Rebuttal to a Review of “In-Service Evaluation of FHWA-Accepted Guardrail Terminals”

Reviewer: Christine Carrigan

Nomenclature: “Report” will be used throughout this rebuttal to refer to the “In-Service Evaluation of FHWA-Accepted Guardrail Terminals.” Also, “Review” will be used to refer to Dr. Carrigan’s review of the Report. “UAB” will be used to refer to the research staff at UAB that completed the study.

Summary of the Report

The Report was commissioned to investigate potential differences in the safety performance of guardrail terminals. The primary focus of this effort was to conduct an initial evaluation of the relative safety of more recent terminal designs to the first terminal to meet NCHRP Report No. 350 performance guidelines, the ET-2000. It is generally accepted that modern safety devices should perform better than older designs. Note that the objective was to identify the statistical significance of safety performance variations between the most widely used guardrail terminals. There was insufficient funding to conduct a comprehensive study of all approved terminals. Therefore, it was necessary to conduct a retrospective study of impact performance.

The most difficult part of any evaluation of safety feature performance is the determination of exposure. The most common approach involves collecting all available accident records and dividing the total number of serious injury and fatal crashes by the total number of recorded crashes. Under this approach, the severity of impact with the feature is measured in terms of percentage of crashes involving serious injury or fatality, %\((A+K)\). The primary problem with this approach is the difficulty in determining how many unreported accidents involve a vehicle striking the system under evaluation. Some authors have utilized repair records to supplement crash data as a means of estimating the number of unreported accidents. A major problem arises when attempting to discern damage attributable to unreported accidents and other types of damage, such as impacts with snow plows or mowing equipment. Because the two states involved in the study do not permanently store photographs of damaged guardrails prior to repair, distinguishing between damage caused by unreported crashes and other types of damage was impossible.

However, it is important to recognize that A+K accidents account for virtually all of the total societal costs associated with ran-off-road collisions. In order to measure the relative safety performance of guardrail terminals for the most important crashes (all A+K), UAB employed a different method for measuring the exposure of the motoring public to each type of guardrail terminal. Exposure was measured using an inventory of guardrail terminals installed along highways where fatal and serious injury accidents occurred. This approach measures the ratio between the number of terminals involved in A+K crashes and the number of terminals installed along the roadway upstream of where the accident occurred. For the purposes of this study, a 10-mile window of highway was chosen for determining exposure of each type of terminal along the
crash route. This approach has been successfully employed to evaluate the severity of impacting roadside slopes. (*)

Upon completion of the first draft of the study, it was observed that only the ET-Plus had sufficient numbers of serious injury and fatal crashes to produce statistically significant comparisons with the ET-2000. The preliminary findings showed that the ET-Plus was nearly 3 times more likely to be involved in a fatal crash than the ET-2000. Also, the ET-Plus was about 1.4 times more likely to be involved in a severe-injury crash than the ET-2000. Both of these results were statistically significant to the 0.03 level.

**Description of NCHRP Report No. 490**

Cost: $720,000

In the Review, NCHRP Report No. 490 [1] was held up as the only acceptable method of performing an in-service performance evaluation (ISPE), citing the fact that it had been peer-reviewed. The Review includes the following quote “The UAB study violates nearly every methodological recommendation in NCHRP Report 490 which provides a peer-reviewed procedure for conducting an ISPE.” In spite of the fact that NCHRP Report No. 490 was reviewed by NCHRP staff and volunteers, the recommended procedures have serious shortcomings.

NCHRP Report No. 490 sets forth a measure of risk for a particular object by calculating the quotient of the number of A+K crashes divided by the total number of crashes, including PDO crashes. In order to capture unreported collisions Report No. 490 incorporated periodic field inspections of all installed safety systems. Thus, this technique is limited to a prospective approach wherein researchers monitor highways to identify and carefully investigate all relevant crashes. Unfortunately, even with the budget of more than $700,000, the research team could not discern between guardrail terminal damage caused by vehicular impact and that caused by highway maintenance and other non-crash related impacts. Faced with an inability to isolate PDO related damage from other sources of damage, the research team elected to assume all damage was due to unreported crashes. The best example of the problems associated with this approach is the study of the performance of Breakaway Cable Terminals (BCT’s) included in Report 490. Classifying all terminal damage as arising from PDO crashes produced a finding that about 90 percent of all BCT-crashes “…are minor collisions that result in little property damage, no occupant damage and are not reported to police.” Further, it was reported that the BCT terminals “…are performing reasonably well in Iowa and North Carolina.”

These findings are quite misleading. The BCT has not met applicable crash test procedures since TRC 191, circa 1976. Thus, the terminal was unable to meet performance guidelines contained in NCHRP Report Nos. 230 [2] and 350 [3], as well as MASH [4], for test levels 2 or 3. In fact, the FHWA published a memo, dated August 18, 1998, that says the BCT should be replaced with terminals that have passed NCHRP Report No. 350 requirements in conjunction with 3R
projects. The reason for this change, they go on to say, was that in a TL-2 test (820-kg car at 70 km/h), there was “unacceptable passenger compartment intrusion indicating clearly that the BCT is too stiff to accommodate end-on hits, even at reduced speeds.” These findings are further bolstered by crash data studies such as (Pigman, et. al., UK, 1991). The failure of the Report No. 490 procedure to identify safety concerns for the BCT illustrates how inappropriate it is for use in evaluating guardrail terminal safety performance.

**Specific Rebuttal to the Review by Section**

**INTRODUCTION**

First and foremost, the Review has missed the point of the Report. UAB conducted an in-service evaluation of terminals by comparing them to each other. Throughout the Review, Dr. Carrigan references the procedures in NCHRP Report No. 490, which has been shown to be inadequate in general for conducting ISPEs for guardrail terminals. It is also not designed for the type of comparison being made in the Report.

Second, a list of questions was posed that the Reviewer says should be answered. Those questions are repeated here and individually addressed.

- **What is the risk of severe or fatal injury for each type of w-beam terminal studied?**
  
  The objective of the study was not to determine absolute risk of severe of fatal injury, but relative risk.

- **What is the risk of severe or fatal injury in each crash type for the terminals studied (i.e., end-on, length-of-need, side impact, angled on the end, etc.)?**
  
  See prior answer.

- **Is the risk substantially different for one type of terminal than another?**
  
  The only terminal found to have significantly different safety performance than the ET-2000 was the ET-Plus.

- **Is the risk substantially different for any variation of a specific type of terminal than another?**
  
  This question was outside the scope of the study.

- **Is the risk substantially different if the terminal is not properly installed or maintained?**
This question is not pertinent to the current study. All guardrail terminals included in the study are/were proprietary. Manufacturers have an obligation to train their customers to properly install their hardware. It is irrelevant to the victims of these crashes whether the A+K accidents arise from improper design or inadequate training. The Reviewer does not seem to recognize the difference in expectations for both impact performance and installation accuracy associated with the higher costs of proprietary safety systems.

- **What are the common installation/repair mistakes and do they affect the risk of severe or fatal injury?**

  See prior comment.

- **Is any particular terminal more likely to be installed or repaired incorrectly?**

  See prior comment.

- **What proportions of crashes are not reported to the police (i.e., crashes where maintenance records indicate a repair not associated with a police reported crash)?**

  The number of unreported crashes is not relevant to this study. As presented previously, improper consideration of unreported damage severely biased findings from Report 490 with regard to the BCT. The author had no interest in repeating that mistake.

- **How does the vehicle type affect the outcome of the crash?**

  Extreme vehicles were excluded; including single-unit trucks (SUTs), semis, and motorcycles. Therefore, included vehicles represent the same range targeted in standardized testing.

- **How do impact conditions affect the outcome of the crash?**

  It was assumed that for the purposes of the study, terminals installed along roadways with similar speed limits would be subjected to similar ranges of impact conditions.

**ISPE APPROACH**

In Table 1 of the Review, the disadvantages of a prospective study are (1) “Collecting sufficient cases may require considerable time or a very large geographic area to obtain enough cases for statistically meaningful results,” and (2) “Successful data collection depends on many parties cooperating.” These studies are cost-prohibitive for any state under today’s economic environment.
**UAB APPROACH**

This entire section is based upon the assumption that a prospective study is needed to obtain statistically significant findings, even though the relative severity of the ET-PLUS and the ET-2000 was found to be statistically significant at the \( p < 0.03 \) level using a retrospective technique.

The Reviewer suggests that proper planning was not done for the Report, in particular stating that this planning would have safeguarded against insufficient sample size. No amount of planning can generate more data than what already exists, all of which was used in the analysis for each state. If this study had been funded at the $700,000 level, like Report 490, it is likely that every terminal crash in the nation over the last 5 years could have been included. Unfortunately, funds were much more restricted.

The Review states that hardware inventory was not estimated. However, that is exactly what the exposure data in the Report did. More importantly, however, was that the Report estimated inventory on highways where A+K accidents occurred and it was collected from the last photolog or Google Earth photos taken prior to the crash.

In regard to the desire to have data to achieve statistically significant results, statistical significance was obtained for the ET-Plus, the terminal most widely used in both states.

The assertion that there is no way to verify the results is wrong. All the crash data and photologs are either publicly available or available for purchase. Anyone can replicate this study.

The reviewer is again indicating that maintenance records are sufficiently detailed to allow a retrospective evaluation of the each of the following items for every historical guardrail repair:

- What type of terminal was repaired/replaced?
- What caused the damage (automobile, heavy truck, snow plow, mower, etc.)
- What is the precise location of the repair?

Unless all of these questions can be reliably answered, repair records are a source of systematic error that must be eliminated from the study. This was a problem with the NCHRP Report No. 490 guidelines is one major reason why they are almost never done.

1. **UAB OBJECTIVES**

Table 3 in the Review seems to suggest that only the 5 systems described in the Report were considered in the analysis. In fact, the language of the Report allows for the inclusion of other systems, if they exist on the highways of subsequent states. In Missouri and Ohio, only the five described in the Report were found, and as such, only those five were studied. Table 3 in the Review also provides an installed inventory for Ohio and Michigan. It is unclear to UAB if these are simply examples (as suggested by the title of the table) or if they are real values. If they are, a citation to the source is missing. Finally, the State of Michigan is used multiple times in the
Review, but it was not included in the Report. In fact, the Report states that there may be regional effects, and each state should first be considered separately. Any argument in the Review with reference to Michigan is specious.

2. **UAB SAMPLINE [sic] PROFILE**

The Review states, “It is questionable why the UAB team limited the study to only most harmful event crashes and further limited to only severe crashes (i.e., A+K).” Again, A+K crashes are the only real measure of the cost to society. Therefore, BCO crashes were excluded. Also, in a severe crash, if the MHE is, for example, an impact with a tree, but the car also struck a guardrail terminal along the way, then the serious or fatal injury would have to be attributed to the tree, generally speaking. This is the definition of a MHE. By ensuring that the guardrail terminal was the MHE, it can be surmised that the terminal caused the injury. Not filtering for this would actually destabilize the analysis.

The Reviewer also said, “However in its analysis of severe and fatal crashes, the UAB team used other methods including photologs to determine terminal heads which could have been used for the less severe crashes as well.” The major restriction in the data for BCO crashes was a lack of information about the crash itself. Without a scene diagram or narrative, it is impossible to know if the crash location was head-on, face-side, or downstream. More severe crashes always had this information, and many included scene photos.

In the Review, an intriguing example was presented wherein two terminals (A and B) were compared, and the respective probabilities for each terminal were calculated. The reviewer clearly has not considered the exposure analysis incorporated into the subject study. The straw man hypothetical situation posed here would have been properly evaluated had the procedures used in the subject paper been employed. Unless there was a fundamental difference in the way the two terminals were deployed, the terminal with 100 PDO crashes would have had roughly 5 times as many installations along the roadway in order to generate the 5 fold higher crash rate. By deliberately ignoring the exposure analysis incorporated in the subject study, the reviewer appears to have lost any pretense of objectivity.

**HISTORICAL CRASH DATA, POLICE AND MAINTENANCE RECORDS**

The Review states, “A well designed ISPE would have used the historical data to determine how long and where to collect prospective crash data to ensure that statistically significant findings could be made.” This Reviewer is clearly of the opinion that only a prospective approach can be used to determine the relative impact performance of two tangent guardrail terminal designs. Prospective studies are extremely costly and time consuming. The study summarized by this Reviewer would have a price tag in the upper six figure range. This type of blind adherence to a flawed report, NCHRP Report No. 490, has had an immeasurably negative impact on the safety of the nation’s highways.
DATA COLLECTION

Exposure

The Reviewer states that the subject study “did not measure exposure but then they continue on and invent their own unconventional measure.” As the term is normally used in the traffic accident analysis field, the definition of “exposed” is the “state of being at risk of impacting or encountering something.” In the true definition of the word, there is no mention of or need to identify the frequency of less severe crashes. When evaluating the societal costs of a roadside safety system, it becomes clear that virtually all crash costs are associated with A+K crashes.

The need to include PDO crashes brings a great deal of random and systematic error to any accident study. Because it is widely accepted that a large proportion of PDO’s go unreported, the need to count these crashes introduces a great deal of random error into the study because estimating the number of unreported crashes is relatively inaccurate. Inclusion of maintenance records only provides an appearance of improved accuracy. It is almost impossible to distinguish between automobile impacts and some types of barrier damage. For example, when a trailer breaks away from a pickup and strikes a guardrail, the resulting barrier damage is often indistinguishable from an automobile impact.

PDOs introduce systematic error when one device is more likely to be installed along low speed roadways. In this case, this device would be involved in a disproportionately high number of low speed and largely PDO crashes.

Finally, the new measure of exposure assures that only highways where a serious crash has occurred are included, which inherently provides a more realistic measure of exposure to A+K crashes than more traditional methods.

The Reviewer says that it is “unclear why this non-traditional measure of vehicle-miles-traveled was used.” None of these proposed data elements overcome the problems inherent with incorporating PDO data in the analysis. If the Reviewer can identify any systematic or random error associated with the unconventional exposure terms, the author would be delighted to explore ways to improve the study and/or explain reasons for not including it. Unconventional is not the same as bad, biased, or flawed. Total effectiveness in this industry has always been fatalities per MVMT. Second, the 10-mile segment was arbitrary. This length of highway proved to be adequate in a study of the safety of roadside slopes (*).

The Reviewer wonders why traffic volume wasn’t incorporated, even if that data was available. However, unless the roadway is operating above capacity for much of the day, traffic volume (and by extension, crash frequency) would not be expected to contribute to crash severity.

Table 6 of the Review presents distorted data pertaining to Ohio and Michigan (the latter of which is meaningless). This data is highly misleading because a vast majority of roads will be in the very low volume category where EATs are not used. This was described in the report.
The Reviewer pointed out a discrepancy in the total mileage of exposure collected and the total expected mileage for 207 A+K crashes. The author appreciates this observation and will correct the error.

The Reviewer states, “It would appear that the UAB team’s sampling method using fatal and severe crashes to identify study locations seriously biased the results to atypical roadways within the study states.” Again, this was not a study of black spots. It was a direct comparison of terminals. The overall terminal density is sure to be lower when smaller functional classes are included.

The Review of this section concludes with, “These more conventional measures of exposure would have improved the study quality and were readily available to the UAB team.” Please explain how the total vehicle miles traveled and the traffic volume helps assess the difference in performance between two terminals. Conventional is not equivalent to “better.”

ANALYSIS

The Reviewer says that UAB assumed that exposure represents a baseline for A+K crash expectancy. If the number of a given type of hazard along a roadway is not strongly correlated to crash frequency, then all crash prediction modeling should be abandoned. The assumption, or null hypothesis, was that no terminal was more or less dangerous than the ET-2000. This was a fair, objective approach, allowing the data to provide the conclusions. This text will be edited to make this point more clearly.

The Reviewer states that guardrail terminals account for only about 10% of crashes. Once again, this study was not a comparative risk analysis. Of course tree collisions are more severe, as an example.

Another quote from the Review, “The UAB team is excluding less severe crashes and then with no basis claiming that the crash frequency for less severe crashes is the same as the fatal and severe frequency!” Nothing about less severe crashes was ever stated other than they do not represent a major component of societal costs. This statement demonstrates that the reviewer is trapped in a traditional mindset where PDO crashes are a necessary component of safety analysis. As described above, there is a strong argument that inclusion of PDO data produces both systematic and random error that would be impossible to avoid. Along those same arguments, “the correct way” to calculate risk is an opinion and has no bearing on the objective of the subject study.

From the Review: “It appears the assumptions about vehicles only driving 10 miles and that the terminals within 10 miles represent all crash severities was made simply to allow for the UAB team to be able to divide the considered A+K crashes by their assumed ‘all crashes’ and arrive at the system performance.” As aforementioned, the 10-mile segments provide a real definition of
exposure. Further, the total number of crashes is not a data point of interest due to the inherent problems associated with unreported accidents discussed above.

Table 7 of the Review provides some example of a distribution of guardrail terminals near the Reviewer’s location. It appears that the purpose was to show how many terminals would be excluded from the UAB study. This reduction is very close to what was observed in the two states. Many removed terminals were turn downs, which are already known to be dangerous. If they weren’t removed, then the turn down results would have overwhelmed the EATs and the SRT. This study was specifically commissioned to study NCHRP Report No. 350 terminals in Missouri and Ohio. Hence, turn downs were removed, among others.

The Reviewer concludes this section by saying less severe crashes and traffic volumes were not used in the analysis. This has been addressed in this rebuttal in multiple locations.

**REVIEWER’S EXAMPLE**

The Reviewer’s example is specious. She constructs a straw man to estimate the probability of striking a terminal and shows that it is much lower than would be expected. However, her prior criticisms clearly demonstrate that she understands that the study never attempted to estimate probability of collision. The 10-mile exposure distance needed only to provide a stable estimate of the distribution of guardrail terminals installed along highways where A+K crashes occur.

In this example, no distinction is made between the ET-2000 and ET-PLUS. This means the Reviewer was conducting an entirely different analysis. This Report used the ET-2000 as a baseline because it was the first EAT. All designers have an obligation to the motoring public to assure that each design iteration should at least maintain the same level of safety as its predecessor. The study was designed to assume that this philosophy was universally implemented. However, it was found that the only statistically significant result was that the immediate successor to the ET-2000 was actually causing harm at a higher rate, not a lower or equal one. No conclusions were drawn to the effect that the ET-PLUS presented an unreasonably high risk to the public, which would have required a comparison with many other potentially problematic objects, such as BCT’s and turndowns. Instead, it was simply concluded that the ET-PLUS presented a higher risk than the ET-2000. State and Federal agencies are encouraged to carefully explore the appropriate action to take.

The study was not limited to 10 miles. It included more than 1,500 miles. Also, Table 8 of the Review is concerning. None of the fatal crashes involved SKTs. This table deliberately misleads the reader into thinking the SKT is dangerous when no evidence exists to support that idea.

**CONCLUSION**

The first sentence of this section states that UAB “did not provide any conclusive research on terminal crash performance.” The results of the report showed a clear difference in performance
between the ET-2000 and ET-PLUS. Also, the Reviewer seems content to credit UAB with defining “exposure.” UAB does not claim ownership of this word. Meriam-Webster has graciously provided that.

The second paragraph is irrelevant. What was relevant was the system involved in the crash and the other systems that were passed by leading up to the crash. None of this changes after the crash has occurred. Also, if newer devices, such as the ET-PLUS, are replacing older devices, like the ET-2000, then the overall percentage calculated for the ET-PLUS should be lower. Despite that expectation, it was 3 times more likely to be involved in a fatal crash than its antiquated predecessor. As a side note, the Reviewer’s note in the discrepancy of mileage reported in Missouri has increased this odds ratio to closer to 4.

The Reviewer states, “The UAB study violates nearly every methodological recommendation in NCHRP Report 490.” NCHRP Report No. 490 is fundamentally flawed. It makes no sense to follow a flawed procedure. In one of its example studies, it showed that the BCT represented no real safety concern. However, in a TL-2 NCHRP Report No. 350 test, the guardrail basically penetrated the vehicle. To make matters worse, this device is found in numerous locations more indicative of TL-3. As such, the FHWA has mandated, since 1998, that the BCT be removed and replaced in all 3R projects. If NCHRP Report No. 490 can be so incredibly wrong on the BCT, a guardrail terminal, then it clearly is not suited for ISPEs of any guardrail terminal.

The Reviewer has digressed into leveling unwarranted accusations against the UAB research team. She says that they, “…skewed the results of the research.” This comment is unprofessional and unwarranted. The analysis was well-documented, and any simple check on the math would reveal that the distance traveled is arbitrary, as long as the exposure data is essentially stable. No attempt to skew the results was made. In fact, the statistical study was designed to use a null hypothesis that said none of the terminals was performing differently than the ET-2000.

The Reviewer advocates that all crashes should have been included, not just the ones where the impact with the guardrail terminal was the MHE. This is an incorrect idea because the relative severity of terminals would be even higher than reported. However, it would have been more indicative of the severity of other objects, like trees, which are known to be more severe than terminals.

The Reviewer maintains that state maintenance records could have been used to precisely identify the crash terminals. However, photos are far more accurate and were relied upon for the subject study. The reference to “approximately” was found in the first chapter of the Report and was included as a general description of the method. In each state, it was made clear that scene photos were used.

The Reviewer reiterates that the statistical analysis used by UAB is biased and statistically unsound. However, no evidence of either was presented. Researchers at UAB stand by the analysis and the tools used therein. When only one independent variable is available in the data...
(the type of terminal), simple descriptive statistics are more than sufficient, along with a check on statistical significance via Fisher’s exact value.

The Reviewer states, “The authors purport that their study shows there is a difference in performance between different terminals, in fact, it only shows that the number of severe and fatal crashes is in the same proportion as the number of terminals installed – hardly a surprising fact.” This makes no sense. The proportion of fatal crashes for the ET-PLUS was about 3 times (now 4 times) higher than the proportion of the ET-2000. Three is greater than one; ergo, the terminals performed differently (which was a statistically significant finding according to Fisher’s exact value).

Another criticism of the Reviewer was that UAB did not utilize a peer-reviewed method. We chose to develop a better approach than the “peer reviewed” NCHRP Report No. 490 because of inherent flaws in that document, as already described.

Finally, another parting shot at UAB is taken when the Reviewer repeats herself and says the analysis was biased and skewed. Any objective reader will quickly identify the obvious bias of this Reviewer, whose husband is being paid by Trinity Industries to defend the performance of its ET-Plus terminal in many tort cases. The Reviewer’s comments are largely based on an unyielding adherence to NCHRP Report No. 490, even though it has never been used to produce meaningful knowledge that has contributed to improved highway safety.

References