Merging waves of the Crime Survey of England and Wales and the British Crime Survey: User Guide

Version 1.2

Dr. Niels Blom

*Department of Criminology & Department of Social Statistics, University of Manchester*

**Citation for the user guide and code:**

Blom, Niels (2023). Merging waves of the Crime Survey of England and Wales and the British Crime Survey. [Data Collection]. Colchester, Essex: UK Data Service.

**Acknowledgement:**

This research was supported by the UK Prevention Research Partnership (Violence, Health and Society; MR-VO49879/1), which is funded by the British Heart Foundation, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Health and Social Care Research and Development Division (Welsh Government), Medical Research Council, National Institute for Health Research, Natural Environment Research Council, Public Health Agency (Northern Ireland), The Health Foundation, and Wellcome. The views expressed in this article are those of the authors and not necessarily those of the UK Prevention Research Partnership. The funder had no role in the study design, data collection, data analysis, data interpretation, or writing of this paper, or in the decision to submit the paper for publication.

I am also thankful to Anastasia Fadeeva, Elouise Davies, Hannah Manzur, and Polina Obolenskaya for helping testing the code and providing feedback on this User Guide.

Contents

[Purpose 3](#_Toc126597490)

[Crucial Points, please pay attention 3](#_Toc126597491)

[Steps to take to merge the files 5](#_Toc126597492)

[Step 1: Organize the data folders. 5](#_Toc126597493)

[Step 2: Create a folder on your computer where files are stored. 5](#_Toc126597494)

[Step 3: Open do-file and adjust file locations. 5](#_Toc126597495)

[Step 4: Specify which files you want to merge. 6](#_Toc126597496)

[Step 5: Run the **full** code 7](#_Toc126597497)

[New and changed variables 8](#_Toc126597498)

[New variables 8](#_Toc126597499)

[Changed variables 9](#_Toc126597500)

[Adding extra years of CSEW after 2019/2020. 10](#_Toc126597501)

# Purpose

This guide accompanies the do-file (CSEW\_merger\_v1-1.do) which enables STATA users to merge multiple years of Crime Survey of England and Wales (CSEW) and/or the British Crime Survey (BCS). The purpose of this user guide is to guide the user on how to use the do-file to merge various years for the CSEW and/or the BCS.

By combining multiple survey sweeps, you are able to look at, for instance, trends in violence. Furthermore, using such a combined file enables you to look at specific offences, population groups, or consequences, that do not have a high enough frequency if you would use only a single year.

In specifying the code, you can decide which files you want to merge. Namely, which years of the Crime Surveys you want to merge and if you want the bolt-on datasets that provide uncapped codes[[1]](#footnote-1), the adolescent and young adult panels, and/or if you want to use the ‘non-white’ panel[[2]](#footnote-2). See also step 4.

Note that we refer to refer to the Crime Survey collected between the years of 1982 to 2001 as “Pre2001” and refer to the surveys from 2001-2002 onwards as “Post2001”.

Currently the code works for the years up to the **CSEW 2023/2024 wave.**

Note that you have to request access to the CSEW and BCS yourselves. This can be done via [UK Data Service](https://ukdataservice.ac.uk/).

It does not merge any waves collected between 2020 and 2021.

Note that you can reduce the length of the code in the do-file by clicking on the minus sign on line 64 in the do-file to collapse the lines between 64 to 1612. This makes the code look a bit clearer but doesn’t change anything on the procedure.

|  |  |
| --- | --- |
| Table 1. Description of Stata files you get after running the code | |
| Post2001\_nvf.dta | Non-victim forms merged for the years of the Crime Survey after 2001 (see Table 2) |
| Post2001\_vf.dta | Victim forms merged for the years of the Crime Survey after 2001 (see Table 2) |
| Pre2001\_nvf.dta | Non-victim forms merged for the years of the Crime Survey before 2001 (see Table 2) |
| Pre2001\_vf.dta | Victim forms merged for the years of the Crime Survey before 2001 (see Table 2) |
| Post2001\_adolescent\_youth\_nvf.dta | Non-victim forms merged for the years that have an adolescent or youth sample. Note that you most likely want to investigate the youth and adolescent samples separately. |
| Post2001\_adolescent\_youth\_vf.dta | Victim forms merged for the years that have an adolescent or youth sample. Note that you most likely want to investigate the youth and adolescent samples separately. |
| Post2001\_emboost\_nvf.dta | Non-victim forms merged for the ethnic minority samples for the years of the Crime Survey after 2001 (see Table 2) that have ethnic minority samples |
| Post2001\_emboost\_vf.dta | Victim forms merged for the ethnic minority samples for the years of the Crime Survey after 2001 (see Table 2) that have ethnic minority samples |
| Pre2001\_emboost\_nvf.dta | Non-victim forms merged for the ethnic minority samples for the years of the Crime Survey before 2001 (see Table 2) that have ethnic minority samples |
| Pre2001\_emboost\_vf.dta | Victim forms merged for the ethnic minority samples for the years of the Crime Survey before 2001 (see Table 2) that have ethnic minority samples |

# Crucial Points

The do-file enables users to merge the raw CSEW/BCS datasets. Consequently, currently this code **does NOT harmonize variables** that change slightly over different years. Considering the measurement of many variables changes over the years, the users of this combined file need to make their own decisions on what operationalisations work best in their situation and for the years they use. Most of the time new variable names are used when a new measurement is used.

However, for a few variables, different measurements seem to be used in different years, but they have the same variable name (for instance for household income variables such as *tothhin2*). In the current code, these variables are treated as being the same. Therefore, please check the variables that you use carefully using the codebooks or questionnaires for the relevant years.

Next, this code to our knowledge does *not* work in the secure researcher environment as provided by UKDS or ONS because the datasets in these environments have different names and/or the structure of the folders is different.

It does not merge any waves collected between 2020 and 2021.

Lastly, to reiterate, we refer to refer to the Crime Survey collected between the years of 1982 to 2001 as “Pre2001” and refer to the surveys from 2001-2002 onwards as “Post2001”, see also Table 2.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.** Datasets, we use “BCS” to refer to the Crime Survey years of 1982 to 2001 and “CSEW” to refer to the surveys collected from 2001-2002 onwards. | | | | | | | |
|  |  |  |  | Available special datasets and add-ons | | | |
|  |  | UK DataService | Name of dataset | Adolescent or youth samples | Ethnic minority boost sample | Bolton add-on | Additional add-on (automatically applied) |
| SN code |
| In this documentation referred to as **"Pre2001"** | Can only select all or nothing. | 1869 | British Crime Survey, 1982: England and Wales Data |  |  |  |  |
| 2077 | British Crime Survey, 1984 |  |  |  |  |
| 2706 | British Crime Survey, 1988 |  |  |  |  |
| 3202 | British Crime Survey, 1992 | X (age 12-15) |  |  |  |
| 3591 | British Crime Survey, 1994 |  |  |  | drugs self-completion, stolen goods |
| 3832 | British Crime Survey, 1996 |  | X |  |  |
| 4081 | British Crime Survey, 1998 |  |  |  |  |
| 4463 | British Crime Survey, 2000 |  | X |  |  |
| 4786 | British Crime Survey, 2001 |  |  |  |  |
| In this documentation referred to as **"Post2001"** | Can be merged, either all, nothing, or a selection. | 4787 | British Crime Survey, 2001-2002 |  |  | X |  |
| 5059 | British Crime Survey, 2002-2003 |  | X | X | technology crimes, social cohesion, mobile |
| 5324 | British Crime Survey, 2003-2004 |  | X | X | technology crimes, social cohesion, mobile |
| 5347 | British Crime Survey, 2004-2005 |  | X | X | technology crimes |
| 5543 | British Crime Survey, 2005-2006 |  | X | X |  |
| 5755 | British Crime Survey, 2006-2007 | X (age 16-24) |  | X |  |
| 6066 | British Crime Survey, 2007-2008 | X (age 16-24) |  | X |  |
| 6367 | British Crime Survey, 2008-2009 | X (age 16-24) |  | X |  |
| 6627 | British Crime Survey, 2009-2010 | X (age 10-15) |  | X |  |
| 6937 | British Crime Survey, 2010-2011 | X (age 10-15) |  | X |  |
| 7252 | Crime Survey for England and Wales, 2011-2012 | X (age 10-15) |  | X (only nvf) |  |
| 7422 | Crime Survey for England and Wales, 2012-2013 | X (age 10-15) |  |  |  |
| 7619 | Crime Survey for England and Wales, 2013-2014 | X (age 10-15) |  |  |  |
| 7889 | Crime Survey for England and Wales, 2014-2015 | X (age 10-15) |  |  |  |
| 8140 | Crime Survey for England and Wales, 2015-2016 | X (age 10-15) |  |  |  |
| 8321 | Crime Survey for England and Wales, 2016-2017 | X (age 10-15) |  |  |  |
| 8464 | Crime Survey for England and Wales 2017-2018 | X (age 10-15) |  |  |  |
| 8608 | Crime Survey for England and Wales, 2018-2019 | X (age 10-15) |  |  |  |
| 8812 | Crime Survey for England and Wales, 2019-2020 | X (age 10-15) |  |  |  |
| 9267 | Crime Survey for England and Wales, 2022-2023 |  |  |  |  |
| 9365 | Crime Survey for England and Wales, 2023-2024 |  |  |  |  |

# Steps to take to merge the files

Please follow the following steps very carefully when you want to merge the CSEW and/or BCS.

## Step 1: Organize the data folders.

The folders with the data pre 2001 and post 2001 should be stored in the same folder. I prefer to use it as downloaded from the UKDataService website, and then ‘unzip’/’extract’ that folder. I renamed them slightly by deleting the seemingly random list 60+ characters of the file name).

After doing this, the folders looks somethings like this:

A screenshot of a computer

AI-generated content may be incorrect.

Within each folder, the structure is the same as when you download it. Thus within the folder, the folder structure will be (as example the CSEW2019/2020 is used): 8812stata -> UKDA-8812-stata -> stata -> stata13\_se -> (folder containing all the data files)

## Step 2: Create a folder on your computer where files are stored.

To be safe and not accidentally delete your files, please create a separate folder from the raw data where data files are temporarily stored and where the final merged datasets will be stored. This can be anywhere on the computer.

## Step 3: Open do-file and adjust file locations.

1. Open the Stata do-file in Stata.
2. On line 21 in the do-file, which starts with global Temp , change the file-location to the folder that you created under Step 2. Remember to end with the \ (backslash) and that the file location is between “ ” (Quotation marks).
3. On line 22 in the do-file, which starts with global data , amend the directory to reflect the location of the folder where all the raw CSEW and/or BCS data is stored (the general folder where the folders from Step 1 are stored).

Remember to end with the \ (backslash) and that the file location is between “ ” (Quotation marks).

1. If you changed the folder structure of the downloaded files, change the file-locations for the CSEW and/or BCS to refer to the specific CSEW and/or BCS waves. This starts at line 26 (which start global CSEW2324 ) to line 56 (which start global BCS1982 ).

You should not include the part of the file directory that you specified in step 3.

* + The CSEW2324 (meaning the CSEW that was collected in 2023 and 2024) should read something like "9365stata\UKDA-9365-stata\" on line 25.

Remember to end with the \ (backslash) and that the file location is between “ and ” (Quotation marks).

* + If you do not use a certain wave of the CSEW or do not use the BCS, you can leave the part between quotations marks blank. Thus, it will be something like global CSEW0102 "" for the CSEW2001/2002 if you do not use this year.

## Step 4: Specify which files you want to merge.

Here you need to specify which files you would like.

1. In line 8 in the do-file ( global Post2001 = "yes" ) specify whether you want to merge some years of the Crime Surveys after 2001 (see Table 2, first wave is 2001/2002, last wave currently is 2023/2024). If you would like to merge any files the Crime Surveys keep this line as global Post2001 = "yes" otherwise, change it to global Post2001 = "no" .
2. In line 9 in the do-file ( global Pre2001 = "yes" ) specify whether you want to merge all the Crime Surveys years before 2001 (see Table 2, first wave is 1982, last was 2001).

This code does not allow you to select only certain years of Crime Surveys before 2001 for merging; if you want to use any these earlier years, you have to use all of them. You cannot just use a selection of them in this code (of course after merging, you can always delete certain waves).

If you would like to merge any files of the pre 2001 keep this line as global Pre2001 = "yes" otherwise, change it to global Pre2001 = "no" .

1. Line 10 concerns the 10 to 15 year old adolescent and youth samples ( global adolescentsample = "yes" ). If you want to merge these files keep this line as global adolescentsample= "yes" otherwise, change it to global adolescentsample= "no" .

Note that you cannot have adolescentsample= "yes" if Post2001 is “no” (line 8).

1. Line 11 concerns the ‘non-white’ files. Those are available between years 2002 and 2006 and includes all non-white respondents of the main design, plus an oversampling of other non-white respondents.

If you want to have the non-white files or ethnic minority boost files line 11 must be global emboost = "yes" , otherwise change to global emboost = "no"

Note that some respondents are both in the main file and in the non-white files, so be careful if you choose to combine the non-white files with the main survey as you will have to delete double observations.

1. Line 12 concerns the bolt-on files which provide the uncapped crime/offence codes for years 2011/2012 and before.

If you want to merge these files keep this line as global bolton= "yes" otherwise, change it to global bolton= "no" .

Note that you cannot have bolton = "yes" if CSEW is “no” (line 8).

1. Specify the years of the Crime Surveys Post2001 that you want to merge in line 15 ( global csewlatest 23 ) and 18 ( global csewearliest 01 ) . This is not applicable for people who use only the BCS.

* The most recent year should be in line 15. For example if the most recent year you want to use is the CSEW of 2005/2006, this line should be changed to global csewlatest 05
* The earliest year you want to merge should be in line 18. For example, if the earliest year you want to use is the CSEW of 2002/2003, this line should be changed to global csewearliest 01 .
* Note that csewlatest cannot be lower or equal to csewearliest
* Note that csewlatest cannot be higher than 23 currently.

## Step 5: Run the **full** code

Lastly, run the full code from the first line to the last line.

The code will run for about 5 minutes to 15 minutes (depending on the quality of the computer).

Note that you can reduce the length of the code by clicking on the minus sign on line 62 to collapse the lines between 62 to 1315. This makes the code look a bit clearer but doesn’t change anything on the procedure.

In the end you will get a maximum of 10 files (dependent on which files you have specified you wanted to merge under step 4):

Post2001\_nvf.dta , Post2001\_vf.dta

Pre2001\_nvf.dta , Pre2001\_vf.dta

Post2001\_adolescent\_youth\_nvf.dta , Post2001\_adolescent\_youth\_vf.dta

Post2001\_emboost\_nvf.dta , Post2001\_emboost\_vf.dta

Pre2001\_emboost\_nvf.dta , Pre2001\_emboost\_vf.dta

In line with the convention, the files ending in \_vf are the victim form files, the \_nvf are the non-victim form files.

# New and changed variables

## New variables

After running the do-file, the datafiles contain variables that are not in the original file. These are:

|  |  |
| --- | --- |
| Variable name | Description |
| *Indicator of main datafiles* |  |
| wave | This represents the first year of data collection of this survey wave. So, for the CSEW wave collected in 2019/2020, it will have the value 2019. |
| PREPOST2001 | This variable indicates whether it is a wave after 2001 (which has the value 1) or a wave before 2001 (which has the value 2). |
| *Indicator of whether extra data is merged* |  |
| bolton | Indicates whether this respondent has variables updated with a bolton dataset (if available for that wave) or the non-victim-form files. This variable labels follow the standard variable labels for merging files in Stata (<https://www.stata.com/manuals/dmerge.pdf> page 4). |
| boltonVF | Indicates whether this respondent has variables updated with a bolton dataset (if available for that wave) for the victim-form files. This variable labels follow the standard variable labels for merging files in Stata (<https://www.stata.com/manuals/dmerge.pdf> page 4). |
| technologyfile | Indicates whether this respondent has variables based from the technology crimes dataset (only available for CSEW 2004/2005 and 2003/2004). Value labels follow the standard variable labels for merging files in Stata (<https://www.stata.com/manuals/dmerge.pdf> page 4). |
| socialcohesionfile | Indicates whether this respondent has variables based from the social cohesion datafile(only available for CSEW 2003/2004). Value labels follow the standard variable labels for merging files in Stata (<https://www.stata.com/manuals/dmerge.pdf> page 4). |
| mobilefile | Indicates whether this respondent has variables based from the mobile file datafile(only available for CSEW 2003/2004 and 2002/2003). Value labels follow the standard variable labels for merging files in Stata (<https://www.stata.com/manuals/dmerge.pdf> page 4). |
| drugsselfcompletedata | Indicates whether this respondent has variables based from the drugsselfcomplete data file datafile (only available for BCS 1994 and 1992). Value labels follow the standard variable labels for merging files in Stata (<https://www.stata.com/manuals/dmerge.pdf> page 4). |
| stolengoodsdata | Indicates whether this respondent has variables based from the Stolen goods datafile (only available for BCS 1994). Value labels follow the standard variable labels for merging files in Stata (<https://www.stata.com/manuals/dmerge.pdf> page 4). |
| *New (renamed) identifiers* |  |
| onspsuidnr | ONSpsuid is renamed to onspsuidnr |
| onspsuidstring | onspsuid is renamed to onspsuidstring |
| *Other new variables* |  |
| whenlftstr | whenlft is renamed to whenlftstr for CSEW 2001/2002 and BCS 2001 |
| hwhenlftstr | hwhenlft is renamed to hwhenlftstr for CSEW 2001/2002 and BCS 2001 |

## Changed variables

The files contain variables that are renamed. The following variables are not available in the merged datafiles (they have different variable names). These are:

|  |  |
| --- | --- |
| Original variable name | Description |
| Rowlabel | In the BCS file only. Because the rowlabel is not unique over the waves, make the ID numbers (rowlabel) unique by replacing adding wave\*10^08 to the rowlabel (first 4 numbers thus refer to the year the survey was collected in the BCS file. |
| ONSpsuid/onspsuid | See onspsuidnr and onspsuidstring above. |
| whenlft / hwhenlft | See whenlft and hwhenlft above. |

1. https://doc.ukdataservice.ac.uk/doc/7252/mrdoc/pdf/7252\_csew\_historic\_bolt-on\_datasets.pdf [↑](#footnote-ref-1)
2. Bolling, Grant, & Sinclair (2007). *2006-07 British Crime Survey (England and Wales): Technical Report Volume I.* UK Data Archive Study Number 5755 - British Crime Survey, 2006-2007 [↑](#footnote-ref-2)