

# BCS Open Data tables User Guide

---

## INTRODUCTION

This document contains information on the content and use of the Home Office British Crime Survey (BCS) Open Data tables. These tables contain BCS estimates related to victimisation and perceptions of crime and the Criminal Justice System broken down by demographic characteristics. These tables are released as part of a new initiative to make government data more transparent and accessible to the public and external researchers and this is the first release of BCS data in this format.

These tables do not contain raw data from the BCS or responses to individual questions from individual respondents. For users wanting to access full BCS datasets, these are available from the UK Data Archive (<http://www.esds.ac.uk/findingData/bcrsTitles.asp>).

It is recommended that prior to using these Open Data tables, users read the User Guide to Home Office Crime Statistics (<http://www.homeoffice.gov.uk/publications/science-research-statistics/research-statistics/crime-research/user-guide-crime-statistics/?view=Standard&pubID=908794>) to familiarise themselves with the context of the data and the scope and limitations of the BCS as a whole.

From April 2012, responsibility for the publication of crime statistics will move from the Home Office to the Office for National Statistics (ONS) (<http://www.ons.gov.uk>). This will also coincide with a change in name from the British Crime Survey to the Crime Survey for England and Wales. ONS will take responsibility for producing future editions of these tables. At present, there are no plans to produce a back series to cover earlier BCS years.

## TABLE FORMAT

The main release consists of six data tables, all with an identical layout. The files have been split by theme, but because all have the same layout they can all be combined into a single table after download.

Tables are laid out in rows and columns with the first row containing column headings. Each row consists of a single BCS estimate, while each column contains information about the estimate. Estimates are principally broken down by respondent characteristics (identified by the **Sex**, **Age**, **HouseholdType**, and **Characteristic** columns). However, other columns provide further information on the estimate, such as the time period it relates to (identified by the **Period**, **Quarter** and **Year** columns) as well as technical details about the dataset that the estimate was taken from (such as the **LFSWeight** column). Further details can be found in the 'Data table specification' section below.

This layout is intended to provide BCS estimates in the most transparent and versatile form for users and allow data to be linked with, for example, ONS population estimates for the calculation of numbers of crimes. Further details can be found in 'How to use BCS Open Data tables' below.

All files are made available in CSV (comma separated values) format which arranges data in rows and columns as outlined above. This type of file can be opened in most data analysis programmes.

In addition to the six main tables, two additional reference tables have been included in this release. See the section on 'Reference data tables' below for more information on these.

## FILE NAMING

The six main tables are contained in files which use the following naming structure:

'[Measurement Type]-[Measurement Subcategory]-[Date].csv'

Eg. household-incidence-2010q1.csv

The [Measurement Type] and [Measurement Subcategory] sections of the filename are used to identify the content of the data file. There are three different values for [Measurement Type] each containing two values for [Measurement Subcategory]. An outline of what these values are and a description of what is contained in the files can be found in Table 1.

The [Date] section of the filename identifies the BCS dataset that the release was taken from. It follows the same convention as the **Year** and **Quarter** columns in the tables themselves (see 'Data table specification' below).

**Table 1: File naming**

Measurement Type	Measurement subcategory	Description	Example of data in file
Household	Incidence	Contains incidence rates for household crime	Number of incidents of burglary per 10,000 households in the 12 months prior to interview
	Prevalence	Contains prevalence rates for household crime	Percentage of households that have been victims of burglary in the 12 months prior to interview
Personal	Incidence	Contains incidence rates for personal crime	Number of incidents of violence per 10,000 adults in the 12 months prior to interview
	Prevalence	Contains prevalence rates for personal crime	Percentage of adults who have been victims of violence in the 12 months prior to interview
Perceptions	CJS	Contains perceptions of the police and criminal justice system	Percentage of adults who thin that their local police are doing a good or excellent job
	Other	Contains other perceptions of crime	Percentage of adults who perceive a high level of anti-social behaviour in their local area

## DATA TABLE SPECIFICATION

Data tables are laid out in rows and columns. Each row contains a single BCS estimate, which each column contains information about that estimate. Below is a description of the meaning behind the values in each column.

SOURCE	
<b>Possible values</b>	BCS
This column identifies the source of the estimate. For this release, this column contains only one value: 'BCS'. This field is provided to allow future releases to include data from different sources.	
PERIOD	
<b>Possible values</b>	12 months
This column identifies the period covered by the data used to generate the estimate. For this release, this column contains only one value: '12 months'. This indicates that the estimate is based on 12 months of BCS interviews. This field is provided to allow future releases to include data from different period lengths.	
YEAR	
<b>Possible values</b>	Various
Combined with the <b>Period</b> and <b>Quarter</b> columns, this column identifies the BCS interview period that the estimate is based on. The combination of <b>Year</b> and <b>Quarter</b> identifies the <i>last</i> quarter of interviews that <b>Period</b> covers. For example, Period = '12 months', Year = '2011' and Quarter = '1' denotes that the estimate is based on BCS interviews between April 2010 and March 2011 – ie. The 12 months ending in quarter 1 of 2011.	
QUARTER	
<b>Possible values</b>	Various
Combined with the <b>Period</b> and <b>Year</b> columns, this column identifies the BCS interview period that the estimate is based on. The combination of <b>Year</b> and <b>Quarter</b> identifies the <i>last</i> quarter of interviews that <b>Period</b> covers. For example, Period = '12 months', Year = '2011' and Quarter = '1' denotes that the estimate is based on BCS interviews between April 2010 and March 2011 – ie. The 12 months ending in quarter 1 of 2011.	
LFSWEIGHT	
<b>Possible values</b>	Various
BCS data are weighted to gross to the national population resident in households, calculated from weighted totals from the Labour Force Survey (LFS). This process is known as 'calibration weighting'. LFS data are periodically re-weighted, and as such BCS data are periodically re-weighted to reflect this. This column identifies the LFS weight that was used in the calibration weighting of the BCS dataset that the estimate was taken from. For many users, this is technical detail that they will not use. However, future releases may revise estimates contained in this release to account for changes in calibration weighting. If this is the case, then new data tables will be released covering the same time period, but with slightly different estimates. A change in this column will signify that this is due to a change in LFS weighting.	

<b>MEASUREMENTVAR</b>	
<b>Possible values</b>	Various (see reference table)
<p>This is the name of the variable that is being measured by the estimate. It could also be considered a dependent variable. In this release it is a crime incidence rate, a crime prevalence rate or a perception of crime. Future releases may contain a wider range of measurements. The variable name is a shorthand way of referring to the measurement in question, therefore the meaning of the values in this column may not be immediately obvious to users. A reference table has been provided with this release that contains more descriptive labels for each of the variable names in this column. For users who have access to the main BCS dataset from the UK Data Archive, the names here match directly with the variable names on the main BCS dataset.</p>	
<b>GEOGRAPHY</b>	
<b>Possible values</b>	England and Wales
<p>This column identifies the geographical area that the estimate relates to. For this release, only estimates for the whole of England and Wales are included. Future releases may provide regional breakdowns.</p>	
<b>AGE</b>	
<b>Possible values</b>	16+; 16-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75+
<p>This column identifies the age of the respondent at the time of interview. Respondents are arranged into banded age groups, with the category of '16+' covering respondents in all age groups combined. This column only contains values where <b>MeasurementLevel</b> is 'Person' (see reference table).</p>	
<b>SEX</b>	
<b>Possible values</b>	All adults; Male; Female
<p>The column identifies the sex of the respondent. The category of 'All adults' covers men and women combined. This column only contains values where <b>MeasurementLevel</b> is 'Person' (see reference table).</p>	
<b>HOUSEHOLDTYPE</b>	
<b>Possible values</b>	All households; Vehicle-owning households; Bicycle-owning households
<p>This column identifies the type of household that the estimate relates to, specifically whether the household owns a vehicle or bicycle. Most estimates relate to 'All households'. 'Vehicle-owning households' and 'Bicycle-owning households' provide an alternative measure for the prevalence of vehicle-related crime and bicycle theft respectively and as such only have a value for these measures. This column only contains values where <b>MeasurementLevel</b> is 'Household' (see reference table).</p>	
<b>CHARACTERISTICVAR</b>	
<b>Possible values</b>	Various (see reference table); Total
<p>This column identifies the variable that is used to provide the full breakdown of estimates by respondent characteristics in the <b>Characteristic</b> column. It could also be considered an independent variable. It is similar to the <b>MeasurementVar</b> column in that it contains variable names that match the variable names on the main BCS dataset. More descriptive labels for the variable names in this column can be found in the reference table for this column. The value 'Total' in this column identifies an estimate that covers all respondents, regardless of individual characteristics.</p>	

Users should be aware that some characteristics only exist at the personal level and some only exist at the household level (identified in the **MeasurementLevel** column – see reference table).

**CHARACTERISTIC**

**Possible values** | Various; Total

This column identifies respondent characteristics that the estimate relates to. Characteristics are grouped by **CharacteristicVar** such that each value of **CharacteristicVar** has a number of unique values for **Characteristic** associated with it. As well as those characteristics included on the main BCS dataset within each **CharacteristicVar**, the Open Data tables also include some combined categories that are not included on the BCS core variables. The value 'Total' in this column identifies an estimate that covers all respondents, regardless of individual characteristics and is the only **Characteristic** for the **CharacteristicVar** of 'Total'.

**ESTIMATE**

**Possible values** | Various

This column contains the estimate for **MeasurementVar** for respondents with the characteristics referred to at **Geography, Age, Sex, HouseholdType** and **Characteristic** from BCS interviews conducted in the period identified by **Period, Quarter** and **Year**. When **MeasurementType** is 'Incidence rate' it is a rate per 10,000 adults/households, where **MeasurementType** is 'Prevalence rate' or 'Perception' it is a percentage (see reference table).

**STANDARDERROR**

**Possible values** | Various

This column contains the standard error of the value of **Estimate**. As the BCS is a sample survey, all estimates are subject to a degree of error, reflected in the size of the standard error. See 'How to use BCS Open Data tables' below for more information. More advanced users may also want to know that as the BCS is based on a complex sample design, the standard errors included in these tables are complex standard errors that take the sample design into account. This means that no further adjustments, such as design factors, need to be applied to calculations involving these standard errors.

**UNWEIGHTEDCOUNT**

**Possible values** | Various

Also referred to as the 'Unweighted base', this column shows the total number of respondents who contributed to the calculation of **Estimate**. Estimates based on a larger number of respondents are generally more reliable, and those based on a very small number of respondents should be treated with caution. For National Statistics publications the Home Office does not show estimates based on fewer than 50 respondents and it is recommended that users refrain from using these estimates as well. They have been included here for reference and in the interests of completeness.

## REFERENCE DATA TABLES

There are two reference data tables included with this release. One each for the columns **MeasurementVar** and **CharacteristicVar**. These provide additional information on the variables included in these two columns in the main data tables. They can be used a lookups to include this additional data in combined tables. The data specification for these tables is below.

<b>MEASUREMENTVAR</b>	
<b>Possible values</b>	Various
This includes all the values in the <b>MeasurementVar</b> column of the main data tables.	
<b>MEASUREMENTLABEL</b>	
<b>Possible values</b>	Various
This column contains a text description of the measure that <b>MeasurementVar</b> relates to.	
<b>MEASUREMENTLEVEL</b>	
<b>Possible values</b>	Person; Household
This column identifies the level at which the measurement applies. For personal crimes and perception measures, this column has a value of 'Person' to indicate that the estimate applies to people (specifically adults aged 16 and over). For household crimes, this column has a value of 'Household' to indicate that the estimate applies to households in England and Wales.	
<b>MEASUREMENTTYPE</b>	
<b>Possible values</b>	Incidence rate; Prevalence rate; Perception
This column identifies the type of measurement. 'Incidence rate' means that the measurement shows the number of crimes per 10,000 adults or per 10,000 households (identified by the <b>MeasurementLevel</b> column) in the last 12 months. 'Prevalence rate' means that the measurement shows the percentage of adults/households (identified by the <b>MeasurementLevel</b> column) who have been victims of crime in the last 12 months. 'Perception' means that the measurement shows the percentage of adults who have a particular perception about crime or the criminal justice system.	

<b>CHARACTERISTICVAR</b>	
<b>Possible values</b>	Various
This includes all the values in the <b>CharacteristicVar</b> column of the main data tables.	
<b>CHARACTERISTICLABEL</b>	
<b>Possible values</b>	Various
This column contains a text description of the measure that <b>MeasurementVar</b> relates to.	

## HOW TO USE BCS OPEN DATA TABLES

At the simplest level, BCS Open Data tables can be used to find BCS estimates for certain demographic groups. The CSV files can be imported into most data analysis programmes for this purpose. A full list of the measurements that are included in these tables can be found in the MeasurementVar reference table. A full list of demographic characteristics that these measurements can be analysed by using these tables can be found in the CharacteristicVar reference table.

The data specification table above should make clear to users that to find the estimate they are interested in requires filtering or searching data across different columns to identify the population group they are interested in. For example, overall national estimates for all adults aged 16 and over can be found by selecting '16+' from the **Age** column, 'All' the **Sex** column, and 'Total' from the **Characteristic** column (for personal crime or perception estimates). At the other extreme of detail, users could find estimates for 16-24 year old married women by selecting the appropriate categories from the **Age**, **Sex** and **Characteristic** columns. Other levels of detail can be found by selecting the appropriate values in these fields (or **HouseholdType** for household crime estimates).

When using these estimates, users should be aware of the fact that the BCS is a sample survey and produces estimates with a margin of error around them. As such, the Home Office recommends that users do not use estimates based on fewer than 50 respondents (identified by the **UnweightedCount** column). This feature of the data should also be considered when comparing differences between groups. Standard statistical tests can be used to identify whether differences between demographic groups are 'statistically significant'. That is, whether the differences seen in the BCS data are due to differences in the population of England and Wales as a whole, or whether they occurred by chance as a result of the random selection of respondents for the BCS sample. Users should familiarise themselves with this concept before using these tables.

When comparing estimates, users should ensure that they are only comparing estimates from independent samples. Although the current set of tables only include data for a single time period, future releases will cover other time periods. When comparing data over time, users should consider the time periods that are being used to ensure that they are not overlapping. This can be identified by consideration of the **Period**, **Year**, and **Quarter** columns in the data tables.

Some users may want to combine data in these tables with data from other sources. One common use of this would be to produce counts of crime and numbers of victims by combining the BCS Open Data tables with population data. Incidence rates in these tables show the number of crimes per 10,000 adults or 10,000 households. By multiplying the values in the BCS Open Data tables by ONS estimates of the total number of adults/households in England and Wales divided by 10,000, it is possible to calculate the total number of crimes that took place over a 12 month period. Similarly, prevalence rates show the percentage of adults or households who were

victims of crime in a 12 month period. If these percentages are multiplied by the total number of adults/households in England and Wales, it is possible to calculate the number of victims of crime. Users should bear in mind that when doing this, it is important to use a population estimate for the same subpopulation as that used in the BCS Open Data tables. So, for example, an incidence rate for men aged 16-24 in England and Wales from the BCS Open Data tables should be multiplied by a population estimate for men aged 16-24 in England and Wales to calculate the number of crimes against men aged 16-24 in England and Wales.

These tables are designed to have many uses. The Home Office would like to hear from any users who have developed applications for these data tables and any suggestions for future releases.

---

Phil Hall

Home Office Statistics

29 March 2012

[crimestats@homeoffice.gsi.gov.uk](mailto:crimestats@homeoffice.gsi.gov.uk)