<u>Final Version</u>

Measures of Uncertainty:

Essential Considerations and Useful Tools for Safety Practitioners and Decision-Makers

Clayton Rudy, E.I.T.
Ourston Roundabout Engineering
(A Member of The Sernas Group Inc.)
45 Vogell Road, Suite 306
Richmond Hill, Ontario L4B 3P6
Phone: (905) 508-3371
E-mail: crudy@ourston.ca

Paper prepared for presentation at the
Implementation of Safety Prediction Methods in the Processes and Procedures of a Transportation Agency Session
of the 2012 Annual Conference of the
Transportation Association of Canada
Fredericton, New Brunswick

Abstract

Safety practitioners are benefiting from an increasing volume of research that quantifies collision prediction so that a single and precise collision frequency can often be calculated. However, decision-makers are often interested in determining two additional qualities of a safety analysis: accuracy, and the worst-case (and/or best-case) scenario. Though the discussion of variability often removes the possibility of a single precise answer, the examination of errors can quantify the confidence associated with an analysis and therefore greatly increase the defensibleness of an analysis and the soundness of any decisions based thereon (particularly in economic applications). Furthermore, where error can be quantified, there is an opportunity for the safety professional to justify departing from the mean predicted values based on their expertise and experience, as well as a deep knowledge of the relevant research and underlying data sets.

This paper presents three methods of discussing uncertainty, or variability, in safety analyses. The application of this information in conducting comparison studies of alternative intersection control is highlighted in a case study that demonstrates how the confidence and defensibleness of safety analyses is improved. These methods present a practical continuation to Highway Safety Manual processes and can serve to enhance an agency's safety policies.

Key words: safety, uncertainty, variability, error, bias

Introduction

Much effort has been expended in developing methodologies for calculating very precise collision frequencies. These efforts enable decision makers to achieve sound, objective, and defensible decisions when investing in public infrastructure in a manner that affects public safety. The recent publication of the *Highway Safety Manual* (*HSM*) (1) has achieved a much-sought-after goal of providing a centralized and standardized safety analysis methodology for several types of studies. The *HSM* has balanced considerations of practicality with scientific and statistical rigour in order to generate a tool that is functionally useful in the majority of situations encountered by safety analysts. Achieving this balance is a challenging endeavour.

Over time, some critical weaknesses of simpler methodologies have been addressed such as the effects of regression-to-the-mean (RTM) and of fluctuations in traffic volume. Estimates of expected and predicted safety performance[1] have therefore become much more reliable. Though many publications include standard errors and dispersion parameters, discuss the concept of confidence intervals, underscore objectivity, and recognize the error inherent in models founded upon historical collision frequencies, no methodology was found that guides analysts through a safety study while including the errors associated with each step of the process. This important element of safety analyses – a pervasive treatment of variability – has yet to be widely embraced in practice or emphasized in central documents.

Decision makers are instinctively aware of the variability uncertainty[2] of calculations of safety performance that are precise but without reporting the associated error. As a hypothetical example, one might ask oneself "What is the chance that there will be *exactly* 26.485 fatal+injury collisions over the next 30 years as predicted by the results of this safety analysis?" Collisions are rare and random phenomenon (AASHTO, 1); there must therefore be some variability around the mean result. But without a quantifiable measure of the range of statistically probable results, there is clear ambiguity regarding the degree of confidence one should have in a single mean result, and thus, often, hesitation (or even refusal) to make important decisions based on estimated future safety performance.

Statistical measures of variability have not been entrenched into practice in part because of the difficulty of accurately modeling a system as complex as safety. Hauer aptly stated that "the statistical interpretation of observational studies is messy, involves ambiguity, may require judgement and, in general, does not provide the intellectual pleasures of clear logic, systematic deduction and incontrovertible proof." (3) Although measures of variance, confidence intervals, etc. give the impression of certain knowledge of the variability, this is not so. The three techniques presented in this paper explore some of these "messy" uncertainties surrounding safety analyses:

[1] 'Expected' refers to results where site-specific historical safety performance has been factored into the analysis though the empirical Bayes method, whereas 'predicted' results are exclusively based on the model itself.

[2] Variability uncertainty is defined as "the uncertainty due to inherent variability," such as expected fluctuations in results that may follow a known probability distribution. This is different from another type of uncertainty that is "the uncertainty due to the imperfection of our knowledge, which may be reduced by more research and empirical efforts," or simply a lack of understanding. (Walker, 2)

1) Distribution of residuals, which can improve the reliability and confidence of results;
2) Consideration of multiple models, which can address situations where no single methodology or model accurately reflects the scenario under consideration; and
3) Use of engineering/expert judgement to incorporate factors that are not clearly recognized, measured, understood, and explained by models.

They are also intended to be easily applicable by only requiring the model attributes of mean and variance (or standard error), which are routinely published, as opposed to more advanced techniques that may require more in-depth knowledge of the underlying data sets and higher-level statistical treatments.

Distribution of Residuals

Many sources of variability exist in safety studies. Quantifiable variability – RTM, known trends, observable model bias, crash cost subsets and variants, etc. – should be explicitly accounted for within a methodology[3].

To build a confidence interval about a mean predicted value, the mean is multiplied by the Multiple of Standard Error (MSE) of the desired confidence interval. Table 3-3 of the *HSM* (3) assigns low, medium, and high levels of confidence as follows:

- Low level of confidence: confidence interval of 65%-70%, MSE = 1
- Medium level of confidence: confidence interval of 95%, MSE = 2
- High level of confidence: confidence interval of 99.9%, MSE = 3

Although a "high" level of confidence is perceptively the ideal, the 95% confidence interval has been accepted in practice as providing a sufficient measure of variability for project-level safety analyses[4]. Moreover, it still communicates an appreciation for the error inherent to the process, and provides a more realistic basis for a worst-case scenario analysis. Another MSE of interest is ±1.5, corresponding to thresholds for LOSS (Level of Service of Safety) analysis, a process proposed for quantifying collision reduction potential (Kononov, 4).

But at which stage of the process should the MSE be applied? There are quantified and published measures of error when applying SPFs (safety performance functions), CMFs (crash modification factors), calibration factors, when applying the EB (empirical Bayes) method, when calculating safety indices, and at other stages of the process. There could be situations where there is a very large standard error with a particular CMF and a well-developed SPF with a small error. If the MSE was applied to the SPF, the large error associated with the CMF would be ignored and thus the confidence interval would be deceptively small. Conversely, if the MSE was applied to the CMF, the error of the SPF would be ignored and the confidence interval

[3] Statistically unquantifiable variability – traffic volumes (how often are AADT values based on a single count day?), effective versus actual roadway characteristics, observed trends with insufficient evidence to quantify the effect, human factors, etc. – should be treated with sensitivity analyses or engineering judgement.

[4] A transportation agency may in fact want to set the MSE based on how sensitive a weight they wish to attribute to the safety performance component of public infrastructure projects at an overall policy level and/or at the project level.

would be deceptively large. The proposed solution is to distribute the MSE equally among the variance of each step of the process, and equally to the variances throughout each subdivision of each step such that the probability of the product of the probabilities of each step equals the desired confidence interval (assuming all events are statistically independent). For example, the product of the three 37% and three 63% confidence intervals yield the 5% and 95% confidence levels, respectively. A list of MSEs by number of steps (or degrees) for confidence intervals and total MSEs of interest can be found in **Table 1**. An example of a 3-degree distribution of residuals is illustrated in **Figure 1.** The process is written detailed below for one common scenario: calculating the 5% and 95% confidence intervals of the expected (EB) collision frequency at a future time.

1. Determine the Multiple of Standard Error (MSE) for the three-degree 95% confidence interval: 0.899. The first step of the distribution of residuals can be referred to as the $(0.05)^{1/3} \cong 37\%$ and $\left(1-(0.05)^{1/3}\right) \cong 63\%$ for lower and upper bounds, respectively. Similarly, the second step approximately equal the 14% and 86% confidence intervals.
2. Calculate the mean (maximum likelihood, denoted as *ML*) predicted collision frequency, N_p at time t_1 at maximum likelihood, $N_{p,t_1,ML}$.
3. Calculate the first degree upper and lower confidence intervals for the predicted model results: $N_{p,t_1,63\%} = N_{p,t_1,ML} + 0.899\sigma_{N_p,t_1,ML}$, and $N_{p,t_1,37\%} = N_{p,t_1,ML} - 0.899\sigma_{N_p,t_1,ML}$
4. Calculate the maximum likelihood of the expected (empirical Bayes) collision frequency at time t_1 using the first degree confidence intervals:
$N_{e,t_1,ML} = (w)\left(N_{p,t_1,63\%}\right) + (1-w)(N_o)$, where $w = \frac{1}{1+(k)\left(\sum N_{p,t_1,63\%}\right)}$ and N_o is the observed crash frequency. The lower confidence interval is similarly calculated, replacing $N_{p,t_1,63\%}$ by $N_{p,t_1,37\%}$.
5. Calculate the upper and lower confidence intervals of the expected collision frequency at t_1, $N_{e,t_1,86\%}$ and $N_{e,t_1,14\%}$, by adding $0.899\sigma_{N_{e,t_1}}$ for the upper limit and by subtracting $0.899\sigma_{N_{e,t_1}}$ for the lower limit (where $\sigma_{N_{e,t_1}} = \sqrt{\left(N_{e,ML}\right)(1-w)}$).
6. Calculate the upper and lower Indices of Effectiveness of the expected collision frequency at t_1, $\theta_{86\%CL}$ and $\theta_{14\%CL}$, by computing the difference between the maximum likelihood of the predicted collision frequency at t_1, $N_{p,t_1,ML}$, and the upper and lower confidence intervals of the expected collision frequency at t_1, $N_{e,t_1,86\%CL}$ and $N_{e,t_1,14\%CL}$, then multiply each by the ratio of $\frac{N_{p,t_2,ML}}{N_{p,t_1,ML}}$.
7. Calculate the upper and lower confidence intervals of the predicted collision frequency at t_2, $N_{p,t_2,ML} \pm 0.899\sigma_{N_{p,t_2}}$.
8. Calculate the 5% and 95% confidence intervals of the expected collision frequency at t_2, $N_{e,t_2,5\%CL}$ and $N_{e,t_2,95\%CL}$, by adding $\theta_{86\%CL}$ for the upper limit and by subtracting $\theta_{14\%CL}$ for the lower limit.

In situations where there are two different traffic volume scenarios and two different predicted mean values of collision frequency, as shown in Figure 1, an additional step is needed (Step 6, above). The Index of Effectiveness, *θ*, is defined as the absolute difference between the upper or lower confidence limit and the mean predicted value (AASHTO, 1). This value is scaled from t_1 to t_2 by the ratio of mean collision frequency of t_2 relative to t_1. This step can be treated without consideration of variability because there is no calculable variance of the above differences and ratios.

There are several assumptions involved in this procedure. Firstly, it assumes that all standard errors are known with complete certainty. The variance of most SPFs is calculated as a function of a constant dispersion parameter, though this constant also has an associated variability at least as a point estimate, if not as a function of a model's covariates, as noted by Miaou and Lord (5). However, accounting for the variance of the variance is beyond the intent of this methodology.

Secondly, for subdivisions of this process, such as multiplying CMFs, it assumes that the terms or factors are independent of each other such that the safety effect of any one is independent of the effect of any other – i.e. there is zero covariance or heteroscedasticity. For example, the installation of a red-light camera and at the same time prohibiting right turns on red could easily have an effect different than the product of their CMFs. Although there is general agreement that CMFs are often not independent, research into covariances of CMFs and other analysis elements is very limited (Gross, 6).

It is helpful to mention some commonly used standard error equations. The standard error of the EB estimate of expected frequency (Hauer, 7) is:

$$\sigma_{N_e} = \sqrt{(1-w)N_e}$$

The standard error of SPF calculations with a single dispersion parameter (Gross, 6) is given by:

$$\sigma_{N_p} = \sqrt{kN_p^2}$$

Despite the foregoing discussion on CMF independence, the standard error of a product of terms is also useful; from Hunter and Schmidt's equation for the variance of a triple product of independent variables (8),

$$\sigma_{abc}^2 = (\bar{a}^2 + \sigma_a^2)(\bar{b}^2 + \sigma_b^2)(\bar{c}^2 + \sigma_c^2) - (\bar{a}^2)(\bar{b}^2)(\bar{c}^2)$$

the general case for the standard error of a product of *m* factors can be extrapolated to be:

$$\sigma_{\prod_{n=1}^{m}} = \sqrt{\prod_{n=1}^{m}(\mu_n^2 + \sigma_n^2) - \prod_{n=1}^{m}(\mu_n^2)}$$

Advantages of this procedure include the ability to account for a variable dispersion parameter and the ability to avoid a bias towards one particular source of error. Disadvantages include the assumption of normal distribution of residuals and the additional computation effort.

Consideration of Multiple Models

Situations exist where there are multiple models with varying strengths and weaknesses, and no single model is clearly superior to others. In this case, choosing to apply one single model, whether due to policy, personal choice, or desire for simplicity, is contrary to the scientific method that considers all valid alternatives. If a safety analyses reports variability, as discussed earlier, one can use the measured degree of variability – i.e. variance – as a measure of relative confidence in the models. It could be appropriate in certain cases that models (including CMFs) that do not have a published measure of error can be discarded from consideration since there is no method of determining the accuracy of these models.

For example, a high-quality calibrated SPF may have been developed for a large jurisdiction that confidently wrote it into policy as the single model for all its safety analyses. One site of concern is a stretch of road in a mountainous area that represents a very small portion of the jurisdiction and that provided little data for the SPF calibration effort. Another jurisdiction had developed a very high-quality SPF specifically for mountainous terrain, though it is obviously not local. Which model is appropriate to use – the local model, or the model that considers the unique terrain? Or, what if the local SPF was able to be calibrated, but its CURE plot (plot of cumulative residuals) shows a considerable bias over the exact range of covariates that are of interest to the analysis, and the non-local model shows very little bias?

This comparison of multiple models is similar to a formal meta-analysis of independent safety studies, where a weighted average can be calculated by using the inverse variance as the weight factor while considering any heteroscedasticity (covariance), such as the methodologies discussed by Persaud (9) and de Blaeij et al. (10). However, the magnitude of covariance would be impractically (or impossibly) difficult to quantify, and thus proper qualifiers should highlight this methodological weakness. Furthermore, the calculations are unlikely to be independent and should not be referred to as meta-analyses to avoid the impression of such a thorough study. Nevertheless, for the practical purpose of restricting the proposed methodology to use of only the widely-available model characteristics of variance and mean, it is assumed that covariates and models are independent, and can therefore be combined with a relatively simple procedure.

The inverse variance is most often used as a weighting factor, though the inverse coefficient of variation may also be appropriate in some situations. Equations for weighted mean and weighted variance are given by Finch (11):

$$\mu = \frac{\sum_{i=1}^{n}(w_i x_i)}{\sum_{i=1}^{n}(w_i)}$$

$$\sigma^2 = \frac{1}{\sum_{i=1}^{n}(w_i)} \sum_{i=1}^{n} [w_i (x_i - \mu_i)^2]$$

Engineering Judgement

Regardless of statistical and methodological rigour, there is still justification to deviate from mean predicted values based on situations such as:

- detailed knowledge of unique characteristics of the site, their effects on safety, and the degree to which they may be reflected in a model (if at all);
- in-depth comprehension of the characteristics of the models and analysis methodologies;
- discarding outliers;
- choosing the appropriate confidence intervals at which to report results;
- assigning different weight factors of a weighted average calculation;
- to increase confidence in situations where there is a significantly demonstrated public and/or political concern;
- worst-case and best-case scenario analysis;
- previous experience with safety studies; and
- to account for known effects that are not explicitly quantifiable.

Strigini (12) states that “In many cases of stringent safety requirements, this form of engineering (or ‘expert’) judgement, i.e., “informal inference from complex evidence,” is the crucial resource for the decision maker, for lack of more solid, objective evidence.” Recognizing the importance of engineering judgement, it details many of the pitfalls associated with these sometimes ambiguous, quasi-rational conclusions.

However, where the variability uncertainty has been quantified, this allows a safety practitioner to apply their experience and judgement to deviate by a non-arbitrary value based on a desired confidence level. This position is advocated in the *HSM* (1), in the *Guidelines for the Screening of Collision-Prone Locations* (Bahar, 13), and by Hauer (3), though it is rare to find published methods for explicitly quantifying variability. It allows an educated choice of an appropriate point estimate when a single figure is to be reported without a discussion of its variability. One practical application example of engineering judgment in this situation is to adjust a point estimate based on patterned bias observed in the CURE (cumulative residual) plot of a safety performance function. See Kononov (14) for a discussion of bias and CURE plots.

As a second, hypothetical example, picture an urban intersection under stop control that is being converted to signal control. A locally-calibrated intersection-level SPF was developed that accounts for traffic volume, setting, and traffic control. The mean predicted collision frequency was calculated with its associated standard error for a future time period, however consideration of historical observed collision frequencies (i.e. applying the empirical Bayes method) is not applicable due to the change of intersection control. The site in question has a relatively large skew angle, heavy left turns, and a much higher than average number of left-turn collisions. Recent research has indicated that large intersection skew angles have a negative safety effect in a rural context, but no model has explicitly quantified this effect in an urban setting. Therefore, because a single point estimate is required for a subsequent analysis, an analyst could suggest that the reported collision frequency be at the 68% confidence interval (the mean predicted value plus one MSE) instead of simply the mean value of the model.

A further application of engineering judgement that is likely far underused in practice is the selection of target collisions when evaluating past collision frequency or applying these observed collisions in the empirical Bayes method. Target collisions are those that can be affected by roadway safety treatments (Izadpanah, 15); they do not include collisions due to vehicle malfunctions such as an engine fire or a blown tire where pavement conditions are good. Including these in an empirical Bayes calculation in a situation with very low predicted collision frequency may significantly distort results. Non-target collisions are unrelated to the safety performance of a facility and should be removed from any subsequent quantitative analysis.

Two groups of causal factors affect safety performance: those that are recognized, measured, understood, *and* can be explained by models; and those that are not recognized, not measured, *or* not understood (15). Applying the full extent of meticulously detailed analysis made possible through decades of research, while only considering one of these two groups of factors, leaves room for much error and bias. Engineering judgement has a valid role in accounting for this second group of elusive factors.

Case Study: Edmonton Traffic Circles

In 2009, the City of Edmonton commissioned a rehabilitation study (16) of four traffic circles that were originally built in the 1950s. These four traffic circles were among the highest collision locations in the City, and their designs differed significantly from contemporary roundabout designs. The sites were investigated for conversion to modern roundabouts as well as signalized intersections, with future safety performance predictions of the two intersection control alternatives being a key factor in the final decision.

At one site, the existing observed collision frequency was approximately 1.5 collisions per week. One alternative was to convert the 2-lane roundabout into a partial 3-lane roundabout – i.e. a ‘3x2’ with two 2-lane entries and two flared 3-lane entries. The original safety analysis applied SPFs for F+I (fatal+injury) and PDO (property-damage-only) collisions for 3-lane roundabouts as derived in *NCHRP Report 572* (Rodegerdts, 17), without correcting for RTM (regression-to-the-mean) or local calibration (due to insufficient data). It predicted a relatively high collision frequency of approximately two collisions per week at the 95% confidence interval, shown in **Table 2**. Results presented a (relatively low) potential for increased collisions. The client did not want to make the significant investment in a modern roundabout if there was a chance, albeit small, of an increase in collision frequency at an already high-collision location.

Table 2
Annual Collision Frequencies

		Collisions/year	
	CI (%)	**F+I**	**PDO**
Existing (2004-08 average)	n/a	19.3	61.8
Original Analysis	95	3.1	109.0
Method 1	95	24.0	102.1
Method 2	95	23.6	76.4
Method 3	95	1.4	11.3
Method 4	95	2.0	12.3
Weighted Results of Methods 1-4	95	**2.6**	**19.2**
	99.7	**3.3**	**24.7**

There were a number of doubts regarding the SPFs: the data used to derive the SPF was taken from only 3 sites in the U.S., none of which met modern roundabout design guidelines, and all exhibited very high collision frequencies; the models had very high dispersion parameters; other large roundabouts built recently did not experience the magnitude of collisions predicted by the models; and a number of other circular intersections operated elsewhere in the City, such that drivers were better conditioned to this category of intersection compared to most other cities. Despite these concerns, no formal methodology existed to suggest a feasible alternative opinion. The safety team attempted to apply engineering judgement and proposed that the 'worst-case' results were unrealistically high, though, understandably, the City did not want to proceed with reconstruction to a modern roundabout if the existing collision problem could not be resolved more convincingly.

In revisiting the analysis, four collision prediction models were applied:

1) The 3-lane roundabout SPFs were applied, as in the original analysis, though corrected for RTM using the EB (empirical Bayes) method. The distribution of residuals method was then applied, as discussed previously. The 95% confidence interval was similar to that of the original study due to the high number of observed collisions and the high dispersion parameters of the models.
2) Given that the reconstructed roundabout would only have two 2-lane entries, the SPFs for 2-lane entries were also applied from the same source, in the same manner, and with similar results for the same reasons as with the 3-lane models.
3) The safety analysis for the signalized intersection alternatives benefited from locally calibrated SPFs specific to an urban setting. The resultant predicted collision frequencies were within the expected range for similar intersections. These were factored by CMFs from *NCHRP Report 572* (17) for conversion from a signalized intersection to a roundabout, and the distribution of residuals method applied. Due to the much lower variance of both the signalized predictions and the CMFs, and to the

robustness of the signalized intersection SPFs, results predicted far fewer collisions than the first two methods.

4) SPFs had also been calibrated for a suburban setting as well as an urban setting; the setting of the site in question could be argued to fall equally well into either category. Thus, the same steps as Method 3 were carried out, and showed similar results.

With four safety analyses completed showing four different outcomes, all models were combined by inverse-standard-error-weighted averaging of the individual results, as presented previously. The overall weighted results were far closer to observed values at other large multilane roundabouts in North America. And although a 95% confidence interval is widely used in practice, the 99.7% confidence interval was highlighted for added confidence. The results of this analysis now indicate a collision frequency of approximately two per month, compared to the 2-3 collisions per week predicted in the original study. Considering that the impetus of the City's examination of this intersection was to improve safety at a high-collision location, the conclusions of this analysis show that collision frequency and severity would decrease with reconstruction to a modern roundabout.

This case study demonstrates that the application of the distribution of residuals method and the consideration of multiple models can serve to enhance safety analyses by quantitatively addressing competing concerns of multiple sources of information. Furthermore, it highlights the substantial effects of variability within and between models, which can obviously have a significant impact to economic analyses of projects when collision costs are included.

Conclusion

This study explored three techniques for decreasing the uncertainty of safety analyses: distribution of residuals, consideration of multiple models, and the use of engineering judgement. A case study highlights the potentially considerable effects of these techniques in practice. They are also intended to be easily applicable by only requiring the model attributes of mean and variance (or standard error), which are routinely published, as opposed to more in-depth knowledge of the underlying data sets. Therefore, these relatively simple tools are easily applied by practitioners in routine analyses. The examination of variability can quantify the confidence associated with an analysis and therefore greatly increase the defensibleness of an analysis and the soundness of any decisions based thereon (particularly in economic applications). Finally, these techniques add a more pervasive treatment of variability throughout a study, thereby increasing the rigour and reliability of safety analysis methodology.

References

1. AASHTO (American Association of State Highway and Transportation Officials). *Highway Safety Manual*. Washington, DC, 2010. pp. 3-5, 3-7
2. Walker, W.E., et al. *Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support.* Integrated Assessment, 2003, Vol. 4, No. 1, pp. 5-17. Taylor & Francis, Great Britain. pp. 13
3. Hauer, E. *Observational Before-After Studies in Road Safety*. Emerald Group Publishing Limited, Howard House, Bingley, UK, 1997. pp. 272
4. Kononov, J., & B. Allery. *Level of Service of Safety: Conceptual Blueprint and Analytical Framework*. Paper No. 03-2112, Transportation Research Record: Journal of the Transportation Research Board, No. 1840, Transportation Research Board, National Research Council. National Academy Press, Washington DC, 2003. pp. 57-66.
5. Miaou, S.-P., & D. Lord. *Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes*. Transportation Research Record: Journal of the Transportation Research Board, No. 1840, Transportation Research Board, National Research Council. National Academy Press, Washington DC, 2003. pp. 31-40.
6. Gross, F., B. Persaud, & C. Lyon. *A Guide to Developing Quality Crash Modification Factors*. U.S. Department of Transportation, Federal Highways Administration (FHWA). Washington, DC, 2010. p. 10
7. Hauer, E., D. W. Harwood, F. M. Council, & M. S. Griffith. *Estimating Safety by the Empirical Bayes Method: A Tutorial*. Transportation Research Record: Journal of the Transportation Research Board, No. 1784, Transportation Research Board, National Research Council. National Academy Press, Washington, D.C., 2002. pp. 126–131
8. Hunter, J. E., & F.L. Schmidt. *Methods of meta-analysis: Correcting for error and bias in research findings*. 2nd Ed., Sage Publications, Thousand Oaks, CA, 2004. pp. 148-149
9. Persaud, B. *NCHRP Synthesis 295: Statistical Methods in Highway Safety Analysis*. Transportation Research Board of the National Academies. National Academy Press, Washington DC, 2001. pp. 73-74.
10. Blaeij, A., et al. *The value of statistical life in road safety: a meta-analysis*. Vol. 35, Accident Analysis & Prevention. pp. 973-86.
11. Finch, T. *Incremental calculation of weighted mean and variance*. University of Cambridge Computing Service. 2009. pp. 4-5
12. Strigini, L. *Engineering judgement reliability and safety and its limits: what can we learn from research in psychology*. SHIP project technical report T030. 1994. p. 1
13. Bahar, G., B. Persaud, & C. Lyon. *Guidelines for the Network Screening of Collision-Prone Locations*. Ottawa, ON, 2011.
14. Kononov, J., C. Lyon, & B. Allery. *Relation of Flow, Speed, and Density of Urban Freeways to Functional Form of a Safety Performance Function*. Transportation Research Record: Journal of the Transportation Research Board, No. 2236, Transportation Research Board of the National Academies, Washington, DC. 2011. pp. 11–19

15. Izadpanah, P., A. Hadayeghi, & H. Zarei. *Before-and-After Study Technical Brief*. Transportation Safety Council, Institute of Transportation Engineers, Washington, DC. 2009. p. 4
16. Ourston Roundabout Engineering (Canada), & Al-Terra Engineering. *Traffic Circle Rehabilitation Study*. Prepared for the City of Edmonton. 2009.
17. Rodegerdts, L., et al. *NCHRP Report 572: Roundabouts in the United States*. Transportation Research Board, National Research Council. National Academy Press, Washington DC, 2007. p. 28

Figure 1
Distribution of Residuals for Present and Future Scenarios

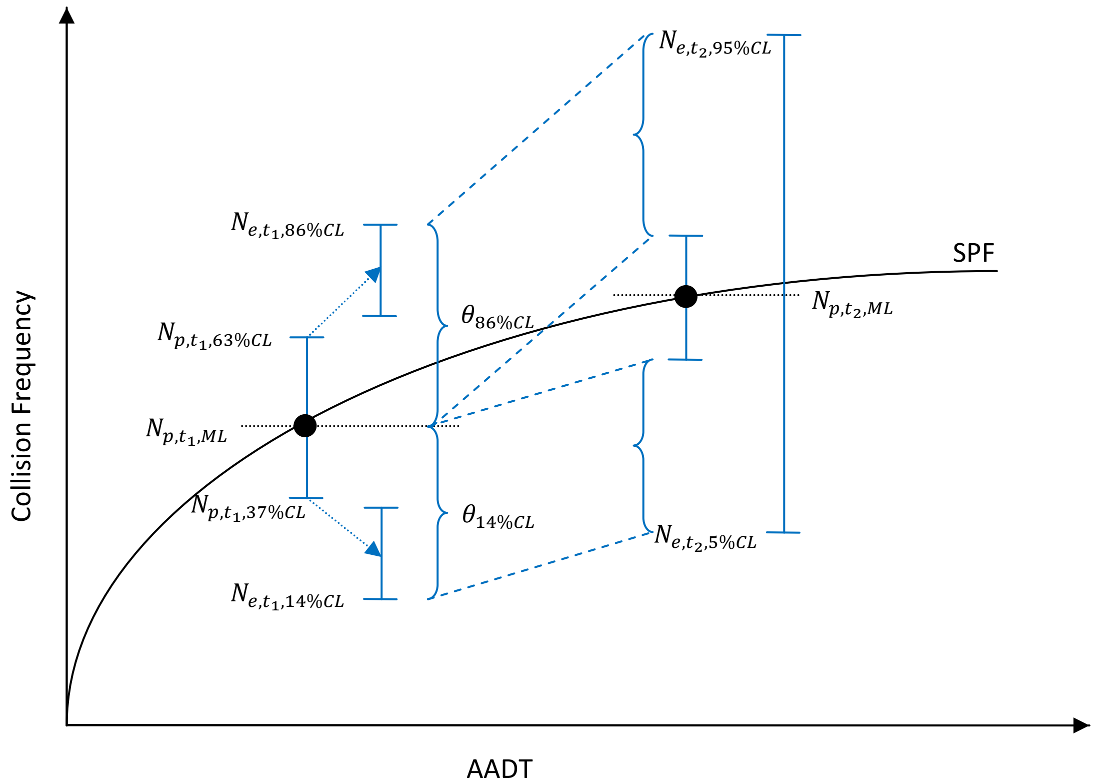

Table 1

Multiples of Standard Error by Number of Steps

Confidence Interval	Total MSE	Degrees of Error	MSE Per Degree
95.0%	**1.960**	1	1.960
		2	1.217
		3	0.899
		4	0.718
		5	0.598
86.6%	**1.500**	1	1.500
		2	0.904
		3	0.656
		4	0.517
		5	0.427
68.3%	**1.000**	1	1.000
		2	0.578
		3	0.410
		4	0.318
		5	0.260
50.0%	**0.674**	1	0.674
		2	0.375
		3	0.261
		4	0.200
		5	0.163