

The Relationship between Driver Acceptance and System Effectiveness in Car-Based Collision Warning Systems: Evidence of an Overreliance Effect in Older Drivers?

Scott Stevens
Volpe Center, U.S. DOT

ABSTRACT

An understanding of the relationship between user acceptance and system effectiveness is necessary for designing collision warning systems for cars and for accurately estimating their safety benefits. To better understand this relationship, an exploration was conducted into the results of a recent naturalistic field operational test of an integrated vehicle-based safety system installed into a fleet of passenger vehicles. Correlation analyses were conducted for system effectiveness, which was measured in terms of near-crash reduction, and for user acceptance, which was measured using responses to a questionnaire that drivers filled out after completing the study. The key findings were as follows: (1) although remaining generally positive overall, user acceptance decreased slightly for drivers as system effectiveness increased. This pattern appears to be due primarily to older drivers (aged 60 to 70), for whom it was strongest; (2) older drivers rated the system higher and were more likely than younger drivers to want more alerts; (3) a possible contributing factor to the above results is false alerts, as older drivers experienced a smaller proportion of false alerts than younger drivers. These results are discussed in the context of existing evidence and recommendations are made for future experiments and safety benefits estimates.

CITATION: Stevens, S., "The Relationship between Driver Acceptance and System Effectiveness in Car-Based Collision Warning Systems: Evidence of an Overreliance Effect in Older Drivers?," *SAE Int. J. Passeng. Cars - Electron. Electr. Syst.* 5(1):2012, doi:10.4271/2012-01-0282.

INTRODUCTION

Designing effective automotive collision warning systems and accurately estimating their potential safety benefits requires an understanding of "user acceptance," i.e., the attitudes that the driver has towards the system in terms of perceived usefulness, reliability, intelligibility and annoyance. These attitudes can have a profound impact on system effectiveness. For example, if drivers have a low opinion of a warning system, perhaps finding it unreliable or annoying, they may ignore warnings or disable the system. On the other hand, drivers may have too high of an opinion of the system if, failing to understand its limitations, they rely on it to alert them of dangers and as a consequence pay less attention to the road [1,2]. A clear understanding of this relationship between user acceptance and system effectiveness is difficult to achieve for numerous reasons: evidence comes from diverse sources, including simulator studies and more naturalistic field tests (in which drivers use test cars in their everyday lives) and drivers may behave differently in each; different studies test different warning systems with different interfaces and different false alert

rates, and the results from one system may not always generalize to all; many studies involve small sample sizes, which may produce spurious results; and, finally, the immense complexity of human psychology results in large amounts of between-subject variation, including effects of age, gender, and experience. When combined with small sample sizes, the high levels of variation can make it difficult to identify important patterns in psychological data of this kind using statistical analyses. Consequently, there is a continuing need for large amounts of real-world data [3]. While there have been large-scale naturalistic driving studies [4,5], few have focused on driver acceptance of warning systems, particularly as it relates to system effectiveness.

In the current study, we take advantage of a recent assessment of an in-vehicle warning system called the Integrated Vehicle-Based Safety System (IVBSS), which provided several different types of crash-imminent warnings to drivers [6]. The assessment consisted of a naturalistic field operational test, in which 108 volunteers in southeast Michigan used IVBSS-equipped cars in their daily routines for 40 days each. In total, the volunteers drove more than

219,000 vehicle miles and data were gathered on over 22,700 alerts. This paper presents the results of a *post hoc* exploration of the IVBSS data set in which we looked for correlations between measures of user acceptance and system effectiveness. Following a description of the methodology, the results are presented and then discussed in terms of other studies and in terms of their relevance to future estimates of system effectiveness.

METHODS

THE IVBSS STUDY

A detailed description of the IVBSS study has been published by the National Highway Traffic Safety Administration (NHTSA) [6]. In short, drivers were provided with a car equipped with the following four warning systems:

1. Forward-collision warnings to prevent rear-end collisions
2. Lane-change warnings to alert drivers of objects in their blind spots
3. Drift warnings to help drivers stay in their lanes
4. Curve-speed warnings to alert drivers entering sharp curves at excessive speed

Warnings were provided in the form of audible alerts and, in some cases, additional messages on a visual display or via haptic feedback (brake pulses and seat vibrations). Volunteers were recruited from three age groups (20-30, 40-50, and 60-70 years old) and were instructed to use the cars for their everyday normal use. For the first 12 days, the system warnings were suppressed in order to generate a baseline of driver behavior, and then the warning system was enabled for a treatment period of 28 days. At the conclusion of the study, drivers filled out questionnaires regarding their experience with the system. The resulting data set consists of both the questionnaire data, which were used to estimate user acceptance, and extensive numerical data regarding the behavior of the car (recorded using over 200 channels), which were used to estimate system effectiveness.

MEASURES OF SYSTEM EFFECTIVENESS

Since actual collisions were extremely rare, system effectiveness (E) was characterized in terms of the percent change in “near crashes” (NC) from the baseline to the treatment period. NC events were defined as situations in which the driver had to perform a maneuver (hard braking or steering) of a certain intensity in order to avoid a crash [7]. NC rates for the baseline (NC_B) and treatment (NC_T) periods were calculated by dividing the number of near crashes that occurred during that period by the number of vehicle miles travelled (VMT) during that period. E was then calculated using the following equations for every driver who experienced at least one near-crash event in both the baseline and treatment conditions:

$$E = 1 - \frac{NC_T}{NC_B}; \text{ if } NC_T \leq NC_B \quad (1)$$

$$E = -\left(1 - \frac{NC_B}{NC_T}\right); \text{ if } NC_T \geq NC_B \quad (2)$$

When the NC rate decreased from the baseline to the treatment period Equation (1) was used, and when it increased Equation (2) was used. E values were then converted into percentages.

Since IVBSS was intended to warn drivers who are not paying attention, the system was designed to suppress alerts deemed unnecessary because the driver was already in the process of reacting (e.g., braking). This makes sense from a safety perspective since unnecessary alerts can degrade user acceptance of the system in the long run by making it more annoying. Of course, a consequence of this alert suppression is that in many near-crash scenarios the system did not issue a warning. From a psychological perspective, drivers may be less aware of the system when it remains silent during an emergency. Since this investigation concerns driver perceptions of the system, a measure of the frequency of events with high impact on driver experience would be useful. Consequently, in addition to calculating system effectiveness using all near crashes (E_{all}), a second measure of system effectiveness was calculated using only those near crashes that occurred within 10 seconds before or 15 seconds after non-suppressed alerts ($E_{alerted}$). Since these “alerted near crashes” only involve events where the driver was not taking any action immediately beforehand, they are more likely to include a higher proportion of scenarios where the driver was taken by surprise. As a result of this surprise factor, alerted near crashes may also be more memorable to drivers and consequently play a large role in drivers' acceptance of the system.

MEASURES OF USER ACCEPTANCE

Drivers' acceptance of the system was quantified using their responses to questionnaires filled out after their experience with the IVBSS-equipped cars was over. These questionnaires were comprised of statements and questions next to which were semantic differential scales on which they could circle their responses. Out of a total of 52 questionnaire items, five were selected *a priori* for individual analysis as being the most indicative of driver acceptance (see questions 1-5 in Table 1). Additionally, in order to summarize a driver's general acceptance of the system, a composite index, termed “usefulness,” was created by averaging the answers to the aforementioned five questions with eight additional questions (questions 6-13 in Table 1). In addition to creating a more comprehensive measure, this composite index serves to reduce within-subject variation, including any effect of inconsistent answers.

Table 1. Questionnaire items used to calculate user acceptance. Individual analyses were conducted on questions 1 to 5 as well as on “usefulness,” which was itself a composite of questions 1 to 13.

1	How helpful were the integrated system’s warnings? Not at all helpful 1 2 3 4 5 6 7 Very helpful
2	Overall, I think that the integrated system is going to increase my driving safety. Strongly disagree 1 2 3 4 5 6 7 Strongly agree
3	Driving with the integrated system made me more aware of traffic around me and the position of my car in my lane. Strongly disagree 1 2 3 4 5 6 7 Strongly agree
4	The integrated system made driving easier. Strongly disagree 1 2 3 4 5 6 7 Strongly agree
5	I always understood why the integrated system provided me with a warning. Strongly disagree 1 2 3 4 5 6 7 Strongly agree
6	Overall, I felt that the integrated system was predictable and consistent. Strongly disagree 1 2 3 4 5 6 7 Strongly agree
7	Overall, how satisfied were you with the integrated system? Very dissatisfied 1 2 3 4 5 6 7 Very satisfied
8	Overall, I received warnings... Too frequently 1 2 3 4 5 6 7 Too infrequently
9	I always knew what to do when the integrated system provided me with a warning. Strongly disagree 1 2 3 4 5 6 7 Strongly agree
10	The integrated system gave me warnings when I did not need them (i.e., nuisance warnings). Strongly disagree 1 2 3 4 5 6 7 Strongly agree
11	Overall, I received nuisance warnings... Too frequently 1 2 3 4 5 6 7 Never
12	The integrated system warnings were... Useful 1 2 3 4 5 Useless
13	The integrated system warnings were... Assisting 1 2 3 4 5 Worthless
	Usefulness (composite score of Questions 1-13) Useless 1 2 3 4 5 6 7 Useful

Not all questionnaire answers were made on a seven-point scale where one was most critical and seven most favorable. These exceptions (questions 8, 10, 12 and 13) had therefore to be converted into an equivalent scale before they could be averaged with the rest. Also, to avoid giving disproportionate influence in the composite score to types of questions appearing more than once, identical questions were averaged with each other before being averaged with the rest; this occurred for questions 10 and 11 as well as for questions 12 and 13.¹

FALSE ALERT RATES

As part of the IVBSS study, video analysis was used to identify false alerts, i.e., alerts issued when there was no corresponding threat visible [6]. In the current study, the false alert rate is the percentage of total alerts that was false.

DATA ANALYSIS

Parametric statistical tests were used except where the assumptions of normality were not met or where non-interval data were involved (including all comparisons using data from the questionnaire). In those cases, equivalent non-parametric tests were used instead. All analyses were conducted using SPSS 19 for Windows and were two tailed. Statistical significance was set at an alpha of 0.05. Correlational coefficients (r in the case of Pearson's analyses, r_s in the case of Spearman's analyses) were considered to be weak when they were between -0.3 and 0.3 ; moderate when they were between -0.3 and -0.6 or 0.3 and 0.6 ; and strong when they were between -0.6 and -1 or 0.6 and 1 .

RESULTS

The average system effectiveness (E_{all}) was an overall reduction in near crashes by 1.4%,² for which the 95% confidence interval (CI) was 9%. In terms of the subset of alerted near crashes ($E_{alerted}$), the reduction was 24.7% (CI = 13%). Driver acceptance of the system was generally positive: on a scale of one to seven, the median scores overall were all above four (Table 2).

Table 2. Median questionnaire responses (scale of one to seven for least to most favorable) with interquartile ranges in parentheses. Refer to Table 1 for question wording.

Question	Overall	By age group		
		Young	Midaged	Older
1 (helpful)	6 (2.0)	5 (1.0)	5.5 (1.0)	6 (2.0)
2 (safety)	6 (2.0)	5 (2.0)	6 (1.0)	6.5 (1.0)
3 (aware)	7 (2.0)	6 (2.0)	6 (1.8)	7 (1.0)
4 (easier)	5 (2.0)	4 (3.0)	5 (1.8)	6 (2.0)
5 (understood)	5 (3.5)	5 (2.0)	5 (3.0)	6 (3.0)
Usefulness	5.5 (1.4)	5.5 (1.5)	5.4 (0.9)	5.8 (1.1)

¹ Questions 12 and 13, being part of a section written to create a van der Laan score, were in fact intended to ask the same thing.

² In the NHTSA report [6], the overall system effectiveness is given as ranging from 6% to 29%. The reason for this discrepancy with the 1.4% used in this study is that in the NHTSA report crash types were omitted when they did not show a statistically significant change from baseline to treatment. This was appropriate given that the aim in that case was to produce a safety-benefits estimate. In the case of this study, though, a characterization of driver experience was needed that aligned with their feedback on the questionnaire, and so all crash types were included because they presumably all contributed to drivers' ratings of user acceptance.

USER ACCEPTANCE AND SYSTEM EFFECTIVENESS

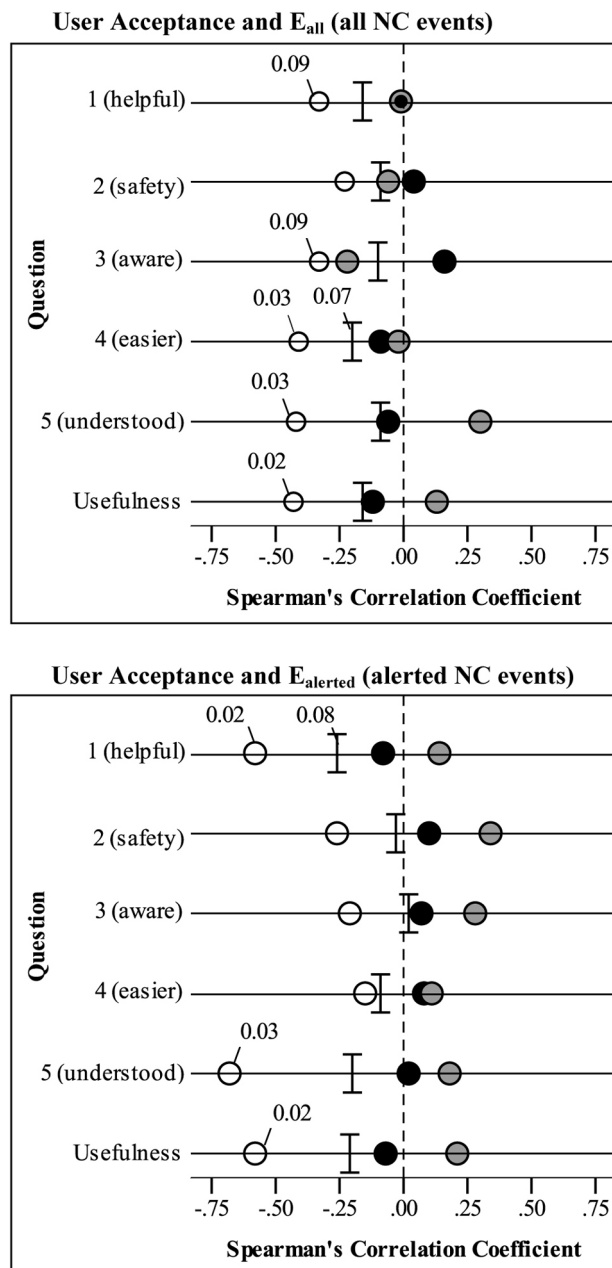


Figure 1. Correlations between system effectiveness and user acceptance. Correlations are labeled with p -values when they were under 0.1; all points not labeled with a p -value were statistically non-significant. Black circles indicate younger drivers, gray circles indicate middle-aged drivers, white circles indicate older drivers, and vertical "I"-shaped brackets indicate all age groups combined. For E_{all} , $n = 91$ (33 young, 30 middle aged, and 28 older), for $E_{alerted}$, $n = 47$ (16 young, 16 middle aged, and 15 older).

Although overall the system appeared to decrease near-crash rates and was rated with general approval, most correlations between system effectiveness and user acceptance were slightly negative. In other words, drivers whose near-crash rates increased tended to rate the system more favorably than those whose near-crash rates decreased. This was true for all five questionnaire items as well as for the combined index, "usefulness" (as indicated in Figure 1 by the vertical "I"-shaped brackets, which includes all ages combined, being to the left of center). Approval of the system was slightly more strongly correlated with the subset of near crashes that were alerted than with all near crashes combined. However, although there were a couple trends (p -values greater than 0.05 but less than 0.1), overall these correlations for all ages combined were weak and not statistically significant.

Another way to check this relationship between system effectiveness and user acceptance is to take the top and bottom 20% of drivers in terms of system effectiveness and to compare their ratings of the system. As with the correlational analyses above, when system effectiveness is characterized using all near crashes there were no significant differences. When just alerted near crashes are used, though, the lowest 20% gave a significantly higher median score than the top 20% for "helpfulness" (question 1) and "usefulness" (Table 3).

Table 3. Comparisons of user acceptance ratings between the top and bottom 20% of drivers in terms of system effectiveness (where system effectiveness is calculated using only alerted near crashes, $E_{alerted}$)

Question	Medians		Mann-Whitney U statistic and p -value
	Bottom 20%	Top 20%	
1 (helpful)	7	5.5	$U = 20.0, p = 0.02$
2 (safety)	6	6	$U = 40.0, p = 0.43$
3 (aware)	7	6	$U = 41.5, p = 0.48$
4 (easier)	5	5	$U = 40.5, p = 0.46$
5 (understood)	6.5	4.5	$U = 28.0, p = 0.08$
Usefulness	6.1	5.3	$U = 22.5, p = 0.04$

Note for the above tests $n_1 = 10$ and $n_2 = 10$

TESTS BY AGE

A series of analyses were conducted to explore the role that age may play in the above correlations between system effectiveness and user acceptance. First of all, when age groups are considered separately, the correlations became stronger, but only among older drivers. Thus, the overall negative correlations appear to be largely due to the older drivers (see the black, grey and white circles in Figure 1). The strongest of these correlations for older drivers were "understandability" (question 5) and "usefulness," both of which were statistically significant with E_{all} values as well as with $E_{alerted}$ values (graphed in Figure 2). Also significant was the negative correlation between E_{all} and ratings of