GETTING ANY BETTER?

1. ONE YEAR AGO

In June 2004, the Department of Transport lifted the embargo on a report by the PA Consulting Group and some UCL traffic experts. The report[1] attempted to elicit what could then be said with three years of data from road traffic accidents in defined neighbourhoods of over 2000 speed and red light cameras. The data had been collected under PA's direction and management by 24 Safety Camera Partnerships. The 15th June press release was technically and legalistically correct to state that 'the number of people killed or seriously injured at sites where safety cameras are in use has fallen by 40%' which 'equates to over 100 fewer deaths a year'. On the same day, Secretary of State Alistair Darling was encouraged to go further and give the BBC the soundbite that 'these figures prove that cameras save lives'.

They did not. Assuming there is some saving from the cameras, an unknown proportion of the fall in casualty numbers has to be attributed to a 'selection bias' otherwise known as 'regression-to-the-mean'. The main grounds for selection of a camera site is the number of people killed or seriously injured in three years on a short stretch of road. Sites where that number is fortuitously above its usual level will be the ones selected — and there is an ethical case for doing that. As far as possible, some allowance must be made for the fact that the fortuitous excess will disappear — and that would happen whether or not a camera has been installed.

Few of us can doubt that the respect that motorists give to yellow cameras results in fewer accidents in the vicinity of most cameras. But, if we accept the need for a realistic cost-benefit analysis of the partnership scheme, it is important either to estimate the rttm selection bias or, if that is not possible with the data collected by PA, to acknowledge that Alistair Darling's numbers (the 40% and the 100 lives saved) overestimate the truth to an unknown extent. If we are to spend more public money (whether from fines or general taxation) on technological options for making roads safer, that should be done on the basis of well-designed projects that would be informative about different options.

2. HEADLINE STUFF

The Department of Transport did not straightaway listen to the voices of those who were critical of the 'spin' (no other word will do) in the 2004 review and its presentation in the media. However, the critics were eventually muted when a press release in June 2005 announced that 'no new speed cameras will be installed for several months, as the government awaits a review of their effectiveness'. A spokeswoman said that it was important to get things right! The department now had a Chief Scientific Advisor, and it was encouraging to hear that it was extending the involvement of university researchers to include some who had been critical of the relative inattention to the question of selection bias.

So it came about that 'The national safety camera programme: Four-year evaluation report'[2] has been published in December 2005. It has probably made a good Christmas read for the critics and aficionados of such publications but, as something of a curate's egg, it has to be said that it is unfit for public consumption. The 'public' is unlikely to get beyond the one-page Executive Summary — attractively coloured to give some sort of emphasis. The statistics that it chooses thus to highlight are what it calls 'high level results' — where by 'high level' I think it means 'headline' as in the phrase 'headline figure' used to introduce comments on results in the body of the report. Among these results we find Alistair Darling's 100 (lives saved) unchanged by the extra year's data and his 40% (saving in killed or seriously injured) going up to 42%. The Summary has the *caveat* that these figures do not allow for 'selection effects (such as regression-to-the-mean)' but the page concludes that 'Overall, this report concludes that safety cameras have continued to reduce collisions, casualties and deaths'.

The Executive Summary arguably leaves a public (and that includes good people who want to help improve road safety but who may well overlook the fine caveat) — with the impression that the 100 lives saved and the 42% saving are bankable. As for those who *have* heard just a little about rttm selection bias (from critics such as Paul Smith of Safespeed) but who think they lack the statistical skill to explore the body of the report, it is difficult to know what they will make of it. They are probably left in a state of exploitable limbo.

3. INSIDE THE REPORT: A GUIDE

Appendix H. The new element in the four-year report[2] is its Appendix H. Two fresh hands, Dr Linda Mountain and Professor Mike Maher, were brought in from Liverpool and Edinburgh and given a free

hand to tackle the pesky problem of regression-to-the-mean (rttm). The basic idea is as follows. Find some way of predicting what the annual accident level for the stretch of road for any particular camera installation would have been had there been no installation. This is done by combining (a) a prediction based on characteristics of the stretch such as speed limit, road type, annual traffic etc (using a prediction formula that is good enough to give good approximations to the average accident levels in the national data-base for each combination of these characteristics) and (b) the accident number that led to the selection of the stretch for camera intallation. The weights used in the combination come from the analysis of national accident data that gave the two researchers their prediction formula. The weights make an automatic adjustment for the rttm selection bias. There may be technical arguments among experts about the precise way in which that should be done. But the method is now well-established and is probably the nearest we can get to a solution of the rttm problem. Its estimates of the effectiveness of the safety camera scheme in accident reduction are the only ones that can be taken seriously.

Unfortunately, although the rttm problem has been recognised by accident analysts for decades, the data collected by safety camera partnerships was not designed with rttm in mind. The data needed for the prediction formula to be applicable were available for only 317 of the more than 4400 cameras on which DfT based its claims. A further reduction came about when Linda and Mike cut the 317 down to the 216 urban cameras, for which speed limits are 30 or 40 mph. With a commendable determination to 'get things right', they made the cut because

(a) 'data concerning speeds are not routinely collected for road sections [stretches] before they have been selected for further investigation and possible remedial treatment',

(b) the prediction formula therefore could not include vehicle speed data among the characteristics of road stretches on which it was based,

(c) the DfT criteria for a camera installation include speed data as secondary selection criteria,

(d) if the installation stretches were above average in speeds (that are known to be correlated with accidents for roads with specified characteristics), the predicted level would be lower than it ought to be, and the camera effect defined as predicted level (after installation) minus observed level would be underestimated,

(e) data for the 216 cameras show that speed was used as a selection criterion for 60mph-limit rural installations but not for 30 mph urban ones (and only to a limited extent for the 16% of urban ones that have a 40 mph limit).

With these five careful considerations and the associated restriction to urban cameras, Appendix H re-

searchers Linda and Mike were confident that there would be little bias in their estimates of camera effect. The headlines would inevitably be about the killed and seriously injured. Appendix H was about accident ('collision') numbers rather than the associated casualty numbers, which is arguably more scientific. (The single collision of a bus and a lorry may well make the casualty case for a camera!) But there is an overall proportionality between collisions and casualties, so that the findings in Appendix H may be justifiably contrasted with the headlines. The findings for the statistic FSC defined as the *average (per camera) annual 'fatal and serious' collisions* are shown in Table 1. It is the 0.59 in this table that is the average predicted value for the after-installation period. The camera effect is the further reduction of $0.59 - 0.48 = 0.11$. As a percentage of the otherwise expected 0.59, the camera effect is therefore $100 \times 0.11/0.59 = 19\%$.

Table 1	FSC	Difference	95% confidence interval
Before installation (observed)	1.05		
		0.10	(0.08, 0.12)
Allowing for trend (estimated)	0.95		
		0.36	(0.30, 0.43)
Allowing for trend and rttm (estimated)	0.59		
		0.11	(0.02, 0.19)
After installation (observed)	0.48		

Main report. The main report, excluding Appendix H, was authored by PA and UCL. It has a lot to say about Appendix H. Noting that there are some significant differences in casualty levels between the subset of 216 cameras (used, I suggest, simply to *demonstrate* the necessity for an allowance for rttm) and the 4400 cameras in the whole study, PA/UCL question the relevance of the Appendix H findings:

> Estimation of reliable adjustment for the effects of regression-to-mean that apply generally would require substantial efforts in purpose-specific study design, monitoring and data collection that are not readily compatible with monitoring a full scale implementation.

In fact, for those killed or seriously injured (KSI) on the mobile camera stretches (76% of the 216 urban ones) there appears to be little difference. The 24% fixed ones do, however, have a KSI level that is 37%

lower. When PA/UCL apply their KSI model to the 216-camera data, they therefore do so separately for the mobile and fixed cameras and get estimates of 0.41 and 0.18 respectively for the analogues of the camera effect 0.11 in Table 1. When the 0.41 and 0.18 are combined with the weights (76% and 24%), the estimated camera effect for all 216 cameras is 0.35.

With this estimate in hand, we can construct the simplified analogue of Table 1 in which KSI predictions are converted to FSC equivalents.

Table 2	FSC	Difference
Before installation (observed)	1.05	
		0.22
Allowing for trend and predictors in the KSI model (estimated)	0.83	
		0.35
After installation (observed)	0.48	

Now, when expressed as a percentage reduction after allowing for trend and other factors but not for rttm, the camera effect is $100\times35/83 = 42\%$. This is presumably the estimate that PA/UCL would have been happy to offer as the 'headline' figure for KSI savings if they had had only the 216-camera data-base to work on. Despite the disparagement of the Appendix H findings in our quotation, they might also see the 42% as confirmation of the figure they gave Alastair Darling.

Questions of bias. Little more need now be said about the manifest bias in figures such as the 42%. I am sure that the Department of Transport will one day apologise for its mismanagement of the speed camera project and for the misrepresentation of what the accident data that the project is generating really tell us. We may have to wait a little longer. The Secretary of State revealed that the ground is shifting when he qualified his earlier claim that 2004's figures *proved* on their own that 'cameras save lives', with a 'whichever way you look at it' comment on the 2005 report. We must hope that the qualification does not mean that DfT is maintaing neutrality between taking account of rttm and ignoring it — in the same way that the Home Office's Fiona McTaggart saw symmetry between the efforts of the Birmingham Rep to maintain freedom of speech and the freedom of others to deny that freedom.

I suspect that DfT's reluctance to be more forthcoming has something to do with its reliance on defective contractual machinery[3] which induces a bias towards cover-up rather than transparency. Even now, more

than a year after the rttm problem hit the media[4], the reluctance shows in the structure of the Four-year Report: a misleading Executive Summary, the corrective work hived off into a technical appendix, and face-saving emphases, even a significant error of logic, in the main report.

Pedestrian victims, especially those of children, are particularly poignant to all except petrol-heads. So quite rightly, the report makes separate analyses of the KSI numbers for these two categories. The error in its comments on the results (a reduction of 29% for pedestrians and 32% for children) is contained in the very plausible assertion:

> The results for pedestrian [or child] ... casualties will be affected less by regression-to-mean than are all ... casualties because these are not an explicit part of the selection rules.

That cannot be. With the reasonable definition of an 'rttm effect' (for any particular category of casualty) as the *difference* between the sum of the observed casualty numbers on the stretches selected and the true (with all chance variation excluded) level of that sum, expressed as a percentage of the sum — breathe now! — the rttm effect for 'all' KSIs is a weighted average of the rttm effects for the different categories into which 'all' may be divided. The average has to be less than the largest of those effects — whose corresponding category (also 'not an explicit part of the selection rules') would therefore violate the plausible assertion. Errors are part of the life-blood of science, providing the raw material for further progress. Nearly all of us make them, as Coleridge recognised so beautifully:

> Though all men are in error, they are not all in the same error, nor at the same time ... each therefore may possibly heal the other ... even as two or more physicians, all diseased in their general health, yet under the immediate action of the disease on different days, may remove or alleviate the complaints of each other.

To put Coleridge's idea to work in the arena of government science would require a good deal of humility all round. It may also clash with the necessarily divergent interests of private sector contracts. But the only alternative to humble pie may be eventual humiliation. Is DfT ready to make the choice?

Mervyn Stone, December 2005.

References

1. Gains, A., Heydecker, B., Shrewsbury, J. and Robertson, S. (2004) The national safety camera programme: Three-year evaluation report. PA Consulting Group and Centre for Transport Studies, London.

2. Gains, A., Nordstrom, M., Heydecker, B., Shrewsbury, J. (2005) The national safety camera programme: Four-year evaluation report. PA Consulting Group and Centre for Transport Studies, London. (www.dft.gov.uk)

3. Stone, M. (2005) Accumulating evidence of malfunctioning contractual government machinery. *Public Money and Management*, **25**, 82-86 (Research Report 254 at www.ucl.ac.uk/stats/research).

4. Stone, M. (2004) Adjudication of the Radio 4 Today Programme Speed Tribunal. (Research Report 245 at www.ucl.ac.uk/stats/research).