SPECIAL REPORT 323:
IN-SERVICE PERFORMANCE EVALUATION OF GUARDRAIL END TREATMENTS

TRB-SASP-14-05
In-Service Performance of Energy-Absorbing W-Beam Guardrail End Treatments: Phase 1

Critical Review of Methodologies for Evaluating In-Use Safety Performance of Guardrail End Treatments and Other Roadside Treatments

BHAGWANT PERSAUD
Persaud and Lyon, Inc. and Ryerson University

Paper prepared for the
Committee for the Study of In-Service Performance
of W-Beam Guardrail End Treatments, Phase 1
Transportation Research Board

2017

Critical Review of Methodologies for Evaluating In-Use Safety Performance of Guardrail End Treatments and Other Roadside Treatments

BHAGWANT PERSAUD
Persaud and Lyon, Inc. and Ryerson University

1. PREAMBLE

The objective of this paper is to inform an in-service performance evaluation (ISPE) study to be designed by the TRB committee that will have the following potential applications for the results:

1. Compare the effectiveness of various guardrail end treatment designs in minimizing injury severity in a crash, while considering factors aside from design that might affect this outcome, such as road conditions, vehicle type, speed, and crash dynamics. Assessment may include quantitative and qualitative elements. The results would aid highway agencies in selecting end treatment designs to be installed at particular locations.
2. Evaluate the impact of installation and maintenance practices and environmental conditions on in-service performance.
3. Develop data capture protocols for in-service performance evaluation for use by state highway agencies in conducting their own evaluations.
4. Provide data to support the improvement of end treatment designs.
5. Estimate benefits and costs of a program of replacing installed end treatments with end treatments that perform better.
6. Validate or refine crash test procedures.

A recent paper by Carrigan and Ray (2016)[1] sees ISPEs similarly in that it states (page 2) that these evaluations are needed "not only to better understand the performance of roadside hardware, but also to improve the crash testing specifications which are used in the design and acceptance of new hardware for use on the nation's highways as well as to improve design guidelines used for warranting and placing roadside hardware in the field."

The literature review for this paper is by no means comprehensive. It is necessarily focused on what can be argued is the most important research need -- better understanding the performance of roadside hardware, which is closely related the first application (#1) identified above. The other five listed applications, as well as the other two ISPE needs identified by Carrigan and Ray, are interlinked with the primary need, so it is natural that methods and data used for better understanding roadside hardware performance may also yield knowledge pertinent to the other needs. For example, the impact of installation and maintenance practices is captured in the typical questions that may be answered in an in-service evaluation of roadside hardware performance as identified by Carrigan and Ray (2016):

---

[1] Carrigan C. and M. Ray. Practitioner's Guide to the Analysis of In-Service Performance Evaluation Data. Paper presented at the 95th Annual Meeting of the Transportation Research Board, January 2016.

- What is the risk of severe or fatal injury for the hardware studied?
- Is the risk substantially different for different vehicles?
- Is the risk substantially different for different speeds?
- Is the risk substantially different for different crash types (i.e., end-on, length-of-need, side impact, angled on the end, etc.)?
- Is the risk substantially different between manufacturers of the same or similar systems?
- Is the risk substantially different if the hardware is not properly installed or maintained?
- What are the common installation/repair mistakes and do they affect the risk?
- Is any type of hardware more likely to be installed or repaired incorrectly?
- What proportions of crashes are not reported to the police (i.e., roadside crash successes)?

This paper is organized as follows. First, there is an overview of the spectrum of methods pertinent to ISPEs that emerge from a scan of the literature. This is followed by more detailed discussion of the methods grouped into logical categories in separate sections. Each section will generally address data needs and availability for applying the various methods. Finally, there is a summary that includes a comparative evaluation of the various ISPE methods, vis-à-vis data needs and suitability of application at various levels of expertise.

2. OVERVIEW OF THE SPECTRUM OF ISPE METHODS

The methods potentially relevant to ISPEs range from simplistic to sophisticated. The more simplistic methods can utilize routinely available data but are limited in terms of the usefulness and credibility of the results. Nevertheless there is value in such methods. For example, as Carrigan and Ray (2016)[1] note "one may also simply gather crash records to determine if there is an unacceptable level of risk within a certain group of hardware (e.g., guardrail, guardrail ends, bridge piers, etc.) and if further analysis is warranted" (page 6). The further analysis may address why issues with performance are observed (e.g., maintenance, installation, hardware design, etc.) by gathering data from maintenance, repair and construction records, as Carrigan and Ray note. Or the further analysis may seek to benchmark performance through a comparative evaluation of various designs, levels of installation and maintenance practices, or manufacturers products. For such a comparative evaluation, methods run the gamut from the simplistic comparisons without controlling for confounders related to application circumstance to various levels of sophistication in controlling for these confounders.

Methods applied in road safety analysis in general are typically borrowed from other disciplines, given the relative novelty of statistical analysis of road safety data. The methods potentially relevant to ISPEs are no exception. For example, case-control methods, which are reviewed in some depth have been widely applied in epidemiology and in general for the study of rare events. And probability severity modeling, which applies the highest levels of sophistication are borrowed from the random utility choice models that have been widely used in the transportation planning field long before they have been applied in road safety. (One of the most referenced original source in both planning and safety applications is a landmark book chapter by McFadden (1974)[2].) These models may employ data massaging techniques, such as defining cut-points for categorizing continuous variables that have been developed for other fields.

Inevitably, it seems reasonable, given the objective of roadside safety treatments, that comparative evaluations should be based on the risk of serious injury, i.e., injuries classified as K or A on the KABCO scale, as suggested in key sources such as NCHRP Reports 350[3] and 490[4] and AASHTO (2009)[5]. This risk can be expressed in an absolute sense in terms of serious injury crash frequency or in a relative sense terms of the probability of such injury given that a crash has occurred. In the former case, studies may calculate the crash rate per vehicle passing, generally using this as a proxy for crash severity distribution; that is, they are assuming that if a device type has a relatively high rate of severe crashes per vehicle passing, it is because crashes with that type tend to be more severe, and not because the device type is struck more often. In the latter case, the propensity for a serious crash involving a roadside treatment given that a crash has occurred can be simply estimated by dividing the count of serious crashes by the count of all crashes of all severities, or as an odds ratio that in effect compares this propensity or an equivalent measure for two different roadside treatments, or by directly modeling the probability with the type of treatment as one of the variables. Various levels of accounting for confounders may be applied in each case, depending on data and analytical resources available.

In the sections to follow, more details on the various methods are provided. The crash frequency methods are first reviewed, followed by the methods to assess risk given that a crash has occurred. Given the objectives of the paper, the focus is, naturally, on the latter.

3. CRASH FREQUENCY METHODS

These methods are attractive, in that they seem amenable to application *at the state level* to get relative if not precise indicators of safety of roadside safety devices, including end terminals. NCHRP Report 490[4] documents two such methods. These are summarized below.

3.1 NCHRP 490 Method 1 – Estimating Raw Injury Collisions Per Unit of Exposure

The summary (page 53) states that “the issue of collision exposure has been a serious shortcoming of most in-service performance evaluations performed in the past. One method for indirectly accounting for unreported events is to base collision rates on the exposure to traffic. If, for example, the number of injury collisions per million vehicle-km traveled past a barrier is calculated, the unreported events are at least indirectly included in the overall exposure to collision events. This requires that basic traffic data be collected as well as police-reported collision data and on-site data.” (Note that “vehicle-km” should be “vehicle-miles”.)

---

[2] McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), Frontiers in Econometrics. New York, NY’ Academic Press.

[3] H. E. Ross, Jr., D. L. Sicking, R. A. Zimmer and J.D. Michie, “Recommended Procedures for the Safety Performance Evaluation of Highway Features,” Report 350, National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., 1993.

[4] M. H. Ray, J. Weir, and J. Hopp. In-Service Performance of Traffic Barriers. NCHRP Report 490, 2003.

[5] AASHTO Technical Committee for Roadside Safety, “Manual for Assessing Safety Hardware,” American Association of State Highway and Transportation Officials, Washington, D.C., 2009.

This method seems reasonable given that traffic data are routinely available these days in most jurisdictions. It seems that it can be extended to apply to point hazards by estimating injury collisions per million vehicles instead of per million vehicle-miles. However, it would be important not to compare the crash rates for features with vastly different traffic exposure. This is because the relationship between crashes and exposure is known to be non-linear, and this is especially the case for run-off road crashes. Run-off-road (ROR) crash frequency per million vehicles will typically decrease with increasing exposure, likely because of decreasing speed and this phenomenon would be even more so, logically, for injury crashes. Thus, a lower injury crash rate observed at a feature with a higher traffic exposure may have little or nothing to do with the crash-worthiness of the feature. Thus, the method can also yield useful conclusions about whether treatment A performs better than treatment B if it has a lower crash rate despite having lower traffic exposure, although the difference in performance could not be meaningful.

It would also be important to control for important confounders in this as for any other method as noted by Carrigan et al. (2017)[6] who suggest that surrogates like road class and posted speed can be used for impact type to restrict the sample to similar ranges of impact conditions. They also note that impacting vehicle type should be considered as well since hardware test levels are limited to certain particular vehicle types. The authors note, for example, that guardrail terminals are typically designed to test level three (TL3) for which tests involve only passenger vehicles such as pick-up trucks and sedans. An example of a study that controlled for these confounders is one by Johnson and Gabler (2015)[7] who used all crashes in the sample and also sub-sets containing crashes on roads of the same access control level, posted speed limit, and road class to evaluate guardrail end terminals and found that car drivers have greater potential for injury in end terminal crashes than light truck/van/sport utility vehicle drivers.

To control for confounders using the NCHRP 490 method would require calculating the crash rates for roadside treatments grouped by categories of these variables. Again, it would be important not to compare across treatments or categories with vastly different traffic exposure. It may make sense in this regard to also categorize by ranges of traffic exposure but that would quickly dilute the sample sizes in each category, so much so that the statistical validity of the comparisons would be questionable.

3.2 NCHRP 490 Method 2 – Estimating Base Injury Collision Rates Independent of Geometry and Roadside Characteristics

Appendix D (pages 81-82) of NCHRP Report 490 presents this method as follows: "From the historical number of collisions, geometric characteristics, and traffic volume of a roadway segment, a base injury collision rate can be back calculated for a roadside device that is independent of geometry and roadside characteristics. Base injury collision rates for different types of hardware can be directly compared since the geometry, roadside characteristics and traffic exposure have been standardized. A base injury collision rate is the number of injury-causing collisions per million vehicle-kilometers traveled past the device on a standard cross-section of roadway. It is expressed by the following equation:" (As noted above, "vehicle-kilometers" should have been stated as "vehicle-miles".)

[6] Carrigan C, Ray R and A. Ray. Evaluating the Performance of Roadside Hardware. Paper offered for presentation and publication at the Annual Meeting of the Transportation Research Board, January 2017.

[7] Johnson N. and H. Gabler. Injury Outcome in Crashes with Guardrail End Terminals. *Traffic Injury Prevention* (2015) **16**, pp. 103–108, 2015.

$$\hat{R}_h = \frac{\sum_{j=1}^{n} C_j}{\sum_{j=1}^{n} AADT(365)\, L_h \times \prod_{i=1}^{m} AMF_i}$$

where $\hat{R}_h$ = base injury collision rate for hardware h on a road section made up of n segments, C_h =

number of injury collisions on segment j,

$AADT$ = annual average daily traffic volume (vehicles/day) on segment j,

L_h = length of hardware h on segment j, and

AMF_i = m accident modification factors for segment j

This base injury collision rate is used to compare two or more features or devices.

The denominator in the equation is the Highway Safety Manual[8] crash prediction algorithm. Issues with applying that algorithm, for example, the multiplicative nature of the AMFs and the unavailability of roadside specific AMFs (the HSM CMF is based on a subjective hazard rating), appear to have limited the application of this method. In one known application of the method[9],BCT and SRT end terminal features in Washington were evaluated using million vehicle passages in the denominator, adjusted by accident modification factors for curvature and shoulder width. The study, which also determined that most BCT features had an acceptable level of correctness in their installations, concluded that the properly installed BCT feature is still a valid end treatment. The BCT and the SRT features were said to be comparable in their severity responses and both are responding within acceptable expectations. According to the report, "not enough data was gathered on the MELT, ET-2000, FLEAT, or SKT terminals to make conclusions" (page 37). The BCT and SRT collision counts were 18/year for BCT and 9/year for SRT, with injury counts of 7 and 3/year, respectively. 90% confidence intervals for the base collision rates were estimated, but the method for doing so was not documented.

Despite the limitations of the methodology, there is promise for future application at the state level in that the denominator can now be more precisely estimated with the methodology being developed in NCHRP Project 17-54 and outlined in a draft TRB 2017 paper[10] that presents a new approach to ROR crash prediction. In this, the theories of the encroachment probability methods and the strengths of the crash-based data are combined to develop an ROR crash prediction methodology. The method has been developed for incorporation of the explicit consideration of roadside features in the HSM and is said to be suitable for use by those wishing to develop jurisdiction-specific models in that it requires minimal data collection by the states; for example, to perform an assessment for a specific longitudinal barrier type, only the geometric data and the percent of that longitudinal barrier present is needed to estimate ROR frequency and severity.

---

[8] AASHTO. Highway Safety Manual. 2010.

[9] Igharo P., Munger E. and R. Glad (2004) *In-Service Performance of Guardrail Terminals in Washington State* WA-RD 580.1

[10] Carrigan C. and M. Ray. A New Approach to Run-off-Road Crash Prediction. Paper submitted for consideration for presentation and publication at the 96th Annual Meeting of the Transportation Research Board, January 2017.

4. METHODS TO ASSESS RISK GIVEN THAT A CRASH HAS OCCURRED

With respect to an evaluation of the safety impacts of, e.g., alternate guardrail end treatments, the focus of the discussion herein, the question to be answered is: *Is a higher severity crash outcome more likely with one type of end treatment over another, all else being equal?*

It seems reasonable to assume that the end treatment design does not itself materially affect the likelihood of a crash occurring, all else being equal. (It is possible, however, that certain end treatment designs are typically located where more (or less) crashes are expected due to other risk factors, whether this trend in placement was intentional or not, e.g. traffic volumes, curvature, urban vs rural location, lighting, driver demographics etc. For example, flared terminals are offset slightly more from the roadway – so fewer crashes may be expected.) Thus, it is not a matter of asking if a certain end treatment causes more crashes, but rather: *given that a crash occurs, is one end treatment more or less likely to result in a more severe crash outcome?*

It could be argued that if an end treatment design was so successful in shifting the distribution of crash severity towards lower severity that some crashes may no longer be reportable. If this were the case then the effectiveness of the end treatment may not be apparent because the remaining crashes are only those with a higher severity and it may appear that crashes with that end treatment type are more likely to be severe. In any case, this outcome may be unlikely or immaterial to the answer to the fundamental question.

Regardless, in answering the fundamental question in comparing two designs, or different installation and maintenance practices, there is a need to control for confounding factors as much as possible to guard against the possibility that different treatments being compared have different application circumstances. For example, if minor crashes are less likely to be reported on rural roads than in urban areas, but severe crashes have more similar rates of reporting, then any device type that is more common on rural roads will tend to have a higher ratio of severe to total reported crashes. Thus, the confounding location variable, urban vs rural, needs to be controlled for. The gamut of controlling for confounders runs from estimating the simple proportion of outcomes that end in injury or serious injury, with limited control for confounders, to matched case-control studies, to sophisticated probability severity modeling that controls for confounders by including them as explanatory variables.

4.1 Simple Proportions

4.1.1 Risk of a Serious Crash with Simple Proportions

In essence, with the right data and with valid assumptions, the propensity for a serious crash involving a roadside treatment given that a crash has occurred can be usefully estimated by simply dividing the count of serious crashes by the count of all crashes of all severities – a measure defined simply as “risk” by Daniello and Gabler (2011)[11] who also computed another simple proportion as an “odds ratio” (Section 4.1.2). This is the recommended method for state and local agencies in NCHRP Report 490 and even in more recent sources such as Carrigan and Ray (2016)[1], which also, importantly, provides a method for estimating the standard error of the proportion for assessing reliability of the estimate and for assessing statistical significance of differences in two or more proportion estimates.

[11] Daniello, A., & Gabler, H. Effect of barrier type on injury severity in motorcycle-to-barrier collisions in North Carolina, Texas, and New Jersey. Transportation Research Record, 2262, 144–151 (2011).

A direct application of the method for end treatment evaluation was a 2004 Kentucky study[12] that evaluated the ET2000 guardrail end treatment based on the percentage of cases where the guardrail performed properly. Proper performance was said to be "related to an interpretation of whether the guardrail extruded and the posts broke away as designed without causing the vehicle to overturn or causing any spearing of the vehicle" (page 5). In this case, performance was judged to be proper in 70 of 80 collisions, a result that was deemed to warrant continued use of the ET2000.

The fundamental problem in applying this method is in making comparisons of this probability across various treatment designs or other objects of comparisons to make conclusions on the relative merits of the designs or applications being compared. First, the denominator in the probability calculation may need to include unreported crashes to avoid the potential for bias against designs and installation and maintenance practices that are so successful that they may result in a higher percentage unreported crashes compared to lesser designs. Second, the comparison needs to be made for similar application circumstances to avoid bias against designs and installation and maintenance practices that are applied in situations more conducive to higher severity crashes, such as higher speed roads or roads with more truck traffic.

Much has been written about overcoming the first difficulty by collecting appropriate data in sources such as NCHRP Report 490. The question is whether or not it is worth the effort. It may not be if the estimated risk based only on reported crashes provides a clear-cut conclusion on a comparison of two designs. The value of estimating unreported crashes will ultimately depend on the research question, and the value of the result. The effort may be minimized by estimating the proportion of unreported crashes over a random sample of applications rather than the entire population of interest. For this work a system needs to be in place for tying maintenance records with crash report and determining unreported crashes from that information (as appears to be the case in Wisconsin). More formally, the question of cost-value of research, and specifically whether investment in such a system is worthwhile, can be addressed with formal methods as suggested by Hauer et al. (2012)[13] who argue that money should be spent on research that promises the most value for the dollar, and suggest a logical and quantitative approach for estimating the dollar value of research that aims to estimate the safety effects of various actions. The methodology can be adapted in principle to research on the relative merits of two roadside treatments, by considering the relative safety of the two treatments as the safety effect of the "action" of interest.

The second difficulty mentioned above can be mitigated in one of two ways. First, if it can be reasonably assumed that the installation policy of the agency is such, and the samples are sufficiently large, that the distribution of application circumstances is similar over the designs, etc., being compared, then the potential bias is minimized. Second, if this assumption cannot reasonably be made then the confounding application circumstances need to be controlled for by matching the treatment types by factors affecting crash risk and controlling for unmatched factors through statistical modeling. These methods are discussed next in some detail given that there is not a lot written about them in the context of the objectives of the paper – compared to the simpler methods.

---

[12] Agent K. Evaluation of the ET2000 Guardrail End Treatment (Agent; January 2004) Research Report KTC-04-1/SPR107(4)-98-2F.

[13] Hauer E., Bonneson J., Council F., Srinivasan R. and G. Bahar. Value of Research on Safety Effects of Actions. *Transportation Research Record: Journal of the Transportation Research Board, No. 2280,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 68–74.

4.1.2 Odds Ratios with Simple Proportions

A related measure that inevitably leads to the same conclusions, but perhaps more meaningful relative comparison statistics, is what is termed an Odds Ratio (OR).

Two OR applications are potentially relevant, one by Daniello and Gabler (2011) and one by Schrum (2014)[14] applied specifically for an ISPE of guardrail end treatments.

Application 1 In the Daniello and Gabler application directly comparing the effects of barrier type on severity of motor cycle crashes, the odds of fatal or severe injury for each barrier was computed using the equation:

Odds of a severe injury = (Proportion of crashes that are severe)/(1-Proportion of crashes that are severe)

The odds ratio (OR) of severe injury was then computed to directly compare two barrier types:

OR = (Odds of Severe Injury for Barrier A)/(Odds of Severe Injury for Barrier B)

A method is provided for computing the 95% confidence interval (CI) for the OR, which would indicate a statistically significant OR if a value of 1 is not included in the CI.

This methodology was applied by Alluri et al. (2014)[15] to evaluate the safety performance of G4 (1S) W-beam guardrails versus cable median barriers on Florida freeways. Their results indicated that guardrails performed better than cable median barriers and the difference was statistically significant.

Application 2 The Schrum application essentially redefined the denominator in the first equation above as "exposure" and used for this the number of similar treatments (guardrail end treatments in this case) in the 10 miles prior to an end treatment that experienced a K or A crash. The numerator used the sum of K or A crashes at the end treatment that experienced a crash. The reason given for not using non-severe crashes in the denominator was that it was often impossible to determine the impact conditions of the crash in the Missouri and Ohio crash reports used since it was very likely that many of the crashes involved downstream impacts with the end terminal or side impacts with the upstream terminal. In addition, the author notes that most crash reports of minor injuries of PDO crashes excluded scene diagrams and almost never included photos or reconstruction reports, a difficulty that did not apply to K and A crashes. Schrum applied the first equation above (with his exposure measure) for several end terminal types and then applied the second equation with the ET-2000 odds in the denominator and each of the other end terminal odds in the numerator. The result was an estimate of "how much more or less likely the end terminal is to be involved in a crash than the baseline terminal, which was the ET-2000" (page 13). The results indicated that the ET-Plus terminals had significantly higher of odds of K and KA crashes than ET-2000 based on Fisher's exact test for statistical significance that was documented in the report.

The Schrum study methodology has been the subject of intense debate and discussion by several researchers, including the author of this paper, who was one of the FHWA contracted

---

[14] Schrum, K. Relative Comparison of NCHRP350 Accepted Guardrail Terminals. UAB School of Engineering Report. 2014.

[15] Alluri P., Gan A., Haleem K. and J. Mauthner; 2014)Safety Performance of G4 (1S) W-Beam Guardrails versus Cable Median Barriers on Florida's Freeways. Journal of Transportation Safety and Security 7:208–227, 2015.

peer reviewers. As such it is documented here, without critique, only for completeness, and to put it in the context of the other methods reviewed. Suffice to say, the same two issues discussed above for risk of a serious crash would be applicable to these odds ratio with simple proportions measures, and the same considerations for resolution or mitigation would apply.

4.2 Case-Control Studies

4.2.1 Elements of a Case Control Study in the Road Safety Context

Case-control studies are frequently used in the field of epidemiology and applied to rare outcomes and, in this sense, may be considered attractive for evaluating in-use safety performance of guardrail end treatments and other roadside treatments. These studies are not to be confused with cross-sectional studies (A good comparison of the two study types is provided in Gross and Donnell (2011)[16].), but they do use cross-sectional data to estimate the increased likelihood of a risk factor being present for a specified outcome. Cases define the outcome, e.g., a location where a crash occurred, and controls are defined by locations where the outcome did not occur. The likelihood of an actual risk factor being present is typically expressed as the odds ratio between two levels of a variable. For example, it may be found that the odds of a crash occurring on horizontal curves with a degree of curvature greater than 15 degrees is 1.5 times the odds of a crash occurring on curves less than 15 degrees. Risk factors may take the form of binary variables (e.g. median barrier, roadway lighting, or guiderail) or multi-level variables such as lane width (e.g. 9, 10, 11 and 12 foot lanes). To illustrate the concept of the odds ratio, consider the data in the table below.

Tabulation for Simple Case-Control Analysis

Risk Factor	Number of Cases	Number of Controls
With	A	B
Without	C	D

The number locations is tabulated, characterized by whether the location has the risk factor present and whether a crash occurred (cases) or a crash did not occur (controls). It is important to note that the sum of crashes is not tabulated for the cases but rather the number of locations where one or more crashes occurred.

The odds ratio is expressed as the expected increase or decrease in the outcome in question due to the presence of the risk factor. An odds ratio greater than 1.0 suggests that the presence of the risk factor increases risk, while a value less than 1.0 would suggest a decrease in risk. Using the notation in the table the odds ratio is calculated as:

$$OddsRatio(OR) = \frac{A/C}{B/D}$$

[16] Gross F. and Donnell E. Case–control and cross-sectional methods for estimating crash modification factors: Comparisons from roadway lighting and lane and shoulder width safety effect studies. Journal of Safety Research 42 (2011) 117–129.

Case-control studies cannot be used to measure the probability of an event (e.g., crash, severe injury, etc.) in terms of the expected frequency. In this regard it should be noted in passing that Schrum (2014), which has been argued by others[17] to be a case-control study, does not meet this requirement in that what were termed as "odds ratios" were calculated based on the sum of crashes for the cases, which is more akin to calculating and comparing a crash rate. In a strict case-control study, by contrast, the cases would be the number of locations with a crash with and without the feature of interest, as noted earlier, not the sum of crashes. In addition, from Scrum's report it is not clear whether any of the controls -- taken to be the number of end treatments of the same type within an area 10 miles upstream -- may have also experienced a target crash and so would not have constituted a control in a classical case-control study. Notably, Schrum himself did not characterize his study as a case-control one, per se.

It is important that the question being asked is considered in defining the cases and controls in a case-control study. This is illustrated through the following three brief reviews of case-control studies in the road safety area.

In a study related to roadway infrastructure, a matched case-control method was employed to estimate the impact on crash frequency for geometric design elements, including combinations of lane and shoulder width for a given total width[18]. In this application, cases were segments of road where a target crash had occurred and controls were segments of road where a target crash had not occurred. Two types of target crashes were defined: all non-intersection crashes, and a subset including only run-off-road, head-on, sideswipe-same-direction and sideswipe-opposite-direction crashes. To account for potential confounding effects, segments were matched by AADT and segment length while other variables were included in a model as covariates. The model shown below was estimated using conditional logistic regression which allows for the consideration of the matching of cases and controls. In essence the model predicts the probability of a given site being a case (crash site) or control (non-crash site) given its site characteristics.

$$Pr(Y) = 1/\left(1 + exp\left[-\left(\alpha_j + \sum_{i=1}^{p} \beta_i x_i\right)\right]\right)$$

where,
Y = 1 if a case; 0 if a control
α_j= the effect of the matching variables for each matched set
β_i= estimated coefficients for explanatory variables
x_i= unmatched explanatory variables included in the model
To interpret this model, the exponential values of the estimated β_s are the odds ratios for the value of the associated explanatory variables compared to its baseline. In this application the baseline conditions for the road cross-section was a pavement width of 36 ft. with 12 ft. lanes and 6 ft. shoulders.

---

[17] Schaefer-Wilson J. The Safety Institute Responds to Critical FHWA Review. http://www.thesafetyinstitute.org/the-safety-institute-responds-to-critical-fhwa-review-3/. January 2015. Accessed September 14, 2016.

[18] Gross, F., Jovanis, P., Eccles, K. and K.Y. Chen. Safety Evaluation of Lane and Shoulder Width Combinations on Rural, Two-Lane, Undivided Roads. FHWA-HRT-09-031. June, 2009.

Data from Pennsylvania and Washington were collected, including variables related to crash events and conditions, roadway geometry and traffic volumes. The model for Pennsylvania included speed limit, lane width, shoulder width, unpaved shoulder width and district as explanatory variables. The model for Washington included speed limit, lane width, shoulder width and presence of horizontal and vertical curves as explanatory variables.

Significant findings showed a declining trend in the odds ratios as the total width increases when controlling for AADT, segment length and the other variables included in the models. Findings for the combinations of lane and shoulder width for a given total width were mixed with some categories of total width indicating an improvement in safety with greater lane width while other categories indicated an improvement in safety with greater shoulder width.

Haworth et al. (1997)[19] applied a case-control approach to collecting data for 222 motorcycle crashes (cases) and 1,195 non-crash involved (controls) motorcyclist trips past a crash site at the same time of day and day of week of crash occurrence. Data collection included detailed information of each crash, a comparison of features of cases and controls, and motorcycle exposure information. The controls comprised three groups. One group included riders who did not stop. A second included riders who stopped and were interviewed roadside. A third group included riders who gave a roadside interview and a follow up interview. Odds ratios were estimated through conditional logistic regression. The approach is termed conditional because cases are matched to their controls by day, time, and location and other confounding variables such as age are included in the model as explanatory variables. Some of the factors found to increase crash risk included age under 25, never married, unlicensed, increased blood-alcohol concentration, use of a side car, motorcycle engine over 750 cc and rider not being the owner of the motorcycle.

Keall et al. (2013)[20] used a case-control study design to quantify fatality risks for motorcyclists based on blood alcohol concentration (BAC). Thirteen case (i.e. crash) data were acquired from police reports and post-mortem data. A total of 194 control data were collected by sampling from a large database of motorcycle riders that had been stopped at the roadside for random BAC testing. A conditional logistic regression model was then estimated using BAC, vehicle type, gender, age, time of night, year and geographical area. Even at low levels of BAC, much higher risk of fatality for motorcyclists was found. At BAC of 0.03, the fatality risk was three times higher; at BAC of 0.08, the fatality risk increased twenty times.

For the evaluation of guardrail end treatments, one could pose a similar question and collect data for guardrail end terminal locations where a crash occurred and similar locations where crashes had not occurred. Since the question of interest is related to injury severity outcomes one would want to collect data for all those variables that may affect injury severity. These could include roadway geometry and traffic control variables such as horizontal curvature and posted speed limits that would affect the angle of crash and vehicle speeds which certainly would impact crash severity, and such variables would be available at both cases (crash sites) and controls (non-crash sites). However, there are many other variables affecting crash severity that would only be relevant where a crash occurred and would not be available for control sites. These would include model and age of vehicle and driver/occupant age for example. Particularly when analyzing what is likely to be a small sample size of data it is important to consider all

[19] Haworth, N., Smith, R., Brument, I. and N. Pronk. Case-Control Study of Motorcycle Crashes. Report Number CR 174, Federal Office of Road Safety, Canberra Australia. 1997.
[20] Keall, M.D., Clark, B. and C. Rudin-Brown. A Preliminary Estimation of Motorcyclist Fatal Injury Risk by BAC Level Relative to Car/Van Drivers. Traffic Injury Prevention. Volume 14 Issue 1. 2013.

differences between the observed crashes that can affect crash severity. In this case, the result would only indicate if locations with a certain end terminal type are more likely to have a crash. Given the need to account for crash-involved variables and working from the assumption that an end treatment itself would not affect the likelihood of a crash occurring, an application of case-control methods is a bit more complicated. A question to be answered could be:

Given that a crash has occurred, what is the increased or decreased likelihood of a K or K+A severity crash for a given end treatment?

In this situation, cases would be end treatment crashes that resulted in a K (or KA) severity and controls would be end treatment crashes that resulted in a non-K (or KA) severity. One would then investigate if the type of end treatment is a significant risk factor for a crash resulting in a K (or KA) severity. In order to control for all of the potential confounding factors one would ideally adopt a matched case-control approach to the extent possible and rely on building statistical models to control for the remaining confounding factors.

To fully understand the effects of different end treatments on all possible crash severities would require several models with a range of definitions of cases and controls, e.g., K vs ABCO; KA vs BCO; KAB vs CO. However, given that the objective of roadside hardware is to reduce KA crashes, and the emphasis on these crashes in studies evaluating roadside safety treatments, this analysis could realistically be confined to KA vs BCO. As well, it seems important to not only look at crashes for all vehicle types combined but also at those for passenger cars only, the object of design for many roadside treatments.

4.2.2 Case Control Study Data Requirements

To conduct a case-control study it would be necessary to carefully define the question to be answered. As was just discussed, in order to fully understand the impacts of different end treatments, given a crash occurs, this may require several definitions of cases and controls, which complicates the analysis somewhat. Working under the assumption that the end treatments do not affect crash likelihood but may affect crash severity, all cases and controls would be locations where an end treatment crash has occurred.

There are many crash, vehicle, driver and roadway variables that may impact the severity of a crash. Sources of these data may include electronic crash data, hard copy crash reports, state roadway inventory data and manual data collection in the field or through video logs and satellite photography.

A study design would need to outline which variables to collect. This will not be done in this review but a sample of potential variables is provided below to illustrate. The intent is not to identify essential variables but simply to illustrate the types of variables that would be required for a case-control study.

Sample Crash Variables

- Most severe injury
- Number of vehicles involved
- Sequence of events

Sample Vehicle Variables

- Vehicle type (e.g. passenger car, light truck etc.)
- Vehicle age
- Vehicle speed

Sample Driver Variables

- Driver age
- Driver sex
- Blood alcohol content

Sample Roadway Variables

- Type of end treatment
- Posted speed
- Horizontal curvature
- Traffic volumes

Determining the number of cases and controls to assemble requires a sample size analysis to be conducted considering the desired level of statistical significance, power and an anticipated detectable difference in risk. An example of sample size determination in a study of roadway infrastructure safety is provided by Gross et al[21].

4.2.3 Methodologies for Case-Control Studies

The goal of a case-control study is to estimate the likelihood of an actual risk factor and is expressed as the odds ratio between two levels of a variable. A matched case-control design is desirable. The primary reason for a matched case-control design is to control for confounding variables. While statistical modeling can be used to control for confounding variables it is desirable to eliminate this need for as many variables as possible to avoid omitted variable bias and other complications associated with estimating such models. Confounding variables include those variables that completely or partially account for the apparent association between an outcome and risk factor. Specifically, a confounder is a variable that is a risk factor for the outcome under study, and is associated with, but not a consequence of, the risk factor in question.

Matching as a means of addressing confounding is accomplished during the selection of controls. Controls are selected so that each matched case-control pair has identical values for the confounding variables (or at least similar values based on a range). For example, continuous variables such as AADT may be categorized into increments of 500 vehicles per day, and controls are randomly selected from the AADT category that corresponds to the matched case. The main advantage of matching during the design stage is direct control of confounders.

In the limit, it may be possible to eliminate the complexity of modeling for a preliminary simplistic categorical analysis by state and local agency analysis by focusing more effort on

---

[21] Gross, F., Jovanis, P., Eccles, K. and K.Y. Chen. Safety Evaluation of Lane and Shoulder Width Combinations on Rural, Two-Lane, Undivided Roads. FHWA-HRT-09-031. June, 2009.

matching. However, that may require the pooling of data by similar states and an assumption (which seems entirely reasonable) that minor differences in design and application practices among states will not affect the outcome.

Refinements to the matching process for both the simplistic categorical analysis and modeling approaches can be accomplished with the application of more sophisticated optimal binning[22] of continuous variables such as AADT and curvature available in software packages such as R and SPSS. More sophistication can be added through propensity score matching[23]. The propensity score considers the conditional probability that a observed entity has the characteristic of interest based on its other characteristics and is commonly estimated using logistic regression models. This has been used in other fields[24], and also in road safety but not strictly for a case-control application. The strength of matching cases and controls based on the propensity score is this is a formal process for ensuring that the potential confounders are balanced between the cases and controls and relies on the predicted value. Attempting to match based on all of the individual variables included in the propensity score model can be very challenging.

For example, Sasidharan and Donnell (2014)[25] applied pair-wise matching based on propensity score when estimating the safety effects of lighting at intersections. The basic approach was to match each treated (lit) intersection with the untreated intersection that had a propensity score value closest to its own. The treatment assigned was presence of intersection lighting and the probability of a site having lighting modeled as a function of minor road stop control, speed limit, presence of depressed median, lanes on major road, indicator of a skewed intersection, indicator of a divided major road, area type, an indicator of absence of a left or right shoulder, major road AADT, minor road AADT.

The nature of the study design (i.e., matched case-control study) requires the analysis to account for the matching process between cases and controls. In a matched case-control design, conditional logistic regression may be used to investigate the relationship between the outcome and risk factor. It should be noted that in a matched design, the effect of the matching variables cannot be estimated because they are used as selection criteria when selecting cases and controls. However, the interaction between matching variables and risk factors may be analyzed (Gross et al., 2009).

Statistical analyses, such as multiple logistic regression techniques, are commonly used to clarify these relationships because they are able to examine the risk associated with one factor while controlling for other factors. Increasing the number of controls will increase the power of the study, especially when there are relatively few cases. Power is defined as the probability that the test will reject a false null hypothesis (Gross et al., 2009).

---

[22] Dougherty, J. Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In Proc. Twelfth International Conference on Machine Learning. Los Altos CA: Morgan Kaufmann, 194-202.

[23] Columbia University, Department of Economics Discussion Paper No. 0102-14. “Propensity Score Matching Methods For Non-Experimental Causal Studies” Rajeev H. Dehejia and Sadek Wahba, February 2002.

[24] Austin P.C. (2008) The performance of different propensity-score methods for estimating relative risks. Journal of Clinical Epidemiology. Volume 61(Issue6) pp. 537-545.

[25] Sasidharan L. and Eric T. Donnell (2014). Propensity scores-potential outcomes framework to incorporate severity probabilities in the Highway Safety Manual crash prediction algorithm. Accident Analysis and Prevention, Volume71, pp.183-193.

4.2.4. Wrap up on Case Control Studies

Case-control studies have traditionally been applied where events are rare, so at first glance that attribute would make this approach an appropriate choice for a safety evaluation of different guardrail end treatments. By selecting data based on outcome/non-outcome a higher percentage of cases is available compared to other studies, e.g. cross-sectional regression, where the unit under study is selected at random. On the other hand, case-control studies do not necessarily reduce data requirements if, by collecting targeted data on a smaller sample, more effort is devoted to collecting a large amount of variables required to account for confounding effects.

Recognizing these important characteristics of the case control methodology, the purpose of this section was to not only review the approach and to point out its merits, but to specifically assess the suitability of case-control studies for a safety evaluation of different guardrail end treatments, in accord with the terms of reference for the paper. The consideration of case-control studies for such an evaluation worked from the following assumption. The type of guardrail end treatment is not likely to affect whether a crash occurs. The chance of a vehicle leaving the intended travel path and striking the end terminal is unlikely to be affected by the end treatment. A caveat to the preceding statement is that if an end terminal treatment could make more crashes of such low consequence that they are unreported. In that case, a study methodology focused on predicting crash frequencies would be more appropriate. This eventuality seems unlikely, as argued earlier. The question for any study to answer is then: given that a crash has occurred is one type of end treatment more or less likely to result in a more severe crash? For this, a methodology focused on severity outcomes, given that a crash has occurred, is more appropriate.

In principle, the case-control approach could be applied to answer this question. To do so cases would be guardrail end terminal crashes with the severity type of interest and controls would be similar crashes of a lower severity. For example, cases are of KA severity while controls are of BCO severity. A case-control study could determine the increased or decreased likelihood of a crash being KA severity if the end treatment type was other than the baseline design. In order to fully understand the increased or decreased severity risk due to alternate end treatments a series of case-control models would ideally need to be developed since it would be of interest to see the effects for all severity categories, i.e. K, A, B, C and O. However, it is recognized that the primary focus could be restricted to KA severity. As well, it is important to not only look at crashes for all vehicle types combined but also at those for passenger cars only, the object of design for many roadside treatments.

Logistic modeling would be required to account for the many potentially confounding variables affecting crash severity that are not related to the end treatment type. A matched case-control design would ideally be used to eliminate as many confounding variables as possible prior to estimating the logistic model. In the limit, it may be possible to eliminate the complexity of modeling for preliminary categorical analysis by state and local agencies by focusing more effort on matching cases and controls. However, that may require the pooling of data by similar states, more intense focus on matching cases and controls and an assumption (which seems quite reasonable) that minor differences in design and application practices among states will not affect the outcome. These challenges can potentially be overcome by considering probability severity models.

4.3 Probability Severity Models

The analysis to fully understand the increased or decreased severity risk due to alternate end treatments could be more straightforward (at least for researchers) by estimating either ordered or unordered probability models of injury severity. Similar types of data used for case-control studies could be used for estimating the probability models. The challenge would be not only in working with small sample sizes, however, but also in properly accounting for all confounding factors. Whether the number of cases available would facilitate this is a complex task and would require further investigation, perhaps in the form of a pilot study. Such a study would consider the desired level of statistical significance, statistical power, likely detectable difference in risk, and the expected number of discordant pairs (i.e., case-control pairs with a different risk factor status).

Given appropriate data, the analysis is more straightforward than for the case-control methodology, per se, in the sense that a single model can be estimated and the increased or decreased likelihood of a severity of any class estimated by the model parameters.

Ordered models may be appropriate when the outcome variable is ordered and discrete, as is the severity types K, A, B, C and O, but not always. A discussion is provided in Washington et al.[26] Ordered probability models can be probit or logit models. Both models are similar but differ by the assumed error distribution of the error term, and both have inherent limitations as noted by Sasidaran and Menendez[27] in proposing a partial proportional odds (PPO) model to "bridge the gap between ordered and non-ordered severity modeling frameworks" (page 330). The PPO model, as they note, "allows the covariates that meet the proportional odds assumption to affect different crash severity levels with the same magnitude, whereas the covariates that do not meet the proportional odds assumption can have different effects on different severity levels" (page 330). Interestingly, their application was to pedestrian crashes, a relatively rare crash type.

There are numerous applications of these models in the road safety context, several specific to evaluating injury outcome probabilities for roadside features[7, 28, 29, 30, 31,32]. The definitive summary of these models, and of the methodological variations as applied in road safety analysis, is provided by Savolainen et al. (2011)[33]. And research is continuing on enhancing the methodology, suggesting that there is promise in using these models for in-service

[26] Washington, S., Karlaftis, M. and F. Mannering. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC Taylor and Francis Group. Boca Raton, Florida, 2011. pp358.

[27] Sasidharan L and M. Menéndez (2014). Partial proportional odds model—An alternate choice for analyzing pedestrian crash injury severities. Accident Analysis and Prevention, Volume 72, pp. 330 -340.

[28] Hu, W., & Donnell, E. (2010). Median barrier crash severity: Some new insights. Accident Analysis and Prevention, 42(6), 1697-1704.

[29] Holdridge, J. M., Shankar, V. N., & Ulfarsson, G. F. (2005). The crash severity impacts of fixed roadside objects. Journal of Safety Research, 36, 139-147.

[30] Zou Y., Tarko A., Chen E., and M. Romero. Effectiveness of cable barriers, guardrails, and concrete barrier walls in reducing the risk of injury. Accident Analysis and Prevention 72 (2014) 55–65.

[31] Johnson N. and H. Gabler. Injury Risk in Frontal Crashes with Guardrail and Guardrail End Terminals. 93rd Annual Meeting of the Transportation Research Board (2014)

[32] Lee J. and F. Mannering Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34 (2002) 149–161

[33] Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis and Prevention, 43(5), 1666-1676

performance evaluation of guardrail end-treatments, which tend to experience fewer crashes than some of the roadside features typically evaluated with these models.

Studies evaluating injury outcome probabilities for roadside features are of particular relevance to the paper. A sample of these studies, as referenced above, is summarized below. These summaries also provide a flavor for the methodological essentials for probability modeling.

Johnson and Gabler (2014)[31] and Johnson and Gabler (2015)[7]

The 2014 analysis examined the effect of a number of factors on injury risk in frontal guardrail crashes using 711 crash records extracted from the National Automotive Sampling System – Crashworthiness Data System from 1997 to 2008. Logistic regression on occupant injury level (and other nominal factors of interest) was performed, but no specific details on the methodology were provided. Among the key relevant findings (extracted verbatim from the conclusions) were:

- LTVs were observed to have odds of rollover 7.4 times higher than cars in guardrail crashes.
- There is significant evidence that frontal crashes to end terminals carry inherent injury odds 5.1 times greater than frontal crashes to the guardrail face.
- Odds of injury in frontal end terminal crashes appear to be between 11 and 19 times lower when the terminal design is compliant with NCHRP 350, compared to non-compliant designs.
- High-risk cofactors are associated with 25 % of all frontal guardrail crashes, yet are associated with 61 % of those resulting in serious injury.
- Odds of an injury crash are significantly elevated when one or more high-risk cofactors are present and this is observed to mask the effect of other factors.

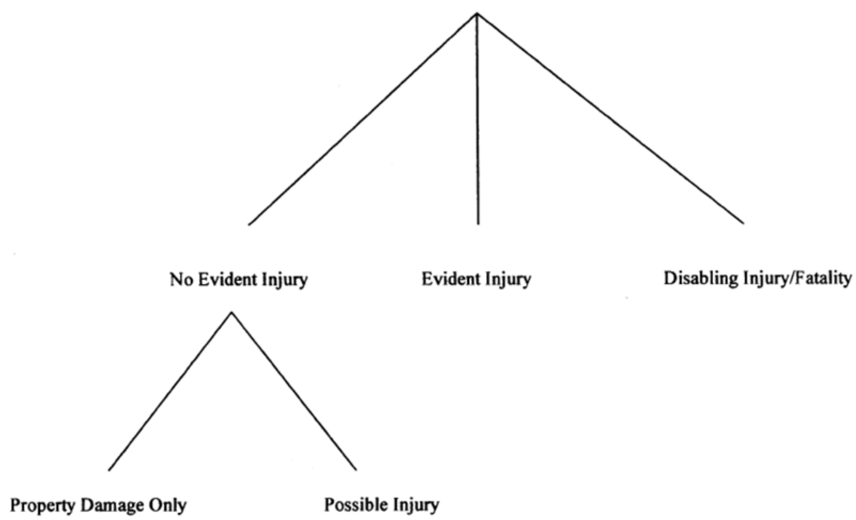
The results suggest that the methodology does seem feasible for conducting ISPEs in general, even with limited data. However, the last two bullets emphasize the importance of properly accounting for confounders in using this or any methodology for ISPEs.

In the later study[7], the same authors used a sample of 451 end terminal crashes from the 2011–2012 Michigan State crash data. Multiple logistic regression was used to test for significant differences in the odds of driver injury and rollover between different terminal types while accounting for other factors such as driver seat belt use, rollover occurrence, terminal orientation (leading/trailing), and control loss. Other confounders were controlled for using subsets of the sample containing crashes on roads of the same access control level, posted speed limit, and road class. Among the findings were:

- No significant difference in moderate to fatal driver injury odds was observed between NCHRP 350 compliant end terminals and noncompliant terminals.
- Car drivers showed odds of moderate to fatal injury 3.6 times greater than LTV drivers in end terminal crashes.
- Rollover occurrence was not significantly associated with end terminal type.
- Rollover and non-use of seat belts carried much larger increases in injury potential than end terminal.

Lee and Mannering (2002)[32]

Using nested logit models and Washington data for two-lane rural roads, statistical models of crash severity were estimated and the findings used isolate a wide range of factors that significantly influence the severity of run-off-roadway crashes. The nested structure, which is shown in the Figure below taken from the paper, is said by the authors to group alternatives with correlated error terms into a nest by estimating a model that includes only crashes with outcomes in the nested group.



The most interesting result of relevance to this paper relates to the effect on severe crashes of the presence of guardrails. In finessing the analysis, the authors note that the presence of guardrails could be influenced by the severity of crashes in those road sections with a history of severe crashes and thus may be more likely to have guardrails installed, creating an endogeneity bias that may result in an incorrect conclusion that guardrails can *cause* severe crashes. To account for this, an instrumental variable was used in model estimation whereby a binary logit model that was estimated using only exogenous independent variables. These estimated probabilities from this model of a guardrail being present were used in place of the one-zero guardrail indicator variable when estimating the model. Even after accounting for this bias, the presence of guardrails in a road section was found to increase the chance of the crash being a disabling injury/fatality. The authors note that "this does not necessarily bring into question the effectiveness of guardrails, but may simply be reflecting the physical complexity of the geography along certain roadway sections" (page 159). This again emphasizes the need for accounting for confounding factors in any ISPE methodology.

Holdridge et al. (2005)[29]

The study evaluated the in-service performance of roadside hardware on the entire urban State Route system in Washington by developing multivariate nested logit models of injury severity in fixed-object crashes, similar to Lee and Mannering (2002). The results show that when controlling for roadway, vehicle, and driver characteristics, several roadside features were found to have significant impacts on crash severities as indicates in the Figure below, taken from the paper (page 146).

The effect of roadside objects on propensities toward injury severities

Object struck	Non-injury	Evident injury	Disabling injury	Fatality
Wood or metal sign post or guide post	↑	↓	↓	↓
Roadway ditch, culvert end, or other appurtenance in ditch	↑	↓	↓	↓
Guardrail or bridge rail Leading End	↓	↓	↓	↑
Guardrail face	↑	↓	↑	↑
Concrete barrier	↑	↓	↓	↑
Rock bank or ledge	↓	↑	↓	↓
Roadway or construction machinery struck	↑	↓	↑	↑
Tree or stump, pole (light, utility, railway, traffic, overhead), or sign box	↓	↓	↓	↑

This Figure indicates, for example, that leading ends of guardrails and bridge rails, along with large wooden poles (e.g. trees and utility poles) increase the probability of fatal injury. Given the contribution of guardrail leading ends toward fatal injuries, the authors suggest that it is important to use well-designed leading ends and to upgrade badly performing leading ends on guardrails and bridges.

Key take-aways from the paper include the importance of accounting for confounding for driver, roadway and vehicle variable in ISPEs. For example, when considering vehicle types, pickup trucks and delivery trucks were shown to be associated with a decrease in the propensity toward non-injury; and when considering road characteristics, it was found that, as speed limits increase, the propensity toward non-injury decreases. Another key take-away is the evident potential for using the methodology for estimating the relative performance of various guardrail end treatments given enough data to create separate variables for each such treatment.

Hu and Donnell (2010)[28]

A nested logit model was estimated with North Carolina data to analyze the effects of median, roadway, driver, environmental, and vehicle factors on four median barrier crash severity outcomes, with K and A categories combined. The authors note that a unique aspect of the data used to estimate the model was the availability of median barrier placement and median cross-slope data, two elements not commonly included in roadway inventory data files. The estimation results indicate that collisions with a cable median barrier increase the probability of less-severe crash outcomes relative to collisions with a concrete or guardrail median barrier; and increasing the median barrier offset was associated with a lower probability of severe crash outcomes. The presence of a cable median barrier installed on foreslopes that were between 6:1 and 10:1 were associated with an increase in severe crash probabilities when compared to cable median barrier installations on foreslopes that were 10:1or flatter.

The key take-aways are similar to those from the Holdridge et al. paper[29]. In addition the importance of capturing and modeling data elements such as placement and slope would suggest that the inclusion of such variables in ISPEs for guardrail end-treatments would enhance the credibility of the results.

Zou et al. (2014)[27]
A binary logistic regression model with mixed effects was estimated for vehicle occupants injuries based on 517 pair-matched homogeneous barrier and non-barrier segments. The segment pairing process and the use of mixed random effects was done to "handle the commonality within the same segment pair as well as the heterogeneity across segment pairs" (page 55). Many results were obtained and of especial relevance to conducting an ISPE for guardrail end-treatment was the ability to compare different types of barriers where some types of barriers were used alternatively. That analysis found that the odds of injury are 43% lower when striking a guardrail instead of a median concrete barrier offset 15–18 ft and 65% lower when striking a median concrete barrier offset 7–14 ft.

5. SUMMARY

The paper reviewed methods for ISPE evaluations grouped into following categories and sub-categories, identified by sections of the paper:

3. Crash Frequency Methods
4. Methods to Assess Risk Given that a Crash has Occurred
   - 4.1 Simple Proportions
     - 4.1.1 Risk of a Serious Crash with Simple Proportions
     - 4.1.2 Odds Ratios with Simple Proportions
   - 4.2 Case-Control Studies
   - 4.3 Probability Severity Models

Some key considerations were revealed in the review that would be pertinent to the application of all of the methods. These are:

- It is important to control for important confounders such as speed, road class and vehicle type. The latter is especially important since terminals are typically designed for passenger vehicles, so conclusions should be formed on passenger vehicles only.
- The capturing and modeling of data elements such as placement and slope in ISPEs for guardrail end-treatments may enhance the credibility of the results.
- If the presence of a specific end terminal could be influenced by the severity of crashes in those road sections with a history of severe crashes, and thus may be more likely to have that end terminal installed, an endogeneity bias may result in an incorrect conclusion that the end terminal can *cause* severe crashes.
- ISPE's should be performed with a view not only to better understand the performance of roadside hardware, but also to improve the crash testing specifications used for new hardware as well as to improve design guidelines used for warranting and placing such hardware.

The Table below summarizes the essential elements of these methods vis-à-vis data needs and suitability of application at various levels of expertise.

Method and reference for application to ISPE	**Level of Sophistication /Ease of Application /Expertise Level**	**Data needs**	**Suitability for comparative guardrail end-treatment ISPE**	**Remarks**
Crash Frequency (4, 9)	Low/Easy/State	Crashes by severity; Exposure (AADT); Variables for applying crash modification factors (ideally)	OK for initial indication where exposure is similar and/or difference in performance is large and/or key confounders can be controlled for	Ideally control for confounders through crash modification factors;
Simple Proportions – Risk (1, 14, 15)	Low/Easy/State	Crashes by severity; information on key confounders, e.g., speed limits, vehicle type	OK for initial indication that further study is warranted especially if key confounders can be controlled for	Ideally include unreported crashes
Simple Proportions – Odds ratios (11, 14, 15)	Low/Easy/State	Crashes by severity; information on key confounders, e.g., speed limits, vehicle type	OK for initial indication that further study is warranted especially if key confounders can be controlled for	Ideally include unreported crashes; same conclusions as "Risk"
Case-Control	Medium-High/Moderately Difficult/State Researcher	Crashes by severity for cases; information on confounders for both cases and controls. See Sec. 4.2.2	OK if cases and controls can be closely matched, which could be a challenge.	Probability models can account for unmatched confounder variables.
Probability Severity (7, 27, 28, 29, 30, 31)	High/Difficult/Researcher with Modeling Skills	Crashes by severity; information on all possible confounders	Most suitable of all methods, but data hungry	Applications for roadside treatments show promise; substantial research on improving modeling techniques for road safety data.

ACKNOWLEDGMENT

The author wishes to acknowledge the contribution of Craig Lyon to writing this paper, as well Malcolm Ray, Clay Gabler and Christine Carrigan, whose brains I picked from time to time. The guidance of the oversight committee was invaluable and is gratefully acknowledged.