

Peer review of the June 2007 VTRC report: “The Impact of Red Light Cameras (Photo-Red Enforcement) on Crashes in Virginia”

Bhagwant Persaud and Craig Lyon

Draft for Review

November 2, 2007

PREAMBLE

There is fairly substantial research that suggests that the safety impact of red light cameras varies across sites, depending on factors such as traffic volumes and the relative proportions of various crash types. The VTRC results are consistent with this suggestion, but do go beyond the common understanding by suggesting further that red light cameras can actually be detrimental to safety. An initial review of the VTRC study by the IIHS and by Persaud and Lyon, Inc. indicated that the conclusions might be different, and more in accord with conventional wisdom, if procedures used for some aspects of the analysis were more robust. This concern prompted the IIHS to solicit a more formal review of the VTRC study by Persaud and Lyon, Inc.

SUMMARY OF THE VTRC STUDY (*taken almost verbatim from the VTRC report*)

To reduce red light running in Virginia, six jurisdictions (Alexandria, Arlington, Fairfax City, Fairfax County, Falls Church, Vienna) deployed red light cameras at some point during the 10-year period when they were permitted under Virginia law.

The report documents the safety impacts of those cameras based on 7 years of crash data for the period January 1, 1998, through December 31, 2004. Consistent with the findings of a previous Virginia study, the study finds that cameras are associated with an increase in rear-end crashes (about 27% or 42% depending on the statistical method used) and a decrease in red light running crashes (about 8% or 42% depending on the statistical method used). The report also shows that there is significant variation by intersection and by jurisdiction: one jurisdiction (Arlington) suggests that cameras are associated with an increase in all six crash types that were explicitly studied (rear-end, angle, red light running, injury red light running, total injury, and total) whereas two other jurisdictions saw decreases in most of these crash types.

It is therefore not surprising that when the comprehensive crash costs for rear-end and angle crashes are monetized, the cameras are associated with an increase in crash costs in some jurisdictions (e.g., an annual increase of \$140,883 in Arlington) and a net reduction in comprehensive crash costs in other jurisdictions (e.g., an annual reduction of \$92,367 in Vienna). When these results are aggregated across all six jurisdictions, the cameras are associated with a net increase in comprehensive crash costs. However, when considering only injury crashes, if the three fatal angle crashes that occurred during the after period are removed from the analysis (the only fatalities that occurred during the study out of 1,168 injury crashes), then the cameras were associated with a modest reduction in the comprehensive crash cost for injury crashes only.

These results cannot be used to justify the widespread installation of cameras because they are not universally effective. These results also cannot be used to justify the abolition of cameras, as they have had a positive impact at some intersections and in some jurisdictions. The report recommends, therefore, that the decision to install a red light camera be made on an intersection-by-intersection basis. In addition, it is recommended that a carefully controlled experiment be conducted to examine further the impact of red light programs on safety and to determine how an increase in rear-end crashes can be avoided at specific intersections.

STRENGTHS OF THE VTRC STUDY

There are a number of strengths of this effort which need to be viewed in the light of the concerns raised in the next section.

1. The study evaluated a relatively large number of treatment sites.
2. The study acknowledged the possibility of spillover effects as it affects the selection of reference comparison sites and the true effect of the programme.
3. The study identified jurisdictional differences in safety impact, a useful step in identifying factors that influence how effective the cameras are.
4. The study examined the differential effects on different crash and severity types and accounted for these in an economic analysis.
5. The study tried to validate the results from the primary methodology by comparing them to those obtained with other methodologies.
6. The study employed the state of the art empirical Bayes methodology for observational before-after studies, following the procedures in Ezra Hauer's book.
7. The study conclusions are carefully guarded in de-emphasizing the negative aspects of the results by recommending that the decision to install a red light camera be made on an intersection-by-intersection basis and that a carefully controlled experiment be conducted to examine further the impact of red light programs on safety and to determine how an increase in rear-end crashes can be avoided at specific intersections.

The researchers should be complimented for making their data publicly available on the web and have been forthcoming with details of their calculations, without which this review would be far from complete. However, despite these positive features of the VTRC study, our main concerns are with the execution of the study, not with the overall research approach which we believe is sound. These concerns/issues are detailed in the next section.

CONCERNS/ISSUES

In illustrating the issues we raise, we have chosen to focus on angle crashes, which is the key crash type typically evaluated by most researchers as the primary targets of red light cameras.

Issue 1: Crash predictions models used in the EB analysis, if calibrated with more conventional model forms may lead to different conclusions

The general form for all of the crash prediction models calibrated by VTRC for the EB analysis is as shown below:

$$y = \alpha \text{ Volume}^{b1} \text{ Speed}^{b2} \text{ Yellow}^{b3} \text{ Trucks}^{b4} (\text{Through lanes})^{b5} (\text{Left lanes})^{b6}$$

This is a highly unusual and debatable model form that ought to be rationalized by the VTRC researchers, since it implies zero crashes if any of the variables have a value of zero. In the absence of that rationalization, the more conventional model form, which avoids this "zero" problem, and which was in fact used by the VTRC researchers for their GLM analysis, might be preferred:

$$y = \alpha \text{ Volume}^{b1} e^{b2(\text{Speed}) + b3(\text{Yellow}) + b4(\text{Trucks}) + b5(\text{Through lanes}) + b6(\text{Left lanes})}$$

We suspect that in the model application, variables with a zero value are set to a very small number, but it is likely that as a result of this apparent mis-specification of the model form, and the relatively small sample sizes overall and for individual jurisdictions, the parameter estimates of many of the models used for the before-after study are suspect. The standard errors of these estimates have not been provided, but we suspect the significance and size of many. For example, for several models, the AADT exponent appears to have been set to 0.00001 or is very, very small (compared to a more conventional value of between 0.6 and 1.2).

As an example, consider the model for angle crashes for all jurisdictions combined. The calibrated VTRC model was:

$$y = \alpha \text{ Volume}^{0.01658} \text{ Speed}^{0.17567} \text{ Yellow}^{0.00001} \text{ Trucks}^{0.30002} (\text{Through lanes})^{0.00001} (\text{Left lanes})^{0.02783}$$

The values of α for each year were:

$$\alpha_1 = 2.64364$$

$$\alpha_2 = 2.45373$$

$$\alpha_3 = 2.55333$$

$$\alpha_4 = 2.50346$$

$$\alpha_5 = 2.20961$$

$$\alpha_6 = 2.57893$$

$$\alpha_7 = 2.15874$$

From our knowledge of existing crash prediction models (See, e.g., Bauer & Harwood (1999)¹ and Lyon et al. (2005)²) and our experience in developing these models, it is evident that the parameter estimate for Volume is unusually low at 0.01658. Regardless of which parameters may or may not have been estimated with an acceptable level of statistical significance, it is the parameter estimate for Volume which is at issue, since this variable is known to explain most of the variation in crash frequency across sites.

To illustrate the difficulty with the possible underestimation of the Volume exponent, we first calibrated a “trial” model using only traffic volume as an independent variable. (Note that a substantial model calibration effort is beyond the scope of this review; this also meant that for our calibration, each site-year was included as an observation without controlling for temporal correlations within sites.) We also included the before period data at untreated sites as the VTRC study did, although we do have reservations about this inclusion, as will be noted later. Our estimated model for angle crashes, with annual multipliers, is:

¹ Bauer, K M; Harwood, D W. Statistical Models Of At-Grade Intersection Accidents – Addendum. FHWA-RD-99-094; Available at: <http://www.tfhrc.gov/safety/99-094.pdf>

² Lyon C., Persaud B., Haq A. and S. Kodama, "Development of safety performance functions for signalized intersections in a large urban area and application to evaluation of left turn priority treatment". Transportation Research Record: Journal of the Transportation Research Board, 1908, pp 165-171, 2005.

$$\text{angle crashes/year} = \alpha \text{Volume}^{0.4518}$$

$$\alpha_1 = 0.0440$$

$$\alpha_2 = 0.0411$$

$$\alpha_3 = 0.0404$$

$$\alpha_4 = 0.0400$$

$$\alpha_5 = 0.0367$$

$$\alpha_6 = 0.0365$$

$$\alpha_7 = 0.0310$$

Note that the Volume exponent of 0.4518, though considerably larger than that in VTRC's model, is still on the low side compared to what might be expected based on the literature; it is possible that this may be a result of not being able to include minor road entering volumes.

Next, we compared the expected change in crash frequency due to traffic volume changes at one treatment site (#2), using the VTRC model and our estimated model. For this site, the volume changed from 54,000 in year 1 to 62,000 in the year following camera installation, a 15% increase. The VTRC model would estimate the expected increase in angle crashes due to traffic volume changes as 0.2% (calculated as $((62,000/54,000)^{0.01658} = 1.002)$). This increase in crash frequency is negligible and seems unreasonable for a volume increase of 15%. By contrast, our model estimates a more realistic 6% increase in the expected angle crash frequency ($((62,000/54,000)^{0.4518} = 1.06)$ for this volume change.

The average change in traffic volumes from the before to after period at the treatment sites is a 10% increase, with some increases being of the order of 50%; and 19 of the 28 treatment sites experienced an increase in volume. If, as we suspect, the VTRC model is underestimating the effect of the traffic volume increase, then it will, in the EB approach, underestimate the expected number of angle crashes in the after period had there been no treatment (which is compared to the crashes actually recorded after treatment to estimate the effect of the treatment). In so doing, it is likely underestimating the reductions in angle crashes that might have been expected and may even estimate an increase in these crashes.

Authors' Response to Issue 1: The Crash Prediction Model is not "Highly Unusual"

In reference to the crash estimation model (CEM), the reviewers write that "This is a highly unusual and debatable model form that ought to be rationalized by the VTRC researchers." We disagree. The form of our model is similar to that suggested by Ezra Hauer in his book *Observational Before-After Studies in Road Safety*, Pergamon, 1997 (see page 242). The difference between our model and that proposed by Hauer is that we have added to this model five variables that, based on a preliminary data analysis, appear to also influence the safety impacts of red light cameras. These variables are (i) truck percentages, (ii) speed limit, (iii) through lanes (iv) left turn lanes, and (v) the difference between the actual yellow time + grace period and the recommended ITE yellow time. Accordingly, we do not consider the form of our crash estimation model to be "highly unusual."

Further, the alleged "flaw" mentioned by the reviewers (the low estimate of the volume exponent) does not affect the conclusions of the study as will be shown later.

It should also be noted that not **all** volume exponent estimates are low and that higher volume exponents do not necessarily indicate that red light cameras will have a more beneficial impact on crashes. For example, Arlington's volume exponent (0.94918) in the CEM was used, in part, to estimate that the cameras are associated with an 89% increase in injury crashes (since $\theta = 1.89$) whereas Vienna's volume exponent for the injury crashes (0.00001) was used, in part, to estimate that cameras are associated with a 59% increase in injury crashes (since $\theta = 1.59$). This shows that our model form which takes other confounding factors into consideration, rather than relying only on volume, disproves the implication in the last paragraph of issue 1 that the EB approach underestimated the expected number of angle crashes without treatment in the after period.

In addition, the reviewers implied that the AADT exponent "appears to have been set." To clarify, the AADT exponent was not "set" to any particular value but rather was determined through the development of the CEMs. It should also be clarified that, for the analysis for individual jurisdictions, minor road volumes were included where there existed a sufficient number of such sites.

The above discussion refutes the "highly unusual" form of the model. In the interests of learning as much as possible from research, however, we have experimented with alternative formulations such as the one the reviewers have suggested, as will be discussed in Table R1.

Issue 2: Crash prediction models used in the EB analysis, if calibrated with data that do not include treatment sites, may lead to different conclusions

For the VTRC study, the before period data at the treated sites were combined with the comparison/reference site data to calibrate the crash prediction models used in the EB analysis. This approach of combining the two sets of data is acceptable if there are insufficient data to calibrate a reliable crash prediction model, but only *providing that the crash frequencies for the two sets of sites are comparable*.

Our analysis of the data provided suggests that this key proviso may not be satisfied. As Figures 1 and 2 show, for angle crashes in all jurisdictions, the mean crash rates (by two different measures: crashes per site-year and crashes per million vehicles on major road) are consistently and substantially higher at the treated sites in the before period than at the reference sites. (Note that, for these plots, the before period at treated sites ends at year 4, since the sample sizes for the remaining years are small).

Figure 1: Mean Right-Angle Crashes per Site-Year by Year #

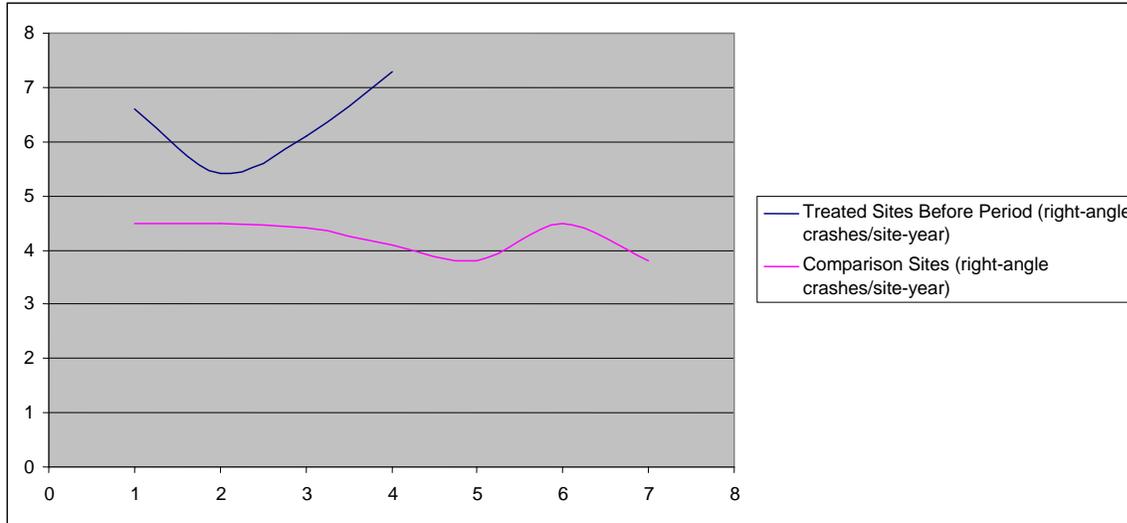
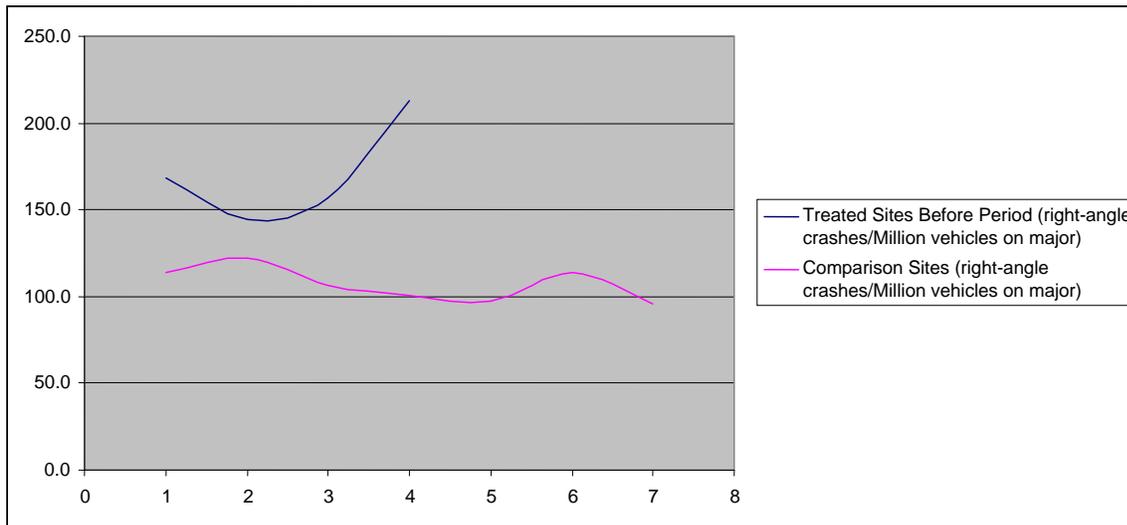
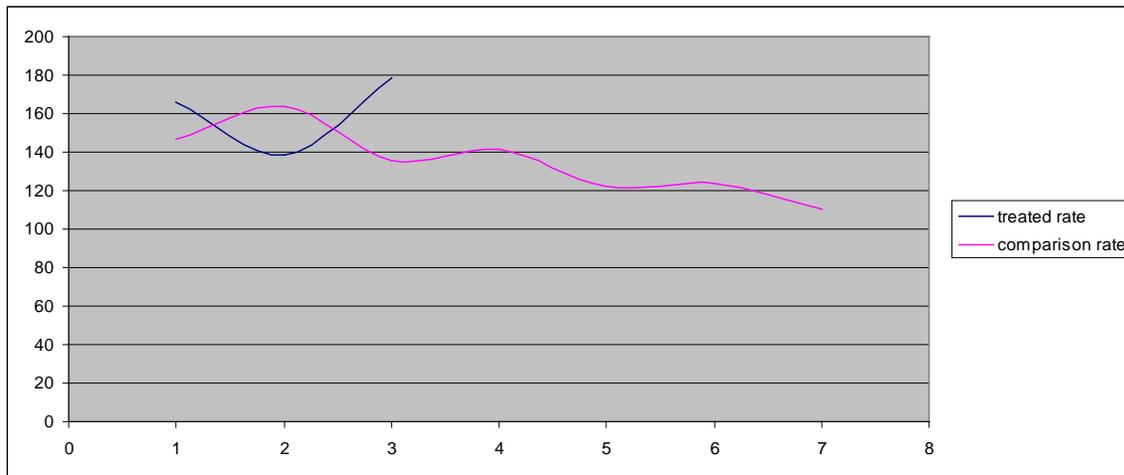


Figure 2: Mean Right-Angle Crashes per Million Vehicles on Major Road by Year #



The higher angle crash rates at the treated sites suggest that these sites must have been selected based on a high frequency of crashes targeted for reduction by red light cameras. This is sound engineering and enforcement practice and is affirmed by a similar comparison in Figure 3 for rear-end crashes – a crash type not targeted by red light cameras – which shows that the treatment and reference sites are much more comparable.

Figure 3: Mean Rear-End Crashes Per Million Vehicles On Major Road



Unfortunately, this application of sound practice means that treatment and reference sites should not be mixed. The upshot of this mixing is that the composite crash prediction model will predict fewer crashes at the treatment sites and more crashes at the comparison sites than is the reality. How much “fewer” and how much “more” will depend on the relative amounts of treatment and comparison site years used. (For the case of angle crashes in all jurisdictions, the fraction of treated sites is 0.39 (=28/72) and for site-years is 0.22 (=90/398)).

If the higher angle crash rates at the treatment sites were in fact genuine, then the use of lower angle crash rate reference sites in calibrating the crash prediction models would underestimate crashes without treatment in the EB analysis.

If, on the other hand, the angle crashes at the treatment sites were randomly high, creating a selection bias, then those sites should not be used in developing models that are used in part to correct for regression to the mean bias. In this case the crash predictions models would overestimate crashes without treatment in the EB analysis. However, this overpredicting model will underestimate the annual time trend factors for the after period with the high crash frequency treatment sites removed in estimating those annual factors. The final result is that the estimates of the expected crashes in the after period without treatment could in fact be underestimated and the treatment benefit underestimated.

Authors’ Response to Issue 2: Crash prediction models, when based exclusively on data from untreated sites, do not necessarily lead to higher values of θ [that is exclusion of before treatment sites from CEM calibration data does not necessarily lead to an underestimation of the treatment benefits.]

The impact of using before-treatment data in the crash estimation model may be considered in two ways.

- **[Angle crashes at treatment sites are NOT randomly high.]** The reviewers write that VTRC’s inclusion of [before] “angle crash rate reference sites in calibrating the crash

prediction models would underestimate crashes without treatment in the EB analysis.” This is not necessarily true. Keep in mind that the reviewers are talking about *before*-treatment sites, which VTRC used—along with comparison sites--to estimate the impact of not treating a site. Logically [and empirically as shown later], inclusion of higher risk before treatment sites in the development of CEMs, contrary to what the reviewers write, means that the VTRC CEMs would thus *overestimate* crashes without treatment. Thus the VTRC CEMs would, if this supposition were true, make red light cameras appear more effective than reality.

- **[Angle crashes at treatment sites are randomly high.]** Consider the reviewers’ second possible supposition that “angle crashes at the treatment sites were randomly high.” Again, the inclusion of data from the before-data from our treatment sites in our CEMs will, assuming before-site treatment crash rates are higher than comparison site treatment crash rates, tend to cause the VTRC CEMs to overestimate crashes without treatment. In short, the use of before-treatment data in the CEMs ensures that the beneficial impacts of red light cameras would not be diminished. This hypothesis is confirmed by the analysis shown in Table R1 that follows. (In fact, the reviewers note in part that “the crash predictions models would overestimate crashes without treatment in the EB analysis”—and they were correct in this part: red light cameras, when analyzed as desired by the reviewers, did not perform as well as they did when analyzed in the report.)

Three issues should be clarified.

First, the authors were not able to select the treatment sites as the decision to install cameras was made well in advance of this study being undertaken.

Second, the reviewers use the term “angle” in their review. To clarify, six different types of crashes were analyzed in the study, one of which were angle crashes and one of which were red light running crashes. The two types of crashes are not identical, as is evident from the definitions given on pages 5 and 6 of the report).

Third, the statement “rear-end crashes – a crash type not targeted by red light cameras” does not adequately address how camera impacts should be assessed. As our analysis showed, cameras have a statistically significant impact on rear-end crashes.

Combined effects of Issues 1 and 2

We suspect that VTRC models may, on balance, be underestimating angle crashes expected in the after period without treatment in the EB analysis, and as a result, the reductions for angle crashes may be underestimated. This is because the models may be underestimating the annual trend factors for the after period trend and may also be underestimating the crash increase that would have occurred with the traffic volume increases that prevailed. Below, we provide an illustration in support of our suspicion. It should be noted that, without a full-blown re-analysis, the precise effects of Issues 1 and 2 are uncertain.

Illustration

Here, unlike the VTRC study, we decided to exclude the before period data at the treated sites from the regression data to avoid using data possibly subject to regression-to-the-mean bias. We then estimated the following model for angle crashes:

$$\text{angle crashes/year} = \alpha \text{Volume}^{0.5266}$$

$$\alpha_1 = 0.0171$$

$$\alpha_2 = 0.0177$$

$$\alpha_3 = 0.0162$$

$$\alpha_4 = 0.0150$$

$$\alpha_5 = 0.0150$$

$$\alpha_6 = 0.0168$$

$$\alpha_7 = 0.0140$$

$$\text{Overdispersion parameter} = 0.3765$$

Using this model we calculated the empirical Bayes expected crash frequency for year 1 and for the after period for two of the sites (Sites 2 and 17) that showed increases in traffic volumes over time. The data for the two sites are in Tables 1 and 2.

Table 1: Site 2 Data

Year	Volume	Angle Crashes
1	54,000	6
2	56,000	8
3	59,000	7
4	61,000	4
5	62,000	5
6	61,000	8
7	61,000	11

Table 2: Site 17 Data

Year	Volume	Angle Crashes
1	63,000	18
2	49,530	11
3	53,069	13
4	52,183	12
5	71,545	19
6	71,583	22
7	71,000	24

Table 3 shows that the EB estimate for the after period from the re-estimated model is roughly 5% higher for both example sites. This difference will vary with the size of the traffic volume differences, the relative lengths of the before and after periods, and on the crash frequencies in the before period.

Table 3: Comparison of EB estimates from VTRC model and re-estimated model

Site	VTRC Estimates	P&L Estimates
------	----------------	---------------

	EB estimate year 1	EB estimate after period	EB estimate year 1	EB estimate after period
2	6.45	16.94	6.18	17.69
4	13.81	48.1	13.80	50.45

Authors' Response to Issues 1 and 2: Reduction in Angle Crashes are not "Underestimated"

The reviewers contend that because of issues 1 and 2, the "reductions for angle crashes may be underestimated" and the reviewers cite Table 3 of their review as an example of such underestimation. It should be noted that the differences in Table 3 of their review are strikingly small. If such differences were observed for all sites in the study, then the conclusions of this work would not change.

In addition, the authors have recalibrated the models as per the reviewers' suggestion to determine the impact on θ as shown below. Given the reviewers comments that

- a different type of CEM should be used (issue 1)
- before data for treatment sites should not be used to calibrate the CEM (issue 2)

we redid the analysis for red light running crashes, rear-end crashes, and angle crashes as per the reviewers' suggestions. That is we

- Used a CEM of the form $\alpha \text{Volume}^{b1} e^{(b2(\text{Speed}) + b3(\text{Yellow}) + b4(\text{Trucks}) + b5(\text{Through lanes}) + b6(\text{Left lanes}))}$
- Excluded data for treatment sites when calibrating the CEM
- Analyzed all sites combined for all jurisdictions

While findings regarding rear-end crashes and angle crashes did not change substantially, Table R1 below suggests that the approaches suggested by the reviewers would have caused red light running crashes to increase slightly. Table R1 shows the results.

Table R1. Impact of Changing the Crash Estimation Model on the Red Light Camera Results

Crash Type	Original VTRC EB Approach	Change the CEM Form as per Issue 1 ^a and remove the treatment sites as per issue 2 ^b
Red light running (RLR)	$\theta = 0.92$ (Table H2 in the report)	$\theta = 1.15$
Rear-end (RE)	$\theta = 1.42$ (Table H1 in the report)	$\theta = 1.48$
Angle	$\theta = 1.20$ (Table H3 in the report)	$\theta = 1.30$

^aContrary to the VTRC EB analysis which used a CEM of the form $= \alpha \text{Volume}^{b1} \text{Speed}^{b2} \text{Yellow}^{b3} \text{Trucks}^{b4} \text{Through lanes}^{b5} \text{Left lanes}^{b6}$ the reviewers suggested a CEM of the form $\alpha \text{Volume}^{b1} e^{(b2(\text{Speed}) + b3(\text{Yellow}) + b4(\text{Trucks}) + b5(\text{Through lanes}) + b6(\text{Left lanes}))}$

^bContrary to the VTRC EB analysis which included before-treated sites in the data set used to calibrate the CEMs, the reviewers suggested removing the before-treated sites from the CEM calibration data set. Thus crashes were calibrated exclusively with data from comparison sites, that is, sites where a red light camera was never installed.

It is evident that the approach suggested by the reviewers would result in red light cameras not performing as well as expected: that is, whereas VTRC had found that they were associated with an 8% reduction in red light running crashes, the method suggested by the reviewers would indicate that cameras are associated with a 15% increase in red light running crashes. Similarly, the red light cameras would, when using the reviewers' methods, lead to cameras increasing rear-end crashes by 48% (rather than 42% according to the VTRC EB estimate) and increasing angle crashes by 30% (rather than 20% according to the VTRC EB estimate).

While our time was limited, we performed two additional experiments just with the red light running crashes from all jurisdictions. We took the original VTRC analysis and then addressed only issue 1: that is, we only changed the CEM form as desired by the reviewers but did not remove the before-treated sites from the CEM data set. The result was that θ *decreased* slightly from 0.92 to 0.90. We then returned to the original VTRC data set and made a different change: address only issue 2 but not issue 1, that is, remove before-treated sites from the CEM data set but do not change the CEM form from that used by VTRC. The result was that θ *increased* to 1.12. We then returned to the angle crashes and did the same type of analysis: the VTRC data set originally had a θ of 1.20; changing the CEM form as desired by reviewers (but including the before-treatment sites in our data set) also kept θ at 1.20; keeping the same VTRC CEM form but removing the before-treatment sites as desired by reviewers raised θ to 1.25). As expected based on the authors' response to review comment 1, performing the reviewers' step of excluding the before-treatment sites from the calibration data set makes red light cameras appear worse than was the case in the VTRC report.

The reviewers noted that the "possible underestimation of the Volume exponent" was also a cause for concern. To determine whether a higher volume exponent would change the conclusions of the report, we performed another analysis with the angle crashes. We returned to the case of angle crashes where both issues sought by the reviewers were addressed and where θ was found to be 1.30 (using the reviewers' form of the CEM and eliminating treatment sites from the CEM calibration data set). We then, contrary to the VTRC report, forced the volume exponent to be at least as big as certain values recommended by the reviewers: 0.6 (their lowest value), 1.0 (a value that presumes crash risk is directly proportional to volume), and 1.2 (the reviewers' highest recommended value). The results were that θ decreased from 1.30 to 1.29 (for the volume exponent as 0.6), 1.26 (for the volume exponent set to 1.0), and 1.25 (for the volume exponent set to 1.2). Thus artificially establishing a high volume exponent can make the red light camera appear slightly better than would be the case with a volume exponent that is based on maximum likelihood estimation, since volumes are increasing (generally) over time. However, overall even performing this step does not cause red light cameras to reduce angle crashes, and in fact, the increase of 25% (based on steps desired by reviewers plus forcing the volume exponent to 1.25) is still larger than the increase estimated by the VTRC EB method (20% in Table H3).

Issue 3: Comparisons between jurisdictions may not be valid

The EB estimates for several jurisdictions are based on small numbers of sites and so comparisons of effects across jurisdictions need to be appropriately guarded. This is compounded by the fact that Issues 2 and 3 are particularly at play when the numbers of

treatment and reference sites are small. That crash prediction models with as many as six variables have been calibrated for use in the EB analysis for as few as six sites appears to be an overextension of the regression analysis procedures. The result is that crash prediction models and EB estimates of safety effects for individual jurisdictions with few sites may not be robust – to a much greater extent than for all sites combined.

Authors’ Response to Issue 3: Interjurisdictional comparisons are indeed “appropriately guarded” and the recommendations in the report are not based on results obtained for individual jurisdictions.

The reviewers write that the “The EB estimates for several jurisdictions are based on small numbers of sites and so comparisons of effects across jurisdictions need to be appropriately guarded.” Unfortunately, it seems that the reviewers missed sections where this was discussed in detail. Listing it as an “issue” means that a reader of only these comments—and not the report—might infer that the report had (i) failed to consider that some jurisdictions have small sample sizes and (ii) drawn conclusions based solely on small data sets. Such an inference is wrong based on explicit and implicit evidence in our report.

- The report explicitly discusses small data sets in several locations. Examples are:
 - Page 32, which states that a small number of sites (6) may give spurious results. Page 32, which discusses the case of only 22 injury red light running crashes in Falls Church.
 - Pages 41 and 42 that give examples why small data sets should be treated with caution.
 - Page 25 which discusses, in excruciating detail, the advantages and disadvantages of reporting statistical findings based on small data sets.
- The report implicitly treats small data sets with caution by *not* drawing conclusions about individual jurisdictions’ programs but rather only drawing conclusions supported by a larger data set. For example, we do *not* try to explain why Arlington County’s program did not have beneficial results compared to other programs. In fact, Conclusion 2 emphasized variation within jurisdictions; page xv in the Executive Summary states that “There is evidence to suggest that this difference might have more to do with variation among intersections than among jurisdictions.”

In short, while we do *report* the jurisdiction by jurisdiction results, we discuss the small data sizes that arose and do not make recommendations based on only a small data set.

Issue 4: *The economic analysis could be more definitive*

The economic analysis was based on naïve before-after crash comparisons, which is somewhat puzzling in the light of the fact that the results of the EB analysis could have been used as was

done by Council et al. (2005)³. This apparent omission is perhaps moot, given the issues with the EB analysis raised earlier. However, compounding this difficulty is that in aggregating crashes per site year and crashes per million vehicle miles over several sites, the sites were weighted equally, even though they varied in lengths of after period. This method of aggregation is inconsistent with the proper method that the VTRC researchers actually applied in the EB analysis, and the difference in results can be non-trivial. The combined effect of these difficulties is that the economic impact of the RLC program may be inaccurate.

Authors' Response to Issue 4: the Economic Analysis is not "Inaccurate"

The reviewers have two separate critiques of the economic analysis: first, why were sites "weighted equally" and second, why were results not based on the EB method?

- Pages 116 and 117 of our report explain that since the after periods and the before periods were of varying lengths, the number of rear-end and angle crashes were adjusted to a per-year basis for consistency. This consistency enables each intersection's impact on the economic analysis to be apparent, as demonstrated by the far right column of Table I3. (Please note that Tables I1 and I2 clarify that for each site, the sum of the before and after years was 7.0. Because a camera was usually installed at some point other than year 3.5, the before period and the after period were unequal. The use of per-year basis enabled a direct comparison at each site even though these periods were unequal.) Sites were also weighted equally because that methodology was consistent with the statistical test associated with the before-after comparison: the paired-sample t-test.
- There are in fact three methods that could have supported an economic analysis: (i) the Before-After approach (Appendices B and C), (ii) the GLM analysis (Appendix G), and (iii) the EB analysis (Appendix H). As noted at the top of page 7 of the report, all three methods have strengths and weaknesses. For this particular data set, the advantage that led the authors to use method (i)—the simple before/after approach—is that this approach is the most transparent of the three approaches. (Completion of the report led to a second advantage of method (i). That advantage is that contrary to methods (ii) and (iii), the angle crashes showed an overall decrease (see Table C12). Thus, for method (i) only it was necessary to compare the impact of the increased rear-end crashes against the decreased angle crashes. Note that the results of methods ii and iii, shown in Tables G3 and H3, showed that angle crashes increased. Thus, had methods (ii) or (iii) been used where rear-end and angle crashes both increased, a "tradeoff" between angle and rear-end crashes is not necessary).

Other issues:

There are few other issues that suggest some inconsistency in the conduct of various aspects of the study that may indicate traits that could be used in assessing the overall quality of the research.

³ Council F., Persaud B., Lyon C. Eccles K. Griffith M., Zaloshnja E., and T. Miller. "Guidance for implementing red light camera programs based on an economic analysis of safety benefits". Transportation Research Record 1922, pp. 38-43, 2005.

Data: Three issues related to data may affect the results. It is possible that the VTRC researchers may have addressed these issues, but we note them here for completeness of our review.

- (i) The inability to account for minor road AADT in much of the analysis is a concern, especially for a treatment like red light cameras which is typically implemented at intersections for which minor road AADT is a key explanatory variable for target crashes.

Authors' Response to Concerns about Not Using Minor AADT

- Unfortunately, it seems that the reviewers missed the sections of the report that addressed this issue in detail. The report (pp. 18-19) discusses the difficult choice the authors had to make when faced with the option of (i) a smaller number of sites with total ADT and (ii) a larger number of sites with major ADT only. When the number of sites is relatively large, the results are similar for both approaches. (For example, Table H2 on page 110 shows that θ was either 0.71 or 0.77 depending on whether 40 sites (with total ADT) or 46 sites (with major ADT) was used.) When the number of sites is small (e.g., 6 or 8), then it is important to choose the option of the larger number of sites, even at the expense of minor ADT, as discussed on pages 40-42 and also on page 32. It should be noted, again, that the recommendations in the report are not based on individual jurisdictions' results.
- (ii) Considerable fluctuations in traffic volumes over time at some intersections raises some questions about the accuracy of some of the traffic volume information.

Authors' Response to Concerns about the Quality of the AADT

Although we agree with this concern about AADT, the research team made exceptional effort to obtain the best data available through substantial emails, phone calls, data verification, follow up efforts, publication of the raw data on the web prior to the analysis being conducted, publication of the draft report on the web, and of course an email to the raw data providers indicating the availability of a draft report based on the data they had provided (see page 5 of the report). These are the best data we could obtain. Certainly data quality is always a concern in this type of research and could be characterized as a national issue.

- (iii) It is unclear whether a driver is more likely to be charged with disobeying a traffic signal with a red light camera in place, and therefore whether there is a bias towards underestimating the effects on RLR crashes.

Authors' Response to Concerns about the Bias on behalf of the Law Enforcement Officer

Certainly that is possible in the sense that anything is possible and we cannot claim to speak for the actions of the investigating officer. Taking into consideration that there was a reduction in red light running crashes, there is no evidence of such a bias.

GLMs: As was case for prediction models calibrated for use in the EB methodology, there are some fundamental concerns:

a) The AADT exponents are often very, very small, and even negative in a few cases (suggesting that increases in traffic volume will result in a decrease in crashes). This is likely due to the use of a large number of variables, many of which are likely correlated.

As noted on page 10 of our report, the objective in developing the GLM models was to obtain a model where “all the parameters are statistically significant... and the model itself has a high Akaike information criterion (AIC) value.” We note that the objective of the GLM was not to determine the impact of volume on crash risk but rather to determine the camera impact on crash risk while properly controlling for other explanatory factors such as left turn lanes. Please see additional discussion on pages 11 and 12 of our report concerning how the AIC was further used to select models.

b) Models are often calibrated with very little data, which is reflected in the quality of the model. Please see authors’ response to comment 3 about small data sets.

Reporting accuracy:

There is some question in regard to the accuracy of the report. For example, the EB methodology seemed to be correctly applied. However, what is reported is incorrect – that the crash prediction model estimate is the expected number of crashes without treatment.

The element of the statement that is “incorrect” is not clear. The Crash Estimation Model (CEM) is formally described as $\hat{E}\{k\}$ (see page 245 of (Hauer, 1997)) which is the expected value of the random variable k , where k is the mean of the number of crashes in our reference population (see page 188 of the same text). [This k should not be confused with the overdispersion parameter k that is used when calibrating the CEMs.] From this CEM, we eventually determine the crashes that would have occurred had cameras not been installed.

(Perhaps the reviewers are asking for the intermediate steps of computing $\hat{E}\{K_{i,y}\}/\hat{E}\{K_{i,1}\}$, the various $m_{i,y}$, and so on that eventually give the crashes that would have occurred as π (see (Hauer, 1997)), but again, the correction sought by the reviewers is not clear.)

Composite aggregation of effects for naïve crash comparisons

As noted earlier, in aggregating crashes per site year and crashes per million vehicle miles over several sites, the sites were weighted equally, even though they varied in lengths of after period. This method of aggregation is inconsistent with the proper method that the VTRC researchers in fact applied in the EB analysis, and the difference in results can be non-trivial.

We do not understand what the reviewers are referring to when they write “crashes per million vehicle miles” as the sites were intersections and thus entering vehicles were used (Tables ES1, ES5, C1, etc.)

Perhaps the reviewers are referring to the issue raised in the first frequently asked question shown on page xix and accompanying Table ES5 on page xx, where we considered two different methods for tabulating the change in crash types in a simple before-after comparison. As shown in the VTRC report, different methods can yield different percentage changes, but these alternatives do not change the conclusions of the report: rear-end crashes increase and red light running crashes decrease.

We also disagree with the reviewers’ assertion that either method is improper: the method favored by the reviewers gives each unit of time equal weight, and the method not favored by the reviewers gives each unit of space (intersection) equal weight. Given that the statistical test associated with the before-after comparison (the paired sample t-test) was consistent with the latter approach, presenting the results in that manner is appropriate.

Negative confidence intervals

The lower confidence limit for the estimates of θ , the index of effectiveness, is often negative, which is theoretically impossible (unless the cameras cause babies to be born!). The difficulty would be avoided if the theory for confidence limits of proportions were to be applied.

When the confidence limits for θ become quite large, the proper approach is to report the limits as indeed quite large rather than choose another method of presentation that artificially shrinks the limits. It is also noted that a negative confidence interval occurred just three times out of the 42 sets of confidence limits for θ presented in Tables H1 through H6.

It is also not clear to us if the reviewers’ suggested approach of using “theory for confidence limits of proportions” is appropriate for our particular situation. To clarify our concern, we have articulated the theory of confidence intervals below and shown why it does not appear to be productive in this particular case.

The theory for confidence intervals for a proportion begins with a sample size (n) and then a proportion of that sample (y/n). For example, n might be the total number of crashes and y might be the number of crashes where an injury was sustained, such that $0 \leq y/n \leq 1$. In such a case, one would write the confidence level for a proportion (y/n) as (Hogg and Ledolter, 1992; Freund and Wilson, 1997)

$$\frac{y}{n} \pm Z_{\alpha/2} \sqrt{\frac{(y/n)(1-y/n)}{n}} \text{ where } Z_{\alpha/2} \text{ is } 1.96 \text{ assuming a } 95\% \text{ confidence interval.}$$

We are not convinced that y/n can be replaced by θ . Certainly, y/n as used in conventional texts is often viewed as a proportion whose maximum value is 1.0, whereas our θ was often greater than 1.0. For example, one of the three instances of a negative confidence limit occurred in Table H2 where θ was 2.09.

Hogg, R.V. and Ledolter, J. *Applied Statistics for Engineers and Physical Scientists*, 2nd Edition, Macmillan Publishing Company, New York, 1992.

Freund, R.J. and Wilson, W.J. *Statistical Methods*, Revised Edition, Academic Press, Inc., San Diego, 1997.

Summary and conclusions

We have found sufficient difficulties with the VTRC study to conclude that a study that resolves these difficulties may likely lead to different conclusions. This finding is based on our assessment that the crash prediction models used in the EB analysis may be underestimating the annual factors for the after period crash experience trend and may also be underestimating the crash increase that would have occurred with the generally significant traffic volume increases that prevailed. The result is that the reductions in angle crashes expected on the basis of published research may be underestimated and, in fact, increases in this crash type have been found in the VTRC study.

That the final conclusions of the VTRC study are guarded and more conservative than the results might suggest supports our belief that the negative results of the study cannot, and should not be cited and used as a deterrent to the implementation of red light camera programs.

We absolutely concur with the VTRC report recommendations that the decision to install a red light camera be made on an intersection-by-intersection basis and that a carefully controlled experiment be conducted to examine further the impact of red light programs on safety and to determine how an increase in rear-end crashes can be avoided at specific intersections. However, we do believe that statistically defensible observational studies, as opposed to experiments, can provide significant and robust insights, as the recent FHWA study⁴ did, and as the VTRC study might have done if the difficulties we raised were mitigated.

Response: A Study that Addresses the Above Concerns would not “lead to different Conclusions”

The crash prediction models used by VTRC were not highly unusual but instead were based on the literature. Further, we have shown that the inclusion of before-treatment site data ensured that the crash prediction models did not underestimate the number of crashes that would have occurred without treatment. As confirmed by the subsequent analysis in Table R1, VTRC’s method did not underestimate the benefits of red light cameras. In fact, taking the reviewers’ approach—analyzing all data as a single set, using the reviewers’ desired CEM, and removing before-treatment data—and then establishing the volume exponent as a favorable 1.2 (rather than estimating the volume exponent from maximum likelihood estimation)—still yielded an angle crash increase of 25%, compared to a VTRC EB estimate of 20%, showing that the model did not underestimate the benefits of red light cameras. Finally, we note that the reviewers

⁴ Council F., Persaud B., Eccles K., Lyon C. and M. Griffith. “Safety Evaluation of Red Light Cameras”. Federal Highway Administration Report FHWA-HRT-05-048; HS-043 838. 2005.

“absolutely concur” with the recommendations of the VTRC report. Those recommendations are being used by the state of Virginia in advising jurisdictions in the selection of treatment sites.

While you are at it, could you clarify the site selection for the spillover and comparison/reference sites. As I understand it the spillover sites are the ones in immediate proximity to the treatment ones and any other site in the jurisdiction could qualify for a comparison/reference site. Are you satisfied that the comparison sites did not experience a spillover effect? Our models for the comparison/reference sites did indicate annual multipliers for angle crashes (p.7 of our draft report) that tended to be lower in the latter years -- a trend that could be consistent with a spillover effect.

Also, could you comment further on the analysis and results for the spillover sites. Was the analysis generally inconclusive and was this why the conclusions are muted in this regard? And was an EB analysis also done for these sites?

There are several questions in this section.

First, as per page 5 of the report, comparison sites were selected based on several criteria: distance from the camera intersection, sites recommended by jurisdictions (which was of value given the jurisdictions’ local knowledge), and the literature. For example, the comparison sites for Fairfax County were identified in a previous report (BMI, 2003).

Within these constraints, we sought to have our comparison sites situated far enough from the camera sites to avoid a spillover effect. We are satisfied that our comparison sites did not experience a spillover effect to the extent that (i) there were comparison sites situated at a reasonable distance from the camera site and (ii) the comparison sites did not experience the significant increase in rear-end crashes nor the significant decrease in red light running crashes that occurred at camera sites (pages 25-26 of the report).