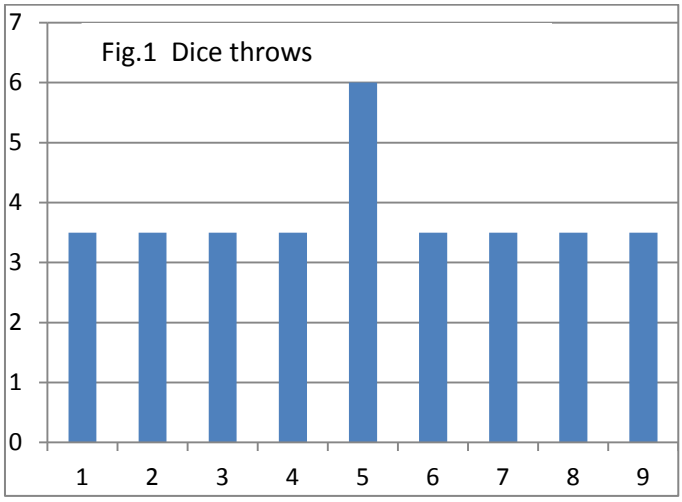
Appendix F - Site Selection Bias (SSB) and Regression to Mean (RTM)

The worst of many serious blunders of the Eight Area Trial of 2000/01 and others since was **treating as insignificant the RTM reductions made inevitable by installing cameras where many collision had recently occurred.** This was all the more astonishing as RTM has been well understood for more than 100 years and as such must be part of statisticians' basic training! As the authors of that and two more reports admitted in their Fourth report of 2005, **that failure resulted in cameras being credited with collision reductions caused in part or in whole by RTM, yet the Hypothecation Scheme that spread cameras across the country was based on those seriously flawed claims.**

Equally bizarre were claims that **no data were available for RTM effects to be estimated.** Should it not have been obvious that as **RTM is the result of biased site selection and nothing whatever to do with the cameras themselves**, it can be estimated using police Stats19 data alone? As indeed this analyst did in 2008.

Understanding RTM, a consequence of the Laws of Chance

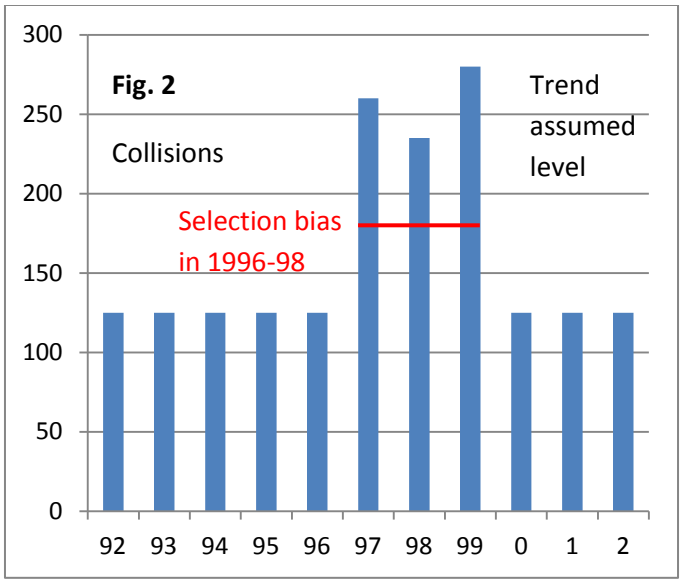


If [say] 6,000 people throw dice at the same time the laws of chance ensure that close to **1,000 score 1, close to 1,000 score 2**, etc. with an average score of (1+2+3+4+5+6)/6 i.e. **3.5.**

If the **1,000** or so who **scored 6 throw again**, chance again ensures that **close to 167 score 1, close to 167 score 2** etc. again with an **average very close to 3.5.**

The odds against **all 1,000** who scored 6 last time doing so again are **6^{167}:1** (a very large number indeed) and the odds against an **average score significantly different from 3.5 are also extremely high.** Fig.13 shows how, for these reasons, the **return to normal after selection is instantaneous.**

Estimating RTM % falls is perfectly easy!



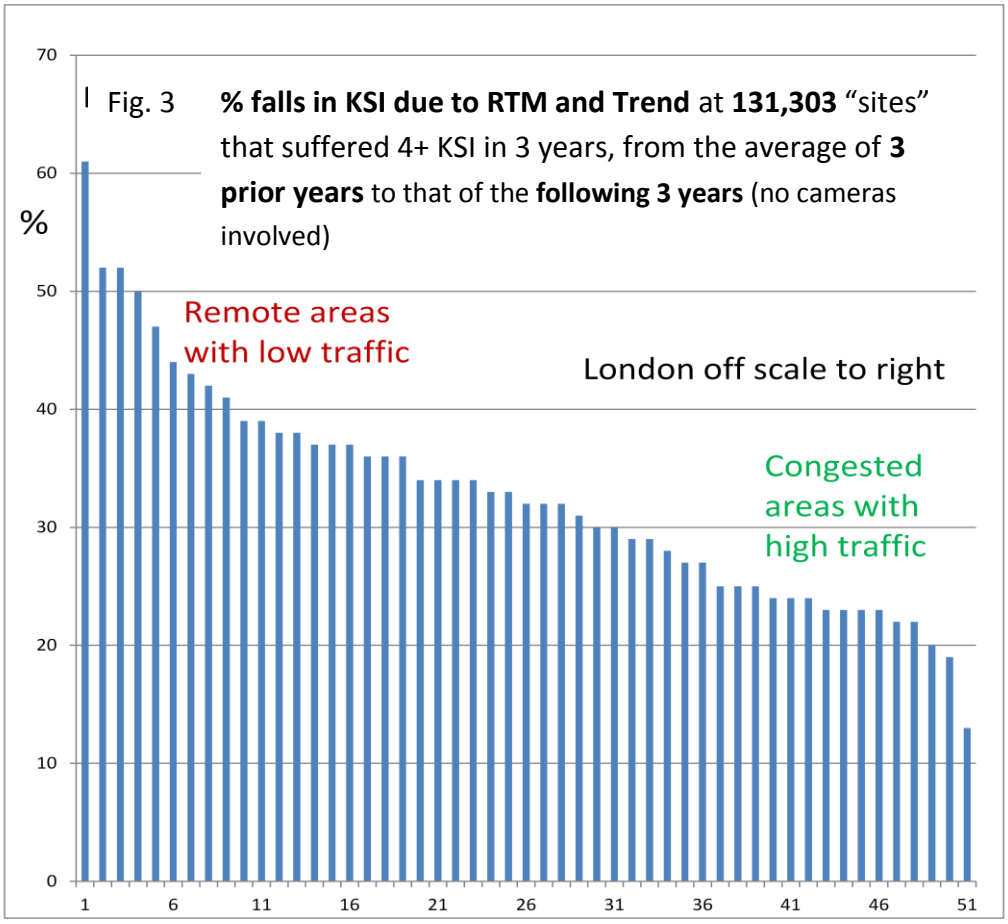
This is vitally important!

Collision totals are not of course determined solely by chance but **precisely where, when and with what severity** each collision occurs **clearly does owe much to chance.** RTM theory applies so (for example) **50 sites normally averaging 2.5 KSI pa but selected for 4+ FSC in 1996-98** would appear as Fig. 14, **the laws of chance again ensuring an RTM fall** to at least close to normal **the moment the site selection period ends**.

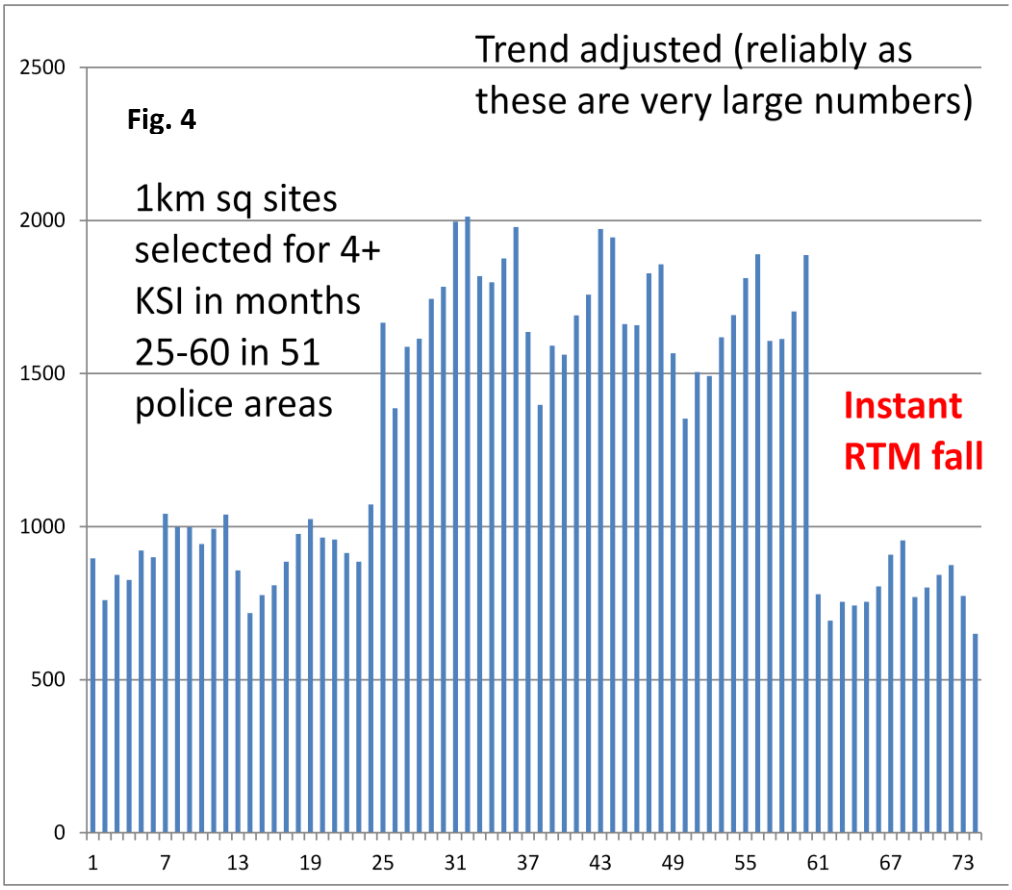
This is vitally important!

In 2008, using 1991-2007 Stats19 data, this analyst identified **131,303 1km sq** locations (approximating to sites) **that had suffered at least 4 KSI in 3-year periods, qualifying for cameras but (obviously) not receiving them:**

Graph of RTM and Trend Effects across 51 police areas



Below – proof that RTM falls are instantaneous and complete



Trend effects being small over an average of only 3 years the **RTM falls** [Fig. 3] are clearly of the **same order of magnitude as reduction often claimed for cameras when RTM was ignored**.

As most real sites cover less than 1sq km their smaller KSI numbers would result in **more bias and higher RTM % falls** than shown here.

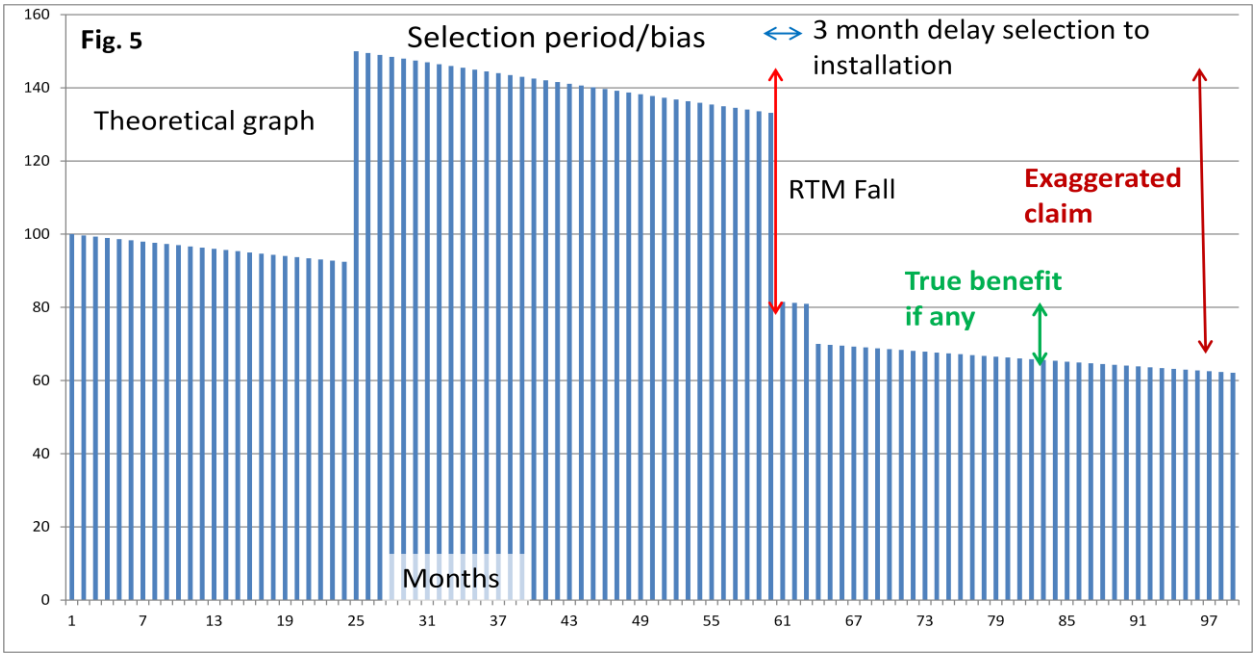
Cognitive Dissonance and Group Think at the DfT

To clarify RTM effects for a 2014 meeting with the foot-dragging Department of Transport, that analysis was repeated using **monthly totals** [Fig.4] to indicate the likely **order of magnitude of RTM falls and also that numbers return to normal as site selection ends.** **This is very important as it allows camera effects to be differentiated from all others on the basis of timing alone.**

However 6 senior DfT officials told this analyst that they could not understand the method or the significance of the results, were too busy to do so and would therefore ignore them unless and until they were peer reviewed and published - all very strange given that the method merely demonstrates that RTM theory understood for 100 years is correct!

Clarification of how ignoring RTM leads to bogus camera benefit

When RTM falls are ignored or miscalculated, cameras are credited not only with any benefit they achieve but also with (usually much larger) RTM reductions, **thus invalidating the results**.



It is impossible to correct accurately for SSB/RTM

This is because the complex relationship between SSB/RTM and the many factors which affect collision numbers include:

- the normal level at each site and hence its volatility
- the selection threshold relative to the normal level
- chance, especially when data volume is low
- varying trends
- differing installation dates
- differing installation delays after site selection periods (as distinct from the point above)

Furthermore, even if a valid relationship between SSB/RTM and those factors could be identified, **accurate correction would still be impossible because:**

- site selection criteria are not used at "public concern" sites
- site selection criteria are applied inconsistently elsewhere
- many of the above parameters are in any case not documented fully, if at all.

It follows that all claims of camera benefit based on the effects of SSB and RTM having been ***rigorously eliminated*** **demonstrate only that the analyst responsible does not understand the problems.**

Indeed, this analyst is astonished that so many supposedly expert analysts wasted so much time, effort and (often public) money **trying to quantify RTM accurately by analysing pre-installation data without realising that (a) it was impossible and (b) there was no need to do it anyway, as all SSB and RTM effects must, by definition, end before camera installation!** end