

**Speed camera transparency data for nine areas as analysed in the original and revised versions of the RAC Foundation Report *Guidance on use of speed camera transparency data* by Professor Richard Allsop of University College London**

The areas for which data were downloaded and analysed are as follows:

Area	Area code	Number of cameras analysed*
Cambridgeshire and Peterborough	CBP	47
Leicester Leicestershire and Rutland	LCR	15
Lincolnshire	LCS	50
Merseyside	MSS	33
South Yorkshire	**	56
Staffordshire and Stoke on Trent – few collisions	STF	42
Staffordshire and Stoke on Trent – many collisions	STM	26
Sussex	SSX	55
Thames Valley	THV	203
Warwickshire	WWS	24

\* For just a few cameras, data mounted on the area website did not allow the camera to be included in the analysis.

\*\* For South Yorkshire there are two area codes SYI and SYC: SYI refers to incomplete data with some omitted records as analysed in the original version of the report, and SYC refers to complete data with the omitted records included, as analysed in the revised version.

For each area with code XYZ, data downloaded from the area website and augmented by total numbers of collisions and casualties for the area is arranged in two Excel spreadsheet files named *XYZcsva.csv* and *XYZcsvb.csv*. In the former, column L is based upon Assumption (a) and in the latter it is based upon Assumption (b) about years for which numbers of collisions may have affected the location of the camera, as discussed in the revised version of the report.

Each data file is in comma-separated format with a header row and up to  $21n$  rows of data, where  $n$  is the number of cameras analysed. For each camera, there is usually one row of data for each of the 21 years 1990-2010 inclusive, which are numbered from 0 in 1990 to 20 in 2010, but for some cameras in some areas the data for some years are missing, and in such cases the rows for those years are omitted. The spreadsheet has the following 16 columns.

Column	Header	Contents
A	areacode	Area code
B	siteid	Local identifier of camera
C	camid	Camera identifier used in analysis
D	year	Year from 0 in 1990 to 20 in 2010
E	fsc	Number of fatal or serious collisions*
F	pic	Number of personal injury collisions*
G	ksicas	Number of casualties killed or seriously injured*
H	allcas	Number of casualties of all severities*
I	datest	Date of establishment of camera <sup>†</sup>
J	speedlim	Speed limit at camera site in miles/h
K	camact	1 if camera was established throughout the year, 0 if not
L	baseline	1 if it is assumed that the number of collisions in the year may have affected the location of the camera, 0 if not

M	cntyfsc	Number of fatal or serious collisions in the area
N	cntypic	Number of personal injury collisions in the area
O	cntyksicas	Number of casualties killed or seriously injured in the area
P	cntyallcas	Number of casualties of all severities in the area

\* recorded in the vicinity of the camera

† dates appearing as ##### can be read by placing the cursor on the cell concerned

The estimate of multiple  $m_c$  for effects of camera establishment on numbers of PIC per year in the vicinity of cameras across partnership area XYZ as given in Table A2.1 of the report is obtained using the software **R** by fitting the model *XYZareaPIC* using the commands shown below. The estimate given in Table 2.2 of the report for the corresponding multiple for numbers of FSC per year across area XYZ is obtained by fitting the model *XYZareaFSC*. The estimated multiples for numbers of PIC in the vicinity of individual cameras in area XYZ are obtained by fitting the model *XYZcamerasPIC*.

The first command in **R** for fitting these models is

```
>DataXYZ=read.csv('XYZcsva.csv', header=TRUE, sep=",")
```

if the models are to be fitted under Assumption (a), and

```
>DataXYZ=read.csv('XYZcsvb.csv', header=TRUE, sep=",")
```

if the models are to be fitted under Assumption (b).

In either case the subsequent commands are:

```
>fcamid=factor(DataXYZ$camid)
```

```
>XYZareaPIC=glm(pic~offset(log(cntypic))+fcamid+camact+baseline,family=poisson(link=log),data=DataXYZ)
```

```
>XYZareaFSC=glm(fsc~offset(log(cntyfsc))+fcamid+camact+baseline,family=poisson(link=log),data=DataXYZ)
```

```
>XYZcamerasPIC=glm(pic~offset(log(cntypic))+fcamid+camact+fcamid*camact+baseline+fcamid*baseline,family=poisson(link=log),data=DataXYZ)
```

The results are listed by means of the commands:

```
>summary(XYZareaPIC)
```

```
>summary(XYZareaFSC)
```

```
>summary(XYZcamerasPIC)
```

The current directory for **R** needs to be the one in which the file *XYZcsv.csv* is located, and work in **R** on area XYZ needs to begin with the appropriate one of the above two first commands, followed by the command `>fcamid=factor(DataXYZ$camid)`. After these two commands, models can be fitted and results of fitted models requested in any order.

In the results for the first two models, the estimates for coefficients *camact* and *baseline* are the  $v$  and  $u$  respectively of Appendix 2 of the report, and the estimates of  $m_c$  and  $m_b$  are  $\exp(\text{camact})$  and  $\exp(\text{baseline})$ . The standard errors of these two coefficients, given in the same lines of the results, enable the estimated confidence intervals for  $m_c$  and  $m_b$  to be calculated as described in Appendix 2 of the report.

From the results from the third model, the estimates of  $m_c$  and  $m_b$  for camera XYZC1 and their confidence intervals are obtained in the same way. But for each other camera XYZCn

the estimates of  $m_c$  and  $m_b$  are obtained from the coefficients  $fcamidXYZCn:camact$  and  $fcamidXYZCn:baseline$  together with the coefficients  $camact$  and  $baseline$  in the forms

$$m_c = \exp(fcamidXYZCn:camact + camact) \text{ and}$$

$$m_b = \exp(fcamidXYZCn:baseline + baseline)$$

respectively. To calculate estimated confidence intervals for these estimates of  $m_c$  and  $m_b$  for cameras other than camera XYZC1, the standard errors of  $fcamidXYZCn:camact + camact$  and  $fcamidXYZCn:baseline + baseline$  are calculated using the formula:

$$\text{Standard Error}(x + y) = [\text{Variance}(x) + 2\text{Covariance}(x,y) + \text{Variance}(y)]^{1/2}.$$

A table of the variances and covariances needed to apply this formula is obtained by means of the **R** command:

```
>vcov(XYZcamerasPIC)
```

The arrays of coefficients and associated numbers produced by **R** can be transferred to an Excel spreadsheet for further calculation by selecting the whole block of lines in the output as displayed by **R**, copying to cell A1 of an empty Excel worksheet, and then clicking on 'Text to columns' in the dropdown menu 'Data' of Excel and clicking 'next' twice and then 'finish' in the panels that appear. Output from **R** can also be copied to text files.

Richard Allsop

November 2013