

Improving estimates of repeat victimisation derived from the Crime Survey for England and Wales

A note outlining some decisions that have been made in relation to implementing changes to the way we estimate repeat victimisation from the Crime Survey for England and Wales.

Contact:
John Flatley
crimestatistics@ons.gsi.gov.uk
+44 (0)20 7592 8695

Release date:
19 October 2017

Next release:
To be announced

Table of contents

1. [Introduction](#)
2. [Background to improvements in measuring repeat victimisation](#)
3. [Issues considered](#)
4. [Expected impact on Crime Survey for England and Wales data](#)

1 . Introduction

Based on the results of our [consultation on improving the way we estimate incidents of repeat victimisation derived from the Crime Survey for England and Wales \(CSEW\)](#), and advice from the National Statistician's Crime Statistics Advisory Committee, we are changing the way we estimate repeat victimisation. The 98th percentile of victim incident counts for each crime type will be used as the maximum number of repeat incidents that are included within the calculation on headline CSEW estimates.

Over the last year we have been undertaking further work to refine the methodology we will use to implement this change and to understand the impact upon CSEW estimates and time series. A number of decisions are outlined in this article, which relate to the way this new approach to measuring repeat victimisation will be implemented.

In summary:

- the 98th percentile will be calculated without the influence of difficult to interpret (“too many to remember”) responses; these responses will then be allocated the value of the 98th percentile for the number of incidents within series of the same crime type.
- three years' worth of weighted data will be used to calculate crime specific 98th percentile values for this purpose.
- the headline categories (such as “all violence”, “all robbery”) will be used to calculate the 98th percentile values for the number of incidents within relevant series.
- minor changes to the weights used to compensate for unequal probabilities of selection will be necessary to reduce volatility in estimates between years.
- our provisional timetable indicates that a back-series at least as far back as the year ending March 2003 could be available by July 2018, with revisions to earlier years to follow.
- uncapped estimates will be published as part of our methodology information.

2 . Background to improvements in measuring repeat victimisation

The Crime Survey for England and Wales (CSEW) was designed as a victimisation survey to estimate the number of victims of crime in the population. It has also been used to estimate the number of times a person is a victim of crime and hence the number of crimes experienced by adults living in households in England and Wales.

Producing such an estimate of incidents of crime is unproblematic for most crime types as the number of repeat victimisations suffered by an individual is usually small and easily recalled. For example, it is unlikely victims will not be able to remember the number of times their car was stolen or their house broken into in the previous 12 months. However, for certain crime types, such as violence in a domestic setting, the victim may suffer repeat victimisation with a frequency that is difficult to quantify over a 12-month period. High order repeat victimisation presents considerable challenges for the CSEW as only a relatively small number of victims yield a high number of victimisations.

Since one of the strengths of the CSEW has been its ability to provide trends for the crime types and population it covers, in cases of repeat victimisation the survey has always only included the first five incidents of a series in its estimate of the total number of incidents of crime in the population.

Only including five incidents within each series is a very effective way to reduce the effects of sample variability from year to year and has enabled us to avoid the publication of incident rates that fluctuate widely between survey years. However, for some crime types such as violent crime, this may result in point estimates being less “reliable” and introduce additional error. An independent review of the methodology for measuring repeat victimisation notes that; “different crime types have different distributions and for some...the cap of five is too suppressive” ([Reporting victimisation in the Crime Survey for England and Wales](#)).

Based on the results of our [consultation on the subject and advice from the National Statistician's Crime Statistics Advisory Committee in September 2016](#), it was decided that:

- the 98th percentile of victim incident counts for each crime type (calculated over a number of years) will be used as the maximum number of repeat incidents for any one respondent that are included within estimates; we will impute the 98th percentile value for any values above that point
- the time series will be revised back as far as possible
- uncapped data will be made available as part of our methodology information; however, as these estimates of total incidents will be subject to considerable volatility from year to year, appropriate caveats will be given around their use

Since October 2016, we have been undertaking exploratory work to help us understand the impact that these changes will have upon CSEW estimates and time series for both adults aged 16 and over and children aged 10 to 15 years. Different approaches have been assessed to consider factors important to our users, such as the level of transparency, the level of volatility introduced into time-series data and the sensitivity of different approaches to measuring changes in repeat victimisation over time.

In completing this work we have unearthed other issues that need detailed consideration, in particular the identification of large variability within the sample design weights. These issues also need to be resolved and we are in the process of assessing proposed refinements to our weighting methodology as a result.

3 . Issues considered

Quantifying the number of incidents experienced for victims who report “too many to remember”

The number of incidents in a series is coded by the interviewer following a spontaneous response from the respondent; interviewers are able to input any number up to 99. Up until the survey year ending March 2016, numbers between 1 and 96 are entered as per the respondents’ response. If the respondent answers otherwise the following conventions are used:

- 97 equals “too many to remember” ¹
- 98 equals “don’t know”
- 99 equals “refused”

The respondent is then asked how many of these incidents happened in different quarters of the previous 12 months and how many were outside of the Crime Survey for England and Wales (CSEW) reference period. If unable to break this down because the respondent is unable to remember the exact number for each quarter, the interviewer can insert a 97 for each or any further questions. However, since all spontaneous initial responses that relate to the full year are restricted to a two-digit number, in incidences in which a 97 is entered for each quarter, these are not summed. The maximum number of incidents in a series, within a year, is therefore treated as 97.

There is a risk of measurement error in interpreting these data because it has been observed in several contexts that survey respondents tend to approximate when the real number is 10 or more² and because it is difficult to determine what “too many to remember” means to each respondent. Feedback from interviewers working on the CSEW has suggested that whilst for some “too many to remember” could mean high volumes of incidences within a year, others may struggle to say exactly how many occurrences of an event there were when thinking about much lower numbers.

In the survey year ending March 2016 a new option became available for interviewers recording the number of incidents in a series. From this time, values 1 to 95 remained the same, 96 became “more than 95” and 97 remained “too many to remember”. In the year ending March 2016, there was only one instance where an interviewer picked code 96 over code 97 (out of a possible five instances). In the year ending March 2017, there were only two (out of a possible 11) instances. This suggests that in most cases “too many to remember” does not equate to numbers higher than 95. In workshops, interviewers have fed back that code 97 can often be used if a respondent is struggling to pinpoint a relatively low number exactly, for example, whether there were 10 or 12 instances.

Based on the evidence which suggests that the majority of code 97s are unlikely to relate to more than 95 incidents, as part of the improvements to measuring repeat victimisation, all 97s will be removed from the analysis so that we can calculate the value that equates to the 98th percentile of series values for each crime type without this additional bias. Once this analysis is completed, all 97s will be replaced with the value of the 98th percentile.

Assessing how many years’ of data should be used to calculate the value of the 98th percentile for any given year

Having assessed a number of options, including using 10-year, five-year, three- year and annual datasets to calculate crime specific 98th percentile values, it has been decided that three years’ worth of weighted data is the optimum amount for this purpose. This allows for sufficient numbers of victims to be able to calculate a 98th percentile for rarer crime types. It also provides sufficient data overlap to avoid any extreme volatility that may occur because of sampling variability in any given year. Calculating the 98th percentile values using three-year rolling datasets , rather than looking at five or more years’ worth of data at a time, also allows for sensitivity to sustained or real changes in repeat victimisation over time.

Of course, this will not be possible for the periods of time in which there is no preceding data. In this instance a slightly different approach will be used. For example, this approach will not work immediately for data collected on fraud, since we are yet to collect three years’ worth of data. In the interim, we will use all previously available data on fraud to calculate 98th percentile values until we reach the point of having three consecutive years.

It is suggested these 98th percentile values are recalculated in advance of each annual bulletin and published in our User Guide alongside our Crime in England and Wales, year ending March releases. Table 1 shows what we might expect these values to be going back to the year ending March 2003³.

Table 1: Crime Survey for England and Wales, 98th percentile values for the number of incidents within a series by crime type, to be applied to years ending March 2003 to March 2017 ¹

	Violence	Robbery	Theft from the person	Other theft of personal property	Fraud ²	Computer misuse ²
Three years ending March 2005	9	2	2	3
Three years ending March 2006	9	3	2	3
Three years ending March 2007	10	4	2	3
Three years ending March 2008	11	7	2	2
Three years ending March 2009	8	5	2	3
Three years ending March 2010	9	5	2	3
Three years ending March 2011	10	3	2	3
Three years ending March 2012	10	4	2	3
Three years ending March 2013	10	5	2	3
Three years ending March 2014	12	5	2	2
Three years ending March 2015	20	5	2	2
Three years ending March 2016	14	3	2	2	3	3
Three years ending March 2017	12	3	2	3	3	3
	Domestic Burglary	Other household theft	Vehicle-related theft	Bicycle theft	Criminal damage and arson	
Three years ending March 2005	3	4	3	2	6	
Three years ending March 2006	4	4	3	2	6	
Three years ending March 2007	3	4	3	3	6	

Three years ending March 2008	4	4	3	3	5
Three years ending March 2009	3	4	3	3	6
Three years ending March 2010	4	4	3	3	6
Three years ending March 2011	3	4	3	3	6
Three years ending March 2012	3	4	3	2	6
Three years ending March 2013	4	4	3	2	5
Three years ending March 2014	4	4	3	2	5
Three years ending March 2015	4	3	3	2	5
Three years ending March 2016	4	3	3	2	5
Three years ending March 2017	4	3	3	2	4

Source: Crime Survey for England and Wales, Office for National Statistics

1. These 98th percentile values have been calculated in advance of the reweighting of datasets (see 'Refinement to the weighting' section). Following the reweighting exercise there may be very minor adjustments to these caps. Final caps will be published alongside the year ending March 2018 publication.

2. Three years of data are not yet available to calculate the 98th percentile value for the number of incidents within series' of fraud or computer misuse crimes, values presented have been calculated based on all available data to date.

These different options have also been assessed against the data from the 10- to 15-years-old element of the survey in considering the previously mentioned decision. The 98th percentile values that would apply to these datasets are displayed in Table 2.

Table 2: Crime Survey for England and Wales, 98th percentile values for the number of incidents within a series reported by children aged 10 to 15 years, by crime type, years ending March 2010 to March 2017

	Violence	Robbery	All theft offences	Criminal damage and arson
Three years ending March 2010 ¹	15	6	5	6
Three years ending March 2011 ¹	15	6	5	6
Three years ending March 2012	15	6	5	6
Three years ending March 2013	16	20	4	5
Three years ending March 2014	15	20	5	3
Three years ending March 2015	15	20	5	3
Three years ending March 2016	17	7	8	8
Three years ending March 2017	15	9	5	12

Source: Crime Survey for England and Wales, Office for National Statistics

1. Three years of data is not available to calculate 98th percentile values to apply to the year ending March 2010 and year ending March 2011 datasets. Values calculated from the 3-year period to March 2012 will therefore be applied back to these to datasets.

Are there any instances in which the current cap of five should be lowered?

Table 1 shows what the 98th percentile by crime type would be (having removed all the 97s from the analysis) when calculated using three-year rolling datasets. For the majority of offences (robbery, personal theft offences, domestic burglary, other household theft and bike and vehicle theft) the 98th percentile for the number of incidents in a series is below the current cap of five. This raises the question as to whether we should also cap these crime types at the 98th percentile to be consistent with our methodology for measuring violence and other crime types with higher levels of repeat victimisation.

However, since these crime types have low levels of repeat victimisation and therefore are much less susceptible to volatility between years, it has been decided not to lower the existing cap of five. By keeping a minimum value of five for the number of incidents in a series we are, in many cases, publishing estimates closer to the 99th percentile for any crime types that do not pose an issue with volatility as part of our main estimates. In our [response to the consultation on repeat victimisation](#), we outlined that we would publish uncapped estimates where possible. Uncapped estimates will be more volatile regardless of the crime type and these will need to be published separately with appropriate caveats to guide users.

Crime types and 98th percentile values

After respondents have answered a series of questions about a crime they have reported experiencing in the last 12 months ([Crime Survey for England and Wales, Questionnaires](#)), a team of specialist coders then assign these crimes a specific offence code, which is designed to closely match the crime code that the police would have assigned (had it been recorded as a crime). Each crime has only one offence code and these codes are used as an important part of the analysis of the Crime Survey.

A list of all the offence codes can be found in our [User Guide](#) (Appendix 2). It is important to note that these offence codes do not match directly to those reported in our annual appendix tables. For example, the offence codes relevant to violence are as follows:

- 11 Serious wounding
- 12 Other wounding
- 13 Common assault
- 21 Attempted assault

We do publish other breakdowns of violence including; breakdowns of violence with injury and violence without injury, as well as a breakdown of whether the violence was classed as domestic violence, acquaintance violence or stranger violence. These breakdowns are derived from additional questions asked about each incident.

The 98th percentile value for the number of incidents in a series of any violent crime was typically between 8 and 20. However, if we look very specifically at violence that has been classed as “domestic”, based on additional questions, we can see the 98th percentile value for the number of incidents in a series typically varied between 15 and 30.

A decision has been made not to calculate the 98th percentile at this lower level after exploration of the effects it would have on the estimates for inter-related offence groupings and the relevant time-series. The option of using the headline category of “all violence” was deemed the most appropriate for a number of reasons, including:

- it means we are able to maintain our current concept of all “violence” and ensure all sub-categories of violence sum to the same total whilst not over-complicating the derivations we use – for example, if we use different 98th percentile values as the maximum values for different types of violence, domestic, stranger and acquaintance violence would sum to one total whilst violence with and without injury would sum to another; avoiding this complicated scenario or the complex derivations that would be needed to resolve this issue is thought to provide better clarity and transparency to users
- by changing the way we measure repeat victimisation, we are inevitably accepting some additional volatility in the estimates; this approach adds a more acceptable level of volatility when compared to the other approaches we tested
- we were able to avoid some specific issues, where individual weight allocations were compounding with high frequency victimisation and impacting estimates for some types of violence; the effects of these types of issues are addressed partly through our proposed adjustments to the weights (Refinement to the weighting methodology), but also by our decision to apply these changes at the headline level for each crime

The same applies to all other offence types, as 98th percentile values will be applied as a maximum number of incidents in a series for the following headline level offence categories; all violence, all burglary, all other household theft, all robbery, all personal theft, all vehicle-related theft, bicycle theft, criminal damage, all fraud, and all computer misuse. The 98th percentile values for incident numbers within these headline offence categories, dating back to the year ending March 2003, are shown in Table 1⁴.

Refinement to the weighting methodology

All CSEW estimates presented in the figures and tables in the Office for National Statistics’ (ONS) crime statistics publications are based on weighted data; that is, results obtained from surveying a sample of the population of England and Wales are scaled-up to represent the entire population. Two types of weighting are used in the CSEW sample. First, the raw data are weighted to compensate for unequal probabilities of selection involved in the sample design. These include: the over-sampling of less populous police force areas; the selection of multi-household addresses; and the individual’s chance of participation being inversely proportional to the number of adults living in the household. Second, calibration weighting is used to adjust for differential non-response.

When reviewing the methodology for improving the way we estimate repeat victimisation, it became apparent some minor changes to the weights used to compensate for unequal probabilities of selection would be necessary to reduce volatility in estimates between years. Calibration weighting will remain unchanged.

Design weights

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

Component weights

The weights are based on a number of components as follows:

- w_1 : weight to compensate for unequal address selection probabilities between police force areas
- w_2 : "address non-response weight" to compensate for the observed variation in response rates between different types of neighbourhood
- w_3 : the dwelling unit weight is simply the number of dwelling units identified at the address – in the vast majority of cases, the dwelling unit weight is one; historically, weight w_3 has been capped at 10 to limit the variance of core household and individual weights
- w_4 : the individual weight compensates for the fact that the probability of any one individual being selected is inversely proportional to the number of adults in the household; the individual weight is therefore simply the number of adults in the household

The two design weights are constructed as follows:

Core household weight equals w_1 multiplied by w_2 multiplied by w_3

Core individual weight equals w_1 multiplied by w_2 multiplied by w_3 multiplied by w_4

When we explored the effects of removing the cap of five from our measure of the number of incidents in a series, there were some instances in which high levels of repeat victimisation (97) coincided with very high weights. In one instance, final weights of more than 6,000 per individual coincided with a series that included 97 incidents of violence. The combined effect of this meant that by uncapping the estimates, one individual was contributing over 582,000 incidents to our annual violence estimates (as compared to the individuals' contribution of just over 30,000 incidents with the cap of five in place).

The component weight that contributed directly to this issue was the dwelling unit weight (w_3). However, analysis of the data indicated the same issue may arise in the future as a result of the individual component weight (w_4), which has similar variability.

A decision has been made to trim the component dwelling unit weight for the calculation of household weights. In calculating the core individual weight the product of the multiplication of the dwelling unit weight and individual component weights will also be trimmed. This aligns with the weighting procedures used for the 10- to 15-years-old element of the CSEW, where a maximum value of four has been applied to the dwelling unit element of the household weight since the year ending March 2016. Although trimming of extreme weights may introduce a small amount of bias this is more than compensated for by the improvement in precision that results.

It is not uncommon for extreme weights to compound with count data to increase volatility. As a result other ONS surveys have typically adjusted weights to account for outliers. [The Living Costs and Food Survey](#) identifies outliers from the weights and removes these cases into different stratum. The end result is to allow for an outlier to represent only itself, giving the other population units that this value would have represented “average values”. The [English Housing Survey](#) use smoothing techniques where appropriate to reduce variability across the weighting classes.

We are in the process of further assessing how these extreme weights are treated in other surveys and are taking advice from survey methodologists at the level at which to trim the weights applied to the CSEW.

Weighting on the aged 10 to 15 years survey

The final weight produced for each case in the 10- to 15-year-old sample is equal to the household weight multiplied by the product of (i) the reported number of 10- to 15-year-olds in the household, and (ii) the inverse of the estimated (conditional) response probability as derived from the logistic regression model (our latest [Technical report](#) provides further detail).

Since the year ending March 2016, the product of component (i) and the dwelling unit component of (w3) has been capped at four to prevent excessive variation in the design weights. Prior to this time, weights allocated to 10- to 15-years-old were trimmed in other manners with similar results. As a result, reweighting of the children's data will not be necessary.

Notes for Issues considered

1. Until the year ending March 2016 survey when interviewers were instructed to enter 96 for cases referred to “more than 95”. Up until this time, 97 meant both “too many to remember” and “more than 95”.
2. See Lauritsen et al (2012), [Methods for Counting High-Frequency Repeat Victimizations in the National Crime Victimization Survey](#). Bureau of Justice Statistics. US Department of Justice.
3. These estimates are subject to change following changes to our weighting procedures as outlined in the “Refinement to weighting” section.
4. Estimates are subject to change, since they have been calculated in advance of implementation of plans to reweighting datasets.

4 . Expected impact on Crime Survey for England and Wales data

Adults aged 16 and over

In advance of reweighting our datasets it has been possible to assess the impact of using the 98th percentile value as a maximum number of incidents within a series (as compared to the current cap of five) on estimates for incident numbers. Owing to the fact that the reweighting process is not yet in place we are only able to give an approximate idea of the impact on estimate numbers at this time.

In the majority of cases, the 98th percentiles for the number of repeat incidents in a series of crimes is lower than five (with the exception of violent offences and criminal damage). Since we have no intention of lowering the maximum number of incidents counted within a series to a level below five, for the majority of crime types the impact of these changes on Crime Survey for England and Wales (CSEW) estimates will be minimal (and will result from minor adjustments made to some components of the design weights as discussed previously).

It is important to note, that as a result of changes to the weighting both prevalence and incidence estimates will change somewhat for all crime types, regardless of whether the level at which we trim counts of repeat incident is increased.

The most pronounced change in published estimates will relate to estimates of incident numbers for violence. Table 3 shows the expected upward impact on estimates of violent incident numbers as a result of implementing the new methodology. All forms of violence will see an upward change in estimate numbers, with smaller changes being seen in stranger violence and violence without injury than in other categories.

Estimates for “all violence” will increase by between 8% to 23% compared with currently published estimates, with some variations in the size of the volume increase that is likely to be seen each year. The largest changes, compared to currently published figures, will apply to violence without injury, domestic violence and acquaintance violence, where percentage changes of up to 30% are likely for some years.

There may also be a relatively minimal upward change in estimates of criminal damage in parts of the time-series as compared to previously published data. The cap on incidents within a series will be raised from five to six for many of the years prior to the year ending March 2012. It is thought this change will be relatively small. Early exploration work suggests estimates are likely to rise by less than 3%.

Table 3: Approximate expected percentage change in number of violent incidents (new methodology as compared to previously published estimates)

CSEW adults aged 16 and over England and Wales, years ending March 2008 to March 2017 ¹

	Violence	Violence with injury	Wounding	Assault with minor injury	Violence without injury	Domestic violence	Acquaintance violence	Stranger violence
	Percentage change							
Year ending March 2008	14	16	18	15	11	25	18	5
Year ending March 2009	8	7	5	9	8	8	10	6
Year ending March 2010	10	11	12	10	10	17	14	4
Year ending March 2011	12	14	13	15	9	19	15	6
Year ending March 2012	14	14	13	14	13	15	21	7
Year ending March 2013	17	20	17	23	18	28	19	8
Year ending March 2014	18	5	3	6	26	22	22	12
Year ending March 2015	23	18	24	12	29	29	30	13
Year ending March 2016	15	7	6	8	16	18	11	20
Year ending March 2017	15	19	26	11	12	23	18	8

Source: Crime Survey for England and Wales, Office for National Statistics

1. The percentage change compared to current estimates displayed above has been calculated prior to planned changes to weights, these figures are therefore subject to change.

Children aged 10 to 15 years

There is much more variability in the 98th percentile values that will be applied to the data coming from the 10- to 15-years-old element of the survey than with the data from the adults. Table 2 shows the caps that will be applied to the headline offence categories for 10- to 15-year-old incident counts. Violence, robbery and theft from the person will be the crime types most impacted by this change in methodology. Additionally, as can be seen in Table 2, for some years in the current (and likely therefore, the future) time-series, 98th percentile incident values may also rise beyond five for other crime types (for example, the 98th percentile for criminal damage rises from below five to eight and 12 for the years ending March 2016 and March 2017 respectively).

The changes in incident numbers will mostly affect violence, robbery and criminal damage offence categories, which are all likely to see consistent upward changes of up to 36%, compared to those already published. Theft offences will predominantly remain unchanged since the level of repeat victimisation for these headline offences is low, with the exception of the year ending March 2016. This is the only year for which the 98th percentile values for these offences surpasses five, though the impact on the estimates is likely to be small in comparison to percentage changes seen for other crime types.

Table 4: Approximate expected percentage change in number of violent incidents (new methodology as compared to previously published estimates), Crime Survey for England and Wales, children aged 10 to 15 years

Years ending March 2014 to March 2017

	Violence	Violence with injury	Wounding	Assault with minor injury	Violence without injury	Robbery	Criminal damage to personal property
							Percentage change
Year ending March 2014	18	15	17	14	25	3	11
Year ending March 2015	34	21	17	22	62	29	16
Year ending March 2016	30	29	34	26	36	21	20
Year ending March 2017	17	17	38	9	14	27	44

Source: Crime Survey for England and Wales, Office for National Statistics

It is important to note, that these values have been calculated based on “broad” as opposed to “preferred” measures of crime. We are already aware of a lot of volatility within the estimates of incident numbers for crimes against 10- to 15-year-olds. Indeed, as a result a similar adjustment to the weighting as that described previously for adults took place in the year ending March 2016.

Implementation

Since reweighting is now considered necessary, we will also need to revise tables that relate to the prevalence of victimisation as well as the number of incidents and this has made the project more extensive than initially anticipated. Implementation of these methodological changes has been provisionally timetabled. We hope to be able to release these estimates in a time series going back to the year ending March 2003 in time for July 2018 (alongside Crime in England and Wales, year ending March 2018). This will be reviewed as we progress with our work and is subject to change. A timetable for revising our non-consecutive calendar year datasets (from 1981 through to 1995) is not yet in place, we hope to release some of this further data alongside our Crime in England and Wales, year ending June 2018 bulletin (published in October 2018).

We have assessed that completing this work for our annual datasets will fulfil the majority of user requirements and we currently intend to work on this basis. However, as we reach our year ending June, September and December 2018 publications, we will also apply these changes to the relevant quarterly datasets for 2017 that will enable us to assess year-on-year change.

We welcome comments and feedback on our intended approach and current timetable.