



**Peer Review of Report:
“Relative Comparison of NCHRP 350 Accepted Guardrail Terminals”**

On October 28, 2014, The Safety Institute released a final report entitled “Relative Comparison of NCHRP 350 Accepted Guardrail Terminals.”¹ The report, authored by Kevin Schrum, examines severe injury and death data involving five different guardrail end terminal designs.

After release of the report, FHWA engaged the John A. Volpe National Transportation Systems Center to manage an independent peer review of the report's methodology and findings. To conduct the peer review, Volpe commissioned individuals with expertise in statistical analysis methods and safety data analysis. Per FHWA's specifications, Volpe ensured that the selected individuals did not have any conflicts of interest, including: no vested interest in the guardrail terminals compared in the report; no affiliation with manufacturers or vendors of guardrail terminals; and no affiliation with any organization involved with the development and testing of the terminals evaluated, or with the study documented in the report.

Volpe commissioned four individuals to conduct the review:

- Ezra Hauer, Professor Emeritus of Civil Engineering, University of Toronto
- Paul Jovanis, Professor of Civil Engineering, The Pennsylvania State University
- Francesca La Torre, Professor of Civil Engineering, University of Florence, Italy
- Bhagwant Persaud, Professor of Civil Engineering, Ryerson University, Canada

All four reviewers raised concerns about limitations or flaws in the study's methodology, which led all the reviewers to question the validity of the study's findings and conclusions. The reviewers cited reasons to doubt assumptions made in the report, expressed concern about potential biases in the approach to selecting the crashes included or excluded from the analysis, and questioned the decision to consider only fatal and severe injury crashes (and not also less severe injury and property damage only crashes). All four reviewers observed that the findings and conclusions are either questionable or invalid. Their comments included:

‘The conclusions are basically invalid because the method is fundamentally flawed.’

‘The findings are not sound and defensible because the methodology is not correct...’

‘Therefore the report provides no basis for forming a judgment about the relative safety of various ET types.’

¹ See: <http://www.thesafetyinstitute.org/final-report-relative-comparison-of-nchrp-350-accepted-guardrail-terminals-released/>

'The validity of the conclusions is questionable because the analytical method used to arrive at these conclusions is inappropriate...'

We have attached each reviewer's complete comments. The reviews are randomly numbered to protect the independence of the reviews.

Template: Review of “Relative comparison of NCHRP350 Accepted Guardrail Terminals” v1

Name:	Reviewer 1
Date:	December 15, 2014

Provide a narrative that addresses the two main review areas. You may use the detailed questions as a guide to your response. Use the boxes below (they will expand) for your response.

1. Provide an objective assessment of the study methodology:

Are the data, statistical methods, and assumptions reasonable and appropriate?

SUMMARY: The report describes an analysis process (including data, methods and assumptions) that has fundamental flaws. Among the most important are:

- a. *Removal of crashes based on the date of visual evidence (e.g. pictures, videologs or from current web sites) may systematically remove crashes from the oldest guardrail system (ET-2000). As a result of this crash removal, the ET-2000 may erroneously appear to be safer than the comparison guardrail systems.*
- b. *The analysis does not consider property damage only or minor injury crashes. This greatly reduces sample size and eliminates from consideration outcomes that may indicate the greatest effectiveness of the systems; i.e. circumstances in which an end treatment may have resulted in property damage only or minor injury outcomes rather than serious injuries or fatalities.*
- c. *By considering only sites that experience crashes, the author systematically ignores those locations where the devices may be the most effective: where no reportable crashes have occurred. While the method includes a count of guardrail devices 10 miles upstream of each crash site, this not a replacement for a more systematic inclusion of locations which have the guardrails and experience no crashes.*
- d. *The method uses accepted safety terminology in confusing ways. This includes the uses of the terms *exposure* and *probability* in confusing or incorrect ways.*
- e. *The analysis ignores a substantial literature on the estimation of the effectiveness of road safety countermeasures. Guardrails and their end treatments should certainly be considered as safety countermeasures; failure to even mention this substantial body of research (e.g. the Highway Safety Manual) is a major omission.*

These and additional issues are discussed in detail in the responses to specific the questions that follow.

What, if any, issues impact the reasonableness and appropriateness of the methodology? The issues are discussed in response to the questions below.

Methodology Assessment:

- Is the research approach appropriate for meeting the study objective: “to evaluate the safety performance of NCHRP Report No. 350-approved guardrail terminals?”

No, the study objective is not met because of flaws in the safety analysis methodology. The flaws are described in the responses to specific questions below.

The author’s approach includes identifying guardrail end treatment crashes through a series of successive filters applied to crash data. The locations of the resulting crashes are then used to initiate a search (by type of end treatment) upstream of each site; the search counts the number of end treatments within the 10-mile section. The frequency of occurrence of the end treatments is referred to as the “exposure” for the treatment. The frequency of crashes with each treatment (i.e. end type) is then divided by the number of times the treatment appears in the “exposure” producing what the author calls a “probability” of crash occurrence. The ratio of the “probabilities” gives what is called the “odds ratio”.

This use of exposure and probability seems incorrect. Exposure is a measure of the opportunity for a crash, usually characterized by some measure of traffic (e.g. Annual Average Daily Traffic, AADT, or perhaps Vehicle Miles of Travel (VMT). Traffic level appears to be available for the crash sites, but the author does not use this fundamental measure, preferring the occurrence of each guardrail end treatment within 10 miles upstream of each crash site. There is no explanation as to how this measure was derived and no explanation of what happens if there are no additional guardrails upstream. The basic relationship of crashes to exposure seems ad-hoc, and not consistent with the state of practice represented by the Highway Safety Manual (HSM; AASHTO, 2010).

A final comment on this approach: it seemingly allows an observation to enter the analysis more than once. Imagine a string of 5 identical end treatments every 2 miles upstream. If crashes occur at all these sites, they enter as crashes and exposure. This is not identified or rationalized. Such multiple entries of a site likely violate the assumption of independence underlying the Fischer exact test for 2X2 tables.

The author ignores property damage only and minor injury crashes and does not discuss the implications of this exclusion. Reliance on scene photos for crash inclusion introduces a bias toward inclusion of more severe events, because scene photos are more likely to be taken if there are injuries and fatalities. The occurrence of crashes with minor and no injuries is, in fact, an important outcome.

The logistic regression is included in the discussion although there are no findings. It is not clear how the logistic regression model was even formulated (i.e. what is the unit of observation and how the crash and exposure data were entered in the model).

The report provides the range of dates used in the crash and exposure analysis. There is virtually no *explicit* recognition of time (e.g. date of crashes) in the analysis. This fails to recognize time-related trends in crashes, especially through changes in traffic levels and more global trends in crash frequency (such as economic downturn). Failure to recognize

these trends may affect both crashes and exposure. These time-related trends are readily addressed in alternative analysis methods.

- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and useful crash data?

No; there are basic and fundamental problems with the data identification, collection, compilation, filtering, and reduction methods for crash data in this report.

There is a potential bias due to the way that older crashes are used in the analysis. The ET-2000 is the oldest of the guardrails being evaluated. As a result its crashes are likely the ones occurring farthest back in time, making them more likely to have occurred prior to the most recent videologs and other visual evidence. Using the author's approach, these older crash involvements would be removed from the analysis. As a result, there would be an underreporting of crash events in Table 11 for the ET 2000; this bias would tend to make the other 4 comparison devices look relatively less safe. This bias is a major threat to the validity of the study method, as it would appear to represent a systematic bias that results in the underreporting of crashes with the ET 2000. If the manipulation of crash records were more carefully tracked and reported in the paper, one would have a better understanding of the magnitude of this potential bias. Information is provided on pages 12 and 15 of the reduction in sample size from original crashes identified to those finally used in the analysis, but there is no detail as to *why* each location was dropped. Thus, we are unable to understand the implications of the assumptions in the method, weakening any attempt to verify the robustness of the findings.

The reviewer does not understand why all crashes other than single vehicle road departures were removed from the analysis. The guardrails protect any driver or passenger who impacts the end treatment, regardless of whether the crash involved multiple vehicles or only a single vehicle. The crash data should include any event that results in an end-barrier impact. There seems no reason to focus on single-vehicle-run-off-road events.

- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and reasonable exposure data?

No; there are fundamental problems with the exposure data.

Exposure is usually defined as reflecting the opportunities for a crash. The study method applies an arbitrary 10-mile upstream data collection distance. It is conceivable, even likely that there are road sections with these devices that have no crashes reported that meet the study criteria. These "safer" sections would be excluded from the exposure measures and thus not included in the device evaluation.

Data on traffic levels at each crash site appear to be available in the data (see page 8). Given the acceptance of traffic data in crash frequency modeling, it is surprising that the author did not use these data in the analysis.

What happens if there are no other guardrails of the types of interest within ten miles upstream of the crash? Does that mean there is zero exposure? Discussion in section 4.2.3 hints at dropping these crash locations from the analysis, but this is not completely clear. Additionally, how does one deal with study crash sites that overlap within the 10-mile

upstream distance? Are crash sites within 10 miles included as exposure for other sites?

- Are the analytic methods and statistical techniques suitable for the types of data given the study objective?

No; statistical analyses are unclear or not fully explained.

On page 15 and elsewhere the author uses the term “probability of guardrail being involved in a crash” when referring to the ratio of crashes to frequency of end treatments 10 miles upstream. The measure used by the author is a ratio, not a probability. The probability of a crash is much smaller as it is normally thought of as per vehicle passing the site, not per end treatment in the area. Probabilities are associated with processes involving uncertainty, when one conceptually repeats the experiment under identical conditions. This may be conceptualized as a trial in terms of repeated passages of vehicles past different end treatment. It is hard to think of the ratios computed in this paper as probabilities so defined.

It is unclear why the logistic regression is even mentioned, since no results are shown and the reader is provided with no results. It is also unclear exactly how the logistic regression was formulated. This method is generally applied when there are discrete outcomes. Crash severity appears to be serving as those outcomes, but it is unclear how the exposure data were entered in the model. Exposure data would have a no injury outcome, which is not described in the paper. One is left to conclude that the logistic regression used each crash event as an observation; the type of barrier was the only predictor, and the outcome was either a fatality or injury. Data from the ET-2000 formed the baseline against which crash odds are estimated for the other end treatments. If properly formulated, the logistic regression would produce odds ratios that could be compared.

The use of 2X2 contingency tables as the Fisher exact test seems correct, but with some qualifications. Data must only enter the table once, and the assumption for entry is that events are statistically independent (verify). It is unclear how the author dealt with multiple crashes occurring at the same location or locations entering as both crash and exposure data. Would this represent one or multiple observations? Fisher exact test not conducted on the odds but the frequencies; there are several places where the author seems to confuse these issues.

- Are there other methods that could have been considered and if so, why?

Any of several HSM-related approaches could be tried in this context. The data would have to be carefully selected and safety performance functions carefully constructed. The advantage is using a method that is well tested, correctly uses the concept of exposure and more fully uses all crash data. Safety performance is usually defined as expected number of crashes by type and severity. The NCHRP Report No. 350-approved guardrail terminals may be considered a safety countermeasure or treatment. There are procedures developed in the recently released Highway Safety Manual (AASHTO, 2010) that describe how such evaluations are to be conducted. The procedures are not mandatory, but represent a consensus of researchers and practitioners of how to conduct such evaluations. There are also numerous additional research papers that describe enhancements to the HSM methods that may be applicable to this end treatment analysis.

A cross-sectional comparison is conducted here comparing across sites. A better design may be to conduct a more traditional safety countermeasure evaluation using the “with-without” comparison. This would entail developing safety performance functions at sites prior to guardrail installation (i.e. without the guardrail), which can be used for subsequent comparison of crashes with the guardrail. The most effective guardrail could then be identified as the one with the greatest reduction in crashes (by severity in particular) compared to a baseline of crashes without the guardrail. This alternative approach would use the full set of crashes (and more because the data would include conditions prior to guardrail installation) and correctly account for exposure as measured level of traffic. There are techniques in the literature that allow usage of crashes with a full range of outcomes (e.g. Ma, et al., 2008; Aguero-Valverde and Jovanis, 2010).

Case-control approaches can also be used if one is concerned with odds ratios associated with safety treatments. Logistic regression has been used with case-control statistical approaches in circumstances where the relative risk or odds ratio of an action are of interest as opposed to an absolute probability of occurrence (see e.g. Haddon, et al., 1961; Jovanis, et al., 2012; or Gross, Jovanis and Eccles, 2009).

Discussion of Table 12 and elsewhere indicate confusion concerning the difference between the value of the test statistic and the level of significance (“p value”) of a statistical test. In this paper, the statistical tests are not conducted on odds ratios but on a null hypothesis of independence of the column totals of frequencies in a 2X2 contingency table. If the logistic regression had been successfully formulated and completed, then tests could be made on odds ratios among the end treatment types.

○

2. Provide an objective assessment of the findings and conclusions of the study:

Are they sound and defensible?

No; if method is incorrect and data are faulty, the conclusions cannot be sound and defensible.

What, if any, issues impact the soundness and defensibility of the findings and conclusions.

See discussion and answers to question 1 above for detailed concerns.

Findings Assessment:

- How would you assess the overall validity of the conclusions? What caveats and conditions would you apply to the results?

The conclusions are basically invalid because the method is fundamentally flawed.

- What additional comments can you provide regarding the study report?

The author does not appear to have a clear grasp of the statistical issues involved with crash data. While this reviewer does not want to appear hypercritical, there are places when the writing seems to reveal some basic lack of understanding.

REFERENCES

1. AASHTO, Highway Safety Manual, 2010.
2. Ma, Jianming, Kara M. Kockelman, and Paul Damien. "A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods." *Accident Analysis & Prevention* 40.3 (2008): 964-975.
3. Agüero, J. and P. P. Jovanis, "Bayesian Multivariate Poisson Log-Normal Models for Crash Severity Modeling and Site Ranking", *Journal of the Transportation Research Board*, No. 2136, (2010): 82-91.
4. Jovanis, P.P., K.F. Wu, C. Chen, Effects of Hours of Service and Driving Patterns on Motor Carrier Crashes, Transportation Research Board, *Journal of the Transportation Research Board*, No. 2231, (2012): 119-127.
5. Gross, F., P.P. Jovanis, and K. Eccles, "Safety Effectiveness of Lane and Shoulder Width Combinations on Rural, Two-Lane, Undivided Roads", *Journal of the Transportation Research Board*, No.2103, (2009): 42-49.
6. Haddon, W. Jr.; Valien, P.; McCarroll, J.R.; and Umberger, C.J., A controlled investigation of the characteristics of adult pedestrians fatally injured by motor vehicles in Manhattan. *Journal of Chronic Diseases* 14(6), (1961): 656-678.

Template: Review of “Relative comparison of NCHRP350 Accepted Guardrail Terminals” v1

Name:	Reviewer 2
Date:	December 15, 2014

Provide a narrative that addresses the two main review areas. You may use the detailed questions as a guide to your response. Use the boxes below (they will expand) for your response.

1. **Provide an objective assessment of the study methodology:**

Are the data, statistical methods, and assumptions reasonable and appropriate? What, if any, issues impact the reasonableness and appropriateness of the methodology?

- Is the research approach appropriate for meeting the study objective: “to evaluate the safety performance of NCHRP Report No. 350-approved guardrail terminals?”
- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and useful crash data?
- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and reasonable exposure data?
- Are the analytic methods and statistical techniques suitable for the types of data given the study objective?
- Are there other methods that could have been considered and if so, why?

Methodology Assessment:

1. **Appropriateness of research approach:** The study objective “to evaluate the safety performance of NCHRP Report No. 350-approved guardrail terminals” might have been more precisely stated and that may have resulted in a more appropriate study design. The real question I think is: Given that there is a run-off-road (ROR) crash, what is the difference in probability of the injury being severe between the two end treatments. This question was not answered by the approach used for arriving at the conclusions, and the logistic approach, which might have had more success, as the author recognizes, was inconclusive and in any case, could not be thorough because of sample limitations. (More discussion in Item 4.)
2. **Crash Data assembly:** This is a positive aspect of the research. It is unfortunate that B and C injury crashes could not be evaluated, though.
3. **Exposure data assembly:** With the author’s definition of exposure, it is a positive that a solid effort was put in to the identification, collection, compilation, filtering, and reduction methods. However, I question this definition of exposure and the use of this measure in the research approach. See below.
4. **Suitability of analytical methods:** Unless I am missing something in the explanations, I think the method used to arrive at the conclusions is inappropriate for the real research question of interest since it does not consider the fact that the frequency and

severity of *encroachments* may be different for the different types of end treatment. To see this, consider that if the “exposure” as defined by the author was the same for the ET-200 and ET-PLUS end treatments but traffic volume and and/or speed were higher at locations with ET-PLUS. Then it is logical that there will be more ROR crashes in general and more severe ones at these locations, and the probability of this system being involved in a crash would be higher – *due to the higher volume and/or speed* – not necessarily because of the treatment. There may also be an endogeneity issue if, as is entirely possible, the newer ET-PLUS treatments were prioritized at locations where severe crashes were more likely to occur.

5. **Are there other (better) methods?** The logistic approach would be an appropriate one but, as the authors found out, a much more substantial dataset would be required. This is the most common approach for probabilistic modeling of crash severity. The outcome would be the probability of a run-off-road crash being KA (or K) given a particular end-treatment. The confounding effects of variables potentially correlated with the presence of a particular end treatment would need to be accounted for in the modeling, and data on these variables would have to be collected. To the author’s credit the importance of including these variables in the model is recognized. There are, of course, other variations of the probabilistic approach that could be explored (nested logit, ordered logit and probit, for example) if a simple logistic formulation is inadequate.

2. Provide an objective assessment of the findings and conclusions of the study:

Are they sound and defensible? What, if any, issues impact the soundness and defensibility of the findings and conclusions.

- How would you assess the overall validity of the conclusions? What caveats and conditions would you apply to the results?
- What additional comments can you provide regarding the study report?

Findings Assessment:

1. **Assessment of validity of conclusions:** The validity of the conclusions is questionable because the analytical method used to arrive at these conclusions is inappropriate, as outlined above. In particular, the conclusions cannot be used as a deterrent to applying ET-PLUS end treatments.
2. **Additional comments:** It is possible that there may be larger and better databases available now or in future for answering the fundamental question of interest. For example, NCHRP Report 655 details these necessary for answering these sorts of questions. I am not sure of the status of the data collection that may have resulted from that effort.

Template: Review of “Relative comparison of NCHRP350 Accepted Guardrail Terminals” v1

Name:	Reviewer 3
Date:	December 14,2014

Provide a narrative that addresses the two main review areas. You may use the detailed questions as a guide to your response. Use the boxes below (they will expand) for your response.

1. Provide an objective assessment of the study methodology:

Are the data, statistical methods, and assumptions reasonable and appropriate? What, if any, issues impact the reasonableness and appropriateness of the methodology?

- Is the research approach appropriate for meeting the study objective: “to evaluate the safety performance of NCHRP Report No. 350-approved guardrail terminals?”
- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and useful crash data?
- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and reasonable exposure data?
- Are the analytic methods and statistical techniques suitable for the types of data given the study objective?
- Are there other methods that could have been considered and if so, why?

Methodology Assessment:

The main conclusion of the report is “that the ET-PLUS placed motorists at a higher level of risk of both serious and fatal injuries relative to its predecessor, the ET-2000.”(Abstract) The question is whether this conclusion is defensible in the light of the data and methods used.

Risk of an ‘End Treatment’ (ET)

What is meant by ‘risk of serious and fatal injury’ that can be attributed to an ET type is not made explicit in the report. It is perhaps implicit in the statement that: “Traditionally the measure of effectiveness of a roadside safety device is the percent of A+K or K crashes relative to the total number of crashes ...” (p.3). This ‘traditional measure of effectiveness’ makes good sense provided that the severity of the collision itself is fixed. Thus, my understanding of the intent behind the traditional measure of effectiveness is that ET’s of type ‘x’ would be considered riskier than an ET’s of type ‘y’ if for the same kinds of collisions (same kinds of vehicles, occupants, speeds, angles etc.) they had a larger probability to result in a K or K+A crash. The ‘same kinds of collisions’ condition has to be added because if the collisions with ET’s of type ‘x’ were usually at higher speeds than collisions with ET’s of type ‘y’, then any difference in percent of A+K or K crashes would be partly due to the different speeds and could not be wholly attributed to the type of ET.

Exposure and its use

The traditional measure of ET effectiveness could not be used in this research because data about the 'total number of crashes' could not be obtained. This is why the report substitutes a different measure (definition) of risk¹. Now the risk of ET of type 'x' is measured by the ratio

R_x

$$= \frac{\text{Number of severe crashes involving an ET of type } x}{\text{Number of ET's of type } x \text{ in 10 miles upstream of severe crash sites with ET of type } x}$$

The denominator is referred to as exposure. Why the author chose to define exposure in this specific manner is not explained.

Thus, e.g., for ET-2000 there were 49 A+K crashes in the numerator, 961 ET's in the denominator and therefore $R_{\text{ET-2000}}=49/961=0.051^2$. This ET type was referred to as the 'baseline system'. The assumption was that if all ET types were equally risky and data was plentiful, then the R's for all ET types should be equal³. Similarly if the R of some system was larger than that for the baseline system it would be deemed riskier. Thus, e.g., for ETs of the ET-PLUS type there were 91 A+K crashes in the numerator, 1200 such ET's in the denominator and therefore $R_{\text{ET-PLUS}}=91/1200=0.0775$. Because $R_{\text{ET-PLUS}}=0.0775$ is larger than $R_{\text{ET-2000}}=0.051$ the author concluded that "*the ET-PLUS placed motorists at a higher level of risk of both serious and fatal injuries relative to its predecessor, the ET-2000.*"(Abstract).

This conclusion, as I show below, is most likely wrong. It is likely to be wrong because it rests on an unreasonable assumption. It would be wrong in many circumstances even if, by some fluke the unreasonable assumption was true.

The unlikely-to-be-true assumption

In the text quoted in footnote 3 the author assumes that if all ET types were equally risky (and if data was plentiful) then the R's for all ET types should be equal. This is why he takes a difference in R's to be the measure of the difference in risk. Were his equal-R's-means-equal-risk assumption incorrect, if the R's of different ET types could differ even if they were equally risky to the motorists colliding with them, then the difference in R's could not be taken to measure the risk inherent in an ET type. In what follows I argue that the author's assumption is incorrect.

The assumption equal-R's-means-equal-risk would be justified if two conditions were met:

Condition a. If the chance of a vehicle to collide with an ET did and the type of ET were causally and statistically independent. This would be true if the decision about what kind of ET to install where was entirely unaffected by the circumstances of the site. It would not be true if the engineers and/or contractors were exercising some judgment about what ET is suitable where. Thus, e.g., condition (a) would not be met if, say, the costlier ET would tend to be chosen for high AADT sites or sharp curves where more vehicles may be expected to run of the road. Unless it can be shown that ET-2000 and ET=PLUS tend to be used in locations where AADT, curvature, and a host of other circumstances affecting the

¹ The author says that: "For each system (i.e. ET type), the number of crashes was divided by the number of those systems in the exposure data. This quotient represented the probability of a system being involved in a crash."(Section 5.1, p. 9)

² See Table 6 on page 14.

³ "Assuming that the exposure data reflects the expected crash frequency, a comparison was made between the observed number of crashes and the expected number of crashes for each identified crashworthy end treatment. It was assumed that this proportion would be equal for all terminals. In other words, no terminal would be more or less dangerous than the baseline terminal". Pages 3-4.

frequency of running-off the road tend to be similar, one may not attribute the difference in their R's to the ET type. The difference may be to an unknown extent due to the circumstances of the site where these two ET types tend to be used.

Condition b. If the characteristic of the collisions with the ET in terms of speed, angle, vehicle type, number and age of occupants etc. did not depend on the type of ET. This would be true if the decision about what kind of ET to use was entirely unaffected by the circumstances of the site and would not be true if those who decide what ET is suitable where exercised judgment. Thus, e.g., condition (b) would not be met if, say, the ET type thought to have superior energy adsorption would tend to be chosen for sites where speed is higher. Unless it can be shown that ET-2000 and ET=PLUS tend to be used in locations where speed and other relevant circumstances tend to be similar, one may not attribute the difference in their R's to the ET type. The difference may be to an unknown extent due to differences in the collision characteristics of the sites where these ET types tend to be used.

In my opinion it is unlikely that ET's are always chosen without considering the circumstances of the site where they are to be installed. If so, if there is some association between ET type and site conditions, the author's assumption that equally risky ETs should have equal R's is incorrect. I conclude therefore that the analysis and the results are predicated on an unlikely-to-be-true and unsupported-by-evidence assumption.

Other paths to failure

The author chose a specific measure of exposure. As shown above, this chosen measure of exposure would fail if conditions a and b were not met. However, it would fail even if in some misanthropic world the two conditions were close to reality and ETs were chosen for completely extraneous reasons, i.e. without any regard of their suitability to site circumstances.

To illustrate, some such extraneous reasons, consider some practices mentioned in the report. The author says that: *"Often times, the contractor has an affiliation with a manufacturer and will use guardrail terminals exclusively from that manufacturer's product line. Other times, the contractor will simply choose the least expensive option."* (p.5).

Consider a scenario such that the guardrail where the severe ET crash occurred was built by the same contractor as the ETs in the 10 miles upstream and the contractor always uses the same type of ET. If n_x is the average number of ETs in the 10 miles preceding the ETs of type x where the severe crash occurred then $R_x=1/n_x$. Thus, if this is how ETs are chosen, then the Rs only reflect the density of ET and have nothing to do with the inherent safety of one ET type or another. With this scenario, if the density of ETs upstream of the severe crash site was the same for type x and type y then R_x would be the same as R_y irrespective of any real differences between the risk to colliding motorists of type x and type y.

Consider another such scenario in which road resurfacing is usually accompanied by other safety improvement, one of which is that of replacing of older ET's by a more modern type. If so, one this more ET would be found on recently resurfaced roads. If the frequency and speed of collisions with ET's is affected by resurfacing one way or another, the R of the modern ET's will be affected. In this case the magnitude of the R would be a function of when the road was resurfaced.

In short, the author's choice of exposure is problematic even in scenarios in which the choice of ET type has nothing to do with site conditions.

2. Provide an objective assessment of the findings and conclusions of the study:

Are they sound and defensible? What, if any, issues impact the soundness and defensibility of the findings and conclusions.

- How would you assess the overall validity of the conclusions? What caveats and conditions would you apply to the results?
- What additional comments can you provide regarding the study report?

Findings Assessment:

Conclusion

The question is not whether the differences in the R's are statistically significant. The question is whether the observed differences between the R's can be reasonably attributed to the safety performance of the corresponding ET types; whether R_x measures the probability of a collision with an ET of type x to end up as a K or K+A crash. To this the answer is: No. The observed differences in the R's can be partly due to differences between the sites where different ET's tend to be installed. Therefore the report provides no basis for forming a judgment about the relative safety of various ET types.

Template: Review of “Relative comparison of NCHRP350 Accepted Guardrail Terminals” v1

Name:	Reviewer 4
Date:	December 15, 2014

Provide a narrative that addresses the two main review areas. You may use the detailed questions as a guide to your response. Use the boxes below (they will expand) for your response.

1. Provide an objective assessment of the study methodology:

Are the data, statistical methods, and assumptions reasonable and appropriate? What, if any, issues impact the reasonableness and appropriateness of the methodology?

- Is the research approach appropriate for meeting the study objective: “to evaluate the safety performance of NCHRP Report No. 350-approved guardrail terminals?” **NO (see below)**
- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and useful crash data? **YES**
- Do the data identification, collection, compilation, filtering, and reduction methods produce accurate and reasonable exposure data? **NO (see below)**
- Are the analytic methods and statistical techniques suitable for the types of data given the study objective? **Theoretically yes but these methods need to be applied to the correct variables (which is not the case now).**
- Are there other methods that could have been considered and if so, why? **See below**

Methodology Assessment:

The research is timely and the data gathering very extensive but there is a **methodology essential flaw** that affects the whole research:

- **the exposure CANNOT be defined independently of the traffic** flowing on the road (see any of the several ROR models available in the literature). In the research the exposure is defined as the number of terminals that ONE DRIVER will encounter in 10 miles. But if there are 5000 drivers the probability of having a crash, if all the other conditions remain the same, is much lower than if you have 50000 drivers. And this is not considered in the paper attributing all the differences IN THE NUMBER OF K+A CRASHES to the terminal type. If, for instance, all the ET-PLUS terminals are all on highly trafficked highways and the ET-2000 terminals are all on roads with a lower AADT this could be the reason for the increased crash probability and it could have nothing to do with the difference in the terminal type;
- the number of K+A crashes is a surrogate measure for the terminal’s effectiveness (as correctly stated by the authors) but it should be handled by comparing equivalent situations. Here you might be comparing a terminal in a straight road with one in a

road with several bends. The higher number of crashes is likely not related to the terminal type but to the bends. **You need to develop a full ROR model accounting for all the different key variables** (at least: curvature, shoulder width, ruble strips etc). Then you will be able to define your baseline with which you can compare different terminal configurations;

- apparently you are combining two-lane two-ways rural roads with divided highways and this is really questionable and can lead to very erroneous results if a ROR model is not used to define the base conditions;
- mixing urban and rural roads is also questionable (ROR models are very different for these roads);
- some additional considerations could be derived by calculating the $K/(K+A)$ ratio and even more if the B crashes could also be included in the database. It is discussed why PDOs are not included (and this is reasonable) but B crashes should be quite well described in police reports (but it clearly depends on local practices);
- crash data collection is well conducted but the following analysis on the collected data needs to be conceptually revised and the dataset needs to be integrated with traffic and infrastructure data.

2. Provide an objective assessment of the findings and conclusions of the study:

Are they sound and defensible? What, if any, issues impact the soundness and defensibility of the findings and conclusions.

- How would you assess the overall validity of the conclusions? What caveats and conditions would you apply to the results?
- What additional comments can you provide regarding the study report?

Findings Assessment:

The findings are not sound and defensible because the methodology is not correct as no comparison can be made between crashes involving terminals on road sections with different traffic volumes and infrastructure features (geometry, lane and shoulder width etc). The exposure needs to be defined as a function of the AADT.