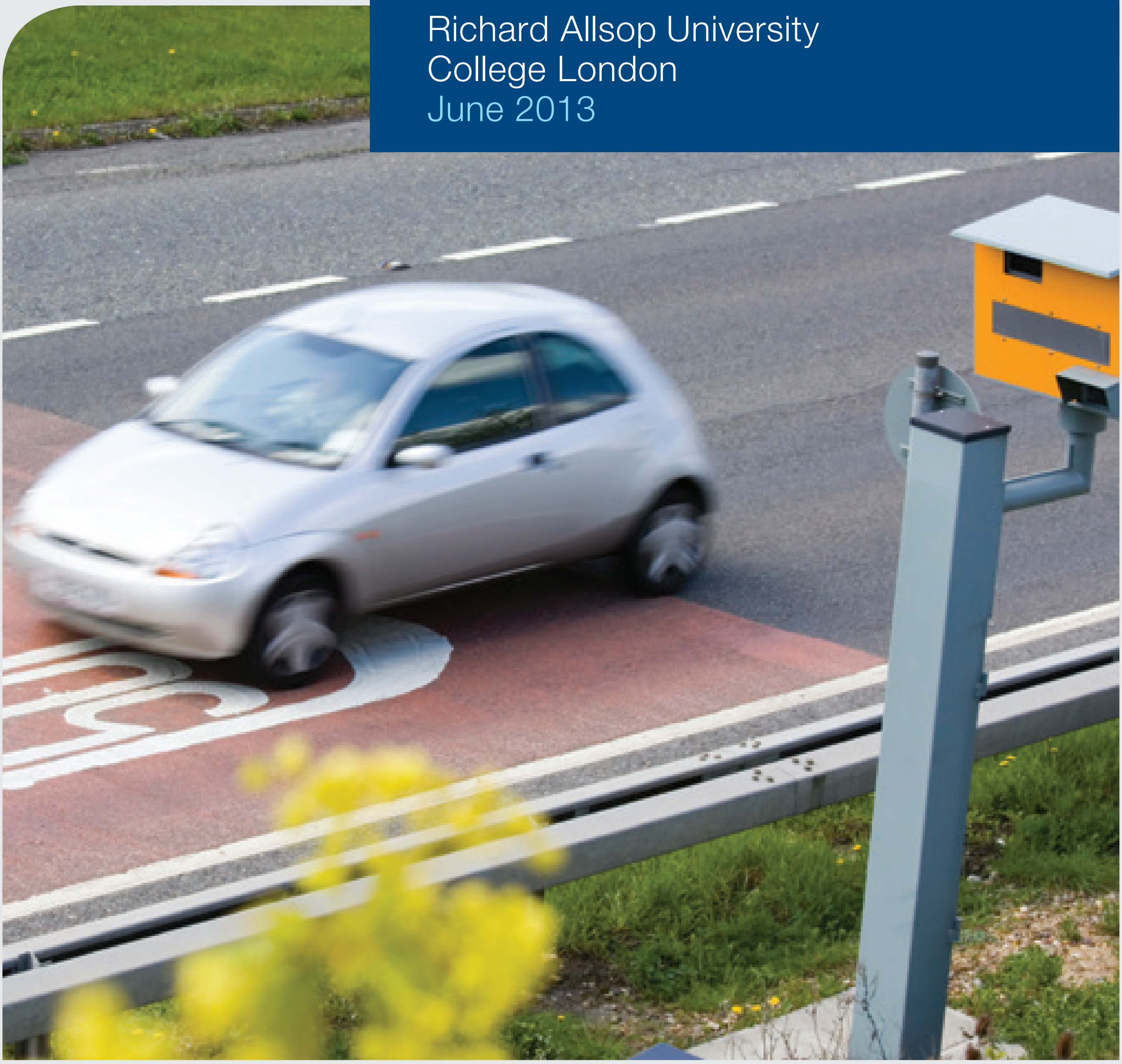
RAC Foundation

Guidance on Use of Speed Camera Transparency Data

Updated November 2013
with Addendum March 2019

Richard Allsop University College London
June 2013



RAC Foundation

The Royal Automobile Club Foundation for Motoring Ltd is a transport policy and research organisation which explores the economic, mobility, safety and environmental issues relating to roads and their users. The Foundation publishes independent and authoritative research with which it promotes informed debate and advocates policy in the interest of the responsible motorist.

RAC Foundation
89–91 Pall Mall
London
SW1Y 5HS

Tel no: 020 7747 3445
www.racfoundation.org

Registered Charity No. 1002705
June 2013 (superseded in November 2013, with Addendum March 2019)

RAC Foundation

Guidance on Use of Speed Camera Transparency Data

Superseded in November 2013 with Addendum March 2019

Richard Allsop University College London

June 2013

About the Author

Professor Richard Allsop has extensive experience of research, training and advisory work on road safety, traffic management and other aspects of transport policy. He has a first in Mathematics from Cambridge, and a PhD and DSc from University College London, where he is Emeritus Professor of Transport Studies, having been Professor since 1976 and Director between then and 1997 of what is now the Centre for Transport Studies.

He has a longstanding involvement in road safety research and policy, including being a Director of PACTS (the Parliamentary Advisory Council for Transport Safety). He is a Board Member of the European Transport Safety Council (ETSC) and leads its European road safety performance index programme PIN. He has also provided inputs to road safety policy in Australia, Hong Kong, Japan, New Zealand and Poland.

He was made an OBE in 1997 for services to traffic management and road safety, is a Fellow of the Royal Academy of Engineering and holds the IHT Award for professional excellence. His report The Effectiveness of Speed Cameras: A review of evidence for the RAC Foundation was given the 2011 Prince Michael Road Safety Award.

Acknowledgements

The analyses on which this report is based could not have been carried out in the time available to the author without extensive data handling to high standards of quality and requiring considerable ingenuity, for which the author is grateful to Dr Reza Tolouei. The statistical modelling was helped by an invaluable suggestion from Jeremy Broughton and encouraged by comments from Jeremy Broughton and Sylvain Lassarre as peer reviewers. Excel spreadsheets of numbers of collisions and casualties in local authority areas in England in the years 1990–2010 were kindly provided by colleagues in Road Accident Statistics in the DfT. Revisions to the report have been informed by comments and analysis from Professor Mike Maher, and comments by Dave Finney and Idris Francis.

Disclaimer

Contents

Foreword

Speed cameras – are they effective? Are they in the right places? What have we learnt from over 20 years of their use?

These are not new questions. Efforts have already been made to provide the answers, RAC Foundation reports included. So why revisit the subject?

In 2011, local authorities were required to provide data on offences, collisions, casualties and speeds at fixed camera sites in response to the Coalition Government's drive for transparency. The information mounted in response to this requirement is now available for the public and professionals to view, analyse and interpret.

Problem solved? Case closed? Controversy over? Well… not quite.

Although the data as mounted is now in the public domain, there is no guidance for the general public, media or practitioners on how to use it in a robust and consistent way.

In the absence of any formal guidance, the RAC Foundation commissioned Professor Richard Allsop of University College London to provide a view on how available speed camera data can be best analysed. His findings are detailed in this report.

Using the available data in a scientifically robust way is no easy task. The paper begins with a non-technical Explanatory Note which explains the issues. This will be helpful to all those interested in this topic and will aid understanding of why naive or selective use of the data could be used to draw misleading conclusions.

The body of the paper and the appendices give a detailed exposition of the subject. It gives a step-by-step guide to obtaining the data and applying a simple manual approach, together with worked examples. It also details a fully rigorous approach (yielding similar results) which will be of interest to readers with a good understanding of statistical analysis.

We commend the work of Professor Allsop as a big step forward in the analysis and development of policy and practice in this important area of road safety.

In the fast-developing world of open data, the wider lesson to government is that while users should be free to make their own analyses and interpretations of data that is made available, objective and non-directive advice about the techniques that can be used for its analysis must be provided.

It is unacceptable that users of speed camera data have been left with no guidance on how to interpret the figures, even though the data have for some years given rise to both controversy and genuine difficulty in interpretation. We would encourage the government to learn from this experience and ensure future releases of large sets of data are accompanied by appropriate guidance and support for users.

November 2013

The first edition of this report, published in June 2013, gave rise to both media attention and extensive technical exchanges with the author. Constructive criticism, debate and revision are essential to scientific progress. Professor Allsop is now able to improve his advice in one respect which will be useful to some users of the data and he has reworked his example calculations.

We are publishing this second edition to incorporate these improvements. Readers will be able to see the differences in the illustrative results: the estimates of reductions in collisions attributable to the cameras are still substantial though the reworking leads to somewhat smaller estimates.

The process of scrutiny and revision in response to comment strengthens confidence in the overall conclusion: the evidence is that on the average speed cameras are effective in reducing collisions and saving lives.

Stephen Glaister

Stephen Glaister
Director, RAC Foundation

Abstract

Since 2011, certain data relating to fixed speed cameras in many parts of England has been available to the public in the interests of transparency. Analysis of these data can provide estimates of changes in numbers of collisions and casualties in the vicinity of cameras, typically on between 0.4 km and 1.5 km of road, and in the speed of traffic there, following the establishment of cameras. This technical report, introduced by an explanatory note, discusses a number of ways in which the data can be analysed, and provides for users of the data practical advice on the scope and nature of the available data and on their analysis and interpretation. Comments are made on lessons to be learnt from the making available of the data. Results of a range of example analyses are presented in the Appendices.

Passages in this edition of the report which indicate the main changes from the June 2013 edition appear like this paragraph. There are many consequential changes in detail in other parts of the text.

Explanatory Note

Since the summer of 2011, certain data relating to fixed speed cameras in many parts of England has been available to the public in the interests of transparency. The data is mounted on the websites of local authorities or road safety partnerships, and a list of the websites can be found on the website of the Department for Transport (DfT). Not as many areas now have road safety partnerships as has previously been the case, but the areas to which the data relates are called *partnership areas* in this report.

The data concerns: year-by-year numbers of collisions and casualties in the vicinity of each camera, typically on between 0.4 km and 1.5 km of road, between 1990 and 2010; observations of the speed of traffic near the camera on certain dates; and information about numbers of offences detected by the cameras and actions taken in respect of the offenders.

As the data began to be published and received some media attention, the RAC Foundation noted the difficulty of interpreting the information provided. Of particular concern was how the numbers of collisions and casualties in the vicinity of speed cameras and the speed of passing traffic are affected by the presence of the cameras. Guidance for the general public, the media and road safety practitioners on interpreting data for individual cameras or cameras in local areas seemed to be called for. This report stems from work by the author, commissioned by the RAC Foundation, to try to meet this need, mainly in respect of the data concerning numbers of collisions and casualties, with limited attention to the data concerning speeds of traffic.

To provide a basis for developing guidance for users of the data, data from nine local authorities or partnerships was downloaded and prepared for statistical analysis. Analysis was undertaken both with and without the use of statistical computer software, and was then subjected to independent peer review. The resulting analyses are outlined in the report and described in some detail in three Appendices. The analysis focuses mainly on estimating changes in the numbers of collisions and casualties in the vicinity of cameras following camera establishment.

This report offers practical advice to users on the nature of the available data and on its analysis and interpretation. Comments are made on lessons to be learnt from making the data available to the public.

The nature of the data

Where DfT recommendations on the coverage of the data have been followed, the data includes, for each camera and for each of the twenty-one calendar years 1990–2010, the numbers of:

- fatal or serious collisions (FSC);
- personal injury collisions of all severities (PIC);
- people killed or seriously injured (KSI) in the FSC; and
- casualties of all severities (CAS) in the PIC.

These numbers are those recorded in the national road accident data system known as STATS19 as occurring in the vicinity of the camera in that year.

The date of establishment of the camera and the speed limit at the camera site are also provided together with, for some but not all cameras, measurements made from time to time of traffic speed in the vicinity of the camera. Numbers of offences are also provided, but these are not discussed in this report. The date of establishment of the camera is the date from which it may at any time have been in operation. Just when the camera has since been in operation has been for the partnership to decide, subject to the occurrence of incidents affecting operation.

Numbers of PIC and CAS are typically up to about ten times the corresponding numbers of FSC and KSI, and numbers of casualties cannot be smaller than the corresponding numbers of collisions. As well as being subject to many systematic influences, including the presence of cameras, all the numbers of collisions or casualties are subject to natural fluctuation known as *random variation*, and this is greater for numbers of casualties than for numbers of collisions. It therefore makes sense to look for systematic influences first by analysing numbers of PIC and FSC, and look to numbers of CAS and KSI mainly for supplementary information about the severity of collisions.

The date of establishment of a camera distinguishes preceding years, with no camera at that location, from subsequent years, throughout which the camera may have been in operation. In comparing these two periods, it is advisable to allow for the possibility that collision or casualty numbers in the vicinity of a camera may have been unusually high in years just before the decision to establish the camera, and so contributed to the choice of location. Where this is so, a subsequent fall in numbers is to be expected through the phenomenon known as *regression to the mean*.

Available by separate extraction from the STATS19 system are the annual numbers of FSC, PIC, KSI and CAS in each local authority area. These provide an indication of how systematic influences affecting the occurrence of collisions or casualties throughout the local area may have affected the year-on-year numbers of collisions or casualties in the vicinity of cameras.

Where observations of speed of traffic in the vicinity of a camera were made before and after camera establishment, estimates can be made of change in speed following establishment.

Practicalities of using the data

The report discusses the practicalities of using the data for any one partnership area to examine changes that have occurred in the vicinity of cameras in that area. Discussion begins with the accessing and organisation of the data and goes on to consider its analysis and interpretation, first for a number of cameras taken together and then for individual cameras.

It is assumed that the user will have use of a laptop or other personal computer with spreadsheet software. Internet access is required for initial downloading of data, but not for its analysis or interpretation. Those who wish to use statistical calculations outside the scope of their spreadsheet software will need additional software such as has been used in the report, but the report also discusses and illustrates analysis without such software.

The method of transferring the data to the user's own spreadsheet software depends on how the partnership has mounted it, ranging from simple downloading of one or a few spreadsheets, through downloading and copying from one or more pdf files, to downloading camera by camera for cameras identified on a map on the partnership website. Suggestions are made for arranging the data for each camera in the user's spreadsheet software, so that the numbers of collisions and casualties for the whole partnership area can be included ready for analysis.

Estimating changes in numbers of collisions

Estimates of changes in numbers of collisions in the vicinity of cameras following their establishment can be made for individual cameras, but the estimate for a camera of interest is hard to interpret without also considering corresponding changes for other cameras in some group to which that camera belongs. This group might be all cameras in the partnership area or some subgroup of these, such as those with a particular speed limit or in a local area of interest. The number of cameras in the group should be at least about ten, and preferably several tens.

Changes in the occurrence of collisions across such a group of cameras relative to changes across the partnership area can be estimated by considering, for each camera and year, the ratio:

$$\frac{\text{number of collisions in the vicinity of the camera in that year}}{\text{number of collisions in the partnership area in the same year}}$$

Changes following establishment of cameras and any excess of collisions in years just before the decisions to establish the various cameras can be estimated in terms of multiples by which this ratio changed. Once these multiples have been estimated, they can then be expressed as estimated percentage changes. Estimates can be made separately for numbers of PIC and FSC.

It is open to the user to adopt any appropriate method for estimating these multiples. The author’s preferred ways of doing so, with and without the help of statistical computer software, are described and illustrated in two Appendices. Results obtained with and without the statistical software were found to match closely in two example comparisons presented in the report.

Figures 1 and 2 show the results of analysing data for 24 cameras in one partnership area (Warwickshire) in these ways to estimate changes in numbers of collisions in the vicinity of cameras following their establishment, first across the partnership area and then for individual cameras.

Interpreting estimates for individual cameras

For PIC the multiples can be estimated, albeit usually with rather wide confidence intervals, for each individual camera in a group. To help in interpreting the resulting values, it is useful to consider how they are distributed. For the multiple representing change following the establishment of the camera, the estimates for nearly all the cameras in the group are typically distributed quite densely over the range from zero up to a value between about 1 and about 2, with larger estimates for just a few of the cameras.

The larger estimates are clearly of interest in view of the possibility that a camera with a large estimated multiple may have been located in such a way that the number of PIC in its vicinity has thereby been increased. For each such camera it is advisable to examine the year-on-year numbers of PIC for features such as particularly small numbers that make it unwise to take the estimate at face value. Where no such features are evident, a large estimated multiple may be taken as indicating that the camera concerned should be checked to see whether its operation may be a source of increased risk of collision in its vicinity. Similar examination is appropriate for cameras with estimated multiples markedly nearer to zero than the lowest of the densely distributed values, to check whether it is wise to take these favourable results at face value.

Values of the estimates within the densely distributed range might be thought to indicate that cameras with estimated multiples appreciably below 1 are doing a good job, those with multiples around 1 are doing neither much good nor much harm, but the future of those with multiples appreciably above 1 should be called into question. But this would be simplistic, because there is no such determinate relationship between the estimated multiple for a camera and the contribution the camera has made to the aggregate change across the group. It is suggested instead that, where the estimated change across the whole group of cameras is favourable, all the cameras with multiples in the densely distributed range should be regarded as having potentially contributed to it.

Where the future of any of these cameras is called into question, this should be on the basis of evidence external to the data that is the subject of this report. The estimate of the multiple for that camera might be cited to corroborate or counter the external evidence, but in doing so the width of its confidence interval should be taken into account.

Where the estimated aggregate change is unfavourable, scrutiny of the operation of the group of cameras is advisable. Consideration of the estimated multiples for the individual cameras may then form part of this scrutiny, but once again the widths of their confidence intervals should be taken into account.

Other analyses

Numbers of CAS, FSC, and KSI can be used together with corresponding numbers of PIC to calculate indicators of severity of collisions and casualties in the vicinity of cameras, such as the ratio:

$$\frac{\text{number of CAS}}{\text{corresponding number of PIC}}$$

which is the number of casualties per collision. Changes in these indicators following the establishment of cameras can be estimated, and example calculations are provided in one of the Appendices.

In a limited analysis, multiples representing changes in the occurrence of PIC in the vicinity of individual cameras are compared with observed changes in average speed of traffic near the same cameras. There was a wide range of changes in average speed, but no relationship is evident between changes in numbers of PIC and changes in speed.

Use of more limited data

A different approach is needed, with associated limitations, to provide analysis and interpretation for partnership areas where little or no data is provided for years before the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the cameras. In particular, this is the case where the only data provided for years before the establishment of cameras is so-called *baseline data*, which relates typically to a period of three years that ends a year or two before establishment of the camera. Assumptions are discussed under which some analyses can be made in these cases with the help of results from areas for which the data goes back to 1990 and from earlier work.

Lessons from this exercise in providing transparency

In the fast-developing world of open data, the experience of mounting of speed camera data points to several practicalities that warrant attention:

- Asking holders of data to make them available in a recommended form does not necessarily result in the data being made available in that form or at all
- Websites and their addresses often change, so any central source of such addresses needs to be robust with respect to such changes
- Users will want to work with data, not just read information on a screen or printout, so data should be mounted in a format that enables use with minimum transcription, that is in a spreadsheet or analogous format
- While users should of course be free to make their own analyses and interpretations of data that is made available, this can be helped by objective and non-directive advice about the nature and characteristics of the data concerned and pointers towards available techniques that are appropriate for application to data of that kind

Example analysis of data from Warwickshire

Warwickshire mounted data for 25 cameras, but one was established only in July 2010, so analysis is possible for only 24 cameras. The data is mounted in a single pdf file. This was downloaded and the data transferred to an Excel spreadsheet and arranged so that it could be copied directly to the statistical software after the annual numbers of collisions and casualties for the county were added. The software was used to analyse the data as described in Appendix 2 of the report.

A statistical model was fitted to represent, for each camera and year, the ratio:

$$\frac{\text{number of collisions in the vicinity of the camera in that year}}{\text{number of collisions in the partnership area in the same year}}$$

This was influenced by two multiples, each common to all cameras: one applying in years throughout which the camera concerned was established; and the other applying in the first three of the last four years before the year the camera was established. These multiples are indicators of change in number of collisions in the vicinity of cameras relative to numbers in the whole county.

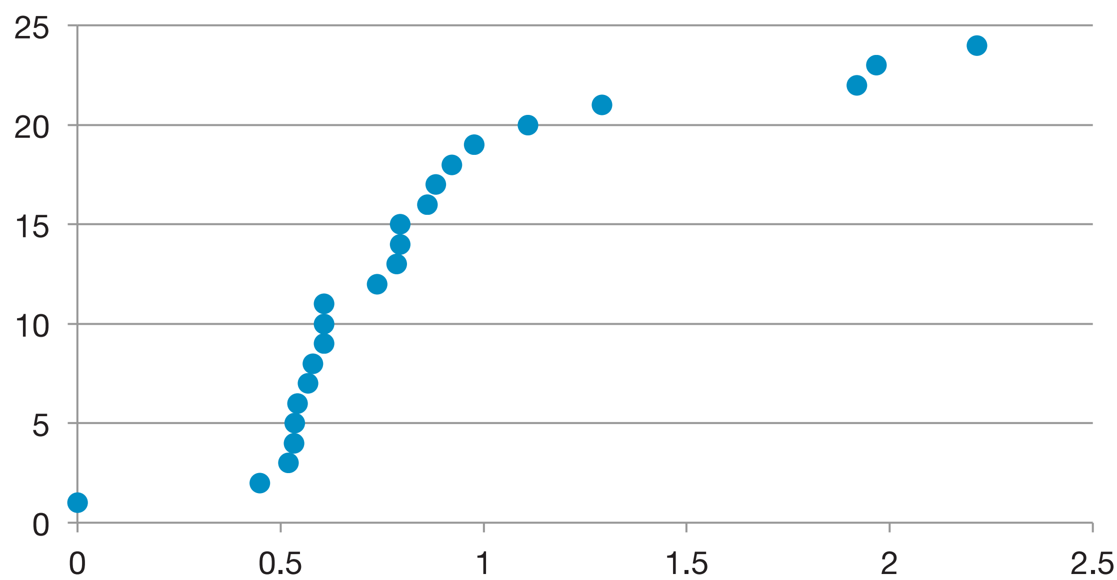
The model estimated the first of these multiples for PIC to be 0.78, with 95% confidence interval (0.67, 0.90). This indicates that, after camera establishment, the numbers of PIC had fallen by 22%, with a 95% confidence interval from a fall of 33% to a fall of 10%, from the numbers more than four years before and just prior to camera establishment.

The second multiple for PIC was estimated to be 1.33, with 95% confidence interval (1.16, 1.53), indicating that, in the first three of the four calendar years before camera establishment, the numbers of PIC had been 33% higher than in other years between 1990 and camera establishment, with a 95% confidence interval from 16% higher to 53% higher.

The corresponding percentages for numbers of FSC were a fall of 32%, with a 95% confidence interval from a fall of 54% to a fall of 3%, and 114% higher with a 95% confidence interval from 59% higher to 164% higher.

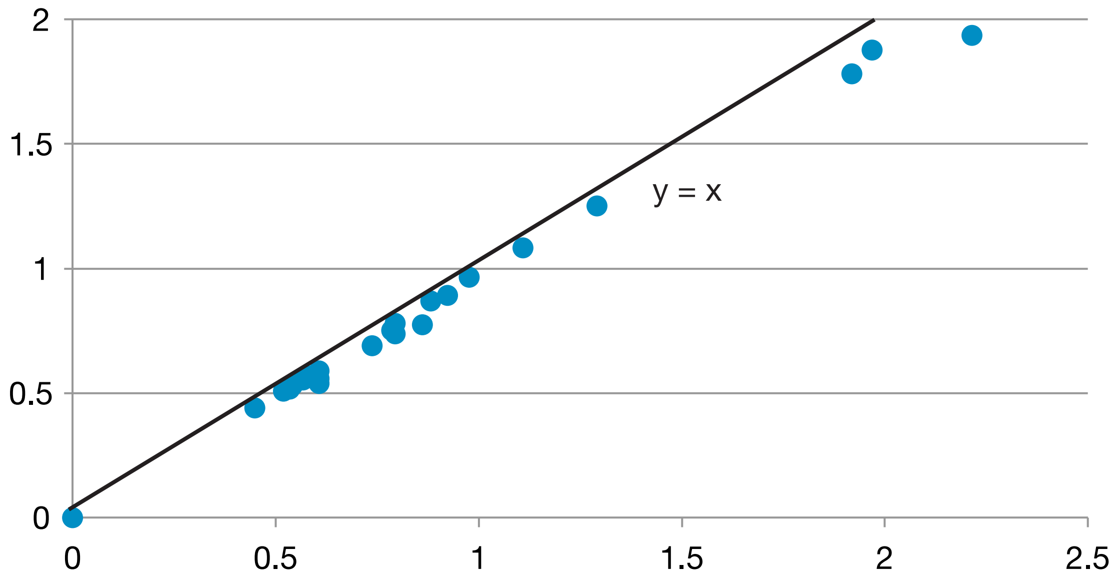
For PIC a further statistical model was fitted, in which separate values of the multiples were estimated for each camera. The 24 values of the multiple representing change in numbers of PIC after camera establishment are shown in Figure 1. The densely distributed values range from about 0.5 to 1. The value zero and two of the five values exceeding 1 are based on only one or two years' data after camera establishment, and should be treated with caution. The other three values exceeding 1 may indicate checking of the effect of the cameras concerned.

Figure 1: Warwickshire PIC: Cumulative distribution of model estimates of 24 camera multiples



Values of the same 24 multiples were estimated manually by a method described in Appendix 3 of the report. In Figure 2, the match between the manual estimates, on the vertical axis, and the ones from the statistical models, on the horizontal axis, provides support for the use of manual calculation where users of the data prefer this.

Figure 2: Warwickshire PIC: Manual vs model estimates of 24 camera multiples



1. Background

On 26 June 2011 the Road Safety Minister announced (DfT, 2011a) that English local authorities would publish figures showing the numbers of accidents and casualties at permanent and fixed speed camera sites, both before and after cameras were installed, and that police forces would publish the number of speeding prosecutions arising from each camera and force-wide information about whether offenders have been fined, completed a speed awareness course or been taken to court. Information about measured speeds of traffic at camera sites was also published. A working group had advised (DfT, 2011b) on what information should be published and how, including recommending that annual collision and casualty data for each camera should usually go back to 1990.

Authorities were required (DfT, 2011c; 2011d) to provide the data, broadly but not in every respect as recommended by the working group, as soon as practicable, on websites whose addresses were to be notified to the DfT by 20 July 2011. The DfT established a central hub (DfT, 2011e) providing links to these local websites. The stated purpose of publication was to improve transparency and accountability to the public. The extent and forms of mounting of data on the local websites by late November 2012 is summarised in Appendix 1.

As the data began to be published and received some media attention, the RAC Foundation noted the difficulty of interpreting the data in terms of how site-by-site numbers of collisions and casualties (and to a lesser extent speed) are affected by the presence of speed cameras. Guidance for the general public, the media and road safety practitioners on interpreting data for single camera sites or sites in local areas seemed to be called for.

In discussion with and commissioned by the RAC Foundation, the author arranged for data from a number of road safety partnerships or local authorities to be downloaded in a format that facilitated statistical analysis, and then undertook a range of analyses as a basis for development of guidance for users of the data. This report outlines the analyses undertaken, providing detail in technical Appendices, and goes on to offer advice on the use of the data. In conclusion, some comments on what might be learnt from this exercise in transparency are offered.

2. Analysis

Attention was concentrated on data that is mounted in accordance with the DfT guidance (DfT, 2011b; 2011d). Collision, casualty and speed data for nine of the partnerships or authorities that have mounted data in this way was downloaded for analysis. This covered a mixture of metropolitan and shire counties and ranged geographically from Warwickshire to Lincolnshire and from Merseyside to Sussex. Not as many areas now have road safety partnerships as has previously been the case, but the areas to which the data relates are called ***partnership areas*** in this report.

Statistical modelling by means of a freely available software package known as **R** (R Development Core Team, 2011)[1] was used to explore the estimation of the changes, following the establishment of cameras, in numbers of fatal or serious collisions (FSC) and personal injury collisions of all severities (PIC) across the cameras in each partnership area and in the vicinity of individual cameras, typically on between 0.4 km and 1.5 km of road. The modelling allows for year-on-year changes in collision occurrence in the partnership area. It also largely excludes the effects of the tendency, known as *regression to the mean*, for collisions in the vicinity of some cameras to be fewer after camera establishment than in recent previous years irrespective of the effect of the camera, as discussed in section 3.2. Numbers of collisions were analysed rather than

1 Other statistical software for fitting of generalised linear models should yield very similar results.

numbers of casualties because, as discussed in section 3.1, this offers greater scope for distinguishing between differences arising largely by chance and the effects of systematic influences such as the presence of a camera.

Because not all those wishing to use the data will have access to statistical software or skills in its use, manual calculations were carried out for two example partnership areas so that the results could be compared with those given by the software. The manual calculations were extended to provide indicators of changes in the severity of collisions.

In a further analysis covering cameras for which there were suitable speed data in eight of the partnership areas, estimated changes in the occurrence of PIC in the vicinity of individual cameras after camera establishment were compared with observed changes in the average speed of traffic in the vicinity of the same cameras.

These analyses and their broad outcome are summarised in the following paragraphs, and details can be found in Appendices 2, 3 and 4.

For each camera, the years 1990–2010 were divided into three sets:

(1) years throughout which the camera had been established, and may therefore have been in operation;

(2) years for which the number of collisions may have been used in selecting the location of the camera; or

(3) years which were neither of these.

To identify the second set of years requires extra information to be requested from the camera partnership concerned. The analyses reported here were made from the standpoint of users who wish to interpret the published camera data without making such requests. They are therefore based on assumptions about this set of years. Because these years are typically three consecutive ones ending a year or two before establishment of the camera, and the available data are for calendar years, two assumptions have been made; these are that the second set of years consist of:

(a) the last three calendar years before the camera was established; or

(b) the first three of the last four years before the camera was established.

As discussed further in section 4.1, Assumption (a) approximates establishment of a camera rather briskly after a period of three years for which collision data are used to select the site, whereas Assumption (b) approximates more typical intervals between the three-year period and the establishment of a camera.

Under Assumption (a), the third set of years comprises the years more than three calendar years before the camera was established and the year during which it was established. Under Assumption (b), it comprises the years more than four calendar years before the camera was established, the year during which it was established and the preceding year. The three sets of years under these two assumptions are shown here in relation to the year of establishment of a camera:

It has been suggested that the year in which the camera was established should be omitted from the third set of years on the basis that it will consist in general of some months before and some months after establishment. In the analyses reported here, however, it is included, because it is appropriate that the months before establishment contribute to the estimation of level of collision and casualty occurrence before establishment, while inclusion of the months after establishment, during which collision and casualty numbers will be subject to any effect of the camera, will on average result in only a small underestimation of that effect, making the resulting estimate on average slightly conservative.

Numbers of collisions in the first set of years were compared with numbers in the third set, each relative to the total number of collisions in the partnership area in the year concerned, to estimate the percentage by which the number of collisions had changed following establishment of the camera. The numbers of collisions in the second set of years were used similarly to estimate any extent to which these numbers may have been unusual compared with other years before establishment of the camera. This allows the effect of regression to the mean to be excluded from the estimation of change following the establishment of the camera, to the extent that the second set of years includes any years for which the numbers of collisions were used in selecting the location for the camera.

Analysis in the first edition of this report (June 2013) was based on Assumption (a), but comments on that analysis and further examination of the data indicate that Assumption (b) is preferable, as discussed in section 4.1. Results quoted in the main text are therefore based on Assumption (b), and results based on Assumption (a) are provided in Appendices 2 and 3 for comparison. In general, the latter results give somewhat higher estimates of the effect of cameras because they exclude less of the effect of regression to the mean.

Consideration of numbers of collisions in the vicinity of the camera relative to the corresponding total numbers in the area allows for trend and other systematic changes in collision occurrence.

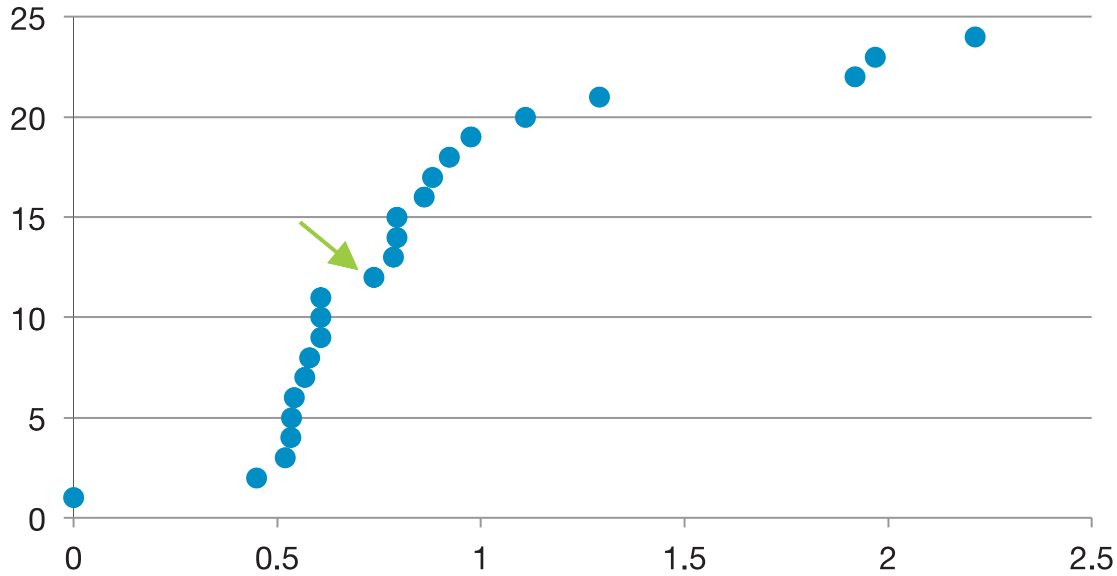
Making these allowances, estimates were made first of the percentage by which the number of FSC and PIC in the vicinity of cameras had changed following establishment of the cameras across each partnership area. For Warwickshire, for example, with 25 cameras, of which one was established only in July 2010 – too late to provide a full year's data after establishment – the resulting estimates were that the numbers of:

- FSC had fallen by 32%, with a 95% confidence interval from a fall of 54% to a fall of 3%, from the numbers more than four years before and just before camera establishment; and
- PIC had fallen by 22%, with a 95% confidence interval from a fall of 33% to a fall of 10%, from the numbers more than four years before and just before camera establishment.

Corresponding estimates for the other partnership areas covered by the analysis can be found in Appendix 2.

The analysis was extended with a view to making corresponding estimates for individual cameras. In the case of FSC, because of the small numbers of such collisions year by year at many cameras, many of the estimates of changes at individual cameras had upper confidence limits several times the estimated change, or which indicated that the change at the camera concerned, taken by itself, could not be determined from the data. This means that it would be unwise to place reliance upon estimates of changes in numbers of FSC at individual cameras. In the case of PIC, with their larger numbers, the confidence intervals for the estimates of changes at individual cameras were rather wide, but not so wide as to prevent useful consideration of the distribution of the estimates for the cameras in a partnership area. Figure 3, for example, shows the distribution for cameras in Warwickshire.

Figure 3: Warwickshire PIC: Cumulative distribution of model estimates of 24 camera multiples



The horizontal axis in Figure 3 shows the estimated ratios of the number of PIC per year after camera establishment to the number before establishment at the 24 individual cameras; in other words, the figure displays the numbers after camera establishment as estimated multiples of the numbers before establishment (after largely excluding the effect of regression to the mean and allowing for general changes in collision occurrence across Warwickshire). For the camera indicated by the arrow in Figure 3, for example, there were 4 PIC in 6 years after camera establishment, or 0.67 per year, compared with 14 PIC in 12 years, or 1.17 per year, up to establishment other than in years in which it was assumed that numbers of collisions may have affected the location of the camera. The corresponding numbers of PIC recorded in Warwickshire as a whole were 10,681 and 27,586, respectively, or 1,780 per year compared with 2,299 per year. So the multiple shown in the chart should be about (0.67/1.17)/(1,780/2,299), which, after a small correction for bias described in Appendix 3, becomes about 0.70. The value estimated by the statistical model is 0.74, and the slight difference arises from the correction for bias.

It can be seen that the estimated multiples for nearly all the Warwickshire cameras are distributed quite densely over the range from about 0.5 to just over 1, with larger estimates for five of the cameras. For nearly 80% of the Warwickshire cameras, the estimated multiple was less than 1, indicating fewer PIC per year in the years after camera establishment than in the years well before and just before establishment. Interpretation both of densely distributed estimates and of smaller and larger estimates, like the zero estimate and those between 1.9 and 2.3 in this case, is discussed in section 4.3.

Results for the other partnership areas covered by the analysis can be seen in Appendix 2. The distributions are qualitatively broadly similar, but the numerical ranges of the densely distributed multiples for individual cameras differ substantially among partnership areas, even among those where the estimated percentage changes in numbers of PIC in the vicinity of cameras across each partnership area as a whole were quite similar.

For two partnership areas – Leicester, Leicestershire and Rutland, and Warwickshire – the same multiples were estimated by manual calculation as described in Appendix 3. The manual estimates are compared with the estimates obtained by means of statistical models using computer software in Figures 4 and 5. The estimates obtained by the two methods are seen to be very similar – as, according to statistical theory, they should be – but the manual estimates are systematically slightly lower because the manual calculation includes a correction for bias which is not made in the statistical models. The larger the numbers of collisions in the years before camera establishment, the smaller is the effect of the correction.

Figure 4: Leicester, Leicestershire and Rutland PIC: Manual vs model estimates of 15 camera multiples

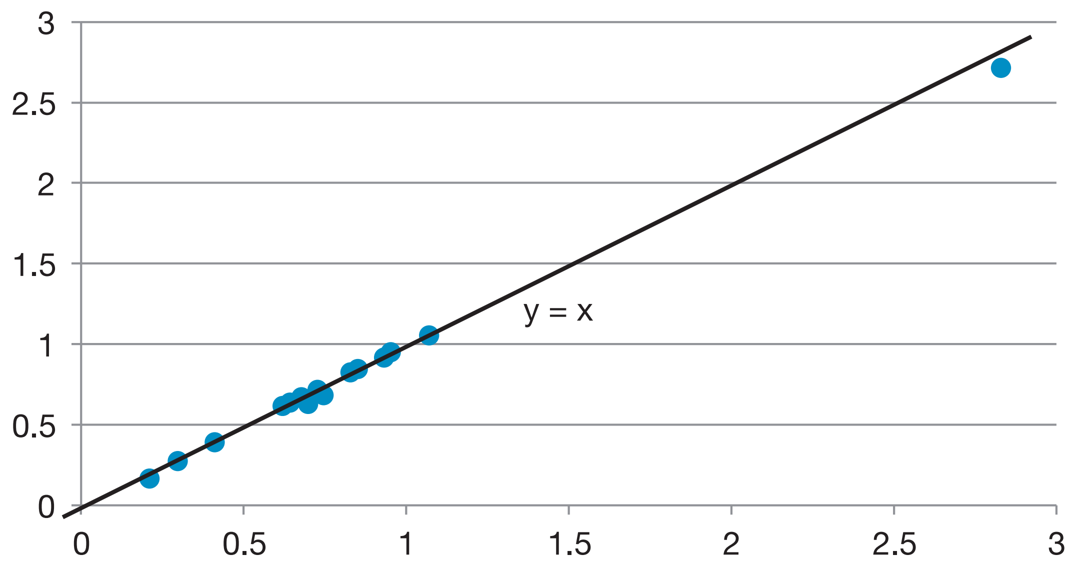
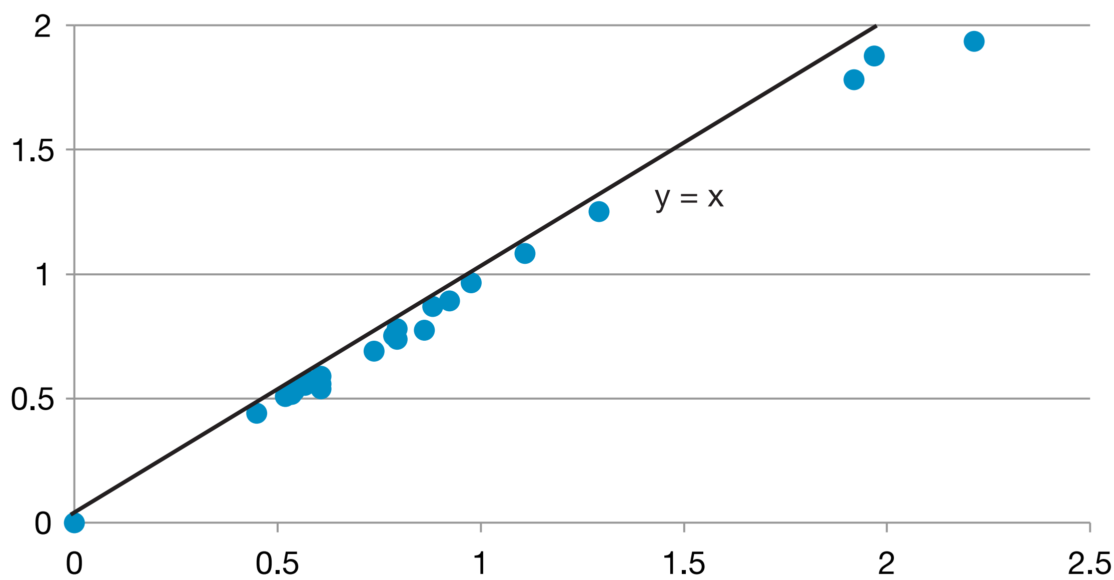


Figure 5: Warwickshire PIC: Manual vs model estimates of 24 camera multiples



In Figures 4 and 5, the match between the manual estimates, on the vertical axes, and the ones from the statistical models, on the horizontal axes, provides support for the use of manual calculation where users of the data prefer this.

The manual calculations were extended to provide estimates of changes in the severity of collisions in the vicinity of cameras across a partnership area as measured by the number of casualties per collision, the proportion of collisions that were fatal or serious, the proportion of casualties that were killed or seriously injured, and the number of people killed or seriously injured per fatal or serious collision. The method is described in Appendix 3, and example results are compared with corresponding estimates derived from the statistical models.

For a total of 132 cameras in eight partnership areas, the estimated changes in numbers of PIC per year in the vicinity of the cameras following the establishment of cameras were compared with observed changes in average speed in the vicinity of the same cameras, as described in Appendix 4. A range of reductions in speed of up to nearly 14 miles/h were observed, but no relationship was found between these and the changes in numbers of PIC per year. This calls for further investigation based on more sites and taking into account other characteristics of all the sites.

Experience and outcomes of these analyses have provided much of the basis for the advice offered in the next three sections of the report.

3. The General Nature of the Camera Data

Where DfT recommendations on the coverage of the data have been followed, the data includes, for each camera and for each of the twenty-one calendar years 1990–2010, the numbers of:

- fatal or serious collisions (FSC);
- personal injury collisions of all severities (PIC);
- people killed or seriously injured (KSI) in the FSC; and
- casualties of all severities (CAS) in the PIC.

These numbers are those recorded in the national road accident data system known as STATS19 as occurring in the vicinity of the camera in that year.

The date of establishment of the camera and the speed limit at the camera site are also provided together with, for some but not all cameras, measurements made from time to time of traffic speed in the vicinity of the camera. Numbers of offences recorded at the camera, usually in a recent year, and the numbers dealt with in various ways by the police are also provided, but these numbers are not discussed in this report. The date of establishment of the camera is the date from which it may at any time have been in operation. Just when the camera has since been in operation has been for the partnership to decide, subject to the occurrence of incidents affecting operation.

Available by separate extraction from the STATS19 system are for each year the numbers of FSC, PIC, KSI and CAS in each local authority area as a whole. Where a partnership area covers several local authority areas, these numbers for the partnership area are obtained by adding together the numbers for these local authority areas.

3.1 The nature of numbers of collisions or casualties

By definition, the PIC include the FSC and the CAS include the KSI. Because every FSC gives rise to at least one KSI and every PIC to at least one CAS, each number of casualties cannot be less than the corresponding number of collisions, and is usually greater when there are more than just a few collisions.

Because the less severe collisions are much more numerous than the more severe ones, the numbers of PIC and CAS are typically up to about ten times the corresponding numbers of FSC and KSI.

As well as being subject to many systematic influences, such as the amount of traffic, the weather over the year and whether a camera may have been in operation over the year, all the numbers of collisions or casualties are subject to natural fluctuation known as *random variation*. This stems from the fact that how often risky situations arise and whether a collision results or is avoided when a risky situation does arise are both down partly to chance.

Because the numbers of FSC and KSI are typically so much smaller than those of PIC and CAS, the effect of random variation is much greater relative to that of systematic influences for FSC and KSI than for PIC and CAS. This means that the numbers of the latter typically contain a good deal more information than do numbers of the former about systematic influences, including the possible effect of a camera having been established, even though typical FSC and resulting KSI are more severe in their consequences – and therefore often of greater concern – than typical PIC and CAS.

The number of casualties arising from a PIC or an FSC is most often 1, but is sometimes 2 or 3 and just occasionally more than 3. This means that the numbers of casualties are in general somewhat larger than corresponding numbers of collisions. And the number of casualties in a given collision is down partly to chance as well as to systematic influences, which means that numbers of casualties are in general subject to greater random variation than corresponding numbers of collisions. Because of this, it makes sense to look for systematic influences first by analysing numbers of PIC and FSC, and to use the numbers of CAS and KSI mainly to provide supplementary information about the severity of collisions: for example, about average numbers of casualties per collision.

3.2 The years before and after establishment of a camera

The date of establishment of a camera allows each of the calendar years 1990–2010 to be identified as:

- a year throughout which the camera was not yet established;
- a year throughout which the camera may have been in operation; or
- the year partway through which the camera was established.

If the occurrence of collisions had played no part in decisions where to establish cameras, it would be natural to investigate how numbers of collisions or casualties changed after cameras were established by comparing numbers in years of the second kind with numbers in years of the first kind.

However, it is well known that occurrence of collisions, especially in the last few years before decisions where to establish cameras, has influenced the locations of substantial numbers of cameras. This means that there is a tendency for substantial numbers, though not all, of the locations at which cameras are established to have had unusually high numbers of collisions in some years preceding the establishment of the camera. For just which cameras, and for such a camera for just which years, this may have been the case could be established only by examining the collision history and process of establishment of each camera.

In the absence of such examination it is advisable to allow in analysis of collision and casualty data for the possibility that collision or casualty numbers in the vicinity of a camera were unusually high in years just before the decision to establish the camera, that is in later years of the first kind.

3.3 Annual numbers of collisions or casualties in each partnership area

Other things being equal, the occurrence of collisions and casualties in the vicinity of cameras in a partnership area is likely to have been influenced systematically in broadly similar ways to their occurrence across the partnership area. The annual numbers of collisions or casualties in the partnership area in the years concerned thus provide an indication of the year-on-year effects of these influences upon the numbers of collisions or casualties in the vicinity of cameras.

3.4 Speed data

For each camera for which speed data is provided on the website, the data comprises either mean or median speed and either 85th percentile speed or percentage of speeds exceeding the speed limit. These are as observed in the vicinity of the camera on typically one or two dates before establishment of the camera and on several dates subsequent to establishment. Dates provided do not always make it clear just which observations preceded establishment of the camera concerned.

Where it is clear that one or more observations were made before establishment of the camera, comparison of speed observations before and after establishment enables estimates to be made of change in speed in the vicinity of the camera following its establishment.

4. Advice on Use of Data Provided in Full

The practicalities of using the data are discussed here in terms of use of data for one partnership area to examine changes that have occurred in the vicinity of cameras in that area. The issues discussed apply with obvious extensions to use of data for a number of partnership areas taken together or in comparison. Discussion begins with the accessing and organisation of the data and goes on to consider its analysis and interpretation, first for a number of cameras taken together and then for individual cameras within that number. For the reasons discussed in section 3.1, the analysis is primarily of numbers of collisions.

4.1 Accessing and organising the data

It is assumed that the user will have use of a laptop or other personal computer with spreadsheet software. Internet access is required for initial downloading of data, but not for its analysis or interpretation. Those who wish to use generalised linear models or other statistical calculations outside the scope of their spreadsheet software will also need relevant statistical software.

Just what is involved in transferring data for cameras in the partnership area of interest to the user's own spreadsheet software depends on how the data is mounted on the partnership website. The possibilities range from simple downloading of one or more spreadsheets, through downloading and copying from one or more pdf files, to downloading from web pages camera by camera for cameras identified on a map on the partnership website. Before embarking on downloading or transcription, it is advisable not only to examine how the data is mounted on the website but also to consider in what ways the user envisages working with the data within their own spreadsheet software.

In any case the user should aim to download or transcribe for each camera at least an identifier for the camera, its location, the date of establishment, the speed limit at the site, and for each year (normally 1990–2010, but the sequence of years' data may be incomplete for some cameras) the numbers of FSC, PIC, KSI and CAS recorded in the vicinity of the camera in that year.

The user may well also wish to download or transcribe data about traffic speed as observed in the vicinity of the camera, including the dates of speed observations, and about offences recorded by the camera (although the data about offences is not discussed in this report).

Before making any analyses of numbers of collisions or casualties or of observed traffic speeds, it is important for both computer-based and manual calculations to identify which of the years 1990–2010 were years:

- throughout which the camera may have been in operation; or
- from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera.

Years of the first kind comprise all calendar years after the date of establishment. To identify years of the second kind it would be necessary to seek information from the camera partnership or local authority, which will need to refer to records of the establishment of the cameras. If these years can be identified, analysis should be based on that information. In the absence of such information, some assumption has to be made about years of the second kind in order to interpret the collision and casualty data. Because these years are typically three consecutive ones ending a year or two before establishment of the camera, and the available data is for calendar years, two assumptions have been made in the calculations in this report; these are that the years of the second kind consist of:

(a) the last three calendar years before the camera was established; or

(b) the first three of the last four calendar years before the camera was established.

These assumptions are based on awareness of what is involved for authorities and partnerships in identifying sites for cameras and moving from site identification to camera establishment, and on identification of the years of the second kind for cameras in some partnership areas. Assumption (a) matches a situation in which cameras are established typically about 6 months after the end of the period for which collision and casualty data is used to help to locate the cameras. Assumption (b) matches a situation where the corresponding interval is 18 months. Assumption (a) was used for all the calculations in the first edition of this report, but informed comments and further analysis indicate clearly that Assumption (b) is preferable in terms of effectiveness in largely excluding the effect of regression to the mean, and in allowing data from the interval between site location and camera establishment to contribute to the estimation of the level of collision occurrence in years before camera establishment other than the years in which collision numbers may have influenced the location of the camera.

No simple assumption, and no analysis of calendar-year data, can match the periods for which collision data was actually used in site identification when some cameras are established as long as several years after the period from which collision and casualty data is used in deciding their location and, moreover, any of the dates concerned may be at any time in the year. But comments on the first edition of this report, and related analysis, have, firstly, improved the exclusion of the effect of regression to the mean by showing that Assumption (b) is preferable to Assumption (a). Secondly, they indicate no other simple assumption preferable to Assumption (b). Thirdly, they have reinforced the author's assessment that assumptions of this kind can be claimed to largely exclude the effect of regression to the mean in estimating changes in numbers of collisions and casualties following establishment of cameras.

Where the camera was established near to the beginning of a year, it may be a helpful approximation to treat that year as being of the first kind, especially if there is data for only, say, one or two other years of the first kind. This approximation was made in the calculations described in Appendices 2 and 3 for cameras established in January. In the first edition of this report a similar approximation was suggested for any camera established near the end of a year, treating it as being established at the beginning of the following year, but this was mistaken because it left no gap between the years of the second kind and establishment of the camera.

It may well be helpful for both computer-based and manual calculations to identify these two kinds of years by means of two indicators that each take the value 1 if the year is of the kind concerned and 0 if it is not, and to enter these values in the spreadsheet in columns alongside the columns of numbers of collisions and casualties. For example, data for one camera established in July 1999, for which the first three of the last four calendar years before the year in which the camera was established were taken to be years of the second kind, might then read as in Table 1.

Table 1: Camera C1 Long Street, established 15 July 1999; speed limit 30 miles/h

Year	FSC	PIC	KSI	CAS	Camera established all year	Numbers affected location
1990	1	11	1	12	0	0
1991	3	8	3	8	0	0
1992	3	1	3	4	0	0
1993	0	10	0	10	0	0
1994	0	6	0	18	0	0
1995	1	3	1	4	0	1
1996	2	13	2	15	0	1
1997	1	5	1	6	0	1
1998	1	9	1	13	0	0
1999	0	3	0	3	0	0
2000	0	4	0	6	1	0
2001	0	10	0	13	1	0
2002	0	4	0	6	1	0
2003	1	8	1	10	1	0
2004	0	16	0	20	1	0
2005	2	9	2	9	1	0
2006	0	3	0	3	1	0
2007	1	9	1	13	1	0
2008	0	6	0	6	1	0
2009	1	6	1	8	1	0
2010	0	7	0	12	1	0

In order to allow for general changes in the occurrence of collisions in the partnership area when estimating changes in occurrence in the vicinity of cameras, it is necessary to obtain for each calendar year the annual numbers of PIC and FSC recorded as occurring in the partnership area. These numbers can be obtained for each local authority area from roadacc.stats@dft.gsi.gov.uk or from the authority itself. Where a partnership area comprises more than one local authority area, the corresponding numbers for the relevant local authority areas can be added together to give the numbers for the partnership area.

4.2 Estimating changes in numbers of collisions across a group of cameras

Estimates of changes following establishment can be made for an individual camera, but these are hard to interpret without also considering corresponding changes for other cameras in some group to which the camera of interest belongs. This group might be all cameras in the partnership area or some subgroup of these, such as those with a particular speed limit, those on roads of a particular kind or those in a local area of interest. The number of cameras in the group should be at least ten, and preferably several tens.

The estimation of changes in the occurrence of collisions across such a group of cameras is therefore discussed first. The following systematic influences upon the year-to-year numbers of collisions in the vicinity of the various cameras can be investigated by use of the available data for each camera:

- influences upon numbers of collisions in the partnership area as reflected in the recorded number of such collisions in the same year;
- whether the year was one throughout which the camera may have been in operation;
- whether the year was one from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera;
- the speed limit at the location of the camera;
- the year in which the camera was established; and
- any observed change in indicators of traffic speed in the vicinity of the camera.

The first influence can be taken into account by considering not the numbers of collisions in the vicinity of each camera, but instead the ratio:

$$\frac{\text{number of collisions in the vicinity of a given camera in a given year}}{\text{number of collisions in the partnership area in the same year}}$$

This will be referred to as *C*.

The influence of whether the year was one throughout which the camera may have been in operation can then be investigated by examining the values of *C* in years of that kind as multiples m_c of its values in years before camera establishment other than years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera.

Similarly, the influence of whether the year was one from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera can be investigated by examining the values of *C* in years of that kind as multiples m_b of its values in earlier years.

Estimation of a single value of each of the multiples m_c and m_b across a group of cameras provides estimates of:

(a) the percentage $100(1 – m_c)$ by which numbers of collisions per year in the vicinity of cameras in the group, taken together, were lower after establishment of cameras than in the years well before and just before establishment. A negative percentage indicates that the numbers were higher; and

(b) the percentage $100(m_b – 1)$ by which numbers of collisions per year in the vicinity of cameras in the group, taken together, were higher in the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera than in earlier years well before and just before camera establishment. A negative percentage indicates that the numbers were lower.

Where percentage (b) is positive, it indicates that the numbers of collisions per year in the vicinity of cameras would have been expected to fall by $100(m_b – 1)/m_b$ % across this group of cameras in subsequent years if no cameras had been established.

All these estimated changes are relative to the numbers of collisions per year in the partnership area as a whole. The multiples can be estimated separately for numbers of PIC and FSC.

The way in which the multiples differ among individual cameras in a group are discussed in section 4.3.

It is open to the user to adopt any appropriate method for estimating these multiples. The author's preferred ways of doing so, with and without the help of statistical computer software, are described and illustrated in Appendices 2 and 3. In working without statistical software, calculation for a group of cameras may well proceed camera by camera, as in Appendix 3, but it is still advisable to complete the calculations for the whole group of cameras before trying to interpret the results for any one camera.

Analysis can in principle be extended to investigate whether the multiples m_c and m_b differ according to the speed limit at the location of the camera. This can be done by estimating separate values of the multiples for each speed limit, or for each chosen group of speed limits (such as 30 miles/h and 40 miles/h taken together and all higher limits taken together). The scope for making such separate estimates is limited by the relatively small numbers of cameras with limits other than 30 miles/h. The data does not allow general differences in the level of occurrence of collisions in the vicinity of cameras according to the speed limit at the location of the camera to be distinguished from differences arising from other features of the camera locations.

Analysis can also be extended to investigate whether the multiples m_c and m_b differ according to the year of establishment of the camera. This can be done by estimating separate values of the multiples for different groups of years of establishment (such as the three seven-year periods 1990–1996, 1997–2003 and 2004–2010; the data is unlikely to be sufficient to distinguish meaningfully between individual years or short periods).

Concerning the influence of changes in indicators of traffic speed in the vicinity of cameras, the analysis could in principle be extended to estimate the multiple m_c in the form $a+bd$, where a and b are estimated constants and d is the difference in mean speed, median speed, 85th percentile speed or percentage of vehicles exceeding the limit. The scope for such analysis is limited by the availability of observations of speed both before and after camera establishment.

All these extensions are within the scope of the kinds of statistical model used in Appendix 2.

Recent publication of advice to Transport Scotland (Maher, 2013) may lead readers to ask whether the Empirical Bayes (EB) method could be applied to the data considered here in order to provide a counterpart to the estimation of the multiples m_c and m_b. The theory behind the EB method is soundly established, but its application requires what is known as a *reference population* of sites where cameras might be established, and if a reference population appropriate to the data considered here could be found, its use would be likely to require substantial data additional to that being discussed in this report. As discussed elsewhere (Allsop, 2004; 2010), the reference population used in the advice to Transport Scotland is subject to appreciable reservations. For these reasons the EB method seems unlikely to be helpful in analysis of the data considered here.

4.3 Changes in numbers of collisions in the vicinity of individual cameras

As described in Appendices 2 and 3, or otherwise, the multiple m_c for PIC can be estimated, albeit usually with rather wide confidence intervals, for each individual camera in a group. To help to interpret the resulting values, it is useful to consider how they are distributed. As illustrated for Warwickshire in Section 2 and for other partnerships in Appendix 2, the estimated multiples for nearly all the cameras are typically distributed quite densely over a range up to a value between about 1 and about 2, with larger estimates for just a few of the cameras.

The larger estimates are clearly of interest in view of the possibility that a camera with a large estimated multiple may have been so located that the number of PIC in its vicinity has thereby been increased. For each such camera it is advisable to examine the relevant year-on-year numbers of PIC – those whose sums are the *b* and *a* of Appendix 3:

b = number of PIC in the vicinity of the camera in all years from 1990 up to and including the year of establishment but excluding the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera; and

a = number of PIC in the vicinity of the camera in all years up to and including 2010 throughout which the camera was established.

Very small numbers of PIC in the years well before and just before camera establishment that make up the total *b*, or there being only one or two full years after camera establishment to provide the total *a*, can each lead to an estimate of m_c for the camera concerned that it would be unwise to take at face value. Examples of these two features of the data leading to high values of the multiple are mentioned in the comments on the diagrams in Appendix 2, which show distributions of the estimated multiples for individual cameras. Where neither of these features is present, a large estimated multiple may be taken as indicating that the camera concerned should be checked to see whether its operation may be a source of increased risk of collision in its vicinity.

Similarly, where the estimated multiple for a camera is markedly nearer to zero than the lowest of the densely distributed values, it should be borne in mind that, if there are only one or two years after establishment to provide the total *a*, then the data should be examined to see whether the low value of the multiple may have arisen from unusually low numbers of PIC in these years. If so, then numbers of PIC in subsequent years may result in a higher estimate of the multiple, so it would be unwise to take the favourable result at face value.

It remains to consider how to interpret the values of the multiple within the densely distributed range. At first sight, these multiples might be thought to indicate that cameras with estimated multiples appreciably below 1 are doing a good job, those with multiples around 1 are doing neither much good nor much harm, but the future of those with multiples appreciably above 1 should be called into question. But this would be simplistic.

It is instead advisable to consider what might have happened to numbers of PIC in the vicinity of each of these cameras, relative to numbers in the partnership area as a whole, if the camera had not been established – bearing in mind that these figures are subject not only to any effect of the camera but also to random variation. Some would have fallen, some would have stayed about the same and some would have risen. And (except perhaps for a few cases where relevant changes in circumstances can be identified), no one could have told in advance which would fall, which would stay the same and which would rise. This means that neither where the estimated multiple is less than 1 can any or all of the reduction in numbers of PIC be attributed definitely to the camera, nor where the estimated multiple is greater than 1 can any or all of the increase be attributed definitely to the camera. Moreover, in the years to come some of the densely distributed multiples that are now less than 1 will move above 1, and some that are now above 1 will move below 1, each as a result of random variation, even if the distribution as a whole, and thus the number of PIC at all the cameras in the group taken together relative to the number in the partnership area as a whole, remain the same.

What can be said is that, in aggregate across the group of cameras, the number of PIC in the vicinity of the cameras is estimated to have changed by the multiple whose estimation was discussed in section 4.2. The estimate of this multiple will typically have a confidence interval substantially less wide than those of the estimates of the multiples for the individual cameras.

Where the estimated aggregate change is favourable, it is advisable to regard all the cameras with multiples in the densely distributed range as having potentially contributed to it. Where the future of any of these cameras is called into question, this should be on the basis of evidence external to the data that is the subject of this report. The estimate of the multiple for that camera might be cited to corroborate or counter the external evidence, but in doing so the width of its confidence interval should be taken into account.

Where the estimated aggregate change is unfavourable, scrutiny of the operation of the group of cameras is advisable. Consideration of the estimated multiples for the individual cameras may then form part of this scrutiny, but once again the widths of their confidence intervals should be taken into account.

4.4 Changes in the severity of collisions

Numbers of CAS, FSC and KSI can be used together with corresponding numbers of PIC to calculate four indicators of severity of collisions and casualties in the vicinity of cameras:

> (number of CAS)/(corresponding number of PIC) = number of casualties per collision;
> (number of FSC)/(corresponding number of PIC) = proportion of collisions that are fatal or serious;
> (number of KSI)/(corresponding number of CAS) = proportion of casualties that are killed or seriously injured; and
> (number of KSI)/(corresponding number of FSC) = number of people killed or seriously injured per fatal or serious collision.

For the values of these indicators to be meaningful, they should be calculated for appreciable numbers of cameras or years or both, so that the numbers in the numerator and denominator are at least well into double figures.

Changes in these indicators following the establishment of cameras can be estimated, as described and illustrated in Appendix 3, by calculating counterparts for CAS, FSC and KSI of the numbers *b* and *a* of PIC defined in Appendix 3 and then calculating and comparing the indicators of severity for the periods corresponding to *b* and *a*, respectively. Where these periods are both of several years and the annual numbers of PIC are not too small, the totals for individual cameras may be large enough to give reliable values of the indicators. In any case, the totals for groups of at least several cameras taken together are likely to be large enough for the purpose. When a group of cameras is considered, the periods corresponding to *b* and *a* for different cameras may comprise different ranges of years, but the numbers of collisions and casualties for these periods can nevertheless be added together in order to estimate the indicators of severity across the group of cameras.

5. Advice on Use of More Limited Data

A number of partnership areas have mounted data in ways that approximate to the DfT guidance, but they either give rise to difficulties in extraction for uses of the kind discussed in this report or are in some respect incomplete in coverage. Where the data goes back to 1990 but is incomplete in some other respect or awkward to extract, the uses discussed in Section 4 are nevertheless largely feasible. Where the data starts in years later than 1990 the uses discussed in Section 4 are feasible only in respect of cameras for which the data covers several years before the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera. For cameras where this is not the case, the advice in the rest of this Section is applicable.

A different approach is needed, with associated limitations to analysis and interpretation, for partnership areas where little or no data is provided for years before the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the cameras. In particular, this is the case where the only data provided for years before the establishment of cameras is so-called *baseline data*, which relates typically to a period of three years that ends a year or two before establishment of the camera, and often comprises aggregate numbers of collisions and casualties for these three years taken together.

For these partnership areas it is first necessary to have regard to the fact that the baseline data or other data relating to years before camera establishment is all data that may have been taken into account in deciding where to establish cameras. This must be allowed for in any estimation of changes following the establishment of cameras, and in the absence of local knowledge about numbers of collisions and casualties in the vicinity of cameras in earlier years it is necessary to make some assumption about the extent to which the available data overestimates the typical numbers of collisions and casualties in years well before and just before the establishment of cameras.

Previous work by the author (Allsop, 2010) and extensive unpublished work by Idris Francis, who has taken a close interest in the use of speed cameras, has indicated overestimation of the order of 9% for numbers of PIC. The combined estimate beneath Table A2.3 from the partnership areas for which data is analysed in Appendix 2 indicates rather greater overestimation, of about 14%. The same earlier work indicated overestimation of the order of 30% for FSC, and this is supported by the combined estimate beneath Table A2.4, which indicates about 33%.

Users of data from partnership areas for which some such assumption is necessary might therefore make analyses of numbers of PIC on the basis that the numbers of PIC in the available data for the baseline period, or for other years for which data may have been taken into account in deciding where to establish cameras, overestimates by, say, 10% or 12% the numbers in years well before and just before the establishment of cameras. For analysis of numbers of FSC, they might assume an overestimation of, say, 30%. A higher assumed percentage is more conservative in terms of estimation of any reduction in numbers following establishment of cameras. When estimates are made on the basis of such assumptions, the assumptions should always be mentioned when the results are quoted or used.

Analysis of numbers of PIC and FSC in these partnership areas can then be undertaken by the methods described in Appendix 3. The fact that some partnerships provide numbers of PIC and FSC after camera establishment in financial years rather than calendar years does not affect the method of analysis.

6. Some Lessons from this Exercise in Providing Transparency

Preparations within DfT for the minister's letter requiring local authorities to mount the speed camera transparency data were extensive and resulted in clear and considered guidance to the authorities including a template for the mounting of data in Excel spreadsheet form and the creation by DfT of a list of websites specified by the local authorities on which data would be mounted.

Follow-up has, however, been less effective. The central list of websites has not been kept up to date, and the response by local authorities in mounting the data has been mixed, ranging from full mounting of data in the recommended form through full mounting in less user-friendly forms to incomplete mounting of data or even mounting of no data on the specified website.

The result is that coverage of the mounted data is incomplete, and in only a few partnership areas can the envisaged users of the data access them in a form that allows them to embark on immediate analysis, as was the apparent intention in recommending mounting in Excel spreadsheet form. In other areas where data has been mounted in full, it needs to be converted to a different format for analysis and, where the data is incomplete, the scope for analysis is accordingly restricted.

It has been left entirely to users of the data to consider how to interpret it, even though the data is of a kind that has for some years given rise to both controversy and genuine methodological difficulty in interpretation.

In the fast-developing world of open data, the experience of mounting of speed camera data points to several practicalities that warrant attention:

- Asking holders of data to make it available in a recommended form does not necessarily result in the data being made available in that form or at all
- Websites and their addresses often change, so any central source of such addresses needs to be robust with respect to changes in them

- Users will want to work with data, not just read it on a screen or printout, so data should be mounted in a format that enables use with a minimum of transcription, that is, in a spreadsheet or analogous format
- While users should of course be free to make their own analyses and interpretations of data that is made available, this can be helped by objective and non-directive advice about the nature and characteristics of the data concerned and pointers towards available techniques that are appropriate for application to data of that kind

7. References

Allsop, R. E. (2004). Impact of speed cameras on safety. *Traffic Engineering and Control*, 45(11): 384.

Allsop, R. (2010). *The Effectiveness of Speed Cameras: A Review of Evidence.* London: RAC Foundation.

DfT (Department for Transport) (2011a). *New Transparency Rules for Speed Cameras*. Press release. Retrieved 20 March 2013 from www.gov.uk/government/news/new-transparency-rules-for-speed-cameras.

DfT (2011b). *Report of a Working Group about the Publication of Information Related to Speed Cameras*. Retrieved 20 March 2013 from www.gov.uk/government/uploads/system/uploads/attachment_data/file/4359/working-group-speed-camera-report.pdf.

DfT (2011c). *Letter from Mike Penning to local authority chief executives outlining requirements for the publication of speed camera information*. Retrieved 5 January 2013 from www.gov.uk/government/uploads/system/uploads/attachment_data/file/4357/ la-letter-penning-20110627.pdf.

DfT (2011d). *Note on the requirements for publication of information relating to speed cameras*. Retrieved 5 January 2013 from www.gov.uk/government/uploads/system/uploads/attachment_data/file/4358/ dft-publication-of-speed-camera-information.pdf.

DfT (2011e). *Fixed Speed Camera Collision, Casualty and Speed Data.* Retrieved 5 January 2013 from www.gov.uk/government/uploads/system/uploads/attachment_data/file/15442/fixed-speed-camera-data-sources.pdf.

Maher, M. (2013). *Statistical Advice: Scottish safety camera programme*. Transport Scotland. Retrieved 28 October 2013 from www.transportscotland.gov.uk/files/images/Safety_Camera_Stats_Report.pdf.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. ISBN 3–900051–07–0. Retrieved 20 March 2013 from www.R-project.org.

Wikipedia (2013). *Generalized Linear Model*. Retrieved 20 March 2013 from http://en.wikipedia.org/wiki/Generalized_linear_model.

Appendix 1: Fixed Speed Camera Data as Published by Local Authorities by Late November 2012

This Appendix summarises the coverage and form of data found by visiting in late November 2012 the websites listed in the DfT's *Fixed Speed Camera Collision, Casualty and Speed Data* of August 2011 (DfT, 2011e). Some of the URLs provided no longer led directly to the data, but most still provided starting points for finding it.

The data sources are grouped in this Appendix according to the extent of the data accessible through the websites and the form in which it can be accessed. For each source named, the number of fixed cameras covered by the data is indicated in brackets.

Unless otherwise stated, for each camera, annual numbers of FSC, PIC, KSI and CAS are provided for the years indicated, and speed data is provided for particular years for each camera for which speed measurements were made. The date of establishment of the camera and the speed limit at the site are also provided. In many cases, data subsequent to 2010 have been added.

The extent to which speed data provides estimates both before and after establishment and for how long after establishment needs camera-by-camera assessment. Where offence data is given, this is usually for 2010 or later, but form and coverage differ among authorities.

A1.1. Data for 1990–2010 provided in a spreadsheet according to DfT guidance

- Cambridgeshire and Peterborough (47)
- Merseyside (33)
- Staffordshire and Stoke-on-Trent (68)
- Surrey (53)

A1.2. Data for 1990–2010 according to DfT guidance but provided in pdf or html format

- Hertfordshire (132) – one pdf per camera
- Lancashire (278) – one html per camera
- Leicester, Leicestershire and Rutland (16) – one html per camera
- Lincolnshire (52) – one pdf covering all cameras
- South Yorkshire (57) – one pdf per camera
- Sussex (52) – one pdf per camera
- Thames Valley (208) – one pdf for each of nine districts
- Warwickshire (25) – one pdf covering all cameras

A1.3. Data approximating to DfT guidance

- Devon and Cornwall (63) – one pdf per camera, but data is shown as graphs rather than in tables, making extraction laborious
- Greater Manchester (240) – collision, casualty and speed data from 1994 only, together with establishment date and speed limit, is displayed by clicking on each camera in turn on a map*
- Norfolk (23) – one pdf giving FSC and PIC collision and KSI casualty data only as aggregates for baseline period and another giving collision and casualty data 1990–2010, but establishment dates and speed data are not evident
- Northamptonshire (40) – casualty data still being checked and establishment dates are not evident; a contact name and phone number for queries is provided
- Nottingham and Nottinghamshire (36) – one pdf per camera, but data is displayed on graphs rather than in tables, making extraction laborious, and start dates for data differ among cameras and are in general later than 1990*
- West Mercia (22) – one pdf for each of four regions, giving collision and casualty data in DfT format; speed limits obtainable from a map
- West Midlands (306) – collision, casualty and speed data from 1997 only, together with establishment date and speed limit, is displayed by clicking on each camera in turn on a map*
- West Yorkshire (135) – one html per camera with collision and casualty data and some speed data

* Later start date may mean that pre-baseline data is limited or lacking for some cameras

A1. 4. More limited data

- Dorset, Bournemouth and Poole (15) – one pdf per camera giving collision and casualty data, but starting only in 1998; no speed data evident
- Essex, Southend and Thurrock (86) – spreadsheet giving aggregate collision data only for baseline period and collision data for financial years thereafter; no casualty data or speed data evident
- Hampshire and Isle of Wight (30) – spreadsheet giving aggregate collision and casualty data only for baseline period and collision and casualty data for financial years thereafter; no speed data evident
- Wiltshire and Swindon (98) – one pdf with total collision data and casualty data by severity as aggregates for baseline period and annually for 2002–8 with some gaps but no speed data; this was available earlier in 2012 but was no longer evident at end November 2012

A1. 5. Very limited data

- Cleveland (3) – one pdf for all cameras, but data starts in 1996 and no establishment dates are evident
- Cumbria (7) – one pdf per camera, but data starts from dates between 2003 and 2005
- Gloucestershire (26) – one html per camera giving collision data for years from about 1999 to 2010, but no baseline because most cameras were established earlier
- Humberside (89) – spreadsheet giving KSI casualty and PIC collision data only as aggregates for baseline period and from establishment to March 2011, with speed data for the same two periods
- Kent and Medway Towns (71) – one html giving only KSI casualty data as aggregates for baseline and for 2009–11
- London (780) – baseline three-year data was available earlier in 2012, but only camera locations were evident at end November 2012

A1. 6. Only camera locations evident so far

- Cheshire (48)
- Derby and Derbyshire (114)
- Northumbria (42)

A1. 7. No data evident yet

- Bedford, Central Bedfordshire and Luton
- Somerset
- Suffolk

A1. 8. No fixed cameras

- Avon and Somerset
- Durham and Darlington
- North Yorkshire and York

Appendix 2: Computer-Assisted Statistical Analysis of Full Collision Data

A2.1 Method

The data for each partnership area was analysed by means of the widely used technique known as *generalised linear modelling* (Wikipedia, current), applied first to the natural logarithm (logarithm to the base *e*, where $e \approx 2.718$) of the ratio:

$$\frac{\text{number of PIC in the vicinity of a given camera in a given year}}{\text{number of PIC in the partnership area in the same year}}$$

This was done with the aim of estimating how the ratio was affected multiplicatively by:

whether that year was one of the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera, and (A2.1)

whether the camera might have been in operation throughout that year (A2.2)

having regard to the general level of occurrence of PIC in the vicinity of that camera and the random variation in the number of PIC in the vicinity of the camera in a year.

These multiplicative relations are expressed as additive ones by taking natural logarithms to give equations (A2.3) and (A2.4) as follows:

Equation (A2.3)

$$\ln p_{ny} = \ln P_y + c_n + ub_{ny} + vc_{ny} + r_{ny} \qquad \text{(A2.3)}$$

is used to estimate single values of effects (A2.1) and (A2.2) across all cameras in a partnership area, where for the partnership area concerned:

the cameras are numbered $n = 1, 2, 3, \ldots, N$

p_{ny} = number of PIC at camera number n in year y

P_y = number of PIC in the partnership area in year y

c_n = fitted indicator of general level of collisions at camera n

b_{ny} = 1 if year *y* was one from which the numbers of collisions or casualties may have been taken into account in deciding where to establish camera *n* and 0 if not

u = fitted indicator across all *N* cameras of the general level of collisions in the vicinity of cameras in years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera relative to the level in other years before the camera was established

c_{ny} = 1 if camera *n* was established throughout year *y* and 0 if not

v = fitted indicator across all *N* cameras of the general level of collisions in the vicinity of cameras in years throughout which cameras might have been in operation relative to the level in years before cameras were established other than the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the cameras

r_{ny} reflects the random variation in $\ln p_{ny}$ resulting from variation (which is assumed to be of a form known as Poisson) in the number p_{ny}

The range of values of *y* is from 0 in 1990 to 20 in 2010 for every camera in the partnership area except where some years' data is missing for some cameras.

The values of the fitted indicators are calculated by the software so that they approximately maximise the likelihood of the recorded numbers of PIC having occurred at each camera and in each year if all influences upon the numbers p_{ny} were represented by Equation (A2.3).

Equation (A2.4)

$$\ln p_{ny} = \ln P_y + c_n + u_n b_{ny} + v_n c_{ny} + r_{ny} \qquad \text{(A2.4)}$$

is used to estimate values of effects (A2.1) and (A2.2) separately for each of the *N* cameras in the partnership area, where:

u_n = fitted indicator of the level of collisions in the vicinity of camera *n* in the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera relative to the level in other years before it was established; and

v_n = fitted indicator of the level of collisions in the vicinity of camera *n* in years throughout which camera *n* might have been in operation relative to the level in years before it was established other than the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera.

The values of c_n and r_{ny} will in general be somewhat different and substantially different, respectively, after fitting Equation (A2.4) from the values they had after fitting Equation (A2.3).

From the fitted values of u , v , the u_n and the v_n as given by the software can be calculated the corresponding estimates exp(u), exp(v), exp(u_n) and exp(v_n), respectively (where exp(x) is the number $e \approx 2.718$ raised to the power x) of the following four multiples:

m_b = number of PIC per year in the vicinity of cameras in the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the cameras as a multiple of the number in other years before camera establishment, estimated across all cameras in the partnership area relative to the number of PIC per year in the area;

m_c = number of PIC per year in the vicinity of cameras in years throughout which cameras may have been in operation as a multiple of the number in years before camera establishment other than years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the cameras, estimated across all cameras in the partnership area relative to the number of PIC per year in the area;

m_{bn} = number of PIC per year in the vicinity of camera *n* in the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera as a multiple of the number in other years before camera establishment, estimated relative to the number of PIC per year in the partnership area; and

m_{cn} = number of PIC per year in the vicinity of camera *n* in years throughout which the camera may have been in operation as a multiple of the number in years before camera establishment other than years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera, estimated relative to the number of PIC per year in the partnership area.

The fitting process carried out by the software provides estimates of the standard errors of the fitted values of *u, v* , the u_n and the v_n . Let *s* denote, for example, the estimated standard error of *u*. Then the 95% confidence interval of the fitted value of *u* is (u – 1.96s, u + 1.96s), and because m_b = exp(u), the corresponding 95% confidence interval of m_b is [m_b/exp(1.96s), $m_b$exp(1.96s)]. The confidence intervals of m_c and the m_{bn} and m_{cn} are calculated similarly from the standard errors of *v* and the u_n and v_n , respectively.

The counterpart of Equation (A2.3) in respect of the ratio:

$$\frac{\text{number of FSC in the vicinity of a given camera in a given year}}{\text{number of FSC in the partnership area in the same year}}$$

is:

$$\ln f_{ny} = \ln F_y + c_n + ub_{ny} + vc_{ny} + r_{ny} \qquad \text{(A2.5)}$$

where:

f_{ny} = number of FSC at camera number *n* in year *y*

F_y = number of FSC in the partnership area in year *y*

The fitted values of c_n, *u* and *v* and the resulting values of r_{ny} now relate to numbers of FSC instead of to numbers of PIC.

Equation (A2.4) has a corresponding counterpart for FSC, but the fitted estimates of u_n and v_n tend to have such wide confidence intervals that it would be unwise to place reliance on them.

A2.2 Results

The models described in section A2.1 were fitted to data from nine partnership areas, for the numbers of cameras shown in brackets:

- Cambridgeshire and Peterborough (47)
- Leicester, Leicestershire and Rutland (15)
- Lincolnshire (50)
- Merseyside (33)
- South Yorkshire (56)
- Staffordshire and Stoke-on-Trent (68)
- Sussex (55)
- Thames Valley (203)
- Warwickshire (24)

The cameras in Staffordshire and Stoke-on-Trent fell so clearly into two groups – 42 at sites with relatively few and 26 at sites with relatively many collisions per year, respectively – that data for these two groups of cameras was analysed separately. Separate results for the two groups are provided throughout the rest of this Appendix.

For the purposes of these calculations, the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the cameras were assumed to be:

either the last three calendar years before camera establishment – Assumption (a);

or the first three of the last four calendar years before camera establishment – Assumption (b).

Numbers of collisions in the vicinity of cameras per full year after camera establishment across partnership areas

The resulting estimates of the multiple m_c for PIC across each of these partnership areas under Assumptions (a) and (b) are shown in Table A2.1, together with their estimated 95% confidence intervals. As discussed in section 4.1, Assumption (b) is preferred.

Table A2.1: Estimates of multiples m_c for effect of camera establishment on number of PIC per year in the vicinity of cameras across partnership areas

Partnership area (number of cameras considered)	Assumption (a)		Assumption (b)	
	Multiple m_c	95% confidence interval	Multiple m_c	95% confidence interval
Cambridgeshire and Peterborough (47)	1.011	0.942, 1.085	1.025	0.955, 1.100
Leicester, Leicestershire and Rutland (15)	0.715	0.648, 0.788	0.745	0.675, 0.822
Lincolnshire (50)	0.913	0.823, 1.013	0.878	0.792, 0.973
Merseyside (33)	1.105	1.034, 1.182	1.115	1.043, 1.192
South Yorkshire (56)*	1.030	0.958, 1.107	1.012	0.942, 1.088
Staffordshire and Stoke-on-Trent – few (42)	0.680	0.638, 0.724	0.686	0.644, 0.731
Staffordshire and Stoke-on-Trent – many (26)	0.774	0.763, 0.786	0.779	0.768, 0.791
Sussex (55)	0.790	0.731, 0.852	0.804	0.745, 0.868
Thames Valley (203)	0.802	0.764, 0.842	0.838	0.797, 0.880
Warwickshire (24)	0.750	0.647, 0.869	0.778	0.671, 0.902

* Figures under Assumption (a) in first edition were based on incomplete data.

Under both assumptions, seven of the ten results in Table A2.1 point clearly to reductions in PIC, ranging under Assumption (b) from about 12% to 31% following camera establishment. Two of the others indicate no change, and one points clearly to an increase of about 12%.

These estimates can be combined by taking their geometric mean weighted by number of cameras to provide a combined estimate m_c = 0.862 with 95% confidence interval (0.840, 0.884), indicating a reduction in PIC of about 14% under Assumption (b), compared with 0.847, (0.826, 0.869) and 15%, respectively under Assumption (a).

Corresponding estimates of the multiple m_c for FSC across each of these areas are shown in Table A2.2, together with their estimated 95% confidence intervals.

Table A2.2: Estimates of multiples m_c for effect of camera establishment on number of FSC per year in the vicinity of cameras across partnership areas

Partnership area (number of cameras considered)	Assumption (a)		Assumption (b)	
	Multiple m_c	95% confidence interval	Multiple m_c	95% confidence interval
Cambridgeshire and Peterborough (47)	0.577	0.477, 0.698	0.588	0.486, 0.711
Leicester, Leicestershire and Rutland (15)	0.468	0.328, 0.666	0.489	0.343, 0.698
Lincolnshire (50)	0.852	0.690, 1.052	0.842	0.682, 1.040
Merseyside (33)	1.047	0.865, 1.270	1.071	0.883, 1.298
South Yorkshire (56)*	0.868	0.728, 1.035	0.916	0.767, 1.095
Staffordshire and Stoke-on-Trent – few (42)	0.559	0.449, 0.696	0.587	0.471, 0.731
Staffordshire and Stoke-on-Trent – many (26)	0.712	0.678, 0.749	0.720	0.684, 0.757
Sussex (55)	0.638	0.535, 0.762	0.656	0.549, 0.784
Thames Valley (203)	0.757	0.667, 0.859	0.859	0.751, 0.983
Warwickshire (24)	0.618	0.429, 0.892	0.681	0.465, 0.971

* Figures under Assumption (a) in first edition were based on incomplete data.

Eight of the ten results in Table A2.2 point clearly to reductions in FSC under both assumptions, ranging under Assumption (b) from about 14% to 51% following camera establishment. The others indicate one smaller increase, which could well have arisen by chance, and one smaller decrease.

These estimates can be combined by taking their geometric mean weighted by number of cameras to provide a combined estimate m_c = 0.777 with 95% confidence interval (0.728, 0.830), indicating a reduction in FSC of about 22% under Assumption (b), compared with 0.728, (0.683, 0.776) and 27%, respectively, under Assumption (a).

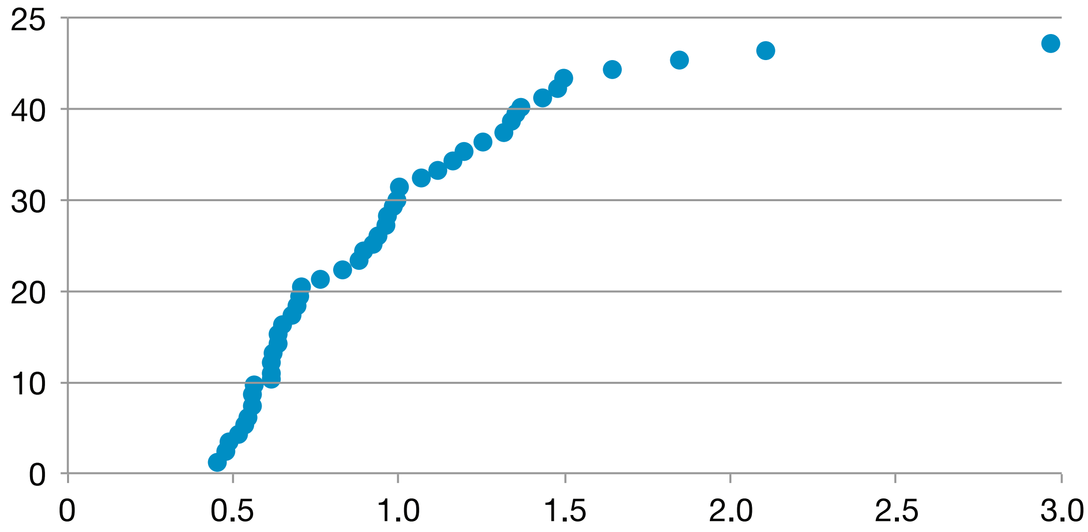
Tables A2.1 and A2.2 and the combined estimates based on them indicate that, as discussed in section 4.1, Assumption (b) leads to lower estimates of reductions in collisions attributable to the cameras than Assumption (a) – lower by about one seventh for PIC and one fifth for FSC. This is probably because of more effective exclusion of the effect of regression to the mean. As to whether there is an appreciable effect of regression to the mean that is not excluded by Assumption (b), this may of course be the case in particular instances, but example calculations assuming still earlier or somewhat longer site selection periods, and others using actual site selection periods obtained from camera partnerships, suggest that this is probably exceptional. The strong indication is therefore that collision reductions of the order of those estimated using Assumption (b) are attributable to the cameras.

Numbers of collisions in the vicinity of individual cameras per full year after camera establishment

The cumulative distributions of the values of the multiples m_{cn} at individual cameras in the same areas, based on Assumption (a), are shown in the following figures, in which the estimated values of m_{cn} are plotted on the horizontal axis and their rankings in increasing order are plotted vertically. Corresponding figures based on Assumption (b) would be somewhat different numerically, but similar qualitatively, as exemplified by comparing the figure for Warwickshire in this Appendix with Figure 1 on page xii, which is based on Assumption (b).

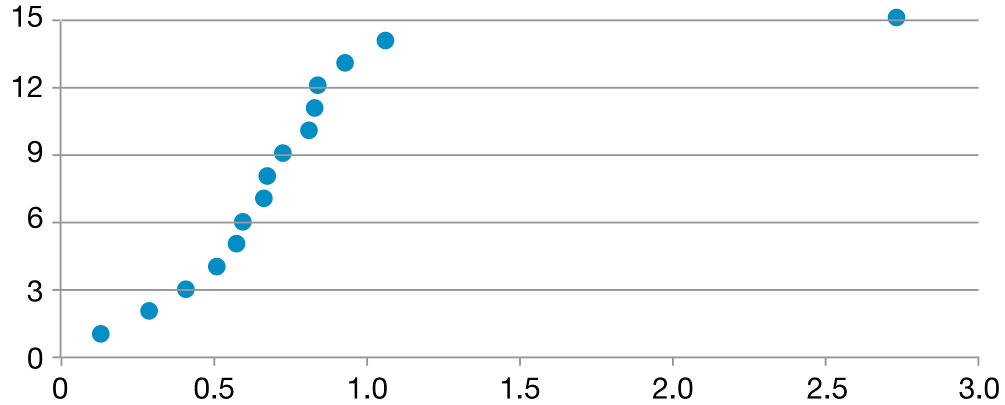
These cumulative distributions have the common feature that the estimated multiples m_{cn} for most of the cameras in the area concerned are distributed quite densely up to a value (different for each partnership area) between about 1 and about 2, while a small number of cameras have much higher values of m_{cn}. Cases where these high values appear to arise from unusual features of the data are discussed beneath each figure.

Cambridgeshire and Peterborough PIC: cumulative distribution of model estimates of 47 camera multiples



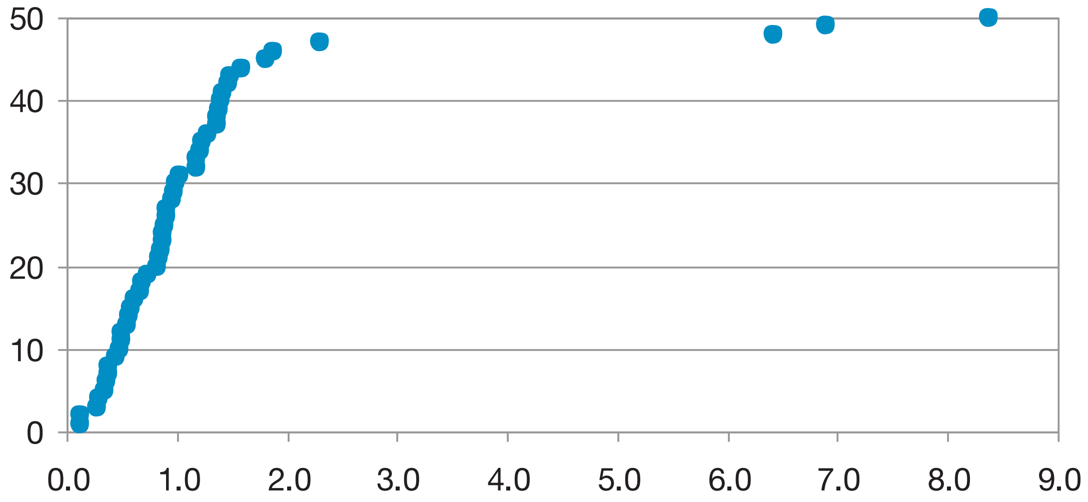
None of the four highest values seems to arise from exceptional features of the data.

Leicester Leicestershire and Rutland PIC: cumulative distribution of model estimates of 15 camera multiples



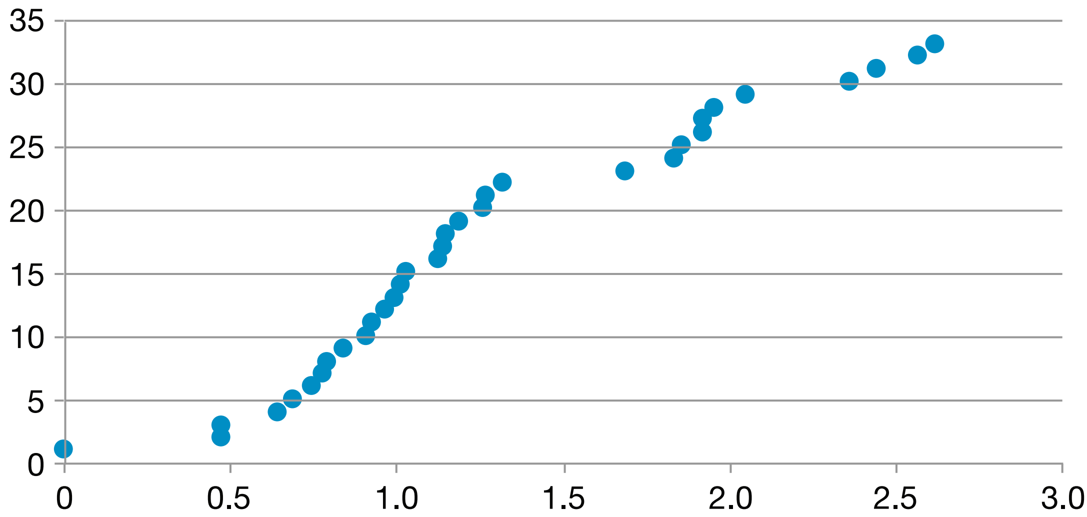
The value of nearly 3 is based on data for only one full year after establishment, and there had been at least as many PIC in each of several years before establishment

Lincolnshire PIC: cumulative distribution of model estimates of 50 camera multiples



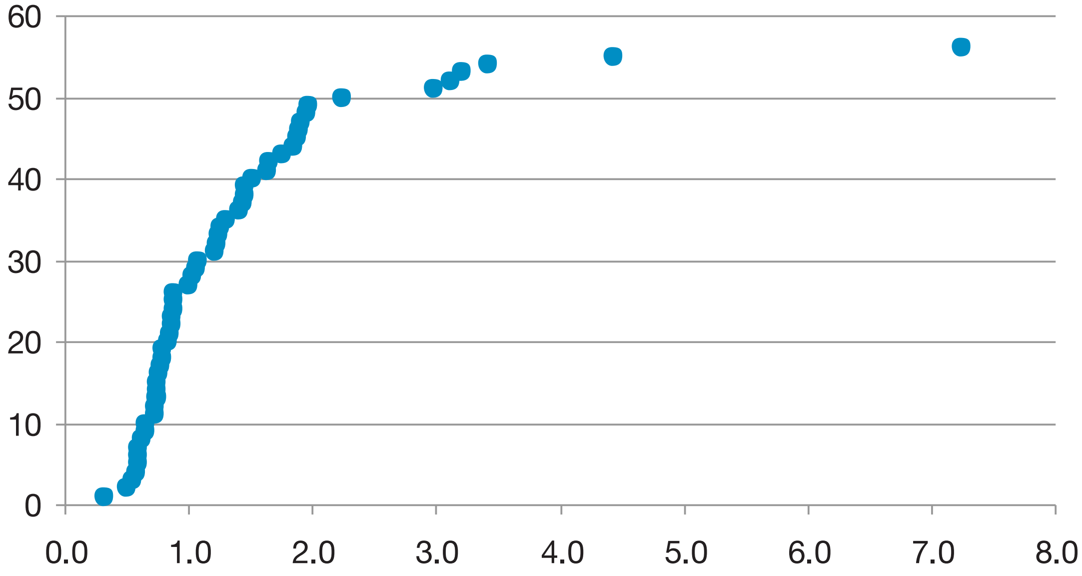
The three values between 6 and 9 arise from cameras where there were hardly any PIC until the last three full years before establishment

Merseyside PIC: cumulative distribution of model estimates of 33 camera multiples



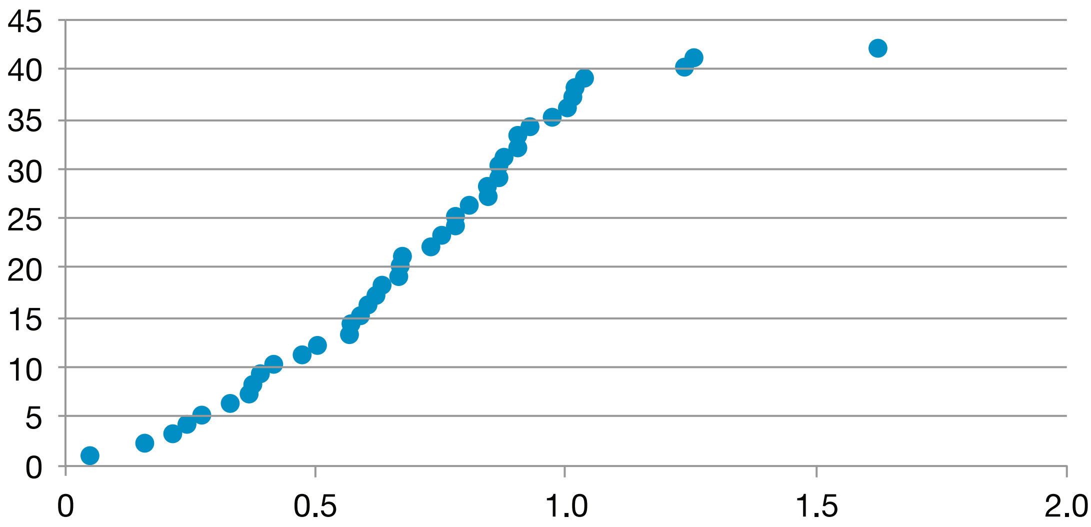
The highest value arises from a camera where there were hardly any PIC until the last three full years before establishment. Of the values just below 2, one results from just one apparently exceptional year with a large number of PIC after camera establishment and another arises from generally very small numbers of PIC both before and after camera establishment. This leaves eight high values for which there seem to be no exceptional features of the data.

South Yorkshire PIC: cumulative distribution of model estimates of 56 camera multiples



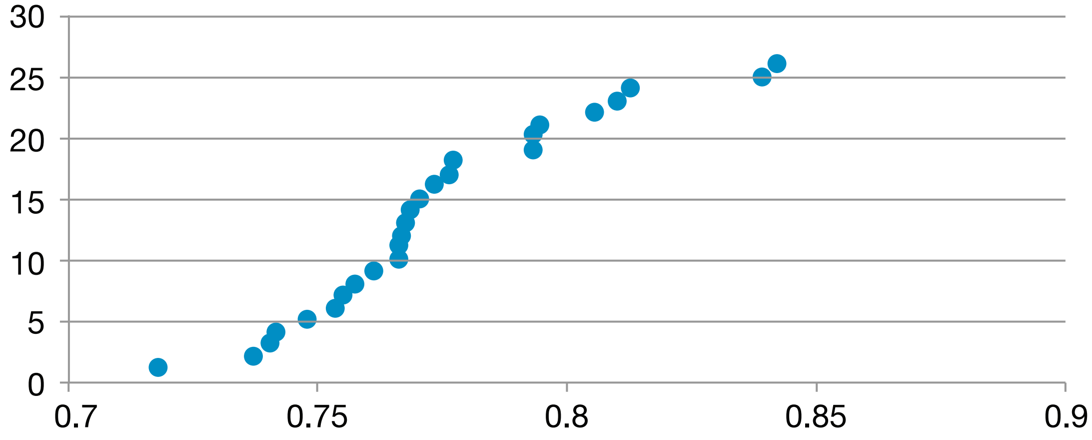
None of the seven highest values seems to arise from exceptional features of the data.

Staffordshire and Stoke-on-Trent few PIC: cumulative distribution of model estimates of 42 camera multiples



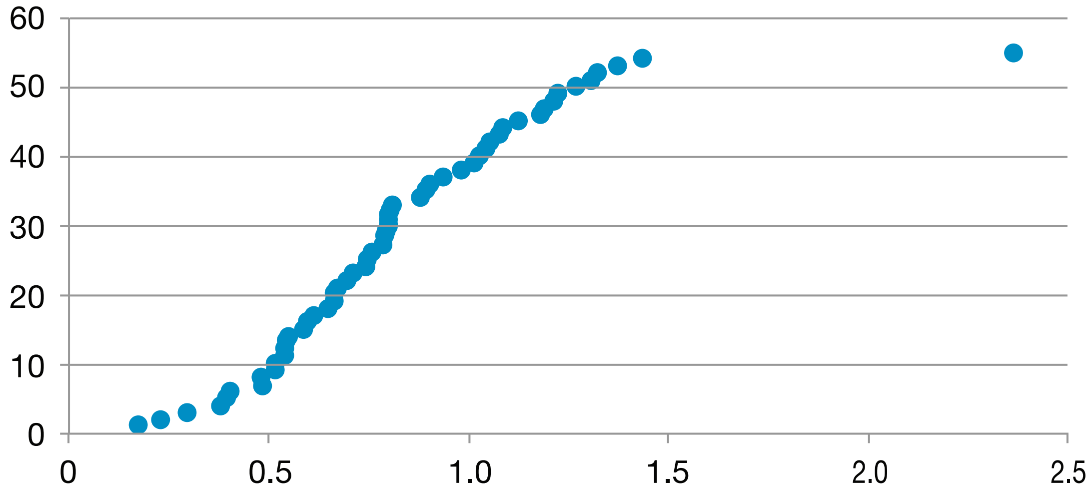
None of the three highest values seems to arise from exceptional features of the data.

Staffordshire and Stoke-on-Trent many PIC: cumulative distribution of model estimates of 26 camera multiples



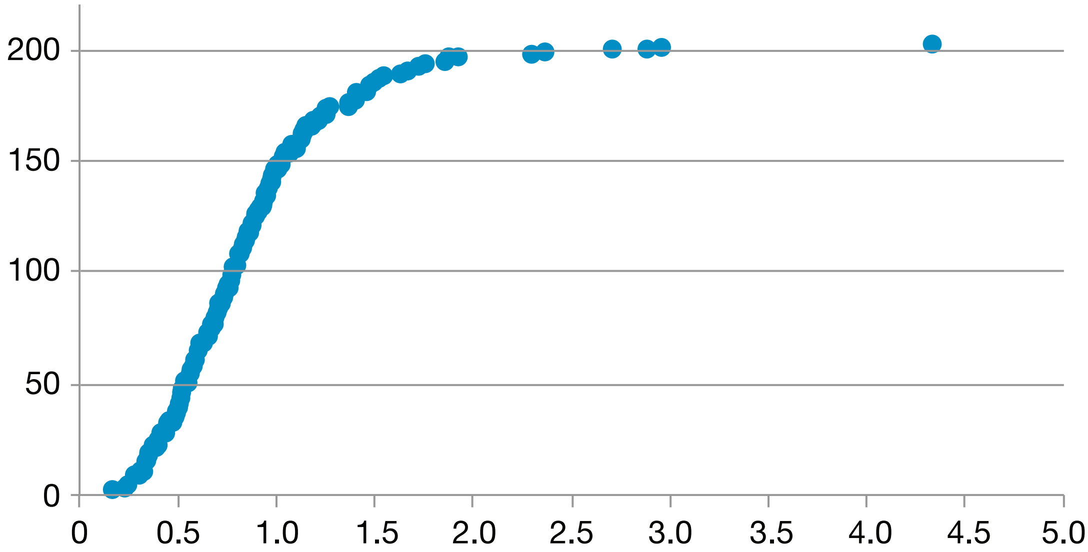
None of the eight highest values seems to arise from exceptional features of the data.

Sussex PIC: cumulative distribution of model estimates of 55 camera multiples



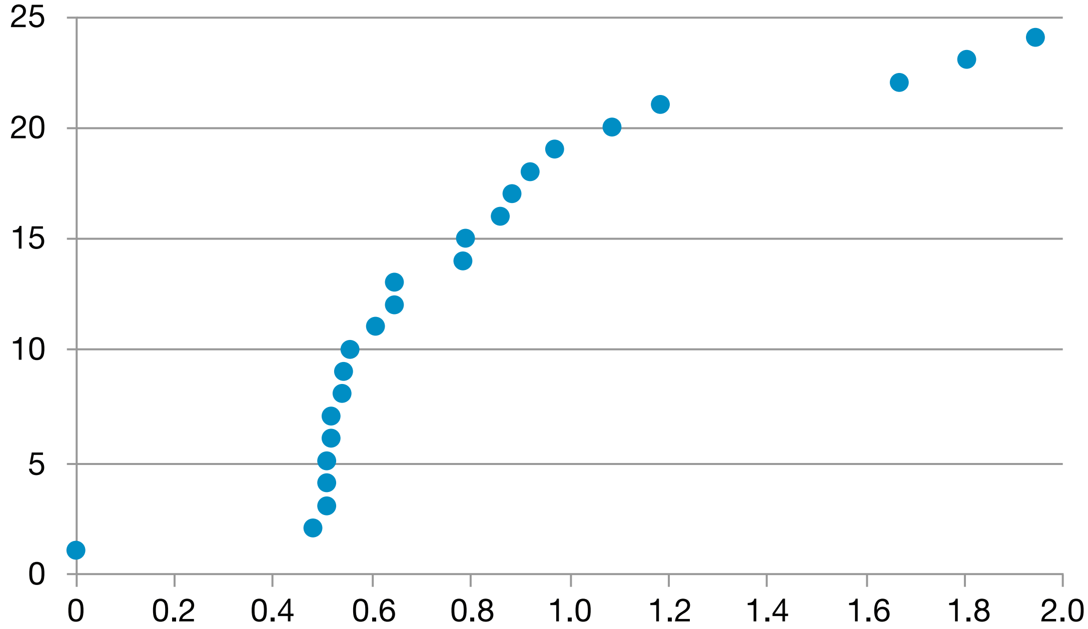
The highest value does not seem to arise from exceptional features of the data.

Thames Valley PIC: cumulative distribution of model estimates of 203 camera multiples



The highest value and two of the values between 2 and 3 are based on only one year's data before camera establishment other than the last three full such years, and another two of the values between 2 and 3 arise from cameras where there were hardly any PIC in just a few years until the last three full years before establishment.

Warwickshire PIC: Cumulative distribution of model estimates of 24 camera multiples



The first and third of the five values exceeding 1 are based on only two and one years' data after camera establishment, respectively

Numbers of collisions in the vicinity of cameras per year in years in which it is assumed that these numbers may have affected where to establish cameras

Estimates of the multiple m_b for PIC and FSC across each of the partnership areas under Assumptions (a) and (b) are shown in Tables A2.3 and A2.4, respectively, together with their estimated 95% confidence intervals. As discussed in section 4.1, Assumption (b) is preferred.

Table A2.3: Estimates of multiples m_b for the number of PIC per year in the vicinity of cameras in years in which it is assumed that these numbers may have affected where to establish cameras vs other years before camera establishment, across partnership areas

Partnership area (number of cameras considered)	Assumption (a)		Assumption (b)	
	Multiple m_c	95% confidence interval	Multiple m_c	95% confidence interval
Cambridgeshire and Peterborough (47)	1.129	1.038, 1.229	1.188	1.093, 1.291
Leicester, Leicestershire and Rutland (15)	0.986	0.881, 1.104	1.149	1.032, 1.280
Lincolnshire (50)	1.547	1.373, 1.743	1.392	1.232, 1.571
Merseyside (33)	1.116	1.036, 1.202	1.150	1.069, 1.274
South Yorkshire (56)*	1.140	1.040, 1.249	1.085	0.989, 1.190
Staffordshire and Stoke-on-Trent – few (42)	0.969	0.903, 1.039	1.011	0.943, 1.084
Staffordshire and Stoke-on-Trent – many (26)	0.869	0.853, 0.885	0.899	0.883, 0.916
Sussex (55)	1.109	1.013, 1.213	1.172	1.072, 1.280
Thames Valley (203)	1.032	0.971, 1.097	1.130	1.063, 1.201
Warwickshire (24)	1.143	0.988, 1.323	1.329	1.158, 1.527

* Figures under Assumption (a) in first edition were based on incomplete data.

Six of the ten results in Table A2.3 point clearly to higher numbers of PIC per year in the years concerned than in the other years before camera establishment under both assumptions, and a further two do so under Assumption (b). Under this assumption, six are higher by between about 8% and 19%, and two higher by over 30%. One indicates little difference between the numbers in the two sets of years and one points clearly to numbers lower by about 10% in the years concerned.

These estimates can be combined by taking their geometric mean weighted by number of cameras to provide a combined estimate m_b = 1.141 with 95% confidence interval (1.107, 1.176), indicating numbers of PIC higher by about 14% under Assumption (b) compared with 1.091, (1.058, 1.124) and 9%, respectively, under Assumption (a).

Table A2.4: Estimates of multiples m_b for the number of FSC per year in the vicinity of cameras in years in which it is assumed that these numbers may have affected where to establish cameras vs other years before camera establishment, across partnership areas

Partnership area (number of cameras considered)	Assumption (a)		Assumption (b)	
	Multiple m_c	95% confidence interval	Multiple m_c	95% confidence interval
Cambridgeshire and Peterborough (47)	1.165	0.966, 1.415	1.241	1.035, 1.489
Leicester, Leicestershire and Rutland (15)	1.233	0.906, 1.677	1.444	1.076, 1.937
Lincolnshire (50)	1.785	1.448, 2.201	1.704	1.382, 2.101
Merseyside (33)	1.452	1.184, 1.779	1.523	1.248, 1.858
South Yorkshire (56)*	1.151	0.927, 1.429	1.350	1.095, 1.664
Staffordshire and Stoke-on-Trent – few (42)	0.988	0.791, 1.233	1.242	1.011, 1.526
Staffordshire and Stoke-on-Trent – many (26)	0.709	0.664, 0.757	0.768	0.720, 0.818
Sussex (55)	1.300	1.075, 1.571	1.404	1.166, 1.691
Thames Valley (203)	0.980	0.831, 1.155	1.272	1.080, 1.497
Warwickshire (24)	1.570	1.192, 2.067	2.139	1.593, 2.645

* Figures under Assumption (a) in first edition were based on incomplete data.

Five of the ten results in Table A2.4 point clearly to higher numbers of FSC per year in the years concerned than in the other years before camera establishment under both assumptions, and a further four do so under Assumption (b). Under this assumption, eight are higher by between about 24% and 70%, one is higher by about 114%, and the other result points clearly to numbers lower by about 23%.

These estimates can be combined by taking their geometric mean weighted by number of cameras to provide a combined estimate $m_b = 1.331$ with 95% confidence interval (1.239, 1.435), indicating numbers of FSC higher by about 33% under Assumption (b) compared with 1.136, (1.054,1.224) and 14%, respectively, under Assumption (a).

Tables A2.3 and A2.4 and the combined estimates based on them indicate that Assumption (b) leads to consistently and appreciably higher estimates of the multiple m_b than Assumption (a), which is consistent with the former assumption being the more effective in excluding the effect of regression to the mean.

A2.3 Discussion

The statistical models presented here show how estimates can be made of the multiples m_c and hence the percentage difference across partnership areas between numbers of PIC and FSC per year in the vicinity of cameras in years well before and just before camera establishment and in years after establishment, largely excluding the effect of regression to the mean and allowing for concurrent general changes in the level of occurrence of collisions. The resulting estimates differ appreciably among partnership areas, but can be combined to provide an overall estimate for all the areas considered taken together.

The ratio of corresponding values of m_c for FSC and PIC provides an estimate of the proportion of collisions in the vicinity of cameras that are fatal or serious after camera establishment as a multiple of the corresponding proportion in years well before and just before camera establishment, and thus of the change in this measure of severity of collisions. These values of m_c together with corresponding values from models for KSI and CAS enable changes in the other three indicators of severity defined in section 4.4 to be estimated. This is illustrated at the end of Appendix 3.

At the level of individual cameras, the estimated values of the percentage difference between numbers of PIC per year in the vicinity of the camera in years well before and just before camera establishment and in years after establishment have wide confidence intervals, but their distributions enable exceptional values to be distinguished from those for most cameras, which are densely distributed over a limited range.

Making separate estimates of the percentage difference across partnership areas between numbers of PIC and FSC per year in the vicinity of cameras in years well before and just before camera establishment, and in the years in which these numbers may have affected decisions about where to establish cameras, provides an indication of the likely contribution of regression to the mean to the change in collision numbers between the latter years and years after establishment. These estimates also differ appreciably among partnership areas, but can again be combined to provide an overall estimate for all the areas considered taken together.

Appendix 3: Manual Analysis of Full Collision Data

A3.1 Method

For each camera, the average number of PIC per year after establishment of the camera, as a multiple of the average number per year before establishment excluding much of the effect of regression to the mean and allowing for concurrent changes in the occurrence of PIC across the partnership area concerned, can be estimated by:

$$m_c = aB/(b + 1)A \qquad \text{(A3.1)}$$

where:

a = number of PIC in the vicinity of the camera in all years up to and including 2010 throughout which the camera was established;

A = number of PIC recorded in the partnership area in the same years in STATS19;

b = number of PIC in the vicinity of the camera in all years from 1990 up to camera establishment, but excluding the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera; and

B = number of PIC recorded in the partnership area in the same years in STATS19.

The addition of 1 to *b* in the denominator is a standard correction for bias.

A convenient measure of the variability of the estimate m_c is the variance var(ln m_c) of ln m_c, the natural logarithm (logarithm to the base *e*, where $e \approx 2.718$) of m_c. If *a*, *A*, *b* and *B* were what are described statistically as Poisson distributed, then unless *a* or *b* was zero, var(ln m_c) would be estimated approximately by $\frac{1}{a}-\frac{1}{A}+\frac{1}{b}-\frac{1}{B}$. In practice $\frac{1}{A}$ and $\frac{1}{B}$ are very small compared with $\frac{1}{a}$ and $\frac{1}{b}$ and *a* and *b* are roughly Poisson distributed. So var(ln m_c) can be estimated for practical purposes by $\frac{1}{a}+\frac{1}{b}$ unless either *a* or *b* is zero.

It follows that the 95% confidence interval of the estimate m_c given by Equation (A3.1) is approximately:

$$\{m_c/\exp[2\sqrt{(1/a + 1/b)}],\ m_c\exp[2\sqrt{(1/a + 1/b)}]\} \quad \text{(A3.2)}$$

unless either *a* or *b* is zero, where for any *x*, exp(*x*) is *e* raised to the power *x* . For an individual camera, this confidence interval will usually be wide compared with the value of m_c.

Estimates of the corresponding multiple across a set of cameras taken together, such as all those in a particular locality or on roads of a particular kind or across a partnership area, will usually have confidence intervals that are less wide.

If the estimate of m_c for each camera in the set is calculated from Equation (A3.1), then the estimated multiple across a set of *n* cameras is given by:

$$m = \exp\{(\Sigma \ln m_c)/n\} \quad \text{(A3.3)}$$

with 95% confidence interval approximately:

$$\{m/\exp[2(\sqrt{\Sigma(1/a + 1/b)})/n],\ m\exp[2(\sqrt{\Sigma(1/a + 1/b)})/n]\} \quad \text{(A3.4)}$$

provided that none of the values of *a* or *b* for any of the cameras is zero, where for any *x*, Σx denotes the sum of the values of *x* for all of the *n* cameras.

Cameras for which either a or b is zero should be omitted from the calculation, with corresponding reduction in the value of n, but the width of the confidence interval will then tend to be underestimated.

Addendum March 2019

Equations (A3.3) and (A3.4) are based on taking a simple average across the sites, giving equal weight to the result from each site. In recent discussion of use of this method in practice, it has been pointed out that the accuracy of the estimates provided by the various sites can differ greatly from site to site, and the average for the local area should take account of this by giving greater weight to results from sites providing more accurate estimates. The author is grateful to his colleague Professor Mike Maher for pointing out the desirability of taking weighted averages and recommends using the following Equations (A3.5) and (A3.6) in place of equations (A3.3) and (A3.4) to achieve this in a statistically standard way.

$$m = \exp\{\Sigma[ab\ln(m_c)/(a+b)]/\Sigma[ab/(a+b)]\} \qquad \text{(A3.5)}$$

$$(\, m/\exp\{2/\sqrt{\Sigma[ab/(a+b)]}\}\,,\ m\exp\{2/\sqrt{\Sigma[ab/(a+b)]}\}\,) \qquad \text{(A3.6)}$$

In the example calculations in Section A3.2 following, the discussion immediately after Table A3.2 has been extended to illustrate the use of Equations (A3.5) and (A3.6) and how the results differ from those given by Equations (A3.3) and (A3.4).

It should be noted that equations (A3.3) and (A3.4) played no part in the example calculations in Appendix 2, in which the computer software used its own appropriate method for estimating the overall effects of cameras across each partnership area.

End of Addendum

Similarly, the average number of PIC per year in years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera can be estimated as a multiple m_b of the average number per year in earlier years by replacing a and A in the foregoing calculations by the numbers of PIC in the vicinity of the camera and in the partnership area, respectively, in all the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the camera.

Corresponding calculations can be made for numbers of FSC, but confidence intervals are likely to be very wide unless quite a large number of cameras, say several tens, are considered together.

Corresponding calculations can also be made for numbers of CAS or KSI, but because these numbers of casualties have greater random variation than the roughly Poisson-distributed numbers of collisions, the estimated confidence intervals should be regarded as approximately, say, 80%, rather than 95%, intervals.

Numbers of CAS, FSC and KSI can be used together with corresponding numbers of PIC to calculate indicators of severity:

(number of CAS)/(corresponding number of PIC) = number of casualties per collision;
(number of FSC)/(corresponding number of PIC) = proportion of collisions that are fatal or serious;
(number of KSI)/(corresponding number of CAS) = proportion of casualties that are killed or seriously injured; and
(number of KSI)/(corresponding number of FSC) = number of people killed or seriously injured per fatal or serious collision.

For the values of these indicators to be meaningful, they should be calculated for appreciable numbers of cameras or years or both, so that the numbers in the numerator and denominator are at least well into double figures.

A3.2 Example calculations for Leicester, Leicestershire and Rutland

Note: In the first edition of this report, these calculations were based on Assumption (a), and collision and casualty numbers for the year of establishment were omitted as a simplification. The calculations that follow are instead based on Assumption (b), and numbers of collisions and casualties in the year of establishment and the preceding year are included in the numbers *B* and *b*, and in the corresponding numbers of CAS, FSC and KSI in Tables A3.1 and A3.3.

This partnership mounted data for 1990–2010 for 16 cameras, the last of which was established only in August 2010, so that data up to 2010 for that camera included no full years after establishment. For the purposes of these calculations, the years from which the numbers of collisions or casualties may have been taken into account in deciding where to establish the cameras were assumed to be the first three of the last four full years before camera establishment.

Table A3.1 shows the values of *a*, *b*, *A*, and *B* and the results of using Equations (A3.1) and (A3.2) to estimate m_c and its 95% confidence interval for the other 15 cameras. The estimates of m_c given by the computer software are shown for comparison, and a graph of the two sets of estimates of m_c plotted one against the other is shown in Figure 4 in Section 2.

Table A3.1: Estimation of m_c for 15 cameras in Leicester, Leicestershire and Rutland

Camera	a	b	A	B	m_c by Equation (A3.1)	m_c from computer software
1	43	88	24,148	35,927	0.719	0.727
2	160	72	42,951	18,009	0.919	0.932
3	15	9	42,951	18,009	0.629	0.699
4	3	24	2,637	59,665	2.715	2.828
5	3	18	17,370	42,877	0.390	0.411
6	85	204	24,148	35,927	0.617	0.620
7	67	117	24,148	35,927	0.845	0.852
8	2	4	42,951	18,009	0.168	0.210
9	85	197	24,148	35,927	0.639	0.642
10	45	164	17,370	42,877	0.673	0.677
11	12	8	35,374	25,266	0.952	0.952
12	93	167	27,789	35,927	0.824	0.829
13	5	12	31,374	25,266	0.275	0.298
14	23	71	14,100	46,545	1.055	1.069
15	9	11	31,553	28,748	0.683	0.745
Total	650	1,166				

Table A3.2 shows the corresponding estimated 95% confidence intervals for the two sets of values of m_c.

Table A3.2: Estimated 95% confidence intervals for the two sets of estimates of m_c in Table A3.1

m_c by Equation (A3.1)	95% confidence interval by Equation (A3.2)	m_c from computer software	95% confidence interval from computer software
0.719	0.495, 1.043	0.727	0.505, 1.047
0.919	0.692, 1.220	0.932	0.706, 1.231
0.629	0.271, 1.462	0.699	0.306, 1.597
2.715	0.798, 9.240	2.828	0.852, 9.393
0.390	0.112, 1.357	0.411	0.121, 1.397
0.617	0.477, 0.799	0.620	0.481, 0.798
0.845	0.622, 1.148	0.852	0.631, 1.150
0.168	0.030, 0.950	0.210	0.038, 1.144
0.639	0.493, 0.828	0.642	0.498, 0.828
0.673	0.481, 0.942	0.677	0.487, 0.942
0.952	0.382, 2.372	0.952	0.401, 2.260
0.824	0.636, 1.067	0.829	0.643, 1.068
0.275	0.095, 0.797	0.298	0.105, 0.845
1.055	0.653, 1.704	1.069	0.668, 1.711
0.683	0.278, 1.678	0.745	0.309, 1.799

Addendum replacement March 2019

Whereas the originally recommended Equations (A3.3) and (A3.4) yield an estimate 0.672 for the value of m, with a 95% confidence interval of (0.542, 0.833), the Equations (A3.5) and (A3.6) now recommended in the Addendum in Section A3.1 yield an estimate of 0.743 with a 95% confidence interval of (0.670, 0.823). The estimate in Table A2.1 given by the statistical software used in Appendix 2 is 0.745, with a 95% confidence interval of (0.675, 0.822). The closeness of these two estimates indicates that the use of Equations (A3.5) and (A3.6) combines the estimates from the individual cameras in a way that is consistent with the technique used in the statistical software.

End of Addendum replacement

Table A3.3 then shows numbers of CAS, FSC and KSI corresponding to the numbers a and b in Table A3.1 to enable the indicators of severity over the years well before and just before camera establishment to be compared with those in the years after establishment in the vicinities of all 15 cameras taken together.

Table A3.3: Numbers of CAS, FSC and KSI corresponding to numbers *a* and *b* of PIC in Table A3.1

Camera	CAS *a*	CAS *b*	FSC *a*	FSC *b*	KSI *a*	KSI *b*
1	52	115	4	6	5	7
2	174	93	6	10	6	12
3	16	20	0	3	0	3
4	5	41	0	9	0	10
5	5	37	3	6	3	9
6	110	279	4	24	5	27
7	94	162	8	12	9	12
8	2	7	0	0	0	0
9	110	271	4	24	5	27
10	55	203	2	28	2	29
11	18	9	0	0	0	0
12	106	191	7	19	7	19
13	7	19	0	3	0	3
14	33	93	2	11	2	11
15	9	15	0	1	0	1
Total	796	1,555	40	156	44	170

For each camera, let *before* denote all years from 1990 up to camera establishment but excluding the first three of the last four calendar years before the camera was established, and let *after* denote all years up to and including 2010 throughout which the camera was established. Then it follows from the column totals in Tables A3.1 and A3.3 that in the vicinity of cameras across this partnership area, the:

- number of casualties per collision was 1.22 after, compared with 1.33 before;
- proportion of collisions that were fatal or serious was 0.0615 after, compared with 0.134 before;
- proportion of casualties that are KSI was 0.0553 after, compared with 0.109 before; and
- number of people KSI per fatal or serious collision was 1.100 after, compared with 1.090 before.

In terms of severity of collisions, this indicates that in the vicinity of cameras in this partnership area, the number of casualties per collision was about 9% lower, the proportion of collisions that were fatal or serious was about 54% lower, and the proportion of casualties that were KSI was about 49% lower after establishment of cameras than well before and just before their establishment. The number of people killed or seriously injured per fatal or serious collision occurring remained the same after as before.

The corresponding ratios of values of m_c for this area from Tables A2.1 and A2.2 and from the counterpart statistical models for CAS and KSI give matching results for the change in the number of casualties per collision and the absence of change in the number of people KSI per FSC. They indicate lesser reductions of only 34% and 27%, respectively, in the proportion of collisions that were fatal or serious and the proportion of casualties that were KSI, but these differences are understandable in view of the confidence intervals associated with both methods of estimation.

Appendix 4: Joint Analysis of Collision and Speed Data

For each camera for which speed data is provided on the relevant website, the data comprises either mean or median speed and either 85th percentile speed or percentage of speeds exceeding the speed limit. These are as observed in the vicinity of the camera on typically one or two dates before establishment of the camera and several dates subsequent to establishment. Dates provided do not always make it clear just which observations preceded establishment of the camera concerned.

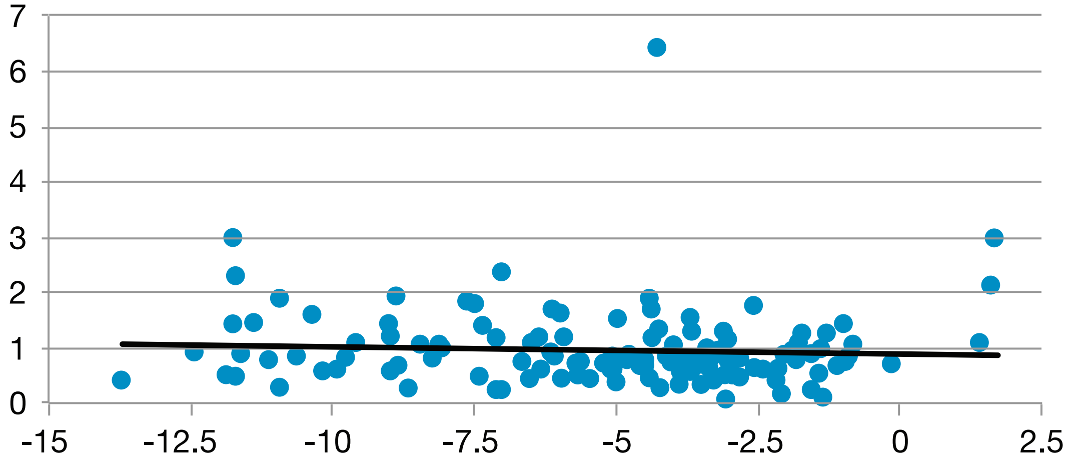
This data was examined for eight partnership areas:

- Cambridgeshire and Peterborough
- Leicester, Leicestershire and Rutland
- Lincolnshire
- Merseyside
- Staffordshire and Stoke-on-Trent
- Sussex
- Thames Valley
- Warwickshire

Where it was clear that one or more observations were made before establishment of a camera and one or more afterwards, the observations of mean speed before and after establishment were each averaged, and the difference between the two averages was taken as an estimate of the change in mean speed in the vicinity of the camera following its establishment. In a few cases, one or more of the observations were clearly out of line (for example, by 10 miles/h or 20 miles/h) with others made at the same site, perhaps because they had been made at another site and wrongly transcribed. Such observations were omitted when taking the averages for the camera concerned.

Changes in mean speed were estimated in this way for 132 cameras in these eight partnerships, and ranged from a reduction of 13.7 miles/h to an increase of 1.7 miles/h. All but three were reductions. The change in collision occurrence at the camera concerned was measured by number of PIC per year in the vicinity of the camera in years throughout which the camera may have been in operation as a multiple of the number in years before camera establishment other than the last three such years, i.e. according to Assumption (a), estimated relative to the number of PIC per year in the partnership area as described in Appendix 2. This is plotted against the estimated change in speed in the vicinity of the camera in the following diagram.

PIC multiple vs change in average speed in miles/h at 132 cameras in 8 partnerships



As is indicated by the fitted line, there is little or no correlation between the estimated changes in number of collisions and in average speed. The correlation coefficient is 0.0037. Omitting the camera with the exceptional multiple of 6.41 leaves the correlation coefficient at only 0.011. At all but 24 of the cameras, the speed limit was 30 miles/h. At the others, the limits were 40, 50, 60 or 70 miles/h in similar numbers, and the locations of the points from these cameras show no particular pattern within the scatter of 132 points. This calls for further investigation involving more sites and taking into account other characteristics of all the sites.

RAC Foundation

The Royal Automobile Club Foundation for Motoring is a transport policy and research organisation which explores the economic, mobility, safety and environmental issues relating to roads and their users. The Foundation publishes independent and authoritative research with which it promotes informed debate and advocates policy in the interest of the responsible motorist.

RAC Foundation
89–91 Pall Mall
London
SW1Y 5HS

Tel no: 020 7747 3445
www.racfoundation.org

Registered Charity No. 1002705
June 2013 (superseded in November 2013, with Addendum March 2019)

Designed and printed by
The Javelin Partnership Ltd
Tel: 0118 907 3494

Produced on paper from a managed sustainable source which is FSC certified as containing 50% recycled waste.