW2



ZREVIEW3



4.3 Scoping stage

As noted, a draft questionnaire for estimating crime via an online survey, including screeners and victim modules, had already been developed and scripted as part of the earlier development work conducted by Kantar (see section 2.4)[33]. However, the earlier questionnaire had been set up to be tested qualitatively and there was not an existing version which could work in a live field setting without an interviewer present. For example, when testing the questionnaire in the previous development work, the researcher manually selected one incident to be covered in a single victim module, but in the field test it was necessary for the transfer between screeners and victim modules to be automated, using a prioritisation algorithm to select victim modules when there were more than six incidents reported (also the limit within CSEW).

As well as producing an online questionnaire which worked in a live field setting, the opportunity was also taken to make some further improvements to the prototype WCSEW questionnaire based on a wider review including residual recommendations made at the conclusion of the previous development stage, and further recommendations following the Stage 4 cogability testing.

As already noted, the very restrictive timescale meant that Kantar was not able to implement all changes that would ideally have been taken forward. For example, it was not possible to implement changes that would have required a radical change to the questionnaire structure, or more complex changes such as those affecting loops or double counting checks (this is discussed further below).

To decide which amendments were feasible to implement within the timescale available, the following scoping tasks were carried out before the development of a specification for the redeveloped questionnaire:

- Firstly, Kantar reviewed recommendations for further changes provided as part of the previous development work to decide which of these could realistically be implemented. This included improvements to wording which could, for the most part be implemented, and some more structural improvements (though as noted, it was not possible to implement all these types of changes due to complexity and timescales).
- Kantar also cross-referenced the existing prototype WCSEW against the questionnaire that was used for the telephone-operated crime survey (TCSEW)

33 https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/redesignofcrimesurveyforenglandandwalescsewcorequestionsforonlinecollection/2018-07-19

during the coronavirus pandemic. The TCSEW was a shortened version of the CAPI questionnaire[34] and is therefore more aligned with CSEW than WCSEW. However, as the TSCEW includes all questions which are essential for offence coding, this task provided a useful check to ensure that all relevant questions for offence coding were included, and to remove any questions which were not essential for this purpose. It was important to ensure the victim modules were as concise as possible, as this time the intention was to include up to six victim modules (in line with the CSEW and TCSEW), to assess the length and feasibility of the survey for victims of multiple crimes.

- Kantar reviewed the screener questions to ensure they included screeners (and associated victim modules) required for calculation of estimates of all CSEW crimes. For example, in the previous version of the WCSEW, some crimes (such as sexual offences and threats) were excluded as these were subject to further development work at the time, but these were now included in WCSEW and new screeners were therefore developed to cover these types of crimes. In fact, there was a more radical redevelopment of the violence, harassment and threats screeners for this version of the WCSEW (see section 4.4.3 below).
- Given that the original online questionnaire only included a manual link between the screeners and victim modules, Kantar developed a far more complex link which carried through all crimes from the screeners to victim modules. However, given the limit of six victim modules, it was also necessary to develop a prioritisation algorithm which identified which six incidents would be selected if more than six were captured in the screeners (see section 4.5).
- Some further changes were then made at a later point as a result of the Stage 4 stage of cogability testing; again these changes were limited to those which were pragmatically possible within the allocated time. The key changes are summarised in the following sections.

4.4 Redevelopment stage

The questionnaire specification was built iteratively over the steps detailed in section 4.3, with Kantar working in close consultation with ONS. The following sections detail the redevelopment of different modules of the WCSEW questionnaire.

All questions were adapted to be suitable for mixed-mode administration (online and telephone).

4.4.1 Introduction and demographics

In summary, the following changes were implemented:

- In the opening introduction screens, more information was added about the topics covered for sensitivity and safeguarding purposes. This included a more explicit ‘warning’ screen that the questionnaire included questions on physical and sexual assault, threats and harassment, and a suggestion that respondents affected by these issues might want to find a private place to complete the survey, or to skip these questions. The respondent was also given the opportunity to skip the victim module for these potentially more sensitive crimes (see section 4.4.5 below).

[34] Broadly based on the short version of the CSEW victim modules.

— In general, introduction screens were condensed as far as possible, given evidence in cogability testing that respondents routinely skimmed or ignored these 'read only' screens.

— Some further minor amendments were made to update the demographics questions and bring them in line with harmonised versions where applicable.

4.4.2 Victimisation screeners

In summary, the following changes were implemented (note that the redevelopment of the violence, threats and harassment screeners is covered separately in section 4.4.3):

— Edits were made to the introduction to the screeners based on recommendations from the previous online development report (see section 2.4) and further cogability testing. For example, one recommendation addressed the issue that respondents who have only experienced, for example, a fraud may feel that the survey, which initially covers household and vehicle crimes, is not relevant to them, risking early drop out. Therefore, the introduction was reworded to more clearly flag the range of crimes asked about, alongside stressing the importance of the survey for non-victims as well as victims. The added wording was *'The questions will start by asking about household crimes such as incidents involving the home and vehicles, then personal crimes, then fraud crimes'*. The sentence emphasising the inclusion of non-victims was moved to an earlier screen: it is just as important to hear from people who haven't experienced any crimes as well as those who have'.

— Minor wording changes to 'yes/no' screener questions to simplify and clarify these where cognitive and usability testing at different stages suggested these could be improved. This included changes to screeners based on people's homes as the definition and boundaries of 'home' was not always clear (covered in more detail in chapter 5). For example, respondents queried whether 'your home' included driveways, annexes, garages etc. and 'outside your home' was in one case interpreted very literally as a café in the town centre. In response to this, the definitions at these screeners were clarified and more specific examples were added. The wording referred to 'elsewhere on your property' rather than 'outside your home' and 'walkway and balcony' were added to the list of examples.

4.4.3 Violence, threats and harassment screeners

The approach for asking about physical and attempted assaults, and sexual assaults, changed more substantially compared with other screeners. In addition, new screeners were added to capture threats and harassment, as these were not covered in the initial online development work (see section 2.4)[35].

Based on a review of wider sources (the CSEW and TCSEW), it became clear that the violence, threats and harassment screeners required further development work, not just online but in all settings (face-to-face and telephone too). Therefore, we used the opportunity of the online CSEW development to trial a different approach. The issues associated with the previous approach for screening for these crimes (across all modes) was as follows:

[35] At the time of the original online development work the threats and harassment screeners were still under review. However, separate development work in 2020-2021 led to new screeners on threats, harassment and intimidation in both CSEW and TCSEW and these were used as the basis to develop online versions of these screeners.

- Work from other strands of CSEW development indicated that there is confusion around what counts as being a victim of 'violence'. In CSEW, the threats question has always created a lot of offence codes for common assault not picked up under the assault screener, as respondents tend to think of these types of incidents as a 'threat' rather than an 'assault', even though the police would legally define these incidents as assaults. In the TCSEW the long-standing threats question asked in the CSEW was expanded to include harassment and intimidation. This was done with the intention of measuring low levels of harassment during the pandemic based on initial speculation that the level of harassment within society might increase during this period. One consequence of this change was that the question captured more offences across a range of offence types, but particularly common assault. The underlying reason for this was that incidents of assault that were typically not included under the original CSEW violence or threats screeners were being picked up instead as 'threats and harassment'[36]. The hypothesis for this is that the original wording of the violence screener (which referred to punching, kicking and use of a weapon) and threats screener was too restrictive. Examples of increased rates of common assault incidents that are now being counted under threats or harassment in TCSEW include spitting, pushing, shoving and road rage, which are all common assault incidents but not necessarily regarded as 'violence' or 'threats'. Therefore, the research team wanted to improve the screeners that cover violence, threats and harassment, so that crime types could be more explicitly defined and differentiated, and to hopefully reduce the potential for double counting in the online version, where there is no interviewer to manage this.
- A second issue was the need to make it clear that incidents of domestic or household violence should be included under the violence screener. This was to avoid having a separate screener as used in the interviewer-administered survey as this would be likely to exacerbate double counting in an online version, without an interviewer to manage overlap.
- A third issue is that in 'cogability' testing there was confusion over whether to include incidents affecting the respondent themselves (which is what the survey aims to collect) vs. incidents affecting other people in the household. The previous version of the online survey used the text '*Has anyone, including people you know or live with, done the following to you'* was found to lead people to think that incidents affecting other household members should also be included.

To tackle these issues, the violence screener was changed from a single 'yes/no' question asking if 'someone had '*deliberately hit, punched or kicked you, or used a weapon of any sort on you'* to a muti-coded question which covered a wider range of types of 'force or violence' or assault. The response list used in this screener also included sexual assaults and sexual attacks, as it was expected that many of these would be captured within 'physical force or violence' and therefore including it within the same list as physical force or violence would avoid the inevitable double counting which would occur if these were also covered in a separate screener. The wording of the question was also amended to attempt to clarify the text relating to 'people you live with'.

[36] For more information on this see https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/comparabilitybetweenthetelephoneoperatedcrimesurveyforenglandandwalesandthefacetofacecrimesurveyforenglandandwales#questionnaire-changes-including-the-effect-of-adding-a-question-on-harassment

Table 4.2 below shows the revised question text for these screeners, alongside the version tested as part of the original online development work.

Although the violence and assault screeners had a different respondent-facing presentation compared to other screeners, for the purposes of the standard loops in the script (incident counts, double counting checks, and allocation to victim module) this was converted back into two 'yes/no' screeners so they could be handled in the script in the same way as all other screeners:

- Physical force or violence (any response at options 1,2,4 selected)
- Sexual assault or attack (option 3 selected)

A similar multi-coded question was included for attempted violence (although this did not cover attempted sexual assault[37]). Two new screeners were also added to cover threats and harassment/intimidation; these were both formatted into more standard ('yes/no') screener questions.

The screeners led to the following five crime types being carried through to victim modules (with their short-hand 'crime description tags'):

- Violence or assault ('physical force or assault')
- Attempted violence or assault ('Attempted physical force or assault')
- Sexual assault or sexual attack ('sexual assault')
- Being threatened ('being threatened')
- Being harassed or intimidated ('being harassed or intimidated')

Table 4.2: Revised online screeners for violence, threats and harassment shown alongside the versions tested in the original WCSEW.

Original WCSEW screeners	*Revised WCSEW screeners*
DISPLAY6 [ASK ALL] The next few questions are more personal in nature. You may wish to find a private place to complete this part of the survey. Please remember that the answers you give are completely confidential. Please click the (>) button below to continue **ZASSAULT [ASK ALL]** This next question is about assaults and attempted assaults. **Since xx,** have any of the following happened to **you personally**. *Please also include assaults or attempted assaults by people you*	**DISPLAY (IntroVio) [ASK ALL]** The next questions are more sensitive and ask you whether you have been deliberately attacked, hurt, sexually assaulted, threatened or harassed in the last 12 months. Please think about attacks, threats or harassment caused by someone you know well, someone you live with, someone you came into contact with through work, or a stranger. You may wish to find a private place to complete this part of the survey. If you would prefer to skip any questions, please select 'Prefer not to say'. **ZASSUALT [ASK ALL]** Since **1st [DATE]**, has anyone deliberately done any of the following to you? *Please include anything that **happened to you**, including if this was done by people you know or live with, as well as by strangers.*

[37] As we felt that attempted sexual assault would be difficult to define and classify in practice

came into contact with through your work, and people that you know. Someone deliberately hit, punched or kicked you, or used a weapon of any sort on you? 1. Yes 2. No Someone **tried to** use physical force or use a weapon of any sort on you? 1. Yes 2. No **SEXATTAK [ASK ALL]** In the last 12 months, **since xx,** have you been sexually assaulted or sexually attacked, either by someone you know or by a stranger? 1. Yes 2. No 3. Don't know 4. Prefer not to say	*Please select all that apply* 1. Used a weapon of any sort on you 2. Used physical force or violence, for example kicked, hit, slapped, punched, scratched or headbutted you 3. Sexually assaulted or sexually attacked you 4. Physically assaulted you in some other way, for example bit you, spat at you, grabbed or pushed you or threw something at you 5. None of these 6. Prefer not to say **ZTRYASST [ASK ALL]** Since **1st [DATE]**, has anyone **tried to do** any of the following to you, but didn't succeed? *Please include anything that **happened to you**, including if this was done by people you know or live with, as well as by strangers.* *Please select all that apply* 1. Tried to used a weapon of any sort on you 2. Tried to use physical force or violence on you 3. Tried to physically assault you in some other way 4. None of these 5. Prefer not to say **ZTHREVIOL [ASK ALL]** Since **1st [DATE]**, has anyone at all, **threatened you** in a way that was intended to **cause you alarm** or distress? *Please include threats made by people you know or live with, as well as by strangers.* *Please include threats made by any means, for example in person, on-line, over the phone or on social media.* 1. Yes 2. No 3. Prefer not to say **ZHARASS [ASK ALL]** Since **1st [DATE]**, has anyone **harassed or intimidated you** in a way that was intended to cause you alarm or distress? *Please include harassment and intimidation by people you know or live with, as well as by strangers* *Please include harassment or intimidation by any means, for example in person, online, over the phone, or on social media.* 1. Yes 2. No 3. Prefer not to say

4.4.4 Crime description 'tags'

As participants may report multiple incidents, short hand 'crime description tags' were used as text substitutions to indicate which incident type was being referred to in follow-up questions on counts and dates, and later in the victim module. Edits were made to these tags to simplify and shorten them where possible and edits were also made in some cases where cogability testing indicated some comprehension problems. Some of the tags were felt to be worded too narrowly, which could cause problems later in the victim module as

respondents felt that the description of the crime in the tag did not accurately reflect what happened to them. This is understandably challenging as each victim's experience is unique and their interpretation of what happened to them may not match official terminology. For example, changes were made to the following tags:

- 'household break-in' was changed to 'entering your home without permission' as there was evidence that some incidents that weren't technically break-ins were included at this screener as it was one of the first screeners the respondent saw, and if this was the case the text substitution of 'household break in' used throughout the victim module felt 'wrong'.
- In cogability testing, the tag 'physical assault' was found to be too narrow a definition; for example, some people do not regard pushing/shoving as 'assault' (see chapter 5 for more detail). The tag was therefore changed to 'physical force or assault' which it was hoped would cover a wider range of incidents. This also links with the issues described above, whereby evidence suggested that physical assaults and threats needed to be better designed and captured in the screeners.

4.4.5 Victim modules

As in CSEW and TCSEW, there are two versions of the victim modules: the original victim module and an amended version for incidents of fraud.

In summary the following changes were implemented:

- Each victim module is linked to a screener. However, the screener could relate to any part of that incident and is not necessarily the 'main' element of that crime, as the victim module is usually linked to the screener which is mentioned first. For example, if a crime involved theft of tools from the shed and an attempted bicycle theft then the victim module would refer to the attempted bicycle theft rather than the actual theft, as this screener is asked first. Given this, and still using the same example, the reference was changed to incidents from '…*the [*Theft from outside your home*] incident'* to *'…the incident that involved [*Theft from outside your home*]*' in an attempt to better describe composite incidents. Obviously, this is only a minor edit which represents an improvement on the previous development work, but there remain more fundamental challenges associated with composite incidents and how these are recorded in the screeners and linked to the victim modules (see section 4.6.1 below).
- The victim module was adapted to incorporate the new physical violence, threats, and harassment incidents captured via the screeners, though it was decided not to follow up sexual offences with a victim module for sensitivity reasons[38]. Some further questions were added regarding the nature of threats and harassment to allow these crimes to be offence coded.
- If the victim module is associated with screeners relating to physical or attempted assault, threats or harassment, the respondent was provided with an opportunity to skip these questions (for sensitivity reasons and to help protect an individual who might be a victim of domestic violence within the household). More information on this is covered in section 7.2.2 and section 8.1.

[38] This would need to be considered more extensively as part of future development work

— Other changes were mainly wording amendments agreed because of changes recommended at the conclusion of the previous testing work or following the Stage 4 cogability testing, or to ensure the inclusion of all questions required for full offence coding.

4.5 Linking the screeners and victim modules and development of a prioritisation algorithm

In the main CSEW and TCSEW surveys, all incidents up to a maximum of six are followed up in more detail via a victim module. Where there are more than six incidents to be followed up, an algorithm automatically works out which six incidents should be followed up based on a priority order (broadly in order of severity, and with non-fraud crimes prioritised over fraud crimes), and then chronologically within each crime type[39]. Where multiple incidents of the same crime type are labelled as 'similar' then only the most recent is followed up in a victim module. This algorithm has been kept consistent since the start of the CSEW.

It was therefore necessary to develop an algorithm for the online CSEW to ensure a smooth, automated transfer between screeners and victim modules, to cap the number of victim modules at six, to prioritise which six to follow up and in which order.

Although the algorithm adopted followed the same broad principles of the CSEW, it was not possible to use the full CSEW algorithm due to the different way the online CSEW screeners are set up, and because a simpler, more pragmatic solution was needed within the timescale available for the live trial. The final approach that was agreed is summarised below.

Step 1: There were 29 screeners/crime types. A prioritisation order (shown below) was agreed in consultation with ONS, from physical assault (highest) to computer virus (lowest) such that where a selection needed to be made, the highest priority crimes were selected first.

[39] In CSEW the first three incidents are covered by a 'long' victim module and subsequent incidents are covered by a 'short' victim module. However, this distinction was not relevant in WCSEW as all victim modules were 'short', focussing only on questions required for offence coding.

Table 4.3: Prioritisation algorithm ranking			
		Crime screeners	***Ranking (1=highest, 29=lowest)***
Household screeners	1	Household break-in	10
	2	Attempted household break-in	11
	3	Theft from your home	12
	4	Attempted theft from your home	13
	5	Damage to your home	23
Vehicle and bicycle screeners	6	Vehicle theft	16
	7	Attempted vehicle theft	17
	8	Theft from a vehicle	18
	9	Attempted theft from a vehicle	19
	10	Vehicle damage	20
	11	Bicycle theft	21
	12	Attempted bicycle theft	22
Outside home screeners	13	Theft from outside your home	14
	14	Attempted theft from outside home	15
Personal crime screeners	15	Theft from the person	5
	16	Attempted theft from the person	6
	17	Theft away from home	7
	18	Attempted theft away from home	8
	19	Damage to personal property	9
	20	Sexual assault	0
	21	Physical assault	1
	22	Attempted physical assault	2
	23	Being threatened	3
	24	Being harassed or intimidated	4
	25	Use of your personal information or account details without permission	24
Fraud or computer misuse	26	Being deceived out of money or goods	25
	27	Attempted deception out of money or goods	26
	28	Theft of personal information online	27
	29	Virus or other interference to computer/device	28

Step 2: Within each incident type, up to a maximum of two incidents were selected as follows:

- If one incident only, then select this incident.
- If 2+ separate incidents, then select the most recent and the next most recent.
- If 2+ incidents of the same type which have been labelled as similar, select only the most recent.
- If a mixture of series and separate incidents, then treat in the same way as similar incidents, and select only the most recent.
- If the crime count is unknown, then the screener is excluded from selection and not allocated a victim module.

Step 3: Allocate the final selection of up to six incidents based on the priority order, with all incidents from higher priority crime types selected before incidents from lower priority crime types. So, for example, if the following separate incidents were recorded: 3 x Physical assault (priority 1), 2 x harassment (priority 5), 4 x household break-in (priority 11), and 2 x bicycle theft (priority 22) then the algorithm would select the 2 most recent physical assaults first, then the 2 harassment incidents, and then the 2 most recent break-ins, and the bicycle thefts would not be followed up in a victim module.

Appendix 3 includes the full priority ranking shown above along with some examples of how this worked in practice.

4.6 Usability and closing questions

At the end of the Live Trial survey (see chapter 8) some additional questions were added to serve more administrative aims:

- To help gain some insight on the user experience of the online survey, and how this varied by complexity of crime experience, for example perceived ease or difficulty of completing the survey and problems encountered (an analysis of these questions is covered in more detail in Chapter 8).
- To provide signposting to support services (such as Victim Support, Crimestoppers, Domestic Violence helplines) for those who indicated they would like this.
- To provide some additional information to help inform another parallel project Kantar conducted on behalf of the Transformation Programme (Transformation Package B – transformation of the child crime survey). Two questions were included for parents of children aged 9-17 on their level of willingness to provide consent for their child to take a part in an online survey about crime and negative online experiences, and the nature of any concerns[40].

The findings of the usability questions are covered in Chapter 8.

4.7 Limitations and recommendations for future development

The timescale for developing the online questionnaire for the live trial was very restricted, and as a result it was not possible to conduct as thorough a redevelopment as would have been ideal. At the end of the previous online CSEW development project, it was concluded

[40] These findings are covered in the report on Work package B which can be found at: https://kantarpublic.com/articles/transforming-the-crime-survey-for-england-and-wales

that some parts of the online questionnaire were not working well (in particular, the part of the script which attempts to manage double counting, see below).

However, to meet the constraints of the timetable, a decision was made to omit some of the more complex changes that would ideally have been made, but which were not possible within the parameters of the current project.

In parallel to this work, Kantar Public has also been working with ONS to consider possible alternative models for an online survey which will build on the current online but could also adopt a more radical restructure. However, this is outside the scope of the current development project.

Focusing here on the online CSEW version that was tested, changes to be considered as part of future development are summarised below. Any updated version of the online questionnaire which includes any of these more complex features would require a considerably longer timescale which builds in more time for scoping, development and iterative user testing before piloting at scale in the field.

4.7.1 Managing double counting of incidents with composite features

The online questionnaire in general works smoothly for people with simple crime profiles, for example those who have experienced up to four incidents of crime which are not linked, and which, most importantly, have not occurred on multiple occasions. The key challenge in the development of an online crime survey is to identify and resolve double counting of incidents when an incident involves composite features. It is known that respondents (understandably) want to report all features of the incident, but without an interviewer to manage and discount double counting, there is a risk that these crimes are counted more than once, and that respondents are taken through multiple victim modules which in fact all refer to the same incident. Although attempts were made in the earlier project to develop a solution for this, it was concluded that this was still not working successfully for respondents with complex crime profiles.

However, this issue was acknowledged to be exceptionally challenging, and it was decided to delay the redevelopment of the script to better manage double counting given that this would require a more radical redevelopment of the questionnaire, with a longer development timetable. Therefore, a pragmatic decision was made to retain the original approach for detecting double counting in the live trial (the ZRELATE AND ZCOUNTCHECK sequence described in section 4.2.1 above) with only minor modifications.

4.7.2 Creation of more amalgamated screeners

Section 4.4.3 documented how the physical violence and assault screeners were changed, as it was felt that the original ‘yes/no’ screeners defined assaults too narrowly. In the future, it may be beneficial to create more of these ‘composite’ screeners to ensure that incidents that tend to be linked are all considered in the same screener. For example, it was found that damage to the home was often linked to damage to vehicle and other personal items. One possibility would be to wrap these into one single question about damage or vandalism (covering homes, vehicles and personal property) with a list that can be multi-coded. The count would then relate to all incidents of ‘damage’ and the victim module could be used to establish the more specific attributes of the incident. The descriptor ‘tag’ could also be tailored further depending on which item(s) in the list were selected.

Similarly, amalgamated versions of screeners could also be considered for thefts both inside the home and outside the home (for example, all incidents relating to a respondent's property) and to group together fraud-related crimes (since testing found that fraud incidents are more difficult to pin down to a specific screener, given that respondents often don't know how the fraud event occurred).

However, this needs to be weighed up against wider evidence that shows that individual forced choice questions lead to higher rates of selection than the same items appearing in a muti-coded list[41].

The U.S version of the crime survey (NCVS) adopts a similar approach involving amalgamated screeners and could be used as a reference approach (see section 3.6)[42].

4.7.3 Definition of series crimes

Where a respondent has experienced more than one incident associated with the same screener, they are asked if these incidents were 'similar' in nature. And if all crimes were similar, they are treated as a 'series' of incidents which means that the respondent is only asked about the most recent incident, and the same offence code is assumed for all earlier crimes in the series.

However, this becomes complex when a respondent has experienced multiple crimes of the same type, with some in a series and others different.

For the live trial, a pragmatic decision was made to treat all cases where there was a mix of both series and separate incidents as series crimes. However, this was a temporary 'workaround' solution and in any main stage, a better solution needs to be developed which is less confusing for respondents.

One possible solution to simplify this would to be to avoid asking about 'similar' and 'different' crimes and to instead apply a cut-off, for example to always treat 1-2 crimes as different, and to always treat 3+ incidents as a series. The NCVS asks respondents whether they know enough details to distinguish between different incidents, which is an alternative way of designating incidents into a 'series' (see section 3.6).

4.7.4 De-duplication of victim modules

In the main CSEW, the interviewer can skip a victim module if it transpires on starting it that the respondent is referring to an incident for which they have already completed a victim module. The likelihood of this situation occurring in an online context is much greater given the increased potential for incidents to be double counted.

One proposal developed, but not implemented due to time constraints, was to add a check question at the start of the victim module: *Just to check, did this happen as part of [the incident/any of the incidents] you have already provided information about?* and if this applied, collect details of which previous incident this related to, and then skip to the next victim module or next part of the questionnaire. However, there were potential complications with this approach and this would require a longer development timescale to allow for more user testing.

[41] Pew Research Center, May 2019, "When Online Survey Respondents Only Select Some That Apply"

[42] See Appendix 5 of https://www.ojp.gov/pdffiles1/bjs/grants/303980.pdf for the full redesigned NCVS questionnaire

4.7.5 Creating a separate ‘block’ within the victim module to cover threats and harassment

The online CSEW included new screener questions on threats and harassment and some new questions were covered within the victim module to capture details of these incidents for offence coding. The online survey is currently divided into four ‘blocks’ grouping questions together that relate to theft, attempted theft, criminal damage, and violence/assault and these rotate such that the victim module starts with the most relevant block of questions, based on the screener which triggered the victim module. For pragmatic reasons, the new questions on nature of threats and harassment were included within the ‘violence/ assault’ block but in the future, it would be better to create a separate ‘Threats and harassment’ block so that the question sequence feels more relevant for people who have experienced these types of incidents. In addition, it would be helpful to expand on victims’ experiences of these incidents to allow them to provide details that will feel relevant to them. For example, questions could be added on motivation of the offender (hate crime, race/homophobia etc.) and whether the incident took place in person or online. Any new questions would draw on the harassment module developed as part of the CSEW.

4.7.6 Creation of an animation/video

One topic Kantar considered, but again did not have time to explore in this work package, is to develop and trial a short animation/video to explain the key concepts to respondents at the start of the screener section, supplemented by ‘avatars’ (or similar) which appear when a further survey definition, concept or reminder needs to be conveyed. For example, the animation could include an avatar to ‘explain’ key concepts at the beginning and then again at appropriate points: for example, a reminder about double counting when a second incident is raised (for example, “*Don’t forget, you only need to tell us about each incident once*”) and to flag the changes in counting rules when introducing the fraud screeners. The animation could also be used to provide reassurance for non-victims that their answers are still important.

4.7.7 Redevelopment of the fraud victim module

At the time of writing, wider work is being conducted to redevelop the fraud victim module within CSEW. Therefore, any redevelopment of the online CSEW would need to keep abreast of wider developments in this area to ensure that the companion questionnaires are aligned as much as possible.

5. Stage 4: ‘Cogability’ testing

5.1 Introduction and methods

‘Cogability’ testing refers to a questionnaire testing approach which combines cognitive testing and usability testing.

Interviews with victims of crime took place via a remote video platform in January and February 2022, with respondents sharing their screen as they completed the survey. Given that previous testing indicated that the online survey worked well for people with straightforward experience of crime, but was much more challenging for people with complex crime profiles, recruitment focused on the latter group.

Recruitment was highly successful in finding participants who had experienced (in some cases, very) complex crime both within and across crime type. Complex crime was defined as having experienced at least three incidents of crime, both across and within crime type (although commonly participants had experienced many more). Additionally, although Kantar did not quota on this basis, a broad range of types of crime was included. Participants were recruited from a spread of regions across England and Wales: London, Cardiff, Manchester, Birmingham, Sheffield and Leeds. Of the 15 interviews conducted, 12 were completed on laptop and 3 were carried out using tablet or smartphone and there were no additional usability issues noted with these devices. Participants received a £40 payment as a thank you for their time.

The recruitment of people who had complex experience of crime was expected to be quite challenging. As such, this was the focus of the quotas, although other demographics were monitored with an aim to ensure a reasonable spread across different types of victims.

The profile of the sample by key characteristics is shown in Table 5.1.

Table 5.1 – Profile of cogability testing respondents	
Crime type	
Household crimes involving the home	9
Vehicle crimes (including bicycle)	13
Theft from person	3
Assault	2
Threats and harassment	4
Fraud	8
Complex aspects of crime	
Crime range	All had experienced multiple crimes, some up to 5-6
Maximum screener loop	11
Mix of series and separate	At least 2
Gender	
Male	7
Female	8
Age range	
Range of ages	Between 18 and 50
Education level/type	
Degree	7
A level	1
GCSE	3
Vocational	4

5.2 'Mental models' exercise and how participants described their experiences

Before asking the participant to complete the online CSEW, they were asked to describe their experiences of crime over the last 12 months in their own words without interjection from the interviewer. The aim of this was to give the interviewer an insight into their experiences and to find out how victims of crime describe these incidents in their own words, outside of the more rigid structure of the CSEW questionnaire.

The main strategy participants adopted here was to take the most serious incident (or series of incidents if that were the case), and then talk about this in detail before moving onto the next incident. A single incident may have involved one or more screeners (for example theft of vehicle, theft from vehicle and vehicle damage). Some participants also included incidents that had happened to friends and family or in the general

neighbourhood rather than to themselves to their household directly. In some cases, participants mentioned salient incidents that happened longer ago than the 12-month time frame but tended to know these shouldn't be included.

When asked about crimes experienced in this open-ended format at the start, not all incidents picked up later in the questionnaire were mentioned at this stage. The next stage of working through the online CSEW screeners showed that the questionnaire tended to prompt respondents to recall other incidents that they didn't mention at the mental model stage. These incidents tended to be more minor, such as vehicle damage, unreported incidents, or incidents where the participant didn't know the details or impact (for example, certain types of fraud) or if they felt it didn't 'count' as a crime. This reinforces the soundness of the approach of asking about each incident type in a separate screener, to help jog memories and emphasise that all these things 'count'.

The mental models exercise was also useful because it allowed the interviewer to get an overall picture of the participant's experience and meant they could check back or verify incidents that came up later during the questionnaire. It also helped build rapport and set the overall tone of the interview in that we'd like to hear from the participant and get them talking, this initial process mirroring the protocol survey interviewers follow when administering the CSEW.

5.3 Key issues

This section provides a brief summary of the key issues that emerged during 'cogability' testing.

5.3.1 Overall structure

Rigid structure

As was already known, the rigid structure of the crime survey questionnaire means that participants don't have the freedom to report their experience in the way they would naturally choose (for example as they were able to in the mental model exercise). Participants naturally want to provide detail about the incident(s) at the point where they are first mentioned, when the screener is coded as 'Yes'. This was particularly so for one participant who coded an attempted assault. This participant was then irritated that the subsequent questions related to other incidents rather than addressing the detail of the more salient incident there and then.

This also raised some issues with the instruction at the introduction screen to the assault, threats and harassment screeners, where respondents are prompted to find a private space in which to answer these questions if they wished. However, sensitive questions about any recorded incidents come up in various places in the survey rather than all being asked within this one section. This participant said they might not answer honestly if it was brought up again at a later point. In the live trial, the initial introduction screens to the survey included clearer reference to the fact sensitive questions would be asked as part of the survey (see section 4.4.3).

'Learning the loop' and satisficing

Taking this further, there was evidence that participants could learn what to expect after saying 'yes' at a screener question, where follow-up questions always follow the same 'looped' pattern of questions about the number of times a type of incident happened and dates. A small number of participants clearly spotted and understood this pattern (for example one said '*if I say yes, you're going to ask me if it's related*'). The risk of this is that participants may self-edit their responses to shortcut their route through the survey by

answering ‘No’ to avoid multiple versions of the loop coming up and the survey taking longer and requiring more effort.

5.3.2 Double counting

As detailed in section 4.2.1, a series of check questions were included in an attempt to detect and correct instances of double counting.

As anticipated from the previous development work, double counting remained the main problem participants experienced when completing the online CSEW independently and this ties in with the findings in the mental models section. That is, participants naturally want to raise things at the point where they feel relevant. For respondents, it feels unnatural to discount duplicated incidents, or delay talking about them, and even where participants understood that they were supposed to discount duplicated incidents, they would still go ahead and select an incident (that is, say ‘Yes’ to a screener) wherever it felt relevant or valid. An example of this is a participant who felt annoyed by the fact that he was only meant to record the incident at one question saying *“‘I would have had that opportunity to get them not only for the bikes but also for breaking the locks on my shed.”*

In some cases, participants would go on to discount an incident at the double counting check question **ZRELATE**, but this could lead to missing incidents.

Missing incidents

There were two types of missing incident:

- **Conceptual error**: Refers to when respondents fail to include an incident they should have done as they don’t think they need to include it (commonly happens where a participant has experienced a more major incident, and as a result more minor incidents get forgotten or the participant doesn’t feel they ‘count’).
- **User error**: discounting incidents in error at **ZRELATE** or entering 0 at **ZCOUNT** because it is at the top of the list (and possibly meant to select 1) (see discussion below).

How the double-checking screens (ZRELATE) worked in practice

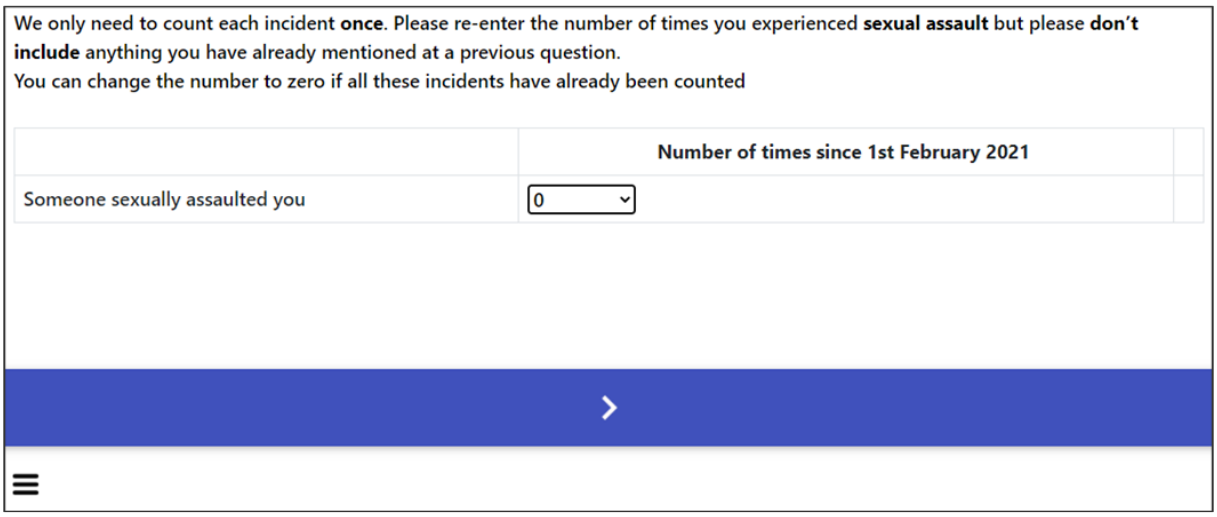
The functionality of the **ZRELATE** double checking screens is described in detail in section 4.2.1.

Cogability testing, building on Kantar’s earlier development work, demonstrated that the **ZRELATE** double counting checking method can work for multiple crimes that are still quite simple, for example comparing two single incidents, or even comparing up to three of four separate incidents, or two within a series with previously mentioned incidents. However, once moving beyond these examples, and larger numbers are involved both across and within crime type (and banded incidents in particular), this approach stops being helpful and can instead add confusion. Also, the length of the ‘crime description tag’ (see section 4.4.4) can add to the confusion; it is particularly long in the Fraud section. One participant said they felt **ZRELATE** was a ‘trick question’, repeatedly checking up on them and they struggled to distinguish between the different iterations of the screen each time it came up.

Even where participants recognised that they had double counted and answered correctly at **ZRELATE** this didn’t mean they necessarily corrected the error at **ZCOUNTCHECK**, especially where they were required to enter 0 as all incidents were double counted. One participant correctly recognised they had re-included three incidents of theft from outside

home, answered Yes at **ZRELATE** but then entered three again at **ZCOUNTCHECK** (possibly because they thought they had made an error initially).

ZCOUNTCHECK



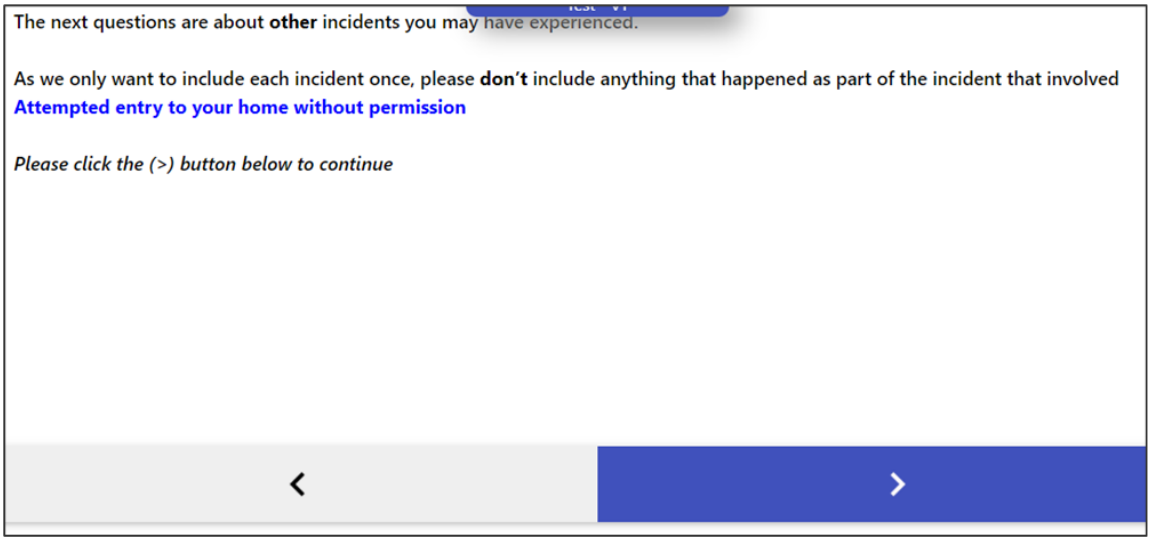
5.3.3 ZREVIEW incident summary review screens

As with **ZRELATE**, participants were often able to correctly interpret the first **ZREVIEW** screen (see section 4.2.1) where there were up to four separate incidents in the list. In the version of the survey that was tested, incidents with a banded count or where the participant didn't know the number of incidents didn't appear in the list and at least one participant questioned why these didn't appear. The second review screen was a key place where confusion and error arose, even where participants appeared to understand it. One suggestion, if this screen is retained in a future version, is to explore ways to change the design to make it clearer where participants can make changes, perhaps by greying out the first column at ZREVIEW2 (see section 4.2.1 for an illustrative example of this screen).

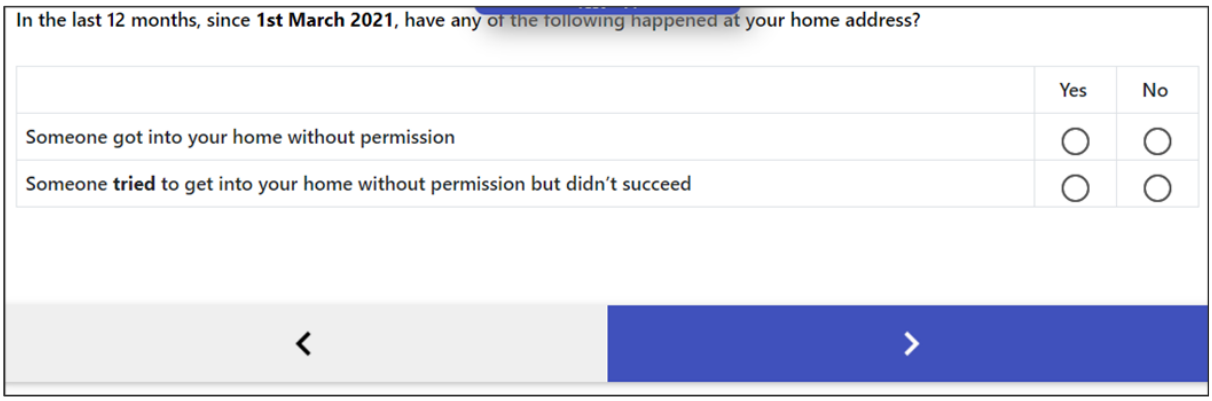
5.3.4 Skim reading/satisficing behaviours

Most, if not all, participants displayed satisficing behaviours such as skim reading text and skipping quickly past text-only screens, meaning they missed key information such as **DISPLAY4**. This is a text-only screen that comes up after the first incident is coded to ask the participant not to count it again and stress that they only need to count each incident once.

DISPLAY4



This also happened at the paired screeners (see example below) where participants were asked about actual crimes alongside attempted crimes. Participants commonly navigated straight to the place where they were required to enter something rather than reading the question text at the top. These findings reinforce that all key information needs to be in the screener text, or at the point where participants are required to type in something, rather than in the question stem.



The issue of skim reading was also a key problem for missed detail in introduction screens.

5.3.5 Deciding whether incidents are similar (a series) or different

Issues familiar from the previous online development project were again uncovered around classification of multiple incidents within the same crime type, as to whether these are similar (a series), different, or a mixture of both. Despite the question wording being simplified as part of the previous work to better suit online administration, there were still

ambiguities among participants where incidents had some features which were similar and others which were different[43].

Participants are then asked whether the incidents were similar or different in nature. Cogability testing found that the reference to the 'nature' of the incident lead to confusion among participants who had experienced multiple incidents of the same crime type. One participant asked, in relation to a theft, whether this meant the nature of the items stolen or the nature in which they were stolen (for example whether it was opportunist or pre-planned). As noted in the review of international crime surveys covered in Chapter 3, Kantar recommends exploring the ways other crime surveys deal with this and reviewing the kind of language they use in expressing these concepts.

5.3.6 Incident count (ZCOUNT)

A small number of participants had trouble in providing a total number of incidents where they felt it to be something persistent or ongoing. For example, one participant said the same person would try to get into their shed each time they walked past, and this was an ongoing occurrence. Using the rationale that this had happened more than once but they were unsure of the total number of incidents, this participant decided to enter two at the count question. A second participant included incidents of attempted vehicle theft that had happened along their road, also affecting other neighbours, but they incorrectly included the one time it had happened to their household as part of a series of seven incidents, that is including all the other neighbourhood incidents too.

5.3.7 Sensitive questions about assault

There were issues for a few participants around the sensitivity of these questions. The wording used in the 'crime description tags' (see section 4.4.4) for 'physical assault' and 'attempted physical assault' created issues for one participant. The 'crime description tag' refers to 'assault' and this participant didn't think of what happened to her in that way. She had picked up on the 'force' wording in the list and felt 'assault' was too harsh and "made it sound worse than what happened". The tag was amended to 'physical force or assault' for the live trial but will need further careful review going forwards.

The wording at these questions asks the participant to include anything that happened to them, 'including if this was done by people you know or live with, as well as by strangers'. One participant interpreted this reference as a request to include people you know who have experienced assault and recorded incidents on behalf of another member of the household. This particular participant rectified the error at the next screen where the questions referred to 'you' but this should be noted that this is a risk (see also section 4.4.3).

5.3.8 Threats and harassment 'route' through the victim module

Participants that came through to the victim module via a threats or harassment screener were filtered through the Force and Violence route which didn't necessarily provide a good 'fit' with their experiences and meant these participants get very few relevant questions here. As discussed further in section 4.6.5, we recommend that threats and harassment is included as a separate 'block' with the victim module (in addition to Theft, Attempted theft, Damage and Force and Violence).

[43] The wording tested as part of the earlier developmental work was 'You mentioned [NUMBER] incidents of [CRIME]**.** Were any of these very similar incidents, where the same thing was done under similar circumstances?' This was simplified from the original CSEW wording which asks 'You mentioned [NUMBER] incidents of [CRIME]. Were any of these very similar incidents, where the same thing was done under the same circumstances and probably by the same people?'

5.3.9 Confusion and overlap around theft from or outside home

An issue that came up several times was around how to classify theft from inside and outside home and what counts as a respondent's 'home'. For example, participants might include places such as a parking shed or garage initially as 'home' at **ZBREAKIN** as these screeners come up first and then double count at theft from outside home at **ZOSTHEFT**. If they have understood that they should only count each incident once, they might say 'No' to these more tailored screeners or resolve the double counting via **ZRELATE**. However, in the latter circumstance, references back to this incident in the victim module can feel incorrect as the script refers to a 'break in'. In essence, even though it doesn't matter where an incident is recorded from a script perspective, if the incident comes through on the 'wrong' screener, the text fill and references can feel wrong.

5.3.10 Issues at the Fraud screeners

Similar issues found in previous cognitive testing were found again here for participants who had experienced fraud, the key finding being that it can be difficult to tell the difference between the types of fraud screeners which all appear to contain similar features. It could feel as though the same question keeps coming up repeatedly. The crime description 'tags' (see section 4.4.4 used to reference earlier incidents at **ZRELATE** could also cause confusion about which incident was being referenced when the script attempts to draw out these differences and remove duplication.

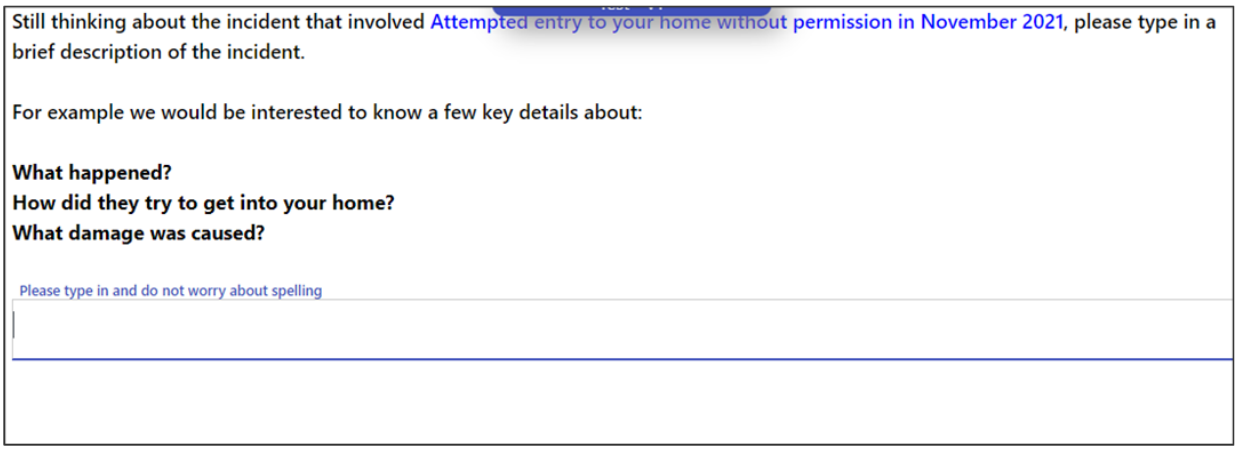
5.3.11 Findings at the open description question (ZDESCRINC)

In the WCSEW, as in the CSEW, participants were asked to provide an open description of each recorded incident.

At the open description question, the script shows three probes which aim to encourage the participant to write as much relevant information as possible about the incident. They are a mix of one generic probe ('*What happened?*') and two tailored probes depending on the screener question from which the victim module was generated. For example, at the break-in screener (ZBREAKIN) the two tailored probes were '*How did they get into your home?*' and '*What was stolen?'*; while at the non-confidence fraud screener (ZNONCON) the bespoke probes were '*How were your personal details or private accounts used?*' and '*Did you lose any money and did you get all or some of it back?*'

As with the previous cognitive and usability testing, the three tailored probes were found to be useful and participants commonly answered these directly, meaning it will be important to ensure the coders have access to these so that they can cross reference. The participants using tablets and smartphones didn't encounter any issues with typing in their responses. One participant commented that they liked that it said, 'don't worry about spelling'. Another said they would give a better description if they were reporting in person compared with this simple description. However, it was acknowledged that the expanding text box might encourage longer responses.

ZDESCRINC – open description



5.3.12 Issues around drawing in multiple incidents within one victim module

A key issue uncovered in the victim module was overlapping of incidents. Some participants did not restrict their thinking to the one incident or series of incidents in question. Even where working through the screeners had been smooth, this didn't necessarily translate over to the victim module and hence there is sometimes a disconnect.

Despite the objective to pin each victim module to a separate incident or series of incidents, where questions seemed to relate to other incidents they had experienced, some participants brought those incidents in too. Taking one participant as an example, their selected incident was an attempted break in. At the outset of the victim module, they started describing the attempted break in at the open description (**ZDESCRINC**). However, later in the victim module, they answered questions thinking about other incidents they had experienced including theft from a vehicle and theft from outside home; this tended to happen where the questions and response options 'fitted' or felt relevant (for example the questions about what was stolen or damaged). This cannot be controlled in a self-completion format and resultingly, as well as misrepresenting the incident in question, this is likely to cause issues in subsequent victim modules where participants feel they've already given the detail requested. These issues ultimately may lead to participants feeling annoyed by the repetition or that they've made a mistake, and there is a risk of break off as a result.

5.4 Summary and conclusions

This stage of cogability testing allowed further exploration of the WCSEW screeners and victim module, this time within the context of (sometimes very) complex experience of crime.

The 'mental models' exercise at the outset of the interview proved very useful in gaining an overall context of the participants' experience over the last year, and acted as a framework to help the researcher view the rest of the interview in context. This exercise demonstrated that the online CSEW successfully prompts participants to recall other (usually more trivial) incidents that were not recalled when asked about crimes experienced without any prompts.

However, the participants included in the cogability testing experienced a wider range of issues when completing the online questionnaire compared with those who took part in the

earlier development work, which is not surprising given the focus here on more complex experience.

Key difficulties were:

- **The rigidity of the questionnaire**: Participants are required to fit their responses into the structured order and flow of the questionnaire, rather than being able to report their experience in the way they would naturally choose.
- **Satisficing and skim reading**: Some participants were able to 'learn the loop', that is understand the pattern of questions which appear each time an incident is reported. Although there was no direct evidence of this in testing, there is a risk that this may lead to under-reporting as it provides an opportunity for participants to avoid a whole set of questions opening up if they answer 'yes'. There was also evidence of participants skim reading text that was not thought to be essential (such as text-only screens), and navigating straight to where they were required to type in or select a response option, missing key information covered in the stem text.
- **Double counting**: The main problem participants experienced was duplication of incidents. The duplication checking screens process (**ZRELATE**) alleviated double counting to some extent but, particularly for those who had complex experience, this could lead to further confusion and inaccurate crime counts. Similar confusion could arise at the **ZREVIEW** incident summary screens; again these 'worked' for some participants but where the experience was very complex these screens could appear overwhelming and confusion could lead to further error.
- **Complexity**: Further elements of complexity included counts, dates and deciding whether incidents were similar (a series) or different. Participants could also be confused by the difference between screeners, for example theft from within vs. outside the home and the different types of fraud screener which covered over-lapping features. This issue was also uncovered within the victim module where some participants 'overlapped' incidents at questions that felt relevant to another incident, rather than restricting thinking to one incident at a time.
- **Sensitive questions about assault, threats and harassment**: particular care needed to be taken for participants who had experienced crimes such as assault, threats and harassment. Additionally, the 'route' participants took through the victim module following a threats or harassment screener was not always felt to be relevant and these questions could be better structured.

6. Stage 5: Live Trial methodology

Usability and cognitive testing can only go so far in assessing a survey instrument. Therefore, a live trial was set up to test the questionnaire at scale. The live trial was conducted on Kantar's Public Voice random sample panel to explore how well the revised screener questions and victimisation module work when implemented with a larger sample of respondents.

This chapter provides an account of the methodology of the live trial, covering the following elements:

- Aims and objectives (section 6.1)
- Sample design (section 6.2)
- Fieldwork approach (section 6.3)
- Fieldwork performance (section 6.4)
- Data weighting (section 6.5)

The main results of the live trial are covered in Chapter 7, and Chapter 8 covers respondent experience of the live trial including usability evaluation, interview length and drop-out rates.

6.1 Aims and objectives

The primary aim of the Live Trial was to assess the reliability of a wholly online version of the CSEW questionnaire. It sought to provide answers to research questions (iii) and (iv) and contribute to the answer to research question (v) (see section 2.5). For reference, these are:

- Would an online victim module, which collects data on the experience of individual crimes, measure all the necessary details for crime estimates reliably?
- Would an online questionnaire with victim modules be able to provide reliable prevalence and incidence rates for all crime types? When would a case become too complex to do this reliably?
- What is the maximum length an online survey can be to ensure answers are reliable? Are there any methods that can be incorporated to ensure reliability of an online questionnaire?

These questions may be repackaged into three critical questions about the screener module and the victim module and how the two relate to each other:

- What difference would an online version of the screener module - as opposed to a telephone version - make to the size and composition of the subset of respondents that qualifies for the victim module?
- Would an online version of the victim module yield the same results as a telephone interview version for all cases, including complex cases?
- Would the information entered by respondents into an online version of the victim module be sufficient for offence coding to be effectively carried out?

6.2 Sample design

The sample for this study was taken from Kantar Public's random sample panel, Public Voice. The Public Voice panel includes c. 23,000 individuals in England and Wales, spread across c. 16,000 households. Most panel members were recruited using a variant of the random sample 'push to web' method, also known as 'ABOS' (Address-Based Online Surveying). This approach involves contacting households by sending them a letter in which they are invited to join the online panel and take part in online surveys.

To meet the objectives outlined above, the Live Trial comprised three elements:

- A large scale randomised controlled trial (RCT) in which the mode of data collection (online questionnaire or telephone interview) was the 'treatment'.
- A repeat follow-up telephone interview of online respondents from the RCT who had completed at least one victim module, to tease out mode effects on an individual basis.
- A web boost which, combined with the online-allocated trial sample, can be used to simulate an online CSEW.

At the time of this survey, there were 14,439 Public Voice panel members eligible for the RCT: resident in England or Wales, aged 16+ and had provided a valid UK telephone number as part of their panel registration or subsequently. These 14,439 panel members were spread across 11,361 households.

Among the 8,679 households with just one eligible panel member, a systematic sample of 7,985 was drawn for inclusion in the trial (a conditional sampling probability of 0.92).

All households with at least two eligible panel members (2,682) were included in the RCT but one panel member from each household was randomly sampled in advance to be included in the trial. Only this randomly sampled individual in each household was then invited to take part in the survey. The eligible panel members in these households each had a $1/N_h$ probability of being sampled for the trial, where N_h is the number of eligible panel members in household *h*.

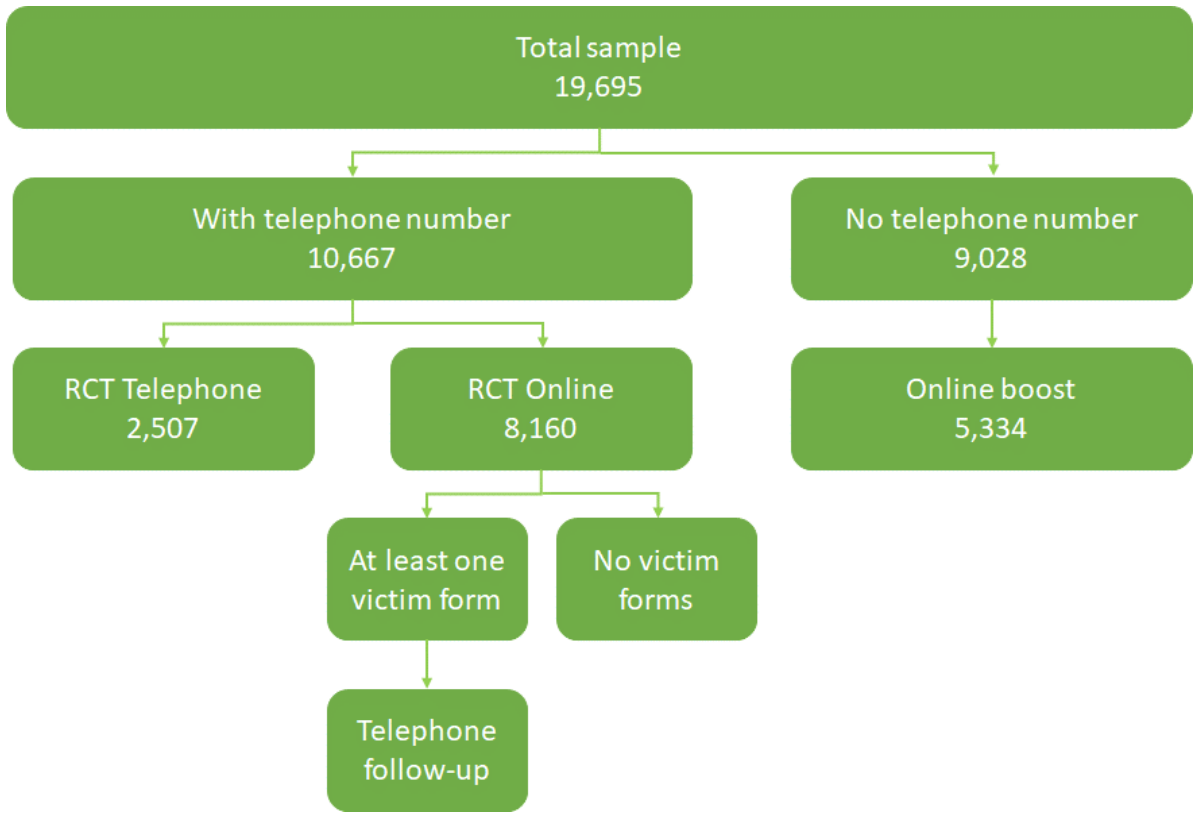
In total, 10,667 individuals were sampled for the RCT (7,985 + 2,682).

Of these 10,667 individuals, 8,160 were allocated to online data collection and 2,507 to telephone interview data collection. This was done using a systematic random sampling method after the sample of 10,667 had been sorted by (i) original recruitment survey stratum, (ii) sex, (iii) age group, and (iv) panel base weight. The panel base weight is derived from the recruitment survey weight and reflects the difficulty of recruiting that 'type' of person into the panel.

There were also 9,028 panel members who were eligible for the survey but not for the randomised controlled trial because they had not provided a valid UK telephone number. These panel members were nevertheless eligible to boost the numbers completing the online survey.

This set of 9,028 was first reduced to exclude 1,831 living in the same household as someone eligible for the RCT. The survey-eligible panel members in the remaining households each had a $1/N_h$ probability of being sampled for the online survey, where N_h is the number of survey-eligible panel members in household *h*. In total, 5,334 panel members were sampled to boost the numbers completing the online survey.

Figure 6.1 – Sample composition



6.3 Fieldwork approach

As outlined in section 6.2 above, different subsets of the Public Voice panel were allocated to different fieldwork plans:

- Telephone interview only
- Online questionnaire only
- Online questionnaire followed by a repeat telephone interview (if at least one victim module was completed in the online questionnaire).

Panellists were sent an advance letter or email (depending on whether an email address was available) inviting them to take part in a survey 'on the subject of crime'. The letter/ email (which can be found in Appendices D, E and F) explained that the study was being conducted on behalf of the Office for National Statistics but did not contain any reference to the Crime Survey for England and Wales. This approach was taken to distinguish this work from the main CSEW and ensure that respondents did not mistakenly believe that were taking part in the CSEW itself.

All respondents were offered a £10 gift voucher in exchange for taking part. Respondents who completed the online questionnaire **and** a telephone interview received a separate voucher for each (£20 in total).

Further details of each plan are provided separately below.

6.3.1 Telephone interview only

Letters/ emails were sent to selected panellists on Thursday 31st March, shortly before the start of the fieldwork period on Monday 4th April (see Appendix 4 for a copy). The primary purpose of the letter/ email was to notify panellists that they were being invited to take part in the survey, and to expect a call from one of Kantar Public's telephone interviewers. The fieldwork period remained open for just over 4 weeks, with calls being made until Wednesday 4th May.

Telephone fieldwork was undertaken by Kantar's UK telephone research operation, which specialises in conducting social research among individuals and businesses. Interviewers worked from home using Kantar's 'CATI@home' set up. This enables interviewers to link directly to the central sample management and auto dialler system and work in a manner similar to working in a central telephone unit. All interviewers can be remotely monitored for quality control purposes.

All interviewers took part in a briefing before the start of the survey, conducted by members of the Kantar Public research team. The briefing focussed on the purpose of the live trial, the composition of the sample, and the questionnaire. In addition to the briefing all interviewers conducted practice interviews before starting work to ensure that they were familiar with the questionnaire.

Supervision and performance monitoring of interviewers was undertaken by existing Telephone (CATI) Shift Leaders and Supervisors. A proportion of interviews were listened to by the supervisory team using undetected listening facilities. Kantar have a systematic method of prioritising and selecting which interviewers are monitored. This is based on their experience, previous performance and regularity of monitoring. Monitoring results were graded using a standard benchmark and interviewers received regular feedback on their performance.

To maximise response, telephone interviewers implemented varied calling patterns, in line with the calling patterns used on the TCSEW; namely enforcing a minimum number of calls and ensuring that calls were spread over a period of time, on different days and at different times.

6.3.2 Online questionnaire

For panellists selected to take part in the online questionnaire, an adaptive contact strategy was implemented to account for the different contact information available for any given panellist. Across this group, three contact methods were used – email, text message and letter.

The full contact strategy for each group is shown in Figure 6.2 below (see Appendix 5 for a copy of the initial email/letter sent).

Figure 6.2 - Web fieldwork contact strategy					
Contact stage ↓	***Contact details available***				
	Date	**Email and mobile phone**	**Email but no mobile phone**	**Mobile phone but no email**	**Address only or address & landline**
1	04/04/22	Email with survey link + supporting text message with no link (24hrs later)	Email with survey link	Text message with survey link	
2	08/04/22	Email with survey link + supporting text message with no link (24hrs later)	Email with survey link	Text message with survey link	
3	11/04/22	Letter	Letter	Letter	Letter
4	15/04/22	Email with survey link + supporting text message with no link (24hrs later)	Email with survey link	Text message with survey link	
5	22/04/22	Email with survey link + supporting text message with survey link (24hrs later)	Email with survey link	Text message with survey link	

Online fieldwork closed on Tuesday 26th April 2022.

Where possible, the initial survey invitation was sent by email. All emails contained individualised survey hyperlinks, so no login details were required. Additional verification was based on panellist birthdate (including year). Where an email address was available, text messages were used only as a supporting communication, sent to non-responders 24 hours after the email and with no embedded survey hyperlink. However, in an effort to maximise the response rate, the last text message (of four) did contain a survey hyperlink.

Where an email address was not available, all text messages contained a survey hyperlink. The contact management system *sendinblue*[44] was used for all email and text message communications.

Eight days after the first email, a letter was sent to (i) all non-responders and (ii) those panellists who have not supplied an email address or mobile telephone number. This letter

[44] https://www.sendinblue.com/

contained the survey website address and unique login details for each respondent, but not a printed individualised survey hyperlink. This meant respondents had to go to the survey landing page and log into the survey using the details contained in the letter.

Copies of the initial contact email/letter for both online and telephone

6.3.3 Online questionnaire followed by telephone interview

For panellists who were eligible for a follow-up telephone interview, the contact strategy for invitation to participate in the online questionnaire was as described above (Section 6.3.2). Respondents who reported experiences of crime were subsequently invited to take part in a telephone interview, in which the same questions were asked. To limit the time elapsed between the online questionnaire and the telephone interview, the sample was divided into three batches. A staggered approach to fieldwork was adopted, whereby batches were started at different points throughout the fieldwork period.

Sample was issued to the telephone survey on the following dates.

- Batch 1: Monday 11th April
- Batch 2: Tuesday 19th April
- Batch 3: Thursday 28th April

Telephone fieldwork closed on Thursday 5th May.

Eligible online respondents were sent an email or letter (depending on whether an email address was available) in advance of being called for the first time. A copy of the letter/email sent can be found in Appendix 6. These reminded the respondent that they had recently completed the online questionnaire, explained that the ONS was 'testing different ways of asking these questions', and asked if they would be willing to answer the same questions again over the telephone.

In the original design, any panellist who completed an online questionnaire that included at least one victim module would be invited to take part in a follow-up telephone interview. However, at the point at which the first sample batch was drawn, the victimisation rate was found to be higher than anticipated.

As a result, the first batch of sample contained only online respondents who had recorded more than one incident type (i.e. answered 'Yes' to two or more screeners). This was done to ensure that respondents with more complex incident profiles would be included in the follow-up telephone survey.

For the second batch, a 1 in 2 selection was carried out among all online respondents who had recorded at least one incident type, including those who had taken part prior to drawing Batch 1.

For the third batch, a 1 in 2 selection was carried out among all respondents who had completed the online questionnaire after Batch 2 selection and had recorded at least one incident type.

6.4 Fieldwork performance

Table 6.1 below shows the number of interviews achieved in each sample group. A total of 7,291 respondents completed a first interview as part of the trial: 6,219 online interviews were achieved and 1,072 telephone interviews. Of the 6,219 online interviews, 4,147 were part of the RCT sample and a further 2,072 were part of the online boost.

A further 436 respondents who initially completed an online survey also completed a second 'paired' interview. This means that a total of 7,727 interviews were achieved with 7,291 respondents.

Table 6.1 – Final achieved interviews by sample group

Sample group	*Issued*	*Target*	*Completes*	*Conversion rate*
RCT online data collection	8,160	4,100	4,147	51%
RCT telephone data collection	2,507	1,000	1,072	43%
Online boost	5,334	2,150	2,072	39%
Telephone follow-up	1,036	410	436	42%

6.5 Data weighting

6.5.1 Randomised controlled trial

For the **5,219 randomised controlled trial respondents**, the weight for each case is approximately equal to 1/p(response | recruitment survey variables). A separate response model was estimated for those allocated to the online survey (8,160) and for those allocated to the telephone survey (2,507). Each fitted value of p(response | recruitment survey variables) was trimmed to ensure it fell within the range of 0.2 to 5 times the median fitted value.

No design weight component was required because there was no interaction between the probability of being sampled for the trial and the probability of being allocated to the selected mode.

6.5.2 Online survey

A separate weight was calculated for all **6,219 online survey respondents** (trial and boost samples), which was considerably more complex than the RCT weight. This weight incorporated the panel base weight, a survey-specific design weight, a non-response weight (p(online response | recruitment survey variables)) and a final calibration weighting step to ensure that the sample profile matched that of a near-contemporary (October-December 2021) weighted UK Labour Force Survey (filtered to include only those aged 16+ and living in England or Wales) on sex and age, region, education level and birth country distributions.

6.5.3 Follow-up telephone survey

No weight has been used to analyse the paired data from **436 panel members who completed an online survey first and a telephone survey second**. This group comprises RCT online respondents who qualified for at least one victim module. However, it omits people who would qualify for at least one victim module if they completed the telephone survey but *would not have qualified for a victim module if they completed the online survey*. This omission prevents the sample from being treated as a probabilistic sample of victims from either survey mode; instead it should be treated as a good quality convenience sample and not weighted to represent a more broadly meaningful population.

7. Stage 5: Live trial results

This chapter describes main findings results of the live trial including of detailed analysis of the following:

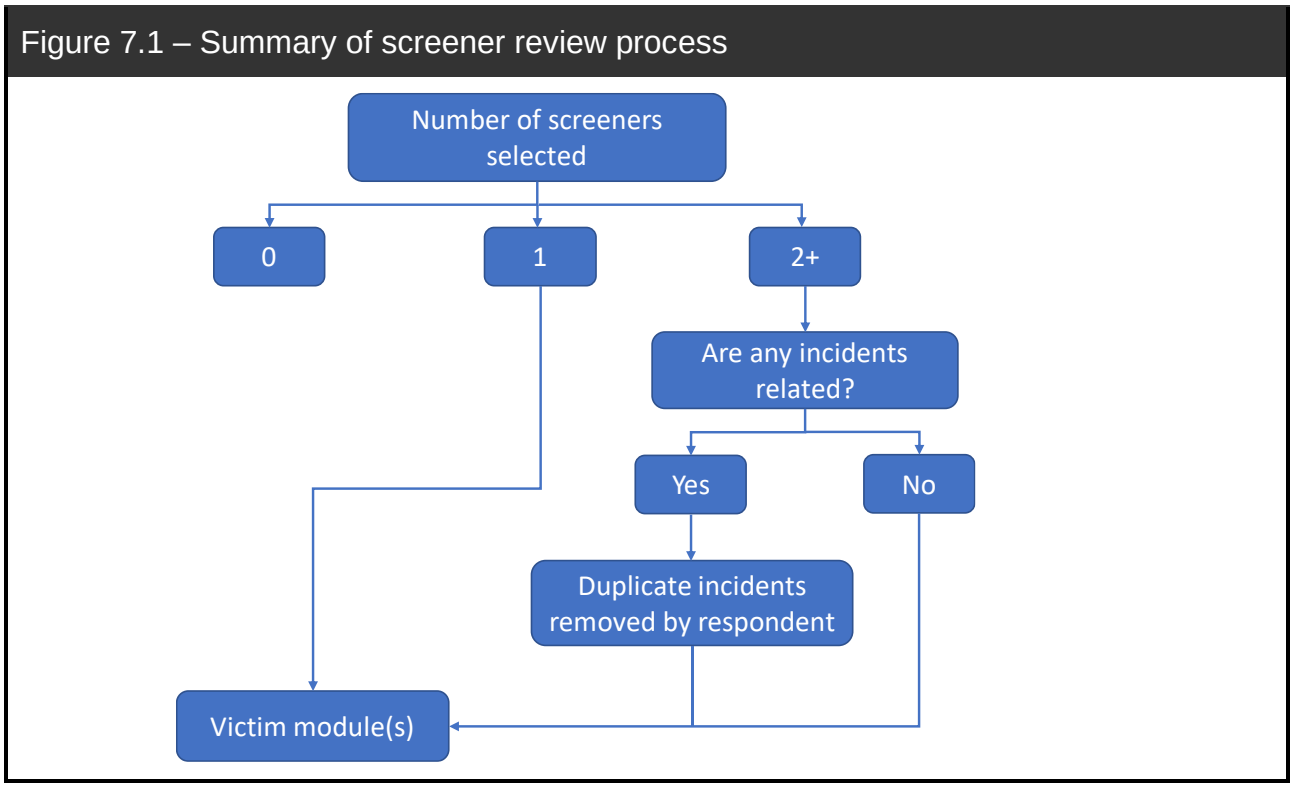
- The screeners (section 7.1)
- Victim modules (section 7.2)
- Telephone follow up (section 7.3)

7.1 Screener questions

As outlined in Chapter 4, any online, self-completion version of CSEW will need to resolve the issue of double counting incidents (which would result in respondents completing more than one victim module for the same incident). In the existing face-to-face CSEW, upon selecting their first screener, respondents see an onscreen instruction at all subsequent screeners to only include new incidents from that point forward ('Apart from anything you've already mentioned, have you…'). However, earlier research[45] has shown this to be ineffective at reducing double counting in self-completion surveys.

The Live Trial attempted to address double counting through the implementation of a review process within the online script, whereby a respondent who answered 'Yes' to more than one screener was specifically asked whether the incidents were related. Figure 7.1 shows a simplified summary of this process (full details can be found in section 4.2.1).

Figure 7.1 – Summary of screener review process



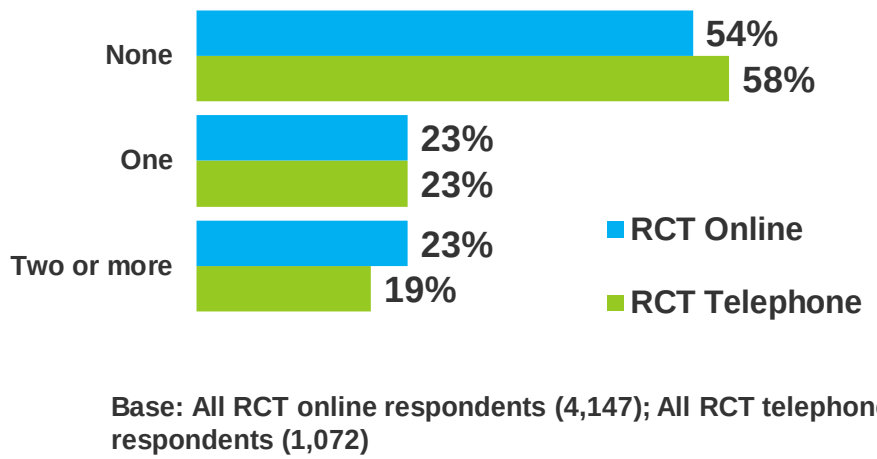
[45] See section 5.8.2 of the 2018 report: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/redesignofcrimesurveyforenglandandwalescsewcorequestionsforonlinecollection/2018-07-19

This process was intended to ensure that, upon completion of the screeners (and prior to completion of any victim modules), each screener selected represented a unique incident.

7.1.1 Recording of incidents at initial screener questions

Figure 7.2 shows the number of screeners initially selected by online and telephone respondents in the randomised controlled trial (RCT). Respondents to the telephone survey were more likely than online respondents to record no incidents at any screener question (58% and 54% respectively). Online respondents were not only more likely than telephone respondents to record at least one incident, they were also more likely to record incidents at more than one screener: 23% of online respondents recorded an incident at two or more screeners compared with 19% of telephone respondents.

Figure 7.2 Number of screeners initially selected (Pre-Review)



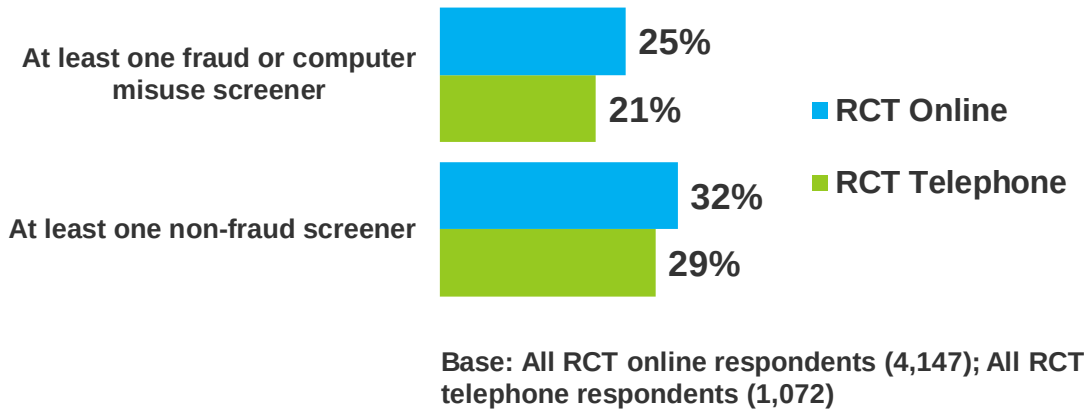
Base: All RCT online respondents (4,147); All RCT telephone respondents (1,072)

The screener questions can be divided into two broad groups – fraud and computer misuse screeners and non-fraud screeners[46].

As shown in Figure 7.3, online respondents were more likely than telephone respondents to have selected at least one fraud screener (25% and 21% respectively). For non-fraud screeners, although online respondents were more likely to have selected at least one screener (32% and 29% respectively), the difference was not statistically significant.

[46] These incident types are: burglary, attempted burglary, theft/ attempted theft from a dwelling, damage to a dwelling, theft/ attempted theft of a vehicle, theft/ attempted theft FROM a vehicle, damage to a vehicle, theft/ attempted theft of a bike, theft/ attempted theft from outside a dwelling, theft/ attempted theft from a person, theft/ attempted theft away from home damage to personal belongings, assault (including sexual assault), attempted assault, threats and harassment.

Figure 7.3 – Type of screeners initially selected (Pre-Review)



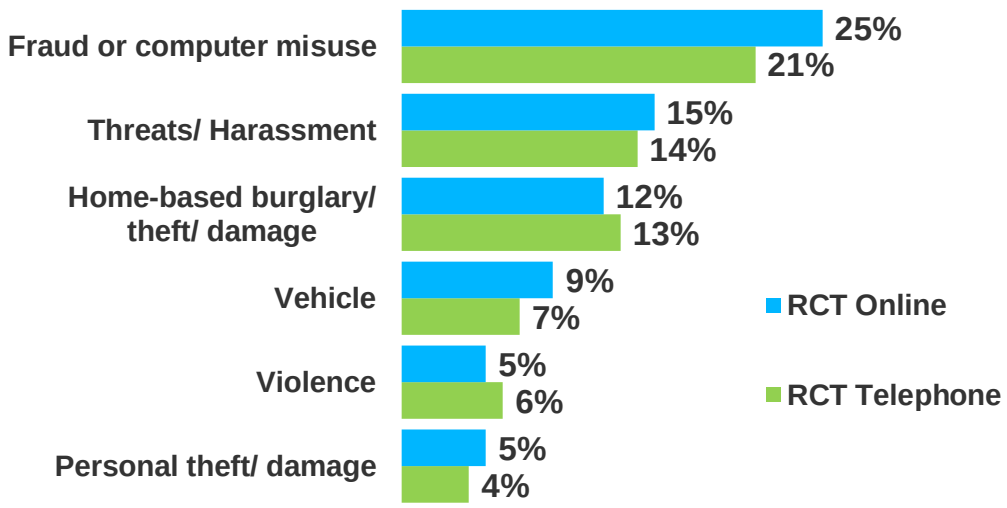
Base: All RCT online respondents (4,147); All RCT telephone respondents (1,072)

Further granularity within non-fraud crimes can be provided by comparing the proportion of online and telephone respondents who initially selected each individual screener type. Given the number of screener questions, these have been aggregated into some broad categories for analysis purposes.

Figure 7.4 shows that the proportion of telephone and online respondents reporting at least one incident in each crime category were broadly similar. Online respondents were more likely than telephone respondents to mention fraud (as noted above) and vehicle crime. However, reporting of violence, threats or harassment, and property crimes (either household or personal) were broadly the same in both groups. Telephone respondents were just as likely as online respondents to report violence at the screener questions (6% and 5% respectively). This is a notable finding: since violence is a more sensitive crime it might have been hypothesised that there would be a lower level of violence reported on an interviewer-administered telephone survey compared with a self-completion online survey.

One notable difference of the screener questions used in the live trial compared with the current CSEW is the inclusion of questions on attempted crimes for break-ins, thefts (both household and personal) and violence. There was no difference between online and telephone respondents in terms of reporting attempted crimes: 9% of respondents in both groups reported at least one incident of attempted crime.

Figure 7.4[47] – Proportion of respondents initially selecting each screener type (Pre-Review)



Base: All RCT online respondents (4,147); All RCT telephone respondents (1,072)

In summary, overall online respondents were more likely than telephone respondents to report at least one incident at the screener questions and were also more likely to record two or more incidents. However, when broken down by crime category or individual screener there were relatively few differences between the two groups. These findings are broadly in line with the international evidence which suggests that self-completion online surveys tend to produce higher reporting levels compared with interviewer-administered surveys.

Based on these findings alone this would suggest that an online survey might produce a higher victimisation rate compared with an equivalent interviewer-administered survey. However, this is before the respondent review process where an effort was made to eliminate any double counting that the respondent may have recorded (discussed further below).

7.1.2 Review of incidents recorded at screener questions

The review process is explained in detail in section 5.3.2 and summarised in Figure 7.1. This section examines how the review process changed the proportion of respondents who reported at least one incident.

Among respondents who had initially selected at least one screener, 33% of online respondents and 21% of telephone respondents changed one or more screener responses

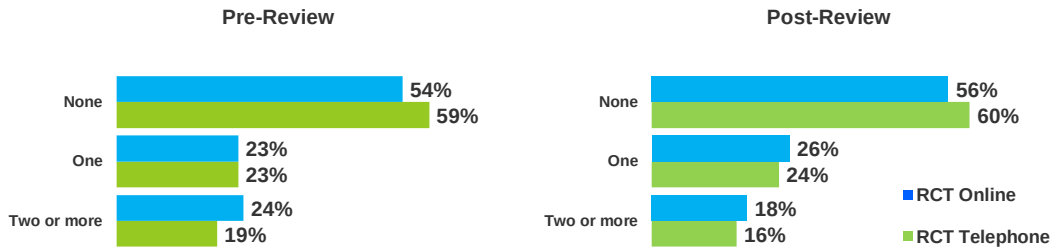
[47] Incident types have been grouped as follows: Personal theft/ damage (Theft/ Attempted theft from person, Theft/ Attempted theft away from home, Damage to personal belongings); Fraud (Non-confidence fraud, Confidence fraud/ Attempted confidence fraud, Theft of personal information, Virus); Threats/ Harassment (Threat, Harassment); Violence (Sexual assault, Non-sexual assault, Attempted assault);Home-based burglary/ theft/ damage (Burglary, Attempted burglary, Theft/ Attempted theft from a dwelling, Damage to a dwelling, Theft/ Attempted theft from a dwelling); Vehicle (Theft/ Attempted theft of a vehicle, Theft/ Attempted theft FROM a vehicle, Damage to a vehicle, Theft/ Attempted theft of a bike).

during the review process. While it is perhaps not unexpected that online respondents were more likely to change their responses than telephone respondents (due to no interviewer being present to provide guidance) the fact that across both modes 30% of respondents who initially reported at least one incident made a change, illustrates the complexity of the task facing respondents.

Figure 7.5 shows the net impact of these changes based on the review process. Following the review process adjustments, telephone respondents were still more likely than online respondents to report no incidents at any screener question (60% and 56% respectively). Online respondents were also still more likely than telephone respondents to report incidents at two or more screener questions (18% and 16% respectively).

Although the overall difference between telephone and online respondents remained, the net impact of the review process was the same for both groups – to reduce the number of incidents recorded at the screener questions. Given that the primary purpose of the review process was to remove any double counting the fact that there was a net reduction in the number of screeners suggests this worked to some extent.

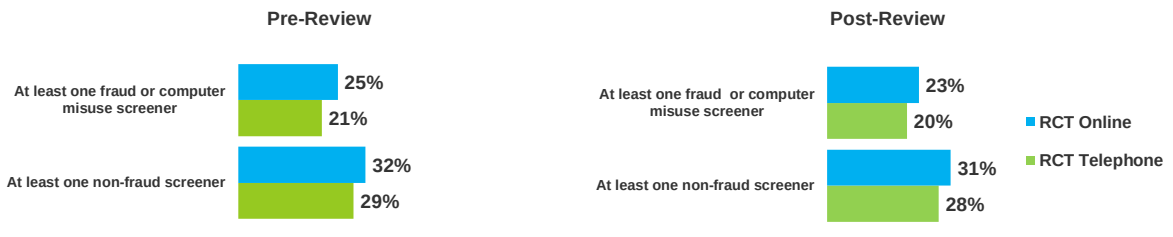
Figure 7.5 – Number of screeners selected (Pre and Post-Review)



Base: All RCT online respondents (4,147); All RCT telephone respondents (1,072)

As shown in Figure 7.6, after the review process online respondents were also still more likely than telephone respondents to have selected at least one fraud or computer misuse screener (23% and 20% respectively) and one non-fraud screener (31% and 28% respectively). For both types of respondents there was a small net reduction in the number of both fraud and compute misuse and non-fraud screeners.

Figure 7.6 – Type of screeners selected (Pre and Post-Review)

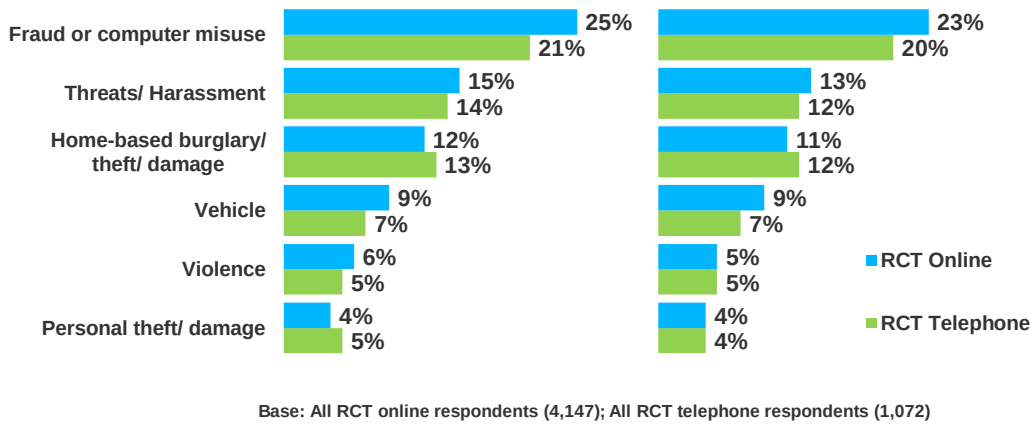


Base: All RCT online respondents (4,147); All RCT telephone respondents (1,072)

Figure 7.7 shows that following the review process there were negligible changes to the proportions recording incidents at each crime category among both telephone and online

respondents. Online respondents were still more likely than telephone respondents to select a fraud or computer misuse screener and a vehicle screener.

Figure 7.7 – Proportion of respondents selecting each screener type (Pre and Post-Review)



Base: All RCT online respondents (4,147); All RCT telephone respondents (1,072)

Although the above analysis suggests that the review process was successful in reducing some element of double counting the net changes were relatively small and the initial differences seen between online and telephone respondents remained. However, the amount of gross change was quite substantial among both groups and so it would be useful to understand changes at the level of individual screener questions. By analysing responses to the online survey in isolation (including respondents from the web boost sample) it is possible to analyse responses to specific, individual screeners in more detail.

As noted above, one of the primary aims of this work was to design a self-completion survey instrument that would, as far as possible, avoid incidents being double counted by building in a review process to the script. For example, a respondent may have experienced a single incident of mugging but answered 'Yes' to both the 'Theft from person' screener and the 'Assault – non-sexual' screener. In this scenario, the respondent would be asked if the two incidents were linked and, if they were, one of the incidents would be discounted to ensure that they only completed one victim module in relation to the incident.

Of course, the preferred scenario would be for a respondent to only answer 'Yes' at a single screener question for each specific, unique incident they have experienced. While the review process is designed to provide a check against double counting, if a lot of respondents are changing their initial answers when prompted this does indicate that the concept of an 'incident' is not fully understood by respondents at the start of the survey when they are answering the screener questions. While some element of double counting will always remain – thus making a review process important - the first objective of the survey instrument should be to minimise the number of respondents who make changes as a result of reviewing their original responses.

Two specific lines of enquiry can provide more insight into which incident types presented the most challenges for respondents when completing the screener questions:

- Which screeners were most likely to be ‘deselected’ at the review stage (i.e. changed from a ‘Yes’ to a ‘No’)?
- Among respondents selecting each screener, what proportion deselected *other* screeners at the review stage?

7.1.3 Screeners most likely to be deselected

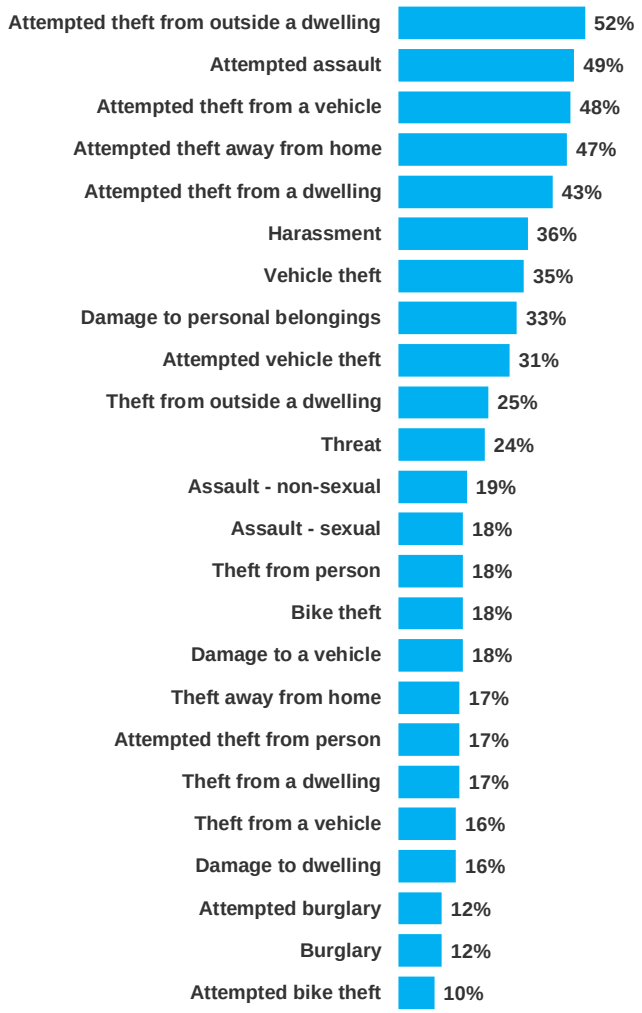
To better understand which screeners were most likely to have been selected by respondents ‘in error’ (i.e., due to double counting), we can look at the proportion of respondents who ‘deselected’ each screener during the review stage. Due to the small base sizes for most individual crimes, it is not possible to compare the RCT online and RCT telephone groups. Instead, this section focuses on the changes made by all online respondents – that is both the RCT group and the online boost sample.

Figure 7.8 shows the deselection figures for each individual screener. The base in each case is all respondents who initially selected the screener and at least one other screener. Respondents who selected only one screener would not be taken through the review process and so had no opportunity to deselect the screener initially selected. As an example, among respondents who initially recorded an incident at ‘Attempted theft from outside a dwelling’ and at least one other screener, just over half (52%) changed their response (or deselected) this screener during the review process.

It is noticeable that the screeners most likely to be deselected were not related to a particular type of crime but were all attempted crimes. The most commonly deselected screeners were attempted theft from outside a dwelling (52%); attempted assault (49%); attempted theft from a vehicle (48%); and attempted theft away from home (47%). However, among other attempted crimes the level of deselection was far lower: for example, 10% of attempted bike thefts and attempted theft from a person (12%).

Fraud and computer misuse crime also had relatively high levels of changes after the review process although this was not the case for non-confidence fraud (3% deselected).

Figure 7.8 – Proportion of online respondents who ‘deselected’ each screener at the review stage



Base: All online respondents (RCT online + online boost) who selected each screener and at least one other screener: Attempted theft from outside a dwelling (62); Attempted assault (125); Attempted theft from a vehicle (51); Attempted theft away from home (42); Attempted theft from a dwelling (32); Fraud - theft of personal information (163); Attempted fraud - confidence (537); Fraud - confidence (287); Harassment (514); Vehicle theft (35); Damage to personal belongings (42); Attempted vehicle theft (74); Theft from outside a dwelling (196); Threat (469); Fraud - virus (130); Assault - non-sexual (194); Assault - sexual (44); Theft from person (44); Bike theft (55); Damage to a vehicle (176); Theft away from home (81); Attempted theft from person (64); Theft from a dwelling (49); Theft from a vehicle (103); Damage to dwelling (213); Attempted burglary (122); Burglary (45); Attempted bike theft (40); Fraud - non-confidence (326)

A closer examination of respondents who deselected attempted incidents gives an indication of the types of incidents these screeners were commonly related to. Among respondents who deselected:

- **Attempted theft from outside a dwelling**, 35% also selected ‘Harassment, 30% also selected ‘Attempted burglary’, and 29% also selected ‘Theft from outside a dwelling’;
- **Attempted assault**, 60% also selected ‘Assault – non-sexual’, 19% also selected ‘Threat’, and 18% also selected ‘Harassment;
- **Attempted theft from a vehicle**, 48% also selected ‘Attempted vehicle theft’, 33% also selected ‘Threat’, and 27% also selected ‘Theft from a vehicle’;

- **Attempted theft away from home**, 35% also selected ‘Attempted theft from person’, 34% also selected ‘Harassment’, and 27% also selected ‘Theft away from home’.

These results suggest that, relative to other screeners, attempt incidents were particularly likely to be recorded as part of an incident that encompasses multiple incident types.

It is perhaps unsurprising that certain incident types were selected together, since the nature of those incidents make it likely that they would occur at the same time (e.g. Attempted assault’ being selected with ‘Threat’ and ‘Harassment’, or ‘Attempted theft from a vehicle’ being selected with ‘Attempted vehicle theft’).

It would also appear that some respondents mistakenly answered ‘Yes’ to both questions in a paired screener (see 5.3.4 for an example of a paired screener). It is notable that, among respondents who deselected an ‘attempt’ screener, this was initially selected in combination with the corresponding ‘actual’ screener. For example:

- ‘Attempted theft from outside a dwelling’ being selected with ‘Theft from outside a dwelling’
- ‘Attempted assault’ being selected with ‘Assault – non-sexual’
- ‘Attempted theft from a vehicle’ being selected with ‘Theft from a vehicle’
- ‘Attempted theft away from a home’ being selected with ‘Theft away from home’.

The fact that in these cases the ‘attempt’ screener was deselected suggests that they were part of the same incident. There are two potential explanations for this double counting. In the case of assault, for example, it is plausible that an assault and an attempted assault genuinely happened as part of the same incident. Alternatively, some respondents might take the view that these incidents happened in sequence, and therefore answer ‘Yes’ to both questions in a paired screener. For example, in the case of theft from a vehicle, one interpretation is that the perpetrator first tried to steal something, and that attempt was subsequently successful. While the ‘attempt’ screeners do include the phrase ‘but did not succeed’, this could be missed by some respondents.

Deselection of *other* screeners

As well as considering which screeners were most commonly deselected, we can explore cases where a respondent initially selected multiple screeners but was ultimately left with just one screener following the review process.

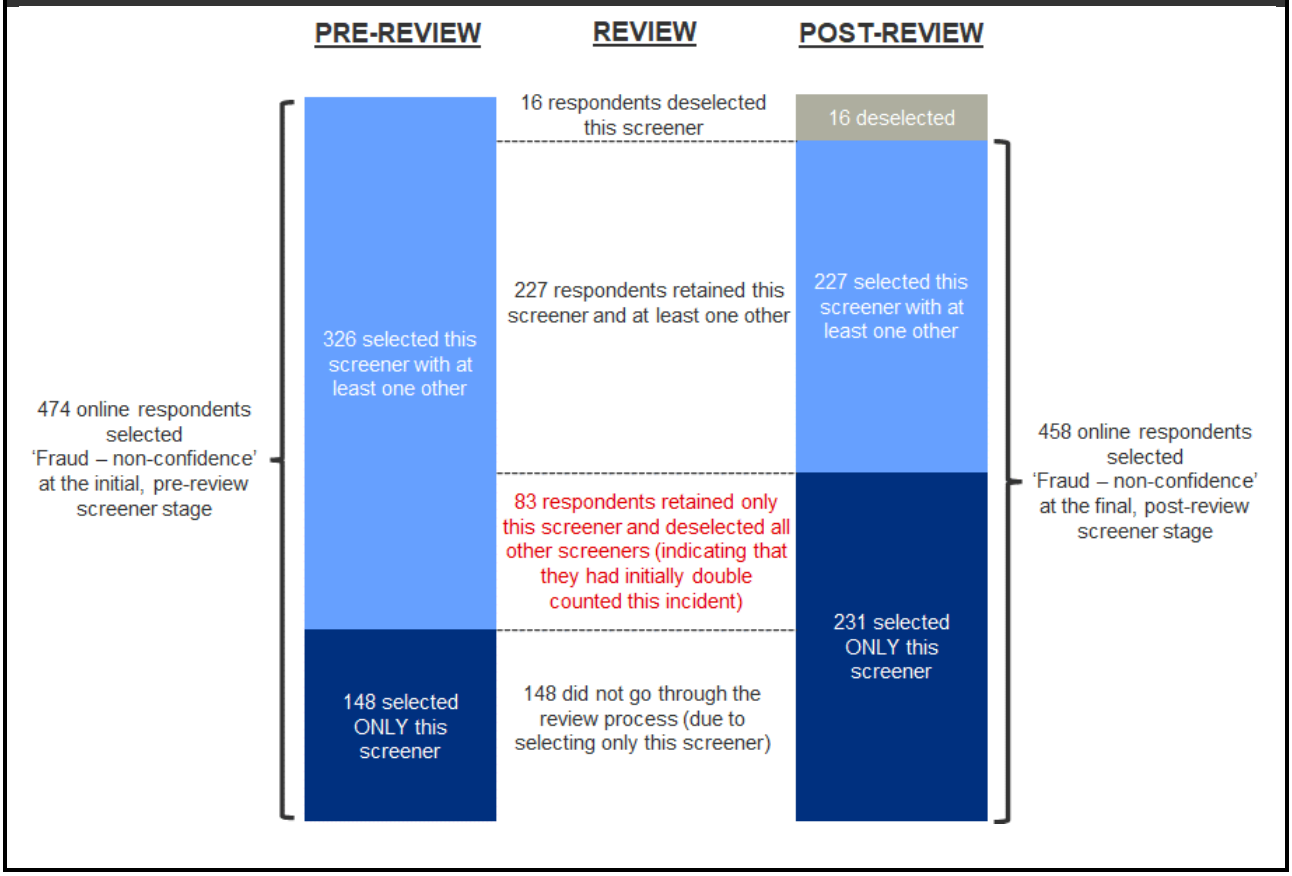
Respondents who initially selected a given screener fall into two groups; (a) those who selected ONLY that screener, and (b) those who also selected at least one other screener. The process for respondents in group (b) was relatively straightforward – they were asked to complete a victim module without the need for any review of their screener responses.

However, group (a) respondents would have gone through the review process (see figure 7.1), resulting in one of three outcomes:

- The screener in question could be deselected
- The screener could be retained along with at least one other screener
- The screener could be retained in isolation, with all other screeners being deselected

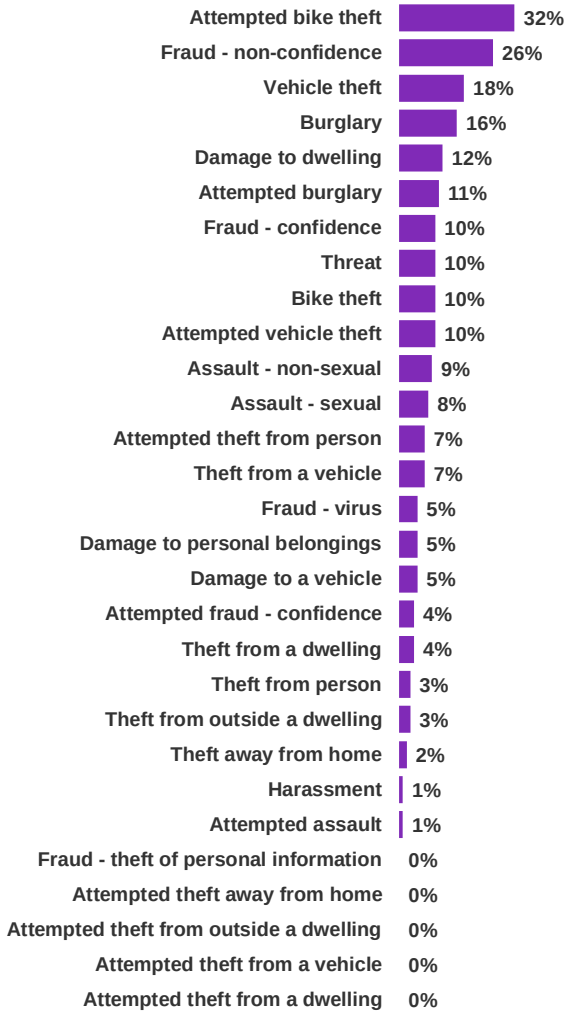
Figure 7.9 illustrates this process, with the responses to the ‘Fraud – non-confidence’ screener used as an example.

Figure 7.9 – Screener review process example: 'Fraud-non-confidence'



In the example above, the group of interest is shown in red: respondents who initially selected multiple screeners but retained only the screener in question post-review. Comparing this group across all screeners (Figure 7.10) provides another indication of which screeners were most likely to feature as part of a 'mixed offence' incident (i.e. one involving more than one type of crime) as opposed to being a standalone incident, and may therefore be more susceptible to being 'double counted'.

Figure 7.10 – Proportion of online respondents retaining each screener in isolation at the post-review stage



Base: All online respondents (RCT online + online boost) who initially selected each screener and at least one other screener: Attempted bike theft (40); Fraud - non-confidence (326); Vehicle theft (35); Burglary (45); Damage to dwelling (213); Attempted burglary (122); Fraud - confidence (287); Threat (469); Bike theft (55); Attempted vehicle theft (74); Assault - non-sexual (194); Assault - sexual (44); Attempted theft from person (64); Theft from a vehicle (103); Fraud - virus (130); Damage to personal belongings (42);Damage to a vehicle (176); Attempted fraud - confidence (537); Theft from a dwelling (49); Theft from person (44); Theft from outside a dwelling (196); Theft away from home (81); Harassment (514); Attempted assault (125); Fraud - theft of personal information (163); Attempted theft away from home (42); Attempted theft from outside a dwelling (62); Attempted theft from a vehicle (51); Attempted theft from a dwelling (32)

The screeners most likely to be retained in isolation were 'Attempted bike theft' (32%) and 'Fraud – non-confidence' (26%). While the base size for 'Attempted bike theft' is too small to allow for more detailed analysis, it is possible to look at which screeners were most commonly deselected by the 'Fraud – non-confidence' group.

Perhaps unsurprisingly, the other screeners most likely to have been selected by this group of respondents at the pre-review stage were fraud screeners. Among respondents who initially selected 'Fraud – non-confidence' and at least one other screener (but ultimately retained only 'Fraud – non-confidence'):

- 63% also selected ‘Fraud – confidence’ at the pre-review stage
- 55% also selected ‘Attempted fraud – confidence’ at the pre-review stage
- 27% also selected ‘Fraud – theft of personal information’ at the pre-review stage

The figures suggest that some respondents found it difficult to differentiate between different types of fraud, leading to double counting. While this could be rectified through the review process, the added respondent burden would likely have an impact on the respondent experience, increasing the length of the survey and making it more complicated to complete. It is also worth noting that there may well be an order effect at play here. The fact that ‘Fraud – non-confidence’ was retained in these cases could be due to it being the first fraud screener to be asked.

7.2 Victim modules

Section 4.1 describes how the current CSEW records participants’ experience of crime via a series of crime screener questions and follow-up victimisation modules based on a set of rules which seek to minimise respondent burden by placing a limit on the maximum number of victimisation modules completed by any single respondent while also prioritising more serious crimes over less serious ones.

As in the main CSEW, a cap of six victim modules was also used on the live trial. While the prioritisation rules used (see section 4.5) were slightly different to those used on the CSEW the intention was to achieve the same purpose of minimising respondent burden and prioritising more serious crimes over less serious ones.

After the completion of fieldwork, all victimisation modules were reviewed by the same team of specially trained coders who work on the main CSEW to determine whether what had been reported represented a crime or not and, if so, what offence code should be assigned to the crime. Unlike the main CSEW, no additional checking was undertaken to quality assure the codes derived by the coders and supervisors.

7.2.1 Number of victim modules completed

Table 7.1 below shows that a total of 3,739 victimisation modules were completed by 2,213 online and telephone respondents in the randomised controlled trial (RCT), with 44% of RCT online respondents and 40% of RCT telephone respondents completing at least one victim module. Online respondents were also slightly more likely than telephone respondents to complete more than victim module: 18% of RCT online respondents completed two or more modules compared with 15% of RCT telephone respondents. However, the proportion of highly victimised respondents (those completing five or six victim modules) was the same in both groups (1.4% and 1.2% respectively).

Table 7.1 Number of victimisation modules completed						
	RCT online respondents			**RCT telephone respondents**		
	N	% of all respondents	% of all victims	N	% of all respondents	% of all victims
Non victims	*2,364*	56		*642*	60	
Victims[48]	*1,783*	44		*430*	40	
No. of victim modules completed						
1	*1,049*	25	59	*265*	25	62
2	*454*	11	25	*88*	8	20
3	*149*	4	8	*44*	4	10
4	*73*	2	4	*21*	2	5
5	*29*	1	2	*5*	<0.5	1
6	*29*	1	2	*7*	1	2
Total victim modules	*3,015*			*724*		
Bases:		*4,147*	*1,783*		*1,072*	*430*

7.2.2 Assessing victim module eligibility

Although online respondents were slightly more likely than telephone respondents to report being a 'victim' of crime neither the responses to the screener questions or the completion of a victim module are used to define a victim of crime in the derivation of the published crime statistics: the definition of a victim is based on having a valid offence code. As is the case in the current CSEW, there are a variety of reasons why a victim module may end up with either an invalid offence code or no offence code. Since the proportion of ineligible victim modules can be seen as a measure of data quality it is useful to examine any differences between online and telephone respondents.

Skipping victim modules

Those who reported an incident of physical or sexual violence, threats or harassment were given the option of skipping the relevant victim module either because recalling details about an incident might cause them distress, or because the respondent might not have sufficient privacy to answer the questions. This skip function is also used in the CSEW survey where incidents of domestic violence or sexual assault can be skipped, either at the request of the respondent or by the interviewer if they judge it is not appropriate to ask the questions.

In the live trial, overall, 22% of respondents who reported one of the incidents mentioned above chose to skip at least one victim module. Online respondents who recorded one of these types of incidents were more likely than telephone respondents to skip at least one victim module (25% and 14% respectively).

Respondents who reported more than one incident of violence or threats may have chosen to skip more than one victim module. In fact, 28% of all victim modules connected to

[48] Victim is used here to refer to a respondent who started at least one victim module.

violent or threatening incidents were skipped by respondents to the online survey compared with 15% of all such modules by respondents to the telephone survey. This represented 8% of *all* victim modules being skipped by online respondents compared with 4% of *all* modules being skipped by telephone respondents.

Table 7.2 below shows that online respondents were more likely than telephone respondents to skip modules for all types of crimes where this was an option. This finding is important because one of the benefits often cited for online self-completion surveys is that respondents may feel more comfortable providing details about sensitive crimes compared with an interviewer administered survey. However, these results suggest that respondents find it easier to skip questions in a self-completion survey when given the option compared with an interviewer-administered survey. If this is true, then an online crime survey may significantly under report sensitive crimes such as physical or sexual violence compared with an interviewer administered survey due to the higher attrition rate.

While it can be hypothesised that a lack of privacy or the sensitivity of the crimes are the reasons for respondents skipping the modules the survey does not actually provide any evidence for why the rates were higher among online respondents compared with telephone respondents. It may simply be that it is easier in a self-completion survey for respondents to skip part of a survey when given the option for reasons which have nothing to do with the content of the questions.

Table 7.2 Proportion of victim modules skipped by type of incident

	RCT online respondents	**RCT telephone respondents**	*Bases: All victim forms for crime type*	
	% of victim modules skipped		*Online*	*Telephone*
(Attempted) physical or sexual assault[49]	32	15	*209*	*52*
Being harassed or intimidated	31	19	*302*	*71*
Being threatened	21	11	*338*	*78*
All violence and threats	**28**	**15**	***849***	***201***

Accuracy of date recall

As noted elsewhere in the report, recall and bounding errors are always a risk as respondents may have a natural tendency to forward telescope incidents[50] from outside the reference period, especially for more salient crimes. As is the case on the CSEW, respondents who reported an incident at the screener question were given the option of recording a date which was more than 12 months ago and therefore out of scope for the survey.

One difference between the current CSEW and the live trial was the positioning of the date questions. In the CSEW respondents are asked about when an incident occurred at the

[49] Due to the small numbers physical assault, attempted physical assault and sexual assault have been combined into a single category

[50] Telescoping is a form of measurement error whereby respondents report events as being longer ago than they were (backward telescoping) or more recent than they were (forward telescoping)

start of each victim module. If they report that it happened more than 12 months ago then the entire module is bypassed. In the redesigned questionnaire respondents were asked about when an incident happened at the end of the screener questions. In practical terms this positioning in the questionnaire should make no difference: cases that are out of scope can bypass being asked questions in the victim module. However, in the trial where an incident happened more than 12 months ago the cases were not automatically filtered out and a victim module was still asked. This was to see whether some crimes were more subject to telescoping than others.

Apart from telescoping, another potential concern is the extent to which respondents can tie an incident down to an exact month. Attempts are made by interviewers to try to help respondents recall the exact month an incident happened if they are having trouble recalling this. However, no such assistance is available for respondents in a self-completion survey and so they may be more likely to say they can't remember when an incident occurred.

Table 7.3 below shows that telephone respondents were more likely than online respondents to report an in-scope incident and to give an exact month: 96% of victim modules in the telephone survey were in-scope and the respondent was able to prove the exact month. This compared with 89% of victim modules in the online survey. Online respondents were twice as likely as telephone respondents to not know the exact month an incident happened (7% and 3% respectively). They were also more likely than telephone respondents to record an incident which happened more than 12 months ago (4% and 1% respectively).

Table 7.3 Proportion of victim modules with out of scope or incomplete date information

	RCT online respondents	**RCT telephone respondents**
	% of victim modules	
Incident in scope (exact month known)	89	96
Incident in scope (exact month not known)	7	3
Incident out of scope (more than 12 months ago)	4	1
Base: All victim modules	*3,015*	*724*

Table 7.4 shows the quality of the date information by screener type. This shows that the proportion of in-scope victim modules where the exact month was known was similar across different types of crimes ranging from 90% to 92%. The proportion of modules that were out of scope by crime type ranged from 2% of threats or harassment to 6% of vehicle crime and 6% of violence. These findings perhaps suggest that forward telescoping is not strongly connected to the type of crime and so is not linked to especially salient crimes such as violence.

There were no differences between online and telephone respondents by crime type: across all types of crime online respondents were more likely than telephone respondents

to report an incident that was out of scope or to not know the exact date of an in-scope crime.

Table 7.4 Proportion of victim modules with out of scope or incomplete date information by crime type				
	% of incidents in- scope	% of incidents in-scope but month unknown	% of incidents out of scope	*Base: All victim modules for crime type*
Home-based burglary/ theft/damage	90	6	4	*676*
Vehicle	91	3	6	*462*
Personal theft/damage	91	6	3	*205*
Violence	92	3	6	*261*
Threats/harassment	92	6	2	*789*
Fraud and computer misuse	90	7	3	*1,346*

Out of scope offence codes

Although respondents are encouraged to report all incidents at the screener questions, when these are coded at a later stage there are several reasons why an incident might be classed as out of scope other than it being outside of the 12-month reference period. These include:

- Incidents outside the survey's coverage: for example, personal crimes that happened to someone else other than the respondent or incidents of fraud where the respondent was not the specific intended victim.
- Incidents where there is not enough information provided to be sure it reaches the threshold of an offence: for example, possible accidental damage rather than criminal damage; possible lost property rather than theft; disputes over goods or services that could be trading standards issues rather than fraud.
- Incidents which cannot be coded for a variety of other reasons: for example, insufficient information provided, duplicate incidents, or incidents recorded under the wrong victim module (i.e. a fraud crime recorded under a non-fraud victim module or vice versa).

Table 7.5 shows that online respondents were more likely than telephone respondents to complete a victim module that was later given a code that was out of scope: 24% of all victim modules completed online which were coded (i.e., excluding skipped modules) were given an out-of-scope code compared with 16% of victim modules completed by telephone respondents. The main reason for this difference was the higher proportion of incidents outside the survey's coverage: 16% of all online victim modules and 9% of all telephone modules. Other reasons for codes being out of scope were similar between the two groups.

At the level of individual offence codes there were few differences between the online and telephone survey in terms of what was out of scope. The overall proportion of victim modules coded as duplicates was low for both groups, although slightly higher in the

online survey compared with the telephone survey (1.7% and 0.5% respectively). Duplicate victim modules may be an indication of double counting that was not resolved earlier in the survey.

The most noticeable difference between the online and telephone survey was the proportion of fraud cases that were outside the survey's coverage: 11% of victim modules in the online survey were fraud cases outside the survey's coverage compared with 6% of modules in the telephone survey.

Table 7.5 Proportion of victim modules with out of scope offence codes

	RCT online respondents	**RCT telephone respondents**
	% of victim modules with out of scope codes	
Outside survey's coverage	16	9
Insufficient information to determine an offence	3	4
Other reason	5	4
All out of scope	**24**	**16**
Base: All victim modules (exc. skipped modules)	*2,795*	*699*

Summary of victim module eligibility

The analysis above has highlighted that there are several reasons why a victim module may not generate an eligible offence code and so contribute towards victimisation rates: because it is skipped, because it is outside of the reference period, or because it results in an offence code that is out of scope.

Taking all these different reasons together a third (33%) of all victim modules generated at the screener stage by online respondents ended up as ineligible for one reason or another compared with 21% of victim modules generated at the screener stage by telephone respondents.

Table 7.6 shows that more ineligible victim modules were generated in the online survey compared with the telephone survey across all crime types. There was considerable variation by crime type in the proportion of ineligible victim modules among both groups with fraud, violence and threats and harassment having higher levels of ineligibility compared to home-based theft and damage or personal theft and damage. However, the figures for violence and threats and harassment are inflated by the fact that respondents could skip these modules which was not the case for other crime types. When skipped modules are excluded the proportion of ineligible modules for both violence and threats/harassment drops significantly, both among online and telephone respondents.

Table 7.6 Proportion of ineligible victim modules by crime type				
	RCT online respondents	**RCT telephone respondents**	*Base: All victim modules for crime type*	
	% of ineligible victim modules		*Online*	*Telephone*
Home-based burglary/ theft/damage	27	23	*523*	*153*
Vehicle	26	15	*383*	*79*
Personal theft/damage	17	4	*166*	*39*
Violence[51]	38 (9)	18 (3)	*209*	*52*
Threats/harassment	35 (12)	21 (7)	*640*	*149*
Fraud and computer misuse	40	24	*1,094*	*252*
All incidents	**33**	**21**	*3,015*	*724*

7.2.3 Victimisation rates based on different definitions

As already noted, in the published statistics a 'victim' of crime is not based on reporting an incident at a screener question or completing a victim module but on having a valid offence code which feeds into the calculation of prevalence rates. It is therefore useful to examine how the victimisation rates change based on different definitions: reporting an incident at a screener question v. reporting an incident which ends up with an eligible offence code

Table 7.7 replicates the figures in Table 7.1 but based only on eligible victim modules rather than all victim modules. Based on this definition a total of 2,596 victim modules were completed by 1,611 online and telephone respondents in the randomised controlled trial (RCT), with 30% of online respondents and 33% of telephone respondents completing at least one eligible victim module.

The proportion of online respondents classified as a 'victim' fell from 44% based on responses to the screener questions (post-review) (Table 7.1) to 30% based only on respondents who reported an incident which resulted in an eligible offence code (Table 7.7). For telephone respondents the drop was from 40% to 33%. Based on this definition telephone respondents were more likely than online respondents to complete at least one eligible victim module and so be a 'victim'. This represents a reversal of the situation based on responses to the screener questions where online respondents were more likely than telephone respondents to be a 'victim'.

[51] Figures in brackets are the percentages of ineligible modules when skipped modules are excluded

Table 7.7 Number of eligible victimisation modules completed						
	RCT online respondents			**RCT telephone respondents**		
	N	% of all respondents	% of all victims	N	% of all respondents	% of all victims
Non victims	*2,885*	70		*723*	67	
Victims[52]	*1,262*	30		*349*	33	
No. of victim modules completed						
1	*796*	19	64	*224*	21	64
2	*300*	7	24	*68*	6	19
3	*94*	2	7	*30*	3	9
4	*33*	1	3	*17*	2	5
5	*22*	1	2	*4*	<0.5	1
6	*17*	<0.5	1	*6*	1	2
Total eligible victim modules	*2,022*			*574*		
Bases:		*4,147*	*1,262*		*1,072*	*349*

It is often hypothesised that respondents are more likely to report incidents in a self-completion survey compared with an interviewer administered survey. This is backed up by the international evidence where crime surveys which have switched modes from interviewer administered to self-completion have generally seen an increase in victimisation rates (see section 3.4). However, no other crime survey codes incidents in the same way as the CSEW and this analysis suggests that although respondents may initially report more incidents in an online self-completion survey this will not necessarily translate into higher victimisation rates once cases are coded.

While filtering out incidents as ineligible at a post-fieldwork coding stage provides a check on the quality of the data and should ensure any published statistics are as accurate as possible this approach does come at a cost. It places a considerable burden on respondents who complete victim modules which are then deemed to be ineligible. The fact that such a high proportion of victim modules were ultimately classed as ineligible in both modes suggests the survey instrument requires further refinement to try and better steer respondents to only reporting relevant incidents. It also highlights the fact that if respondents are given an option to bypass certain modules in a self-completion survey this could lead to significant loss of information. Balancing this loss of information against privacy and safety concerns will be an important issue for any future online crime survey.

[52] Victim is used here to refer to a respondent who started at least one victim module.

7.2.4 Association between screener questions and offence codes

Another measure of how well the survey instrument worked is to examine the association between the screener question from which a victim module was generated and the offence code attached to that module once all ineligible modules have been excluded.

Table 7.8 shows the mapping between crime types based on the screener questions and crime types based on offence codes. It is worth noting that while some crimes (such as vehicle crime or fraud and computer misuse) are relatively easy to map, other crimes are not as straightforward: for example, robbery codes have been mapped as personal theft/damage although robbery is considered a violent crime.

Table 7.8 Mapping between screener questions and offence codes[53]		
Screener question	**Crime type**	**Offence codes[54]**
Break in	Home-based burglary/theft/ damage	50,51,52,53,54,55,56,57, 58,**59**,65,66,73,80,83,84
Attempted break in		
Theft from dwelling		
Attempted theft from dwelling		
Damage to dwelling		
Theft from outside a dwelling		
Attempted theft from outside a dwelling		
Vehicle theft	Vehicle crime	60,61,62,63,64,71,72,81,82
Attempted vehicle theft		
Theft from a vehicle		
Attempted theft from a vehicle		
Vehicle damage		
Bike theft		
Attempted bike theft		
Theft from a person	Personal theft/damage	41,42,43,44,45,**48**,**49**,67, **68**,**69**,85,86,**87**,**88**,**89**
Attempted theft from a person		
Theft away from the home		
Attempted theft away from the home		
Damage to personal property		
Physical assault	Violence	11,12,13,**19**,21,31,32,33, 34,35,**39**
Sexual assault		
Attempted physical assault		
Threat	Threats/ harassment	91,92,93,94,**95**,**97**
Harassment		
Non-confidence fraud Confidence fraud	Fraud and computer misuse	200,201,202,203,204,205, 206,207,208,210,211,212, **219**,320,321,322,323,324, **329**
Attempted confidence fraud		
Theft of personal information		
Virus		
Other ineligible codes	n/a	**1**,**2**,**4**,**5**,**96**

Table 7.9 shows that across all crime types there was a high degree of association between the screener question where an incident was recorded and the final offence code given to the incident. This association held for both the online and telephone survey

[53] All offence codes shown in bold are ineligible codes

[54] For a full list of offence codes used on CSEW see Appendix I in: https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/crimeandjustice/methodologies/crimeandjusticemethodology/202021csewtechnicalreportvolume2v1.pdf

although it is noticeable that for every crime type there was a higher matching rate in the telephone survey than the online survey. Thus, for example, 84% of incidents in the online survey recorded at a vehicle screener question ended up with an offence code related to vehicle crime compared with 90% in the telephone survey. For personal theft and damage crimes, 79% of incidents in the online survey recorded at a personal theft or damage screener question ended up with a similar offence code. For the telephone survey the equivalent figure was 87%.

The crime type with the lowest matching rate was violence for both the online and telephone surveys: 77% of incidents in the online survey recorded at a violence screener question ended up with a similar code, while 18% were coded as threats or harassment. For the telephone survey, the equivalent figures were 81% and 15%. This perhaps suggests that, irrespective of mode, some ambiguity remains in the wording of the screener questions to enable respondents to clearly distinguish violence from threats or harassment.

Table 7.9: Association between crime type recorded at screener question and final offence code

		Crime type based on screener questions									
		Home-based/ burglary/ theft/ damage		Vehicle crime		Personal theft/ damage		Violence		Threats/ harassment	
		RCT Online	RCT Telephone	RCT Online	RCT Telephone	RCT Online	RCT Telephone	RCT Online	RCT Telephone	RCT Online	RCT Telephone
Crime type based on offence code		%	%	%	%	%	%	%	%	%	%
	Home-based/ burglary/ theft/damage	87	93	11	7	9	5	3	4	3	1
	Vehicle crime	7	3	84	90	5	2	-	-	<0.5	-
	Personal theft/damage	4	4	3	-	79	87	3	-	2	-
	Violence	<0.5	-	<0.5	-	2	3	77	81	6	8
	Threats/ harassment	2	-	1	3	5	2	18	15	89	91
	Base: All eligible victim modules	*383*	*120*	*287*	*66*	*139*	*37*	*133*	*44*	*427*	*119*

Note that fraud and computer misuse crimes are not shown in the table above because all eligible fraud and computer misuse offence codes screeners are associated with a fraud and computer misuse screener (i.e., a 100% association). However, it is possible to look at how the individual fraud and computer misuse screener questions are related to the fraud and computer misuse offence codes.

Table 7.10 shows the association between the screener question where an incident was recorded and the final offence code given to the incident for eligible fraud and computer misuse cases. Due to small base sizes the numbers are not shown split by online and telephone survey although there were no obvious differences between the two groups.

This table illustrates that the association between the fraud screener questions and what offences they are coded to is not as strong as for non-fraud crimes. This is perhaps not surprising for two reasons. First, the mapping itself is not so clear cut: while the screener questions try to distinguish between confidence and non-confidence frauds, offences are coded by type of fraud such as bank and credit card fraud (primarily a non-confidence fraud), advance fee fraud (primarily a confidence fraud which includes a lot of attempted frauds), and consumer and retail fraud (a mix of both). Secondly, it is often difficult for respondents to distinguish between different types of fraud in the same way as they can distinguish between other types of crime: for example, it is difficult to clearly explain in a single screener question the difference between having your personal information used to obtain money or goods (non-confidence fraud) and someone deceiving you out of money or goods (confidence fraud).

Table 7.10: Association between fraud crime type recorded at screener question and final offence code

		Fraud type based on screener question				
		Non-confidence fraud	Confidence fraud	Attempted confidence fraud	Theft of personal information	Virus
		%	%	%	%	%
Fraud type based on offence code	Bank and retail fraud	67	20	15	6	4
	Advance fee fraud	6	6	37	4	2
	Consumer and retail fraud	17	69	29	8	6
	Other fraud	6	6	7	1	-
	Hacking or theft of personal information	3	-	7	77	10
	Virus	1	-	6	4	77
	Base: All eligible victim modules (RCT online and telephone combined)	*547*	*237*	*231*	*142*	*141*

7.3 Telephone follow-up

As explained in section 6.3, a subset of respondents who reported experiences of crime in the online survey were subsequently invited to take part in a telephone interview, which consisted of the same questions being asked.

The purpose of this part of the study was to help us understand the individual-level overlap in what would be reported in an online survey and what would be reported in a telephone interview, given a fixed experience of crime. The much more substantial RCT data is perfect for assessing the *net* effects of using a different survey mode but cannot tell us anything about the strength of the correlation between the two. Although this sample is smaller and less representative, it can do that.

In total, 436 of those allocated to the RCT Online group completed both an online survey first and a telephone interview second. Only those who had completed at least one victim module in the online survey were put forward for a telephone interview. As noted in section 6.5.3, this sample of 436 should be treated as a good quality *convenience sample* rather than a probabilistic sample from a meaningful population. Nevertheless, it can provide indicative information about the cross-mode sensitivity of the questionnaire.

There are two principal statistics of interest:

$P(T_s=1|O_s=1)$

$P(O_s=1|T_s=1)$

T_s = telephone screener question *s* (response options: 1 = yes, 0 = no)

O_s = online screener question *s* (response options: 1 = yes, 0 = no)

$P(T_s=1|O_s=1)$ is the probability that telephone screener question *s* yields the answer 'yes', given that online screener question *s* has yielded the answer 'yes'. There are 29 screener questions.

$P(O_s=1|T_s=1)$ is the probability that online screener question *s* yields the answer 'yes', given that telephone screener question *s* has yielded the answer 'yes'.

For many of the screener questions, the base size is too small for separate analysis but the weighted[55] average value for $P(T_s=1|O_s=1)$ is 47% and for $P(O_s=1|T_s=1)$ it is 59%.

In other words, if the online screener question yielded the answer 'yes', there was a 47% chance that the telephone screener question yielded that answer too. If the telephone screener question yielded the answer 'yes', there was a 59% chance that the online screener question yielded that answer too.

Given the relatively small proportions *overall* answering 'yes' to each screener question (an average of 5% for the 29 telephone screener questions and 6% for the 29 online screener questions), these proportions (47% and 59%) are high, reflecting a modestly strong correlation between the online screener question data and the telephone screener question data. These correlations range from around +0.3 to around +0.7 where the sample size allows a reliable enough calculation.[56]

However, a 'modestly strong correlation' is short of ideal because it raises the potential that each mode produces a different victimisation covariance structure even if it produces

[55] Weighted by √n.

[56] Matthews Correlation Coefficient, calculated where there are at least 30 'yes' responses from at least one of the modes.

similar prevalence and incidence rates for each type of victimisation. To assess this, a correlation matrix was produced for the full set of 29 screener questions, one for online data and one for telephone data. The average absolute difference in the correlation coefficient was 0.13 points on a scale from 0.00 to 1.00, but differences of more than 0.30 points were observed.

Table 7.11 below shows some statistics for the seven screener questions with at least 30 'yes' responses in both modes (i.e. the data is most reliable for these types of victimisation). It illustrates the fact that the between-mode correlations vary substantially by victimisation type. For example, if the online screener question on harassment yielded a 'yes' response, there was only a 29% chance that the telephone screener question yielded that answer too (overall correlation = +0.32). In contrast, if the online screener question on (non-confidence) fraud yielded a 'yes' response, there was a 72% chance that the telephone screener question yielded that answer too (overall correlation = +0.71).

Table 7.11 – Key correlational statistics for the seven most prevalent types of victimisation

Screener	Probability that telephone screener will be 'Yes' if online screener is 'Yes'	Probability that online screener will be 'Yes' if telephone screener is 'Yes'	Probability that telephone screener will be 'Yes'	Probability that online screener will be 'Yes'	Overall correlation between telephone and online responses
Harassment	29%	46%	11%	17%	+0.32
Attempted fraud - confidence	32%	72%	12%	28%	+0.43
Fraud - confidence	52%	62%	8%	10%	+0.57
Theft from outside a dwelling	57%	65%	8%	10%	+0.61
Threat	58%	58%	19%	19%	+0.51
Damage to a dwelling	71%	69%	10%	9%	+0.69
Fraud - non-confidence	72%	76%	17%	17%	+0.71

8. Stage 5: Live trial respondent evaluation, interview length and survey dropout

This chapter covers various aspects of respondent experience of the survey including:

- Evaluation of usability (section 8.1)
- Usability questions – open text responses (section 8.2)
- Questionnaire length (section 8.3)
- Drop out rates (section 8.4)

8.1 Evaluation of usability

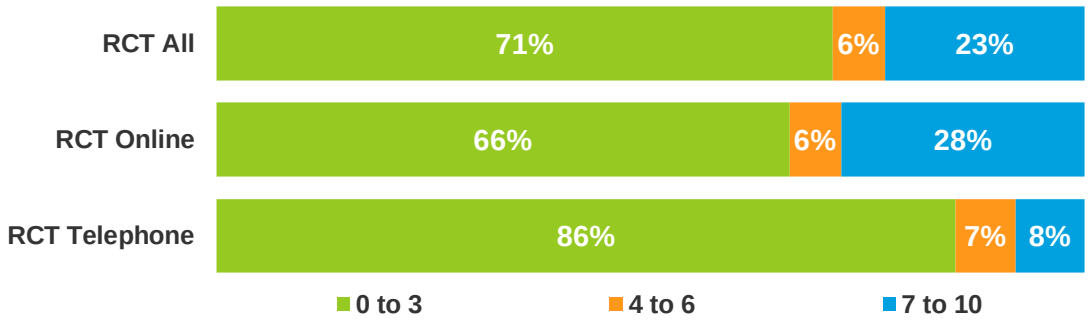
Additional questions were included at the end of questionnaire, asking respondents to give their impressions about general aspects of the survey from a usability perspective. Responses to these questions will help gain some insight on the user experience of the survey, how this varied by mode, as well as complexity of crime experience.

The first question asked respondents how easy or difficult they found completing the survey, selecting their response on a scale from 0 (extremely easy) to 10 (extremely difficult). Ratings were categorised into ‘easy’ (score of 0-3), ‘moderately difficult’ (score of 4-6) and ‘very difficult’ (score 7-10). The results from this question are shown in Figure 8.1.

Overall, seven in ten RCT respondents (71%) gave a score of 0-3, indicating that they had found the questionnaire easy to complete. Conversely, nearly a quarter of respondents (23%) gave a score of 7-10 indicating that they had significant difficulty.

Online survey respondents were more likely than respondents who took part in a telephone interview to find it difficult to complete: 28% of online respondents gave a score of 7-10 compared with 8% of telephone respondents. This may reflect the additional cognitive burden of answering a self-completion survey without the support or guidance of an interviewer.

Figure 8.1 – Level of difficulty experienced when completing the questionnaire (0-10 Scale)



WPEASE. On a scale of 0 to 10, where 0 is extremely easy and 10 is extremely difficult, how easy or difficult did you find completing the questionnaire?

Base: All RCT respondents (5,219); RCT online respondents (4,147); RCT telephone respondents (1,072)

However, when asked whether they had experienced a range of specific issues in completing the survey, telephone respondents were generally more likely than online respondents to say that they had (Figure 8.2). Most notably, a third of telephone respondents (33%) said that the questions felt repetitive (compared with 18% of online respondents) and a similar proportion (32%) said that the questions didn't fit with their experiences (compared with 13% of online respondents). Telephone respondents were also more likely than online respondents to say they found the survey too long (10% and 3% respectively).

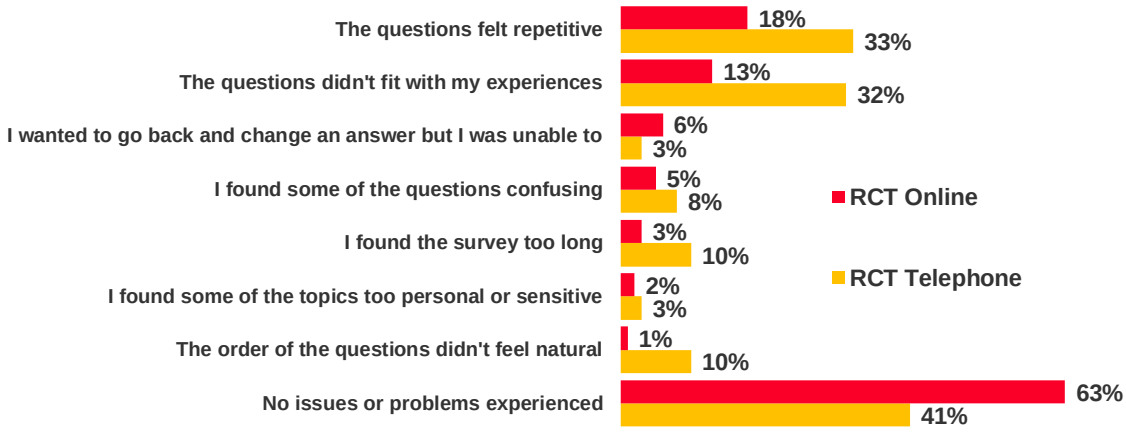
Some of these differences may partly be explained by the way in which online surveys and telephone surveys are administered, which is true of all surveys and not specific to this one. For example, in self-completion surveys, respondents often quickly become familiar with the basic questionnaire structure and will often skim read the questions. By contrast, in telephone surveys where all the questions need to be read aloud the pace of the interview is dictated more by the interviewer rather than the respondent. These differences alone may make a telephone interview seem more repetitive and longer to a respondent compared with the same online survey.

One interesting finding is that a relatively small proportion of respondents felt that the topics were too personal or sensitive with no difference between telephone and online respondents (3% and 2% respectively). Those who reported an incident of either violence or threats and harassment were more likely to feel the topics were too personal or sensitive: 10% of those reporting violence and 7% of those reporting threats or harassment felt this to be the case. Again, there was no difference between online and telephone respondents.

While these findings are relatively encouraging it is important to note that those who reported an incident of violence or threats and harassment were given the option of not completing a victim module. Overall, 22% of respondents who reported one of these incidents chose to skip the victim module. However, online respondents who recorded violent or threatening incidents were more likely than telephone respondents to skip the victim module (25% and 14% respectively). This highlights a potential weakness of an online self-completion approach which is that it is easier for respondents to miss our

questions or topics compared with an interviewer-administered survey (see section 7.2.2 for further discussion on this).

Figure 8.2 – Issues experienced when completing the survey



WPISSUES. Did you experience any of the following issues when completing the survey?

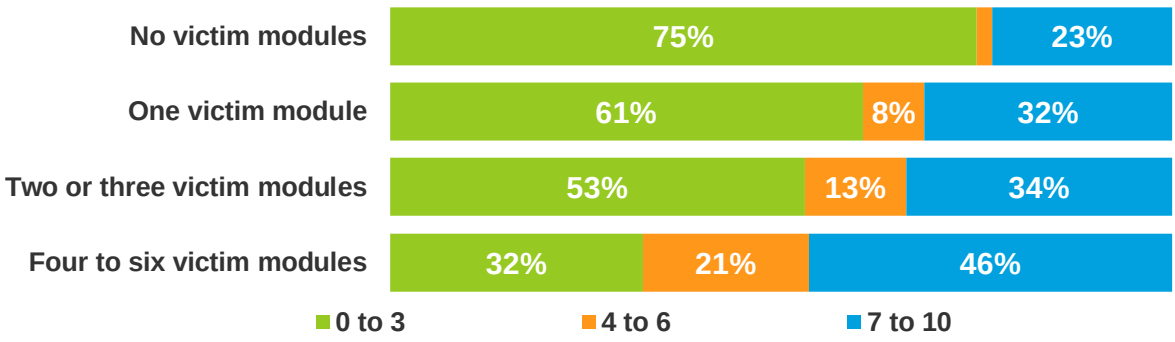
Base: All RCT online respondents (4,147); All RCT telephone respondents (1,072)

By analysing responses to the online survey in isolation (including respondents from the online boost sample), it is possible to explore responses to these questions in more detail.

Perhaps unsurprisingly, the level of difficulty experienced varied depending on the number of victim modules completed. (Figure 8.3). Among online respondents who had not completed any victim modules, three-quarters (75%) gave a score of 0-3 suggesting they found it relatively easy to complete while around a quarter (23%) gave a score of 7-10 suggesting they had found it difficult to complete. The proportion of online respondents who rated the survey as 'easy' declined with increasing victim modules completed: 61% of those completing one module found it 'easy' compared with 53% of those completing two or three modules, and only 32% of those completing four to six modules. Respondents completing four to six modules were more likely to rate the survey as difficult: 46% giving a score of 7-10.

Given the association between number of victim modules and overall interview length it is difficult to be sure whether those with multiple victim modules (especially four or more) find the survey difficult because they cannot easily distinguish between different incidents and so find the questions more challenging to answer or whether the greater perceived difficulty is more a function of interview length which leads to cognitive fatigue.

Figure 8.3 - Level of difficulty experienced when completing the ONLINE questionnaire by number of victim modules completed (0-10 Scale)

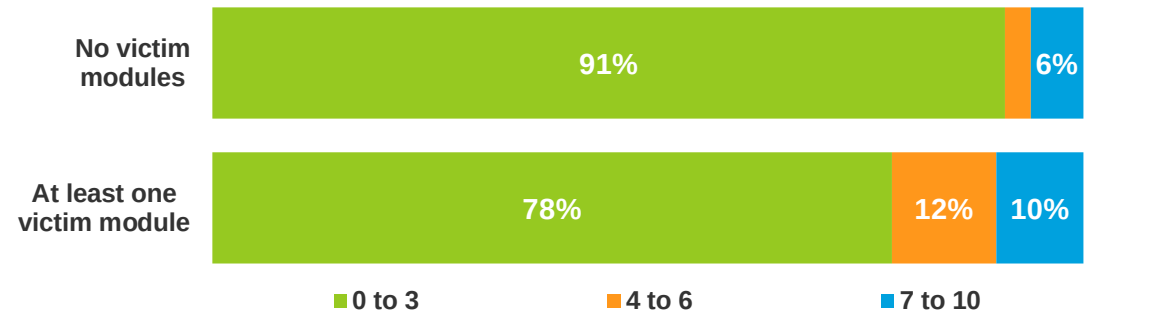


WPEASE. On a scale of 0 to 10, where 0 is extremely easy and 10 is extremely difficult, how easy or difficult did you find completing the questionnaire?

Base: All online respondents (RCT online + online boost) – No victim modules (3,547); One victim module (1,569); Two or three victim modules (900); Four to six victim modules (203)

A similar picture emerges among RCT telephone respondents (Figure 8.4), albeit this group were more likely than online respondents to have found the survey easy to complete. Among respondents who had not completed any victim modules, nine in ten (91%) gave a score of 0-3 while less than one in ten (6%) gave a score of 7-10. Among respondents who had completed at least one victim module, just over three quarters (78%) gave a score of 0-3 while one in ten (10%) gave a score of 7-10.

Figure 8.4 – Level of difficulty experienced when completing the TELEPHONE survey by whether completed any victim modules (0-10 scale)



WPEASE. On a scale of 0 to 10, where 0 is extremely easy and 10 is extremely difficult, how easy or difficult did you find completing the questionnaire?

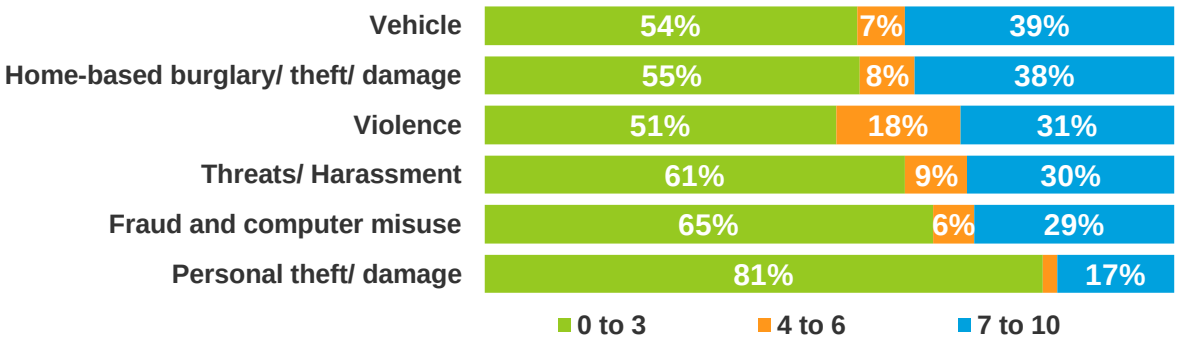
Base: All RCT telephone respondents – No victim modules (642); At least one victim module (430)

The type of incident experienced also appeared to have some impact on the level of difficulty experienced in completing the survey (Figure 8.5). Among online respondents who completed only one victim module, those who experienced personal theft or damage were the most likely to feel that the survey was easy to complete: 81% found it easy compared with 17% who found it difficult. Conversely, those who had experienced a home-based incident or vehicle incident were more likely to find completing the survey difficult:

38% of those experiencing one home-based incident and 39% of those experiencing one vehicle incident found the survey difficult to complete. Among those who experienced violence, 31% said they found the survey difficult to complete.

Again, further investigation would be needed to be sure what is driving these differences. Respondents may find household crimes more difficult because they are less aware of the details of the incident if they have not been personally involved (e.g. if they are not the actual owner of a vehicle involved in an incident). However, with violence the reasons for finding the survey difficult to complete might be related to the sensitivity of the incident or it may be related to the fact that violence (especially domestic violence) is often part of a series of incidents. There is some suggestion from the data that respondents found completing a victim module involving a series of similar incidents (but represented by the most recent incident in the series) more difficult than a victim module for a single incident.

Figure 8.5 – Level of difficulty experienced when completing the ONLINE survey by type of incident experienced (0-10 scale) [57]



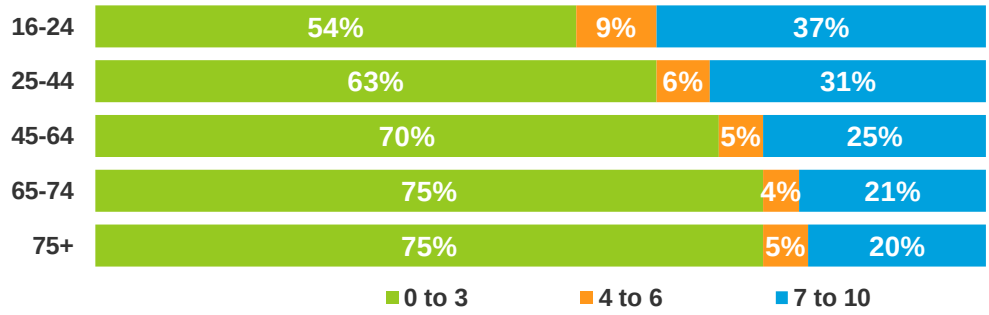
WPEASE. On a scale of 0 to 10, where 0 is extremely easy and 10 is extremely difficult, how easy or difficult did you find completing the questionnaire?

Base: All online respondents (RCT online + online boost) who completed one victim module only – Vehicle (218); Household (247); Violence (72); Threats/ Harassment (271); Fraud and computer misuse (700); Personal (61)

Perceived level of difficulty also varied by different age groups (Figure 8.6), with younger respondents more likely than older respondents to have found the online survey difficult to complete. This may be linked to the factors discussed above: for example, young people are less likely to be householders and so may find it more difficult to answer questions on certain types of crime or they may experience more incidents and so have multiple victim modules to complete. Given these findings it seems unlikely that a lack of digital skills was a reason why respondents found the online survey difficult.

[57] Incident types have been grouped as follows: Personal (Theft/ Attempted theft from person, Theft/ Attempted theft away from home, Damage to personal belongings); Fraud (Non-confidence fraud, Confidence fraud/ Attempted confidence fraud, Theft of personal information, Virus); Threats/ Harassment (Threat, Harassment); Violence (Sexual assault, Non-sexual assault, Attempted assault); Household (Burglary, Attempted burglary, Theft/ Attempted theft from a dwelling, Damage to a dwelling, Theft/ Attempted theft from a dwelling); Vehicle (Theft/ Attempted theft of a vehicle, Theft/ Attempted theft FROM a vehicle, Damage to a vehicle, Theft/ Attempted theft of a bike).

Figure 8.6 - Level of difficulty experienced when completing the ONLINE survey by age (0-10 Scale)

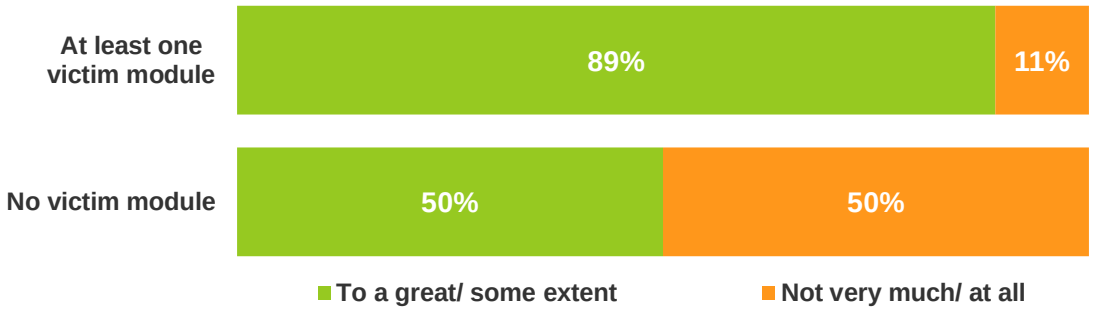


WPEASE. On a scale of 0 to 10, where 0 is extremely easy and 10 is extremely difficult, how easy or difficult did you find completing the questionnaire?

Base: All online respondents (RCT online + online boost) – 16-24 (482); 25-44 (2,177); 45-64 (2,025); 65-74 (1,044); 75+ (468)

Overall, two-thirds of respondents (67%) felt that the questions asked were relevant to their circumstances to at least some extent (Figure 8.7). Respondents who had completed at least one victim module were much more likely to feel this way (89%) compared with those who had not completed any victim modules (50%).

Figure 8.7 - Extent to which the ONLINE questions were relevant to your circumstances by whether completed any victim modules

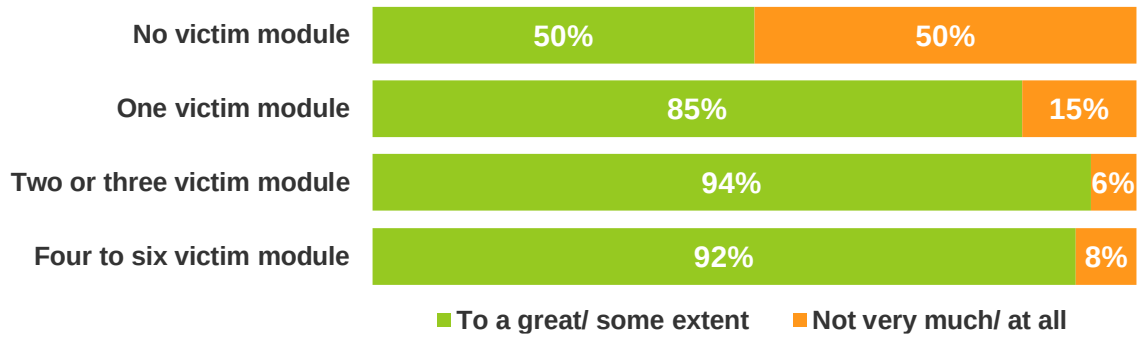


WPREL. To what extent do you feel the questions were relevant to your circumstances?

Base: All online respondents (RCT online + online boost) – At least one victim module (2,672); No victim modules (3,547)

Respondents who had completed two or more victim modules were more likely than those who only completed one victim module to feel that the questions were relevant to their circumstances (Figure 8.8).

Figure 8.8 - Extent to which the ONLINE questions were relevant to your circumstances, by number of victim modules completed

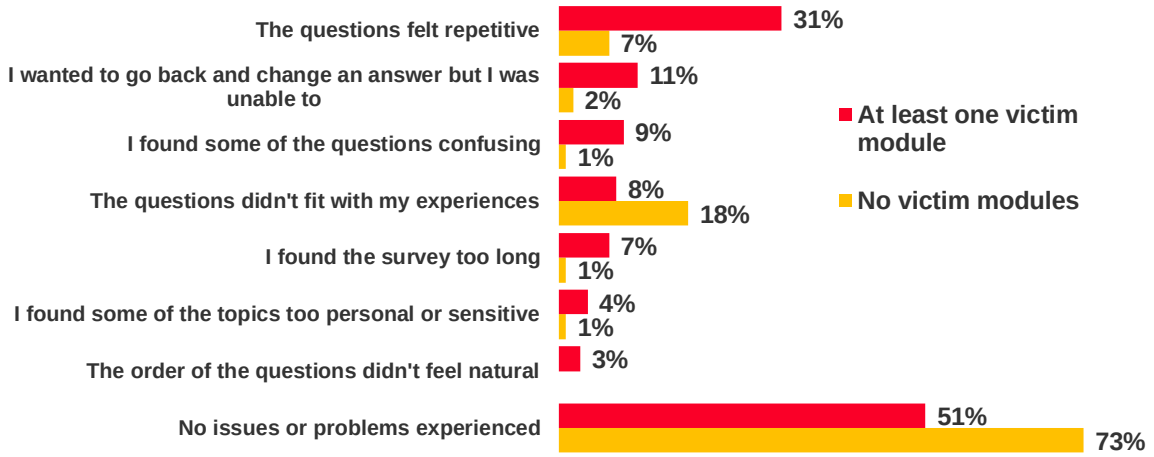


WPREL. To what extent do you feel the questions were relevant to your circumstances?

Base: All online respondents (RCT online + online boost) – No victim modules (3,547); One victim module (1,569); Two or three victim modules (900); Four to six victim modules (203)

Respondents who had completed at least one victim module were also more likely to say that they had experienced specific issues when completing the survey (Figure 8.9).

Figure 8.9 - Issues experienced when completing the ONLINE survey by whether completed any victim modules



WPISSUES. Did you experience any of the following issues when completing the survey?

Base: All online respondents (RCT online + online boost) – At least one victim module (2,672); No victim modules (3,547)

8.2 Usability questions – open text responses

Where participants had selected one or more issues at **WPISSUES**, they were asked to expand on these in their own words. Online respondents were routed to an open text box and asked to give further detail on the issues they had experienced when completing the survey. The wording is shown below:

WPOPEN **[ASK IF WPISSUES=ANY PROBLEMS]**

Please can you briefly tell [CAWI: us] [CATI: me] in your own words about some of the issues or problems you found when completing the survey?

Although open responses tended to be more detailed among telephone participants, for both modes responses generally mapped across to the eight answer categories at **WPISSUES** with some additional comments about the practicalities or technical aspects of completing the questionnaire. Below are some illustrative examples of the kinds of comments participants gave.

The questions felt repetitive

Participants commented that the questions felt repetitive, or that the topic or incident had already been covered and they were going over it repeatedly. Some participants said they understood why the repetition was needed but it meant the survey took longer overall and that there was scope to streamline or for questions to be skipped. Examples of comments participants made are shown below.

As a lot of the questions were on a similar theme, there was a lot of repetition (Online).

The question threads following each incident were long and repetitive, finished each set of questions feeling like I had not been entirely clear (Online).

I felt a little under pressure to report all experiences however trivial (unsettling incidents but no loss or damage to report), but was then asked the same questions repeatedly, and wish I'd just not mentioned them at all (Online).

I didn't think it was necessary to keep repeating the date all through. You could have just said all these questions are since 1st April 2021 (Telephone).

I suppose because I hadn't been affected by crime, in some ways, I didn't feel I had much to contribute. Some of it felt too repetitive, even though they were slightly different issues but to me they were similar (Telephone).

Some telephone participants found the paired screener wording repetitive where the wording was repeated for **actual and attempted incidents**, with confusion noted by one around the 'succeed' and 'didn't succeed' wording. For example, at the DWELTHEFT screener the actual wording is 'Someone with permission to be there stole from your home' and the attempted wording 'Someone with permission to be there tried to steal from your home but didn't succeed'.

I think it's just when you said the first thing, and then read the question again but didn't succeed. Repetitive (Telephone).

The questions did not fit with the participants' experiences

Participants within both modes commented that, if they hadn't experienced any crime over the last 12 months, the questions didn't always feel relevant. Some participants commented on the level of crime in their area being low for various reasons. While this

lack of relevance wasn’t seen as a problem, it could mean that participants felt their responses were not useful to us. Examples of comments are show below.

> *Just repetitive and not very conducive to my experience (Online).*
>
> *No problems, the questions were just not relevant to any recent experiences I've had. (Online)*
>
> *Some of the questions I felt were aimed at more serious crimes which thankfully I have not experienced. Overall though survey was concise easy and simple to complete (Online).*
>
> *I don't think I was able to help you with a lot of the questions There were only some where I had had experiences that were relevant (Telephone).*
>
> *As a 16 year old I don't think some of the questions were quite relevant to me (Telephone).*
>
> *Time frame of the survey – i.e. last 12 months meant it did not fit my experiences of crimes (Telephone)*

Some participants also commented at the end of the questionnaire, that their experiences hadn’t been covered within the questionnaire.

> *I just believe that my personal experiences did not really fit the survey (Online).*
>
> *In one section none of the three possible answers matched my experience (Online).*
>
> *Some of the questions had specific scenarios but didn't cover everything (Telephone).*
>
> *[There] needs to be questions about someone enter[ing] your surrounding property (Telephone).*

Participants could be surprised that the scenarios also covered relatively minor incidents that wouldn’t necessarily be reported to police and wondered about the value of these within the research as shown by the comment below.

Because the incidents reported were minor and ones which happened on the internet, I'm not convinced in what way they might contribute to research. I would also say the questions on whether incidents [were] reported to the police, got me thinking on whether I should of reported it. It would be good for the public to know whether incidents like that should be reported to the police (Telephone).

There was also a feeling among participants within both modes that experiences were more complex than the instrument allowed and in relation to the screeners in particular, that what had happened was too nuanced for a simple Yes/No answer.

[I] think open questions would've been easier than yes or no for a lot of it - it's also hard to identify what's a threat a lot of the time (Online).

The questions asked were difficult to answer yes or no only to, or they were a close fit but not quite the answer I wanted to give (Online).

I think some of the answers where I have to just say yes or no but they are not actually yes or no answers (Telephone)

[The questionnaire] did not allow enough room or opportunity for more explanation (Telephone)

There were also topics participants would have liked to see included in the questionnaire; these tended to be focused on general crime rather than personal victimisation.

Boring subject so a bit tedious to complete. no mention about fly tipping, anti-social behaviour, littering, which also concern me (Online).

The area where I live is plagued by antisocial behaviour which is not addressed by regular street policing. This contributes to general property damage within our village on a weekly basis. Though we have not experienced a personal incident in the last year, there has been damage and vandalism to the green/council property just outside of our home...I don't think the questions - as asked in this survey - provide the full picture of local and personal crime (Online).

I live in [area of London where] we have Anti-Social Behaviour with drug dealers. It's a blight everywhere. The survey didn't touch on that. Although you did ask about vandalism, people on the estate selling drugs is also ASB (Telephone).

[I] would have liked questions to be expanded to include questions about public risk for example dangerous areas to walk, speeding cyclists (Telephone).

Among participants who recognised that they should limit their answers to their own personal experiences, there were some comments that they would have liked to include incidents that had happened to family or other members of their household but which felt relevant to them, especially where the incident had been quite serious.

> *Things my family experienced would be good on the questionnaire - knife crime anti-social behaviour and bike theft (Telephone).*

With the fraud questions there were some comments that these did not fit with what the respondent had experienced as illustrated by the comments below.

> *Questions on fraud and deception were solely concentrating on pure web or phone scams (Online).*
>
> *As most people I regularly receive scam emails, texts and phone calls which I ignore and indeed said so halfway through the survey hence the remaining questions seemed pointless and trivial (Online).*
>
> *I didn't actually have any money taken as I was aware of the scam. None of the questions asked that (Online).*
>
> *It interpreted it as paying money for a service I didn't receive so not a fraud but a problem with their system (Telephone).*
>
> *It's hard to define what damage has been done by me opening an email (Telephone).*
>
> *A telephone interviewer had noted: separate out scam emails from normal cybercrimes; sees scam emails as not much of a crime and kind of normal (Telephone)*

The order of the questions did not feel natural

There were comments from both online and telephone participants that the question order felt illogical as illustrated by the comments below.

> *Felt out of order, when more than 1 situation applies (Online).*
>
> *Questions did not seem to follow logical order with some repetition (Online).*
>
> *Questions were sometimes confusing and did not fall in the order of what happened to me. That said I was okay to answer them (Online).*
>
> *It was a little odd to be asked about all the types of crime then go back to answer detail about each one. It would have probably felt more natural to answer the questions about each incident at the time of recording that it happened (Online).*

A further example of a comment that fell under this category related to wanting to describe the incident in their own words and the narrative be pieced together afterwards in the format required. This fits with the idea of fitting the questions to participants' mental models primarily and making sense of /organising the data during the 'back end' processing stage.

> *Could have been described what happened and after the call, answers could have been put together (Telephone)*

Some of the questions were confusing

There were a variety of reasons participants found some of the questions confusing.

It could be **challenging to recall** whether any of these incidents had taken place over the last 12 months.

> *Trying to remember the answers (specifically thinking back to whether events happened prior or during the time window for the survey (Online).*
>
> *It was a bit difficult to remember things within that time frame (Online).*
>
> *Just trying to remember when things happened was a bit hard (Telephone).*
>
> *I needed to recall all the problems that happened in the last year…prepared questions in advance would help (Telephone).*

Participants for both modes also mentioned the **impact of covid** both on recall over the last 12 months and how it could affect experience of crime.

> *My issues happened at work but felt the survey was aimed at people who were at home. People have changed so much due to Covid (Online).*
>
> *Fortunately I have not experienced any crime, I have not been out much due to Covid. However, I am getting more scam calls, texts and emails (Online).*
>
> *Due to covid, there has been little movement compared to other years and potentially less exposure to crime (Online).*
>
> *I'm just thinking it just goes back from the year and we've only just come out of the pandemic so I've kind of been isolated for most of the time. If it hadn't been for the pandemic, I'd probably have been out more and more people would be out and there'd probably be more crime and maybe I'd have been more subject to crime (Telephone).*
>
> *The survey was only over a year span and where we are only now coming out of Covid. If we had had more time to go out and mix in that time we would probably had a few more things going on (Telephone).*

The **wordiness** of some questions caused issues for some participants

> *Too many wordy questions (Online).*
>
> *Because of the caveats at the beginning and at the end of the questions such as since April 2021 and having only occurred [in] England and Wales this made it difficult to understand the key point of the question at some point[s] (Telephone).*

Telephone participants also commented that it could be **difficult to understand the questions when they were read out** and that they sometimes needed to ask the interviewer to repeat certain questions. Three examples of difficulties that some participants had were:

- The distinction between actual and attempted crime
- How to answer where the intention of the potential offender is unclear, for example:
- Questions that didn't seem to 'learn' from previous answers, linking to the idea of them being repetitive

> *Could have been clearer on actually versus tried (Telephone).*
>
> *The] question asking about 'tried to be rob' for example, it's hard to know whether someone has tried to or not (Telephone)*
>
> *It was the question about the child in the home that was a bit confusing (whether I had any children under 18 that did or didn't live with me when I'd already said I did have a child) (Telephone).*

Some participants found the survey too long

Participants from both online and telephone modes commented that the survey was too long. This was sometimes linked with the repetitive nature of the questions and that the survey could be streamlined overall. Examples of comments participants made are shown below.

> *Survey seemed to me to be far too long. Maybe you could divide it into short – for people who had only 1 or a few small experiences of crime, and long – for people who had many and a range of experiences of crime (Online).*
>
> *The survey is long, confusing, and repetitive (Telephone).*

Some participants found some of the topics too personal or sensitive

There were comments from participants across both modes that fell within this category and some participants commented on feeling pressured or reminded of difficult or negative experiences. There could also be the difficult situation that participants see what has happened to them as 'harassment' or 'assault' for the first time, not having previously

labelled it as such. The questionnaire could also act as a reminder of the dangers around for everyone. Examples of things participants typed or said are shown below.

> *The questions were very personal and was not explained how the information gathered will be used i.e. for what purposes and by whom (Online).*
>
> *There is not sufficient content warning that there will be questions about physical trauma, abuse and/or assault. This warning should occur at the beginning so respondents have opportunity to prepare to take the full survey in a private location. Currently the survey is designed so that respondents would have to move to a secure location midway through survey, with questions/phrases already onscreen that could cause an unsafe situation (Online).*
>
> *The survey asked a whole bunch of personal questions that had no relevance to the later parts of the survey. Going through a whole list of incidents was fine, going back to collect details about the incident just felt strange and disorganised. One follow up question was confusing. Either response could have fitted. Much of the later part of the survey was pointless and asked about stuff I had already answered (Online).*
>
> *I guess some things which are a little bit personal like abuse was a bit too personal – that's about it really (Telephone).*
>
> *I only remembered the incident of harassment and intimidation happened when you mentioned it. It did bring up some issues for me, but I appreciated that I had the option not to answer (Telephone).*
>
> *It reminded me of the dangers and insecurity of the internet reminded vulnerable using the web (Telephone).*

Participants were unable to go back and change an answer

Although surveys would normally allow a respondent to go back and forth during an interview, in the WCSEW prototype this was not possible due to logistical scripting complexities. Aa result, the inability to go back and correct earlier answers was a source of frustration for both online and telephone participants, especially where they realised an earlier error meant they needed to answer a lot more questions. Examples of comments are shown below.

> *I clicked on the wrong button and couldn't see how to correct/go back (Online).*
>
> *After I made a mistake about the answer, I could not go back and this affected all the following questions (Online).*
>
> *I got confused by some of the wording as it didn't quite fit the circumstance and didn't realise I was talking about another incident to the one being asked about and couldn't go back and change it ((Online).*

I can't find a button to go back to previous questions and change my answers (Online).

I said yes to the bicycle question but it actually belongs to someone else in my household and couldn't go back to correct (Telephone).

It was disappointing and frustrating that i couldn't go back and change my answer to one question because I misunderstood it. So it made me lose my concentration (Telephone).

Only where we couldn't go back when I remembered I had had someone try to take money out of my account (Telephone).

To pick up on a specific example in a telephone interview, despite the participant saying that they didn't find any issues during the survey, the interviewer had added the following note.

Respondent misremembered date of incident then realised it was more than a year ago but there was no way of getting out of the section by that point so had to enter random answers to get through it as quickly as possible and move on (Telephone).

Other issues or problems

Other issues that emerged related to the **logistics or practicalities** of completing the questionnaire. **Technical** issues came up in relation to both online and telephone modes: internet connectivity issues for online completion and call or connection quality for telephone. Some telephone participants commented that they don't like using the telephone or that they had hearing issues and so the interview could be challenging. Some online participants commented that they had difficulties with reading (for example dyslexia) or understanding the wording. There were also some comments that completing on a small smartphone screen could be more challenging than using a larger device.

There were also some comments among telephone participants that they would have preferred to complete the survey online.

When you do it online, it's a lot easier, you just look at it and click it. You don't get too much gasbagging in your earhole (Telephone)

Not knowing when the pause will come when to give the answer, because of the way it was, this survey would have been better online to tick boxes. [There is] a lot of info you're giving and people might not take it all in

I would prefer to attempt it online rather than have receiving a call [although] this is a preference not an issue.

The only issue I have is I find it easier reading than being read because it was something on the list which was actually further down on the list.

Finally, it is worth noting that a few participants did give positive feedback about the questionnaire.

> *It was very interesting to talk about these incidents the crime is very rare and its very nice to talk to someone about the crimes happened to me.*

8.3 Questionnaire length

Average (median) questionnaire and section lengths are shown in Table 8.1.

Interestingly, online respondents completed the screener questions in almost half the time of telephone respondents, even though online respondents were more likely to report incidents at the screener stage. While questions will naturally take longer when read out by a telephone interviewer (compared with respondents reading them from a screen), this could also indicate that some respondents are engaging in 'skim' reading, and not taking in the question in full.

Table 8.1 – Average questionnaire length (in minutes & seconds)

Section	Online			Telephone		
	All	Victim module	No victim module	All	Victim module	No victim module
Demographics and vehicle ownership	02:18	02:15	02:20	03:17	03:23	03:11
Non-fraud screeners	03:34	04:10	03:07	06:47	07:53	05:40
Fraud screeners	01:19	01:39	01:04	02:15	02:46	01:44
Victim module 1	-	05:02	-	-	06:40	-
Victim module 2	-	03:45	-	-	05:40	-
Victim module 3	-	04:07	-	-	05:05	-
Victim module 4	-	03:40	-	-	06:09	-
Victim module 5	-	04:13	-	-	06:01	-
Victim module 6	-	04:45	-	-	06:48	-
Usability questions	01:14	01:18	01:11	02:44	02:45	02:44
Total length	**12:05**	**18:02**	**07:44**	**20:53**	**28.20**	**13:19**

8.4 Drop-out rates

An important aspect of the live trial was testing whether an online-led CSEW would result in a higher number of partially completed questionnaires. Self-completion surveys can, if poorly designed, prompt respondents to 'drop-out' mid-way through the questionnaire, limiting the overall response rate and compromising the representativeness of the data. Typically, a respondent will abandon a survey if it is excessively long, insufficiently engaging, includes confusing language/ questions or is generally difficult to complete. All these factors were considered when designing the survey instrument (see chapter 4).

Table 8.2 shows the cumulative number of drop-outs at various points in the questionnaire. The proportion of drop-outs are broadly similar across both modes. Among respondents who started the survey, 2.9% of online respondents and 2.5% of telephone respondents dropped out before the end the questionnaire. It should be noted that, while a drop-out rate of 2.9% is relatively low for an online survey, this may be due, at least in part, to the composition of the sample. Respondents for this survey were recruited from Kantar's Public Voice panel, making them accustomed to taking part in research of this type. A general population survey administered via an ABOS[58] methodology would probably experience a higher drop-out rate.

Table 8.2 – Drop-out rates

Question	***Online***		***Telephone***	
D1 (1st Question)	6,399	0	1,546	0
ZBREAKIN (1st Screener)	6,337	62	1,529	17
ZNONCON (1st Fraud Screener)	6,301	98	1,521	25
WPEASE (1st Usability Question)	6,225	174	1,510	36
ZHELPSUPP1 (Signposting Question)[59]	6,219	180	1,508	38

[58] Address-Based Online Survey.

[59] Respondents were required to reach this question for their interview to be classified as 'Complete'.

9. Stage 6: Follow-up depth interviews

Once the live trial survey was in the field, ten depth interviews were conducted with participants of the Public Voice live trial, in two iterations of five interviews. This chapter outlines the administration of, and findings from, these interviews.

The objective of these interviews was to follow up with 'real life' participants to qualitatively explore their experiences in detail, retrospectively comparing these accounts against the information recorded in the online survey. As such, these interviews fell somewhere between a 'respondent debriefing' interview, picking up on the answers provided in the survey and an exploratory depth interview, further exploring participants' experiences of crime.

9.1 Identifying potential respondents

As with the 'cogability' testing, the focus was on participants who had complex experiences of crime. In identifying potential participants from the live trial, the research team designed a set of parameters that fell into two categories (although these were later adjusted as detailed below):

- **Experienced complex crime:** participants who had recorded multiple instances of crime (across and within crime type), given a 10+, banded or 'don't know' response at a count question or a mix of series and separate incidents, completed 3 or more victim modules
- **Reported difficulty completing the online survey at the usability questions:** participants who gave a high score for level of difficulty and selected three or more types of problems experienced

Taking these factors into account, a flag system was developed which gave each participant an overall score. While participants did not need to meet the full range of requirements, this flag system provided a starting point in targeting a subset of potential participants for depth interviews. For these, all flags were considered holistically, the data for each participant reviewed on a case-by-case basis. Five participants were interviewed remotely in the first round, each receiving a £40 gift voucher as a thank you. The research team also ensured that the interviews reflected a good range of crime types including serious crimes such as assault, threats and harassment as well as some cases of fraud and online crime.

Due to the complex nature of participants' experience, particularly where the count was banded or unknown and there was a lot of overlap within the screeners, it could be challenging to try to untangle what had happened both in preparation for and during the interview. As a result, while again those with a more complex crime profile were targeted, the requirement for number of incidents within crime type was lowered to between 4 and 9 for the second round of interviews so that a broader range of experiences could be reviewed in depth. The requirements for a high difficulty score at the usability questions were also relaxed, as at Round 1 it was found that these weren't always good indicators of complex experience; participants could say they found the questionnaire easy and experienced few issues but there were still a range of inconsistencies or errors throughout the questionnaire which were useful to explore. Some of the more complex indicators arose naturally in the cases flagged but they were not a requirement for their selection.

Table 9.1 shows the full list of flags.

Table 9.1 - Flags used for selection of follow-up depth interviews		
Flag	**Indicators used at Round 1**	**Indicators used at Round 2**
Level of difficulty	7-10 (of 0-10 scale)	
Issues experienced	3 or more issues	
Number of screeners selected	3 or more screeners	3 or more screeners
Mix of crime type	Mix across crime types Non-fraud and Fraud screeners selected	Mix across crime types Non-fraud and Fraud screeners selected
Number of instances within crime type	4-9 incidents at **ZCOUNT** 10+ incidents recorded at **ZMCOUNT** Banded or DK answer given at **ZBCOUNT**	4-9 incidents at **ZCOUNT**
Similar and different crimes within crime type	Mix of series and separate	
Mix of crime across and within type	Multiple crime types and multiple incidents within crime type	
Number of victim modules	3 or more victim modules	3 or more victim modules
Double counting	Coded **ZRELATE** at least once Changed answer at **ZCOUNTCHECK**	

9.2 Interview preparation

Prior to each interview, the researcher reviewed in detail the data collected in the live trial for that participant. To make this task easier, answers to key questions were exported into Excel from the SPSS data. Data exported included answers to all screeners, prioritisation data from the victim module algorithm, the open description at **ZDESCRINC**, key check questions from each 'mini module' plus any additional contextual verbatim data provided at the end of each victim module. While this summary helped in terms of having all open data in one place to review and refer to during the interview, it was still necessary to review each case in the SPSS dataset to fully explore and understand count information and where the respondent may have 'coded out' or amended incidents at **ZRELATE** and **ZREVIEW/ZCOUNTCHECK**. Depending on the complexity of cases this process could be lengthy in untangling what appeared to be the 'story' within the data.

9.3 Depth interview structure

There were four main sections to the interviews:

— Initial 'top of mind' feedback on the online survey.

- An updated version of the mental models exercise included at the outset of the Stage 4 'cogability' interviews where participants were asked to describe their experiences of crime over the last 12 months in their own words. This account was then compared against what had been reported in WCSEW.
- Review of participants' responses to the usability questions in detail, showing examples of the key questions on the screen to help with recall.
- Time permitting, a discussion of some of the key question layouts in the online questionnaire.

The interview topic guide is contained in Appendix 7. The topic guide content became steadily more ambitious throughout the interview; the initial stages were found to be easier (top of mind thoughts and mental models), but it was very challenging for participants to remember how they had answered or what they thought at some of the specific screens. For example, the participant might remember answering the date questions and the layout of the paired screeners but recalling the more complex **ZRELATE** and **ZREVIEW** aspects was much more difficult.

As there was a longer time lag between the second round of depth interviews and the initial completion of the live survey, the structure of the second round interviews was simplified. The review of the usability questions was replaced with a simpler 'catch-all' question asking whether they could remember experiencing any of the key issues on original completion of the survey. The final section on key layouts wasn't included.

Depth Interviews took place remotely in May 2022. Table 9.2 shows key information for participants at both rounds.

Table 9.2 – Profile of depth interview respondents

Indicator/Characteristic	Round 1 (across 5 participants)	Round 2 (across 5 participants)
Number of flags (0-20)	10-14	4-7
Number of screeners (0-29)	6-14	4-7
Number of victim modules (0-6)	2-6	3-6
Male	2	2
Female	3	3
Age range	20-58	37-71

9.3.1 'Top-of-mind' feedback on the online survey

Following initial introductions, the first section asked participants for their **'top of mind'** thoughts about the questionnaire. There were two main themes that emerged, both of which reinforced findings from cogability testing at Stage 4:

Finding the survey repetitive and complex

Many participants felt the questionnaire repeatedly covered the same ground and was inflexible in places. Where experience had been easier to untangle, participants realised how the questions loops were structured and that each referred to a different incident. However, those with more complex experiences, where incidents were more difficult to differentiate, felt they were putting in identical information repeatedly. This related to

repetition both within the screeners and the victim modules where similar, multiple incidents were difficult to separate out. Below are some quotes to illustrate this finding.

> *"Was there a section where I had to put in everything that had happened to me in the last year?…I had to keep doing new ones".*
>
> *"It was very easy and then it steadily...became more and more complex and repetitive…it was just like filling in a standard form…like applying for something…then mid-way through it…suddenly started to become worse and…I started to feel as though I've been answering the same question over and over again and it was quite frustrating".*
>
> *[You go] "over and over and over it...you go into this maze and go round in circles". [The questionnaire] "tries to trick you by asking the same question over and over again".*
>
> *"It was kind of like overlapping; it was hard to put individual incidents in when something has happened and it…connected together and got a bit confusing".*
>
> *"The good points were that I understood that it had to be structured in a certain way because you were looking to obtain some detail. But whilst doing the survey, I felt it was a little bit firm and not as flexible [as I would have liked]. For example, if I selected between 'Yes' and 'No' then I would have thought it may have taken me in a slightly different direction; the next question would have been more connected."*

Being unable to resolve their own errors or amend answers

Participants often recognised that there was a 'pattern' to the survey set up: selecting 'Yes' to a screener meant an extra set of questions came up. Participants said, retrospectively, that if that hadn't answered 'Yes' to as many questions, they wouldn't have had to answer all the extra questions. Participants could find it frustrating that they were unable to go back and amend their answers, particularly where they ended up with multiple victim modules about the same incident.

> *"I shouldn't have said Yes, wouldn't have got some many questions...gone over it so many times".*

Ideally, a 'back' button would have been included in the questionnaire, to allow participants to amend earlier questions or check previous answers. This was not possible for the live trial as navigating backwards and forwards within the instrument would interfere with the victim module prioritisation algorithm and the way the ZRELATE check screens were set up. This is something that would need to be considered in future development of the survey.

Only one specific example of confusion came up during the opening discussion which also echoed findings from cogability testing (see chapter 5). This related to interpretation of the 'theft from outside home' screener (**ZOSTHEFT**) and what this includes.

> *"I wasn't sure whether they meant the plants or the stuff on your doorstep or did they mean people nicking your purse while you're on your way to the supermarket"*

9.3.2 'Mental models' exercise and how participants described their experiences

Next, the interviews moved to the 'Mental models' exercise where, in a comparable way to the cogability interviews, participants were asked to summarise their experiences of crime over the last 12 months.

There were two key strategies that participants used:

'Nature of local crime' approach: In this approach, respondents described the nature of crime occurring in their local area and how it affects them (and sometimes others), in a general way. Respondents did not necessarily restrict their thinking to the last 12 months. This approach was also observed during cogability testing. The types of issues identified included wider societal problems such as anti-social or drug-taking behaviours which, although fit within the context of crime and are included in other CSEW modules, are not included in the screeners as they are not incidents of personal victimisation (and are therefore out of scope). One example of this was a participant who mentioned calling the police as a large group of teenagers were drinking and playing football outside their house in the late evening and had concerns about the ball hitting cars and homes.

Where this approach was taken, the mental models exercise tended to make sense in terms of the 'story' within the live trial data for that case. However, it clarified that, in some cases, where incidents involved multiple crime types, there was repetition in the data across all or some screeners that felt relevant to the incident(s). For example, one participant had experienced ongoing anti-social behaviour within his local area, affecting him and his neighbours with a range of crime types taking place including damage to property (**ZHOMDAM**), theft from outside home (**ZOSTHEFT**), attempted break in (**ZBREAKIN**) and even attacking people (**ZASSAULT**, **ZTHREAT**, **ZHARASS**) and trying to steal from them in the street (**ZPERSTHEF**). It is difficult to quantify these in the way required for the survey, and incidents were included at all or some relevant screeners which resulted in a lot of double counting.

'List of crimes' approach: A second strategy used by participants was to list the things they could recall that had happened to them over the last 12 months. While some participants were able to restrict their thinking correctly to the time frame, other participants reported incidents that had taken place longer ago (similar to strategy 1 above), particularly where they were salient and memorable (for example, an attempted break-in that had been quite frightening). One participant commented that thinking back over the last 12 months was challenging due to the impact of COVID, where the concept of time has 'gone out of the window'. As a result, people have fewer events on which to anchor incidents and relate back to. Crimes that had happened to other people were also included in the lists and in the live trial data, for example theft of a flatmate's bike and theft of a neighbour's car. This shows the importance of the question **BELONG** in the theft/attempted theft modules of the victim module, in ensuring crimes that happened to other people are filtered out as out of scope. However, it should be kept in mind that, as this would not occur until after the prioritisation algorithm had allocated the victim modules, other in-scope crimes such as fraud and online crime may not be allocated a victim module when they technically should be if other incidents were out of scope.

At Round 1, almost exclusively, if not for every participant, the 'list of crimes' did not match up to the data looking at it from both perspectives:

- There were crimes that were missing from the mental models exercise that were captured in the data.
- And there were crimes that were mentioned in the mental models exercise that weren't captured in the data

At Round 2 (R2), where participants tended to have a less complex crime profile than Round 1 (R1) participants, there was more symmetry between the mental models and incidents recorded in the interview.

Incidents that fell into category **a)** are not of concern to us as they were included in the data after being prompted for at the screeners. These tended to be less salient incidents such as online crime (**ZTRYCON**, **ZNONCON**), and incidents that had not been reported to the police, although there were exceptions to this pattern. In-scope incidents that had been included in the online survey but omitted from the mental models exercise are listed below. It is interesting that many of these are related to threats and harassment, suggesting that harassment may not be thought of as a relevant incident until this is specifically prompted for in the screeners:

- Threats of violence from a former housemate who was evicted as a result (R2)
- Threatening behaviour when out in the local area (R1)
- Threatening behaviour when asked someone to move their car blocking driveway (R2)
- Harassment while on a trail hunt (R1)
- Harassment from someone demanding money for a taxi not booked by participant (R2)
- Ongoing threatening behaviour on social media (R1)
- Attempted break-in to home (R1)
- Vehicle damage (car tyre stabbed on driveway) (R2)
- Hacking/virus on smartphone which subsequently needed replacing (R2)
- Fraudulent payments taken from bank account each month (R2)

This further supports existing evidence that the screener approach 'works' in terms of jogging participants' memories so they include more incidents than if they were asked more generally about their experience.

In some cases, participants were surprised on discovering they had omitted incidents in their mental model recap as illustrated by the quote below.

> *"the disparity...I remembered more when I was filling in the questionnaire than I did just now…I hadn't remembered that but something in your survey made me remember it".*

Incidents that fell into category b) (mentioned in the mental models exercise but not included in the online survey data) were of greater concern as these are potential crimes

that were excluded, for whatever reason, during the online survey. Interestingly, this only happened among Round 1 participants who had more complex crime profiles, suggesting the more complex the experience, the more likely participants are to miss relevant incidents. Where participants were unsure of the number of incidents within a certain crime type, one or more could be left out. It might be that participants struggle to distinguish between incidents or simply take one or a few to use as examples of the type of thing that happens regularly and even become commonplace. This can mean that a victim module can be a more general representation of a range of incidents, rather than being reflective of the most recent. One participant who experienced workplace assault, attempted assault, threats and harassment made the observation below.

> *"[these incidents are] an occupational hazard…there might be some things that happened that are so menial I just forget about them…it just happens all the time…I end up in a lot of these situations at work".*

While most incidents that had been excluded from the online survey had been excluded correctly as they had taken place longer ago than 12 months or very recently since the survey was completed, or had happened to other people (for example, neighbours or family members), there were some in-scope incidents that had been missed. These were:

- Workplace attempted assaults left out as so many
- Theft of rent money/fraud
- Attempted theft from person
- Attempted bike theft

9.3.3 Usability questions

The third stage of the depth interviews was to return to the usability questions and discuss the rationale for responses in cases where the participant could accurately recall their thoughts. However, in many cases, the online survey completion had been completed too long ago and was not salient enough for participants to remember the reasons for their answers to these questions. Therefore, findings are included below only where the participant felt certain of their answer. As such there were two key response options in the list at **WPISSUES** to highlight:

Finding some of the topics too personal or sensitive

Some participants made comments on the sensitive and personal nature of the questionnaire. It could act as a reminder of unwelcome and upsetting memories and mean they had to relive them. Below are some quotes to illustrate how the questionnaire made some participants feel.

> *[It was] "hard, raw, repetitive, difficult, challenging and daunting. [You] "try to put it to the back of your mind, put it to one side" [but now you have to] "tell a third party".*
>
> *"It is a scary thing to relive some of the scary things, so it wasn't always fun".*
>
> *"…it was sexual harassment and I'm used to talking about it, but it could be quite triggering to someone…how else do you find out what happened really".*

One participant commented on it being possible to skip most of the victim module for incidents that came through on assault, threats or harassment screeners.

> *"not really because you've got the option to skip the answers anyway so if someone finds it too sensitive, they should just skip it".*

Questions not fitting with individual experiences

The participant who raised this issue felt that the questionnaire didn't cover general issues as well as focusing on her own personal experience. She explained this as illustrated below.

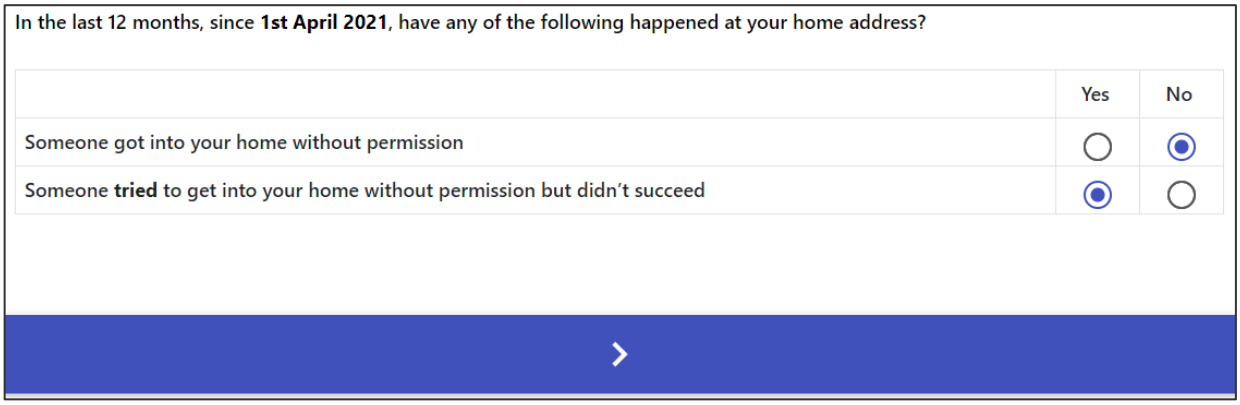
> *"things that aren't necessarily considered crime but are social injustices…I'm more worried about prejudice and violence and hate and that sort of thing…it was relevant to what was going on in the time frame but not necessarily relevant to my general experience of crime".*

9.3.4 Key question layouts

The last part of the interviews (at Round 1 only) involved showing the participant screenshots of key parts of the online questionnaire. This was the most challenging area in terms of recall and participants typically said they couldn't remember the specific questions or why they had answered in a certain way. Offering a visual prompt of the questionnaires wasn't particularly helpful except for a few specific issues that stood out. This mainly involved the date screens (**ZDATE**) and it was found that participants struggled to recall the double counting and checking screens (**ZRELATE**, **ZREVIEW** and **ZCOUNTCHECK**). Specific findings for each of the key question layouts are included below.

Screeners (paired approach)

While generally well received, one participant pointed out that you don't always know if a screener should be answered as 'yes' as details of the incident are unknown, for example whether damage was deliberate or whether this was part of an attempted theft. This also links to issues around selecting an earlier screener rather than the 'right' one or the best fit and then not being able to go back and correct it.



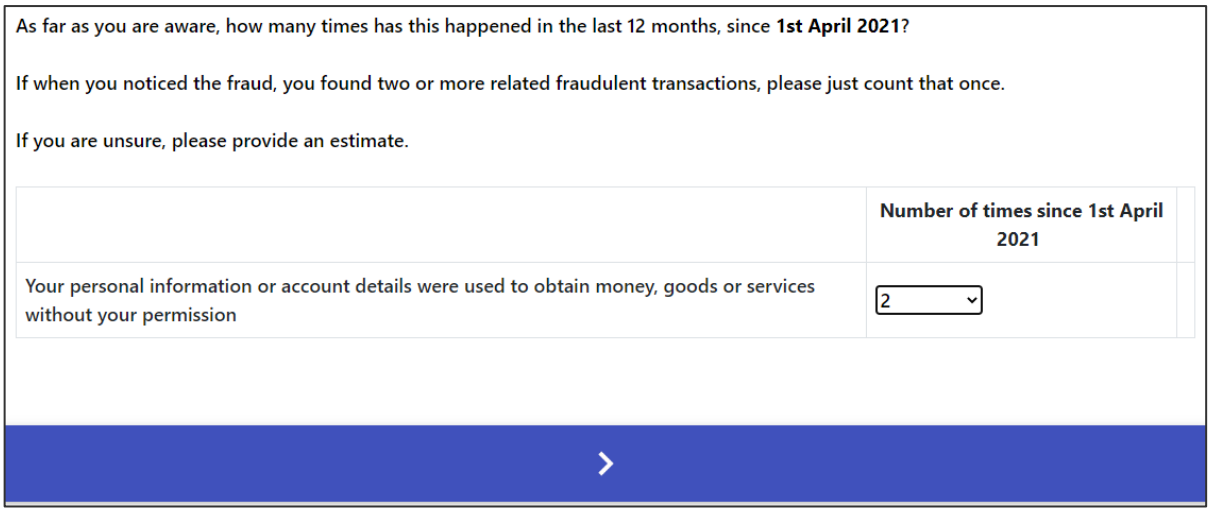
Counting screens

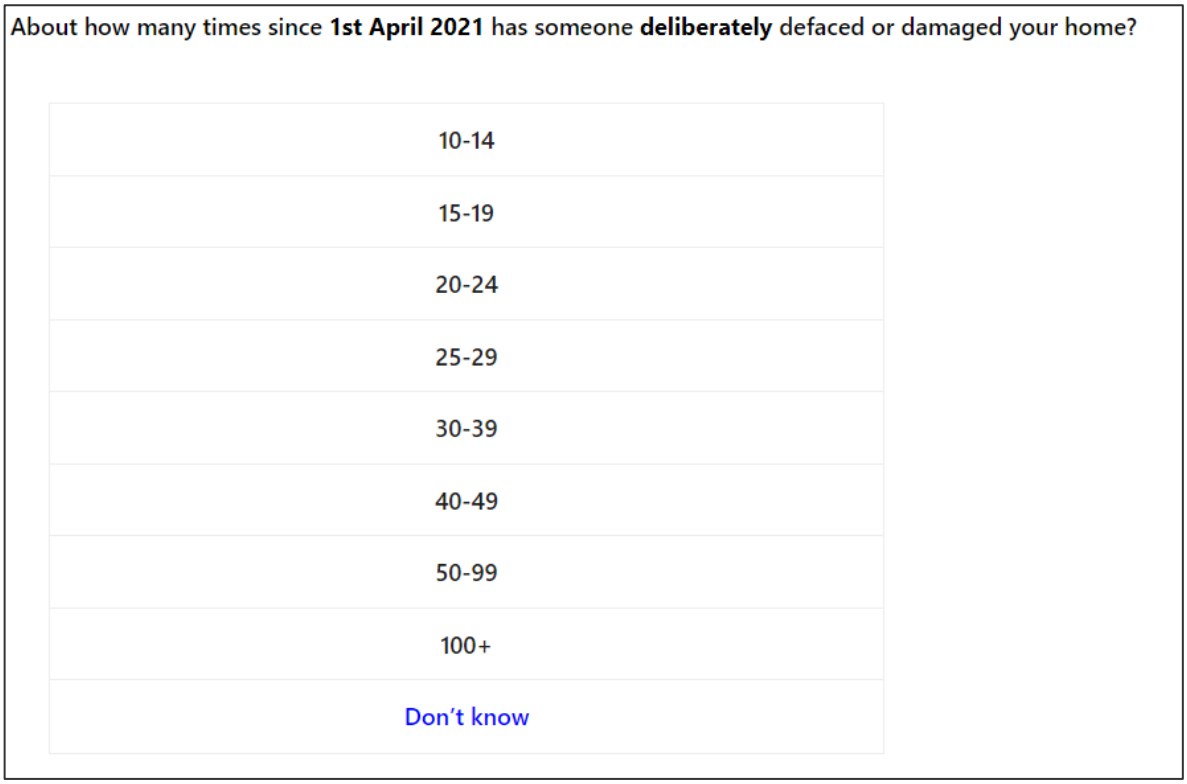
ZCOUNT

As noted, incident counts were challenging where something has repeatedly happened or where there has been a mix of series/separate and there is a convoluted process of going through three screens to enter don't know. On seeing these screens participants made the comments below.

> *"that's the bit I was complaining about, I don't know how to explain it, it's just not great".*
>
> *"the numbers, I struggle with numbers and time frames anyway and some of it was…I was doing a lot of don't know".*

The three counting screens are shown below.



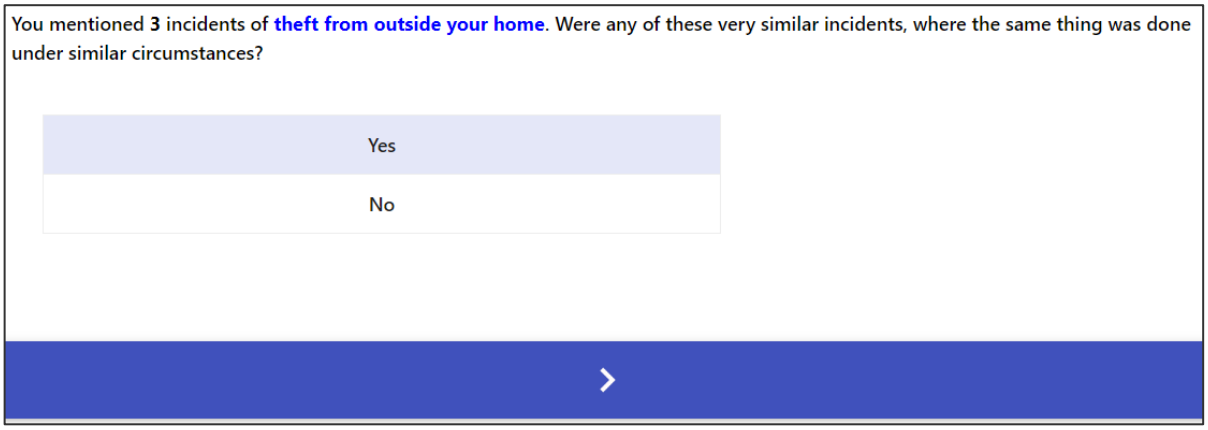


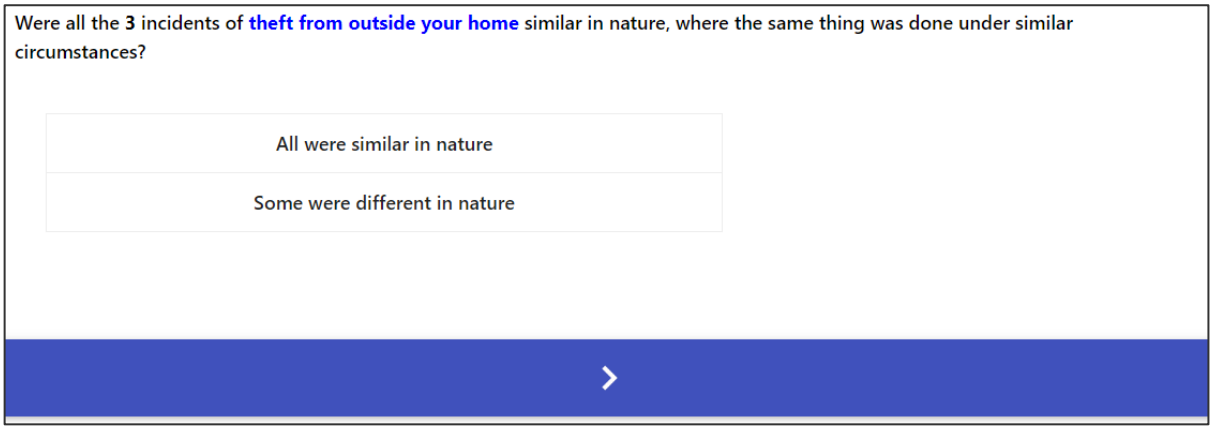
ZSIMILAR and ZSEPARATE

It was also difficult to recall these count screens as one participant commented.

> *"it ended up again a bit confusing as you're trying to link incidents and you're not sure if they fit into these different categories or not".*

The series and separate question screens are shown below.





Participants are then asked how many of the total number of incidents are similar and how many were different.

Double counting screens

ZRELATE

Our overall sense is that, at best, ZRELATE 'works' for some participants some of the time. However, this can mean a crime gets the 'wrong' 'crime description tag', or one that isn't the most appropriate or best fit. As the script relies on this tag to reference back to the incident later in the questionnaire, this can lead to confusion about which incident is being referred to, especially when there are multiple ones. It could also mean that participants are not given the opportunity to skip the victim module for crimes of a personal or sensitive nature where they have been given a different tag (as they are only offered the opportunity to skip incidents that have come through on assault, threats or harassment screeners) (as mentioned in section 4.4.5).

To illustrate this, one participant said:

> *"for some I put the same, some were different… [it made overall sense but] I don't think it was done in the right way".*

ZCOUNTCHECK

Participants couldn't recall this screen and, because it appeared complicated, wanted to move on from it. To illustrate this one participant said:

> *"I think so, but at that point it was just too much brain power and I just pressed a number and moved forward".*

ZREVIEW

As mentioned, participants struggled to recall these review screens. They were seen as complicated and there was evidence that participants would satisfice and answer yes to get past them. Comments included:

> *"I can't remember what I answered for that"*
>
> *"At that point, I think I might have figured out the pattern and just said no".*

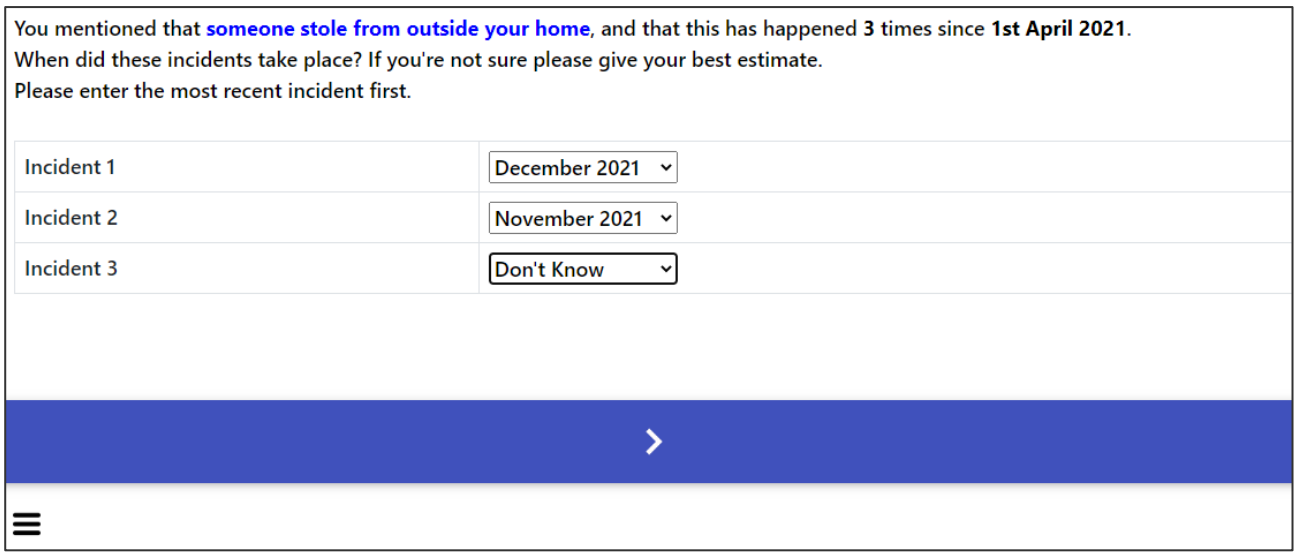
Date screens

ZDATE

Participants felt that it was generally difficult to give dates and that a lot of incidents wrap together, meaning some dates were estimations rather than exact. It was also more difficult in cases where incidents weren't reported to the police. To illustrate this, one participant commented as below.

> *"that was the hardest bit, because a lot of these dates I'm giving are approximates as I can't actually remember the month it happens as stuff happens all the time and I just…put a month but the thing is I never report it to the police so I don't have the actual log of the date".*

The screen which asks for three most recent dates (where three or more separate incidents were reported) was felt to be even more challenging. It was easier to provide a date for incidents that were more salient or that were pinned to specific events.



Open description

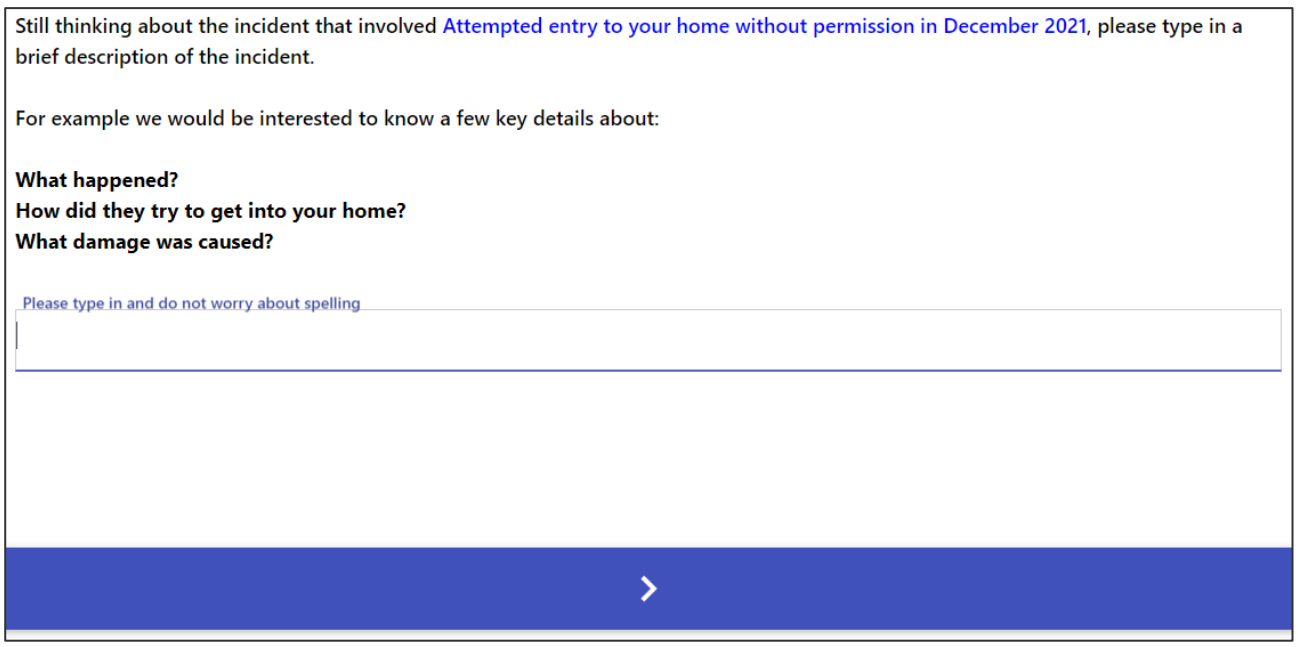
ZDESCRINC

Finally taking the open description, this was generally well received although the length and quality of responses varied widely. Being presented with an open text box could be off-putting and one participant said they would have preferred to answer closed questions. Regarding the logistics of typing in on a device, this is now a common everyday practice across laptops, tablets and smartphones that people are familiar with. Two examples of participant feedback on the open description are shown below.

> *“I do remember there being white boxes to fill in, to provide additional detail. I didn't feel I wanted to go into too much detail. However, if there had been questions there that said "Do you feel x, y, z..' as choices, rather than me having to fill in a box, I would have preferred that. Having to think back and provide details rather than just a high level description...I didn't want to do that, really, because I didn't really want to think about it anymore. It was in the past and I'd moved on. It might go back and think about something that I didn't really want to think about, so I was kind of pushing out of my mind.”*
>
> *"I was alright, I'm on my phone all the time so typing isn't really an issue".*

Another participant said that she is dyslexic and dyspraxic and used her laptop to complete the survey, she wouldn't choose to use a smartphone for this purpose as she prefers a basic keyboard. Kantar’s scripting template optimises across device so participants can have the choice of what device suits them best. Generally, there weren’t any issues raised about functionality.

Only one participant had needed help from a family member with completing the open descriptions. This participant did not routinely use a device for typing a lot of text; they said they were ok when giving single word answers but found the verbatim descriptive questions more difficult.



9.4 Summary and conclusions

While challenging to conduct, the depth interviews gave us the opportunity to explore the online questionnaire in a more 'applied' setting, using evidence, examples, quotes and case study detail within a 'real-life' context to back up what has until this point been qualitative evidence collected in more 'artificial' settings. Ultimately, the evidence collated, and drawn together in comparison with the live trial data for those specific cases, further supported what has already been learned about the limitations of the online instrument in relation to multiple experiences of crime, mainly in cases with multiple incidents of the same crime type, rather than providing new insights. Four key summary findings are:

- **The online CSEW is unsuited to very complex experience of crime**: The existing online CSEW questionnaire doesn't suit multiple instances of crime, in particular where a) participants don't know the total number of incidents or give a banded count; and b) participants' experience is vague and hard to describe (for example ongoing trespassing, anti-social behaviour or damage to the outside of your home). These experiences don't fit neatly into the instrument and a flaw of the current tool is that, due to time constraints and the complexity of the instrument, the script doesn't follow up with a victim module for incidents with a banded or don't know count. This is an important weakness of the online CSEW as not following up incidents with a banded or don't know count with a victim module means cases of very complex crime are missed. In turn this means incidents that are perhaps less important to the participant are picked up in victim modules. These restrictions around which incidents go through to victim modules mean participants don't always feel their experience is accurately captured.
- **Crime description 'tags' unable to distinguish between incidents:** A further issue relating to high frequency crime is being able to distinguish between these in a meaningful way when referring back to them. To give an example, one participant had experienced three separate incidents of attempted assault in March 2022. However, when referring back to these within both the screens and victim modules, due to the complexity involved with scripting this in the time available, the current instrument didn't differentiate between them as it only highlights the type of crime and the month. Future work should explore ways to better differentiate incidents which occurred in the same month while not over complicating or lengthening the crime description 'tag'.
- **Counts for mix of series and separate incidents don't match overall total**: It can be very hard to untangle what has happened when there has been a mix of series and separate incidents, and regularly the sum of the similar and different incidents didn't add up to the original count.
- **Further evidence of systemic double counting**: In line with what has been discovered in earlier phases of work, while asking each screener separately can help recall, there was considerable evidence that this method can, despite efforts to reduce this through ZRELATE, lead to extensive double counting and overlap of incidents. This leads to the range of associated problems that are already known, including duplicate victim modules, participant burden and potential for break off.