U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**

NHTSA
www.nhtsa.gov

DOT HS 811 470a

August 2011

Crash Warning Interface Metrics

Final Report

Technical Report Documentation Page

1. Report No. DOT HS 811 470a	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Crash Warning Interface Metrics: Final Report		5. Report Date August 2011
		6. Performing Organization Code
7. Author(s): Neil Lerner, James Jenness, Emanuel Robinson, Timothy Brown, Carryl Baldwin, & Robert Llaneras		8. Performing Organization Report No.
9. Performing Organization Name and Address Westat 1600 Research Boulevard Rockville, MD 20850		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The Crash Warning Interface Metrics (CWIM) project addressed issues of the driver-vehicle interface (DVI) for Advanced Crash Warning Systems (ACWS). The focus was on identifying the effects of certain warning system features (e.g., warning modality) and on methods and metrics that may be generally applied for evaluating DVIs in different vehicles. The project did *not* have the goal of proposing standard interfaces for particular warning functions. However, it did consider where there may be issues related to DVI variability, since even systems that test adequately by themselves may suffer problems in actual application, because users face problems due to the differences in DVI among vehicles. The project included both analytical activities and five experiments. The outcome of these efforts led to discussions of ACWS display characteristics and evaluation methods. Considerations for DVI evaluation methods included driving scenarios, research participant characteristics, pre-familiarization with the warning system, the distraction task, the participant's task and associated expectancies, accommodating user settings and options, the use of comparison benchmarks, and issues in the treatment of data.

17. Key Words Driver-vehicle interface (DVI), advanced crash warning system (ACWS), forward collision warning (FCW), lane departure warning (LDW)	18. Distribution Statement Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 74	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Acknowledgements

The authors thank Eric Traube, the NHTSA Task Order Manager for this project, and Dr. Eric Nadler with the Volpe National Transportation Systems Center, for their guidance and insight throughout the project. NHTSA's Vehicle Research and Test Center (VRTC) conducted test track research closely related to portions of the effort within the present project, though conducted under separate contract. The authors thank Garrick Forkenbrock and his colleagues at VRTC for assistance in our research and for their insights and opinions in interpreting the findings.

The Crash Warning Interface Metrics project involved a collaborative effort among Westat, The University of Iowa, George Mason University, and the Virginia Tech Transportation Institute. Numerous individuals at each of these institutions made substantial contributions to the project. Given the large number of contributors, only the primary researchers from each institution are formally included as Final Report authors. This does not diminish the roles of other individuals on the various project tasks. We specifically acknowledge the efforts of the following:

Westat: Jeremiah Singer, Richard Huey, Michael Gill, Jeremy Walrath, Shawn McCloskey, Diane Snow

The University of Iowa: Nicole Hollopeter, John Lee (now with the University of Wisconsin), Chris Schwarz, Daniel McGehee, Omar Ahmad, Cheryl Roe.

George Mason University: Chris Monk (now with the Federal Highway Administration), David Kidd, Daniel Roberts, Erik Nelson, Andre Garcia.

Virginia Tech Transportation Institute: Justin Owens.

Table of Contents

List of Figures

List of Tables

Executive Summary

The Crash Warning Interface Metrics (CWIM) project addressed issues of the driver-vehicle interface (DVI) for Advanced Crash Warning Systems (ACWS). ACWS have the potential to improve driver performance and reduce the frequency and severity of common crash situations. ACWS are systems that use sensors to assess potential or emerging hazard situations and provide crash warning information to the driver. In some cases the system may also initiate some vehicle control action. Examples of ACWS include forward collision warning (FCW) systems and lane departure warning (LDW). ACWS are increasingly common in passenger vehicles and the characteristics of these systems vary considerably among vehicle manufacturers.

The success of any ACWS will depend in part on the properties of the DVI. The DVI refers to the displays and controls through which the driver and the system interact. Display attributes include the warning mode (e.g., visual, auditory, haptic), warning display content, signal conspicuity, display location, timing, reliability, active intervention in vehicle control aspects, and relation to (and integration with) other systems and displays. System controls may allow the driver to activate or deactivate the system or select certain operating characteristics. The DVI also includes information about the status of warning systems, so that the driver comprehends what safety functions are present in the vehicle and what their current operational status is. The DVI design issues are complex and various products providing similar functions may perform quite differently. If ACWS are to fulfill their potential for improving safety, it is essential to have an effective DVI.

The focus of the CWIM project was on identifying the effects of certain warning system features (e.g., warning modality) and considering methods and metrics for evaluating the interfaces in different vehicles. The evaluation methods and metrics can be applied to whatever specific interface a given vehicle uses for a particular warning function, such as FCW. The project did *not* have the goal of proposing a standard interface for that function. However, it did consider where DVI variability between vehicles might limit usability or effectiveness, since even systems that test adequately by themselves may suffer problems in actual application if some users have incompatible expectations.

The project included both analytical and empirical activities. Analytical activities included review of research and standards literature, assessment of crash scenarios, the development of a taxonomy of potential DVI evaluation measures, the acquisition of stakeholder feedback (including both automotive industry and government input), and analysis of methods and results from a major Department of Transportation program (Advanced Crash Avoidance Technologies, or ACAT). The empirical portion consisted of five experiments (data collection for two of these experiments was conducted under independent contracts, but planning of the research and interpretation of the findings was done in collaboration with this project). Three of the experiments compared a variety of active and passive warning modes for LDW or FCW applications. These included two experiments in the National Advanced Driving Simulator (NADS-1) at the University of Iowa and one test track study at the NHTSA Vehicle Research and Test Center site. Another experiment was a simulator study on the potential negative transfer effects of auditory FCW alerts that may come about when signals differ from vehicle to vehicle. The final experiment was an assessment of the ability of people to comprehend vehicle status displays associated with ACWS.

The outcome of these efforts led to a discussion of human factors considerations for ACWS displays and a discussion of evaluation methods for assessing the DVI for FCW or LDW applications. Problems of negative transfer and user difficulties in comprehension of displays indicated that a lack of common, well-understood DVI features may limit driver comprehension and performance. The discussion and recommendations for CWIM assessment methodologies included consideration of dealing with warning system context for the DVI, accommodating user settings and options, driving scenarios, research participant characteristics, pre-familiarization with the warning system, the distraction task, the participant's task and associated expectancies, the use of comparison benchmark conditions, and issues in the treatment of data.

1 CWIM Overview and Objectives

1.1 Background

This report summarizes the methods and findings of the project "Crash Warning Interface Metrics (CWIM) Phase 2." This project is part of a broader NHTSA CWIM program that evaluates alternative driver-vehicle interfaces (DVIs) for crash avoidance functions and methods for evaluating such interfaces. The project continues work begun under a predecessor project (titled "Development of Driver Performance Metrics for Advanced Collision Prevention Systems"). References to the "project" in this document refer to the Phase 2 project, although a number of the review and analytic activities were begun under the predecessor project. It is anticipated that NHTSA will continue work under the CWIM program with additional projects.

The project examined the potential beneficial aspects of and concerns with Advanced Crash Warning Systems (ACWS), with a particular focus on the driver-vehicle interface (DVI). Examples of ACWS include forward collision warning (FCW), lane departure warning (LDW), lane change/blind spot warnings, adaptive cruise control, and curve speed warnings. ACWS have the potential to improve driver performance and reduce the frequency and severity of common crash situations. They use sensors to assess potential or emerging hazard situations and provide crash warning information to the driver. In some cases the system may also initiate some vehicle control action.

ACWS are increasingly common in passenger vehicles and the characteristics of these systems vary considerably among vehicle manufacturers. The magnitude of actual safety benefits of various ACWS in production vehicles have yet to be established. It is not known how effectively the systems will result in drivers taking appropriate corrective actions in potential crash events. Furthermore, systems conceivably may generate new problems, such as driver confusion, inappropriate responses, distraction, automation complacency, or poor user acceptance. Therefore it would be valuable to have the capability to evaluate and predict the safety consequences of ACWS.

The focus of the project was on the effects of certain warning system features (e.g., warning modality) and on common methods and metrics that may be generally applied for evaluating the interfaces in different vehicles. The evaluation methods and metrics would be intended to apply to whatever specific interface a given vehicle uses for a particular warning function, such as FCW. The project did *not* have the goal of proposing a standard interface for that function. However, it did consider where there may be benefits of common DVI features across vehicles, since even systems that test adequately by themselves may suffer problems in actual application, because users face problems due to the variability in DVI among vehicles.

1.1.1 DVI Evaluation

The success of any ACWS will depend in part on the properties of the DVI. The DVI refers to the displays and controls through which the driver and the system interact. Display attributes include the warning mode (e.g., visual, auditory, haptic), warning display content, signal conspicuity, display location, timing, reliability, active intervention in vehicle control aspects, and relation to (and integration with) other systems and displays. System controls may allow the driver to activate or deactivate the system or select certain operating characteristics. The DVI also includes information about the status of warning systems, so that the driver comprehends what safety functions are present in the vehicle and what their current operational status is. The DVI design issues are complex and various products providing similar functions may perform quite differently. If ACWS are to fulfill their full potential for improving safety, it is important to have an effective DVI.

While it is recognized that it is important to have effective DVIs for ACWS functions, a consensus means of *evaluating* a given system does not exist. The field lacks a valid, practical, consensus method for determining the efficacy of a DVI for a particular ACWS application. A set of specific research methods, dependent measures, and analysis methods is required in order to provide valid, reliable, and repeatable assessments. Such a consensus set of methods is what is meant by Crash Warning Interface Metrics (CWIM). The CWIM considered in this project are directed at the evaluation of operational (commercial or prototype) ACWS, rather than as techniques to be used in earlier design stages. The metrics might be applied in various ways, such as evaluating the performance of the ACWS DVI (quantitatively and/or against established criteria), comparing the performance of alternative systems, providing a basis for consumer information (for example, the type of information useful for the New Car Assessment Program, NCAP), or supporting regulatory or safety actions.

The process of developing and establishing consensus for CWIM is complex for a number of reasons. Since various manufacturers may use different modalities and display types, a common metric must be able to encompass any type of interface. Since a particular ACWS may be integrated as part of a system of warnings, the method must have a reasonable means of testing a given function in isolation, without penalizing the system by removing important context. Not all nominally similar safety functions operate in the same manner; for example, some warnings may only operate within certain speed thresholds. Some vehicles may provide advance information or alerts, prior to the situation in which the actual crash warning occurs; the means of incorporating this aspect into a test protocol is not obvious. Some ACWS include limited active intervention in some aspect of vehicle control (e.g., partial braking, counter-steering). This complicates the use of vehicle control or driving outcome measures as indices of the effectiveness of the DVI. Any evaluation method will have to specify the driving scenario(s) in which the warning occurs, yet the relative effectiveness of two interfaces may depend on the specific scenario used. Finally, it must be remembered that the metric is intended to be applied to operational (commercial or prototype) systems, and these may not be readily available or may employ proprietary algorithms not easily adapted to test methods such as driving simulators. Thus while there are important advantages to a common evaluation method, there are challenges in accomplishing this.

1.1.2 Active Warning Systems

An important emerging issue (alluded to above) related to ACWS effectiveness is the inclusion of some active automatic vehicle control intervention as part of the warning system. An “active” warning system, as defined here, is where there is automatic partial control of a vehicle’s behavior (e.g., direction, speed) through steering/braking. This automatic action may itself serve as a warning cue, and may directly promote driver responses that aid in crash avoidance, in addition to any direct safety effects from the vehicle response itself. However, little is known about actual driver response to such active interventions. Furthermore, current commercial examples of such systems are typically very moderate in terms of how aggressive the vehicle control action is, and they appear intended as aids to driver actions rather than as autonomous vehicle control of the situation. Active warnings are of particular interest both because of their potential to promote improved driver response and because of the possibility that they may induce inappropriate driver reactions or poor consumer acceptance of warning systems. Examples of inappropriate driver reactions include overcorrection in steering, strong lateral acceleration, severe deceleration, startle responses, and driver confusion. In order to devise CWIM that remain appropriate as active systems evolve and become more common, it is important to gain some better understandings of how drivers respond to these types of ACWS.

1.1.3 DVI Variability Considerations

One objective for developing CWIM is to ensure that systems in new vehicles perform adequately, to at least some basic standard. The intent is to achieve this through the development of proven, repeatable, and efficient test metrics. Even systems that test adequately by themselves, however, may suffer problems in actual application because users face problems due to the variability in DVI among vehicles. Drivers may come to be familiar with the DVI in their personal vehicles. But as ACWS become more ubiquitous, drivers may confront unfamiliar interfaces when they use rental vehicles, share vehicles, or acquire a new vehicle. They may have false assumptions about vehicle functions and displays or may react slowly or inappropriately to emergency events. The concerns related to variability among DVIs therefore will become more prominent as diverse vehicles with such systems proliferate.

One approach to address this concern might be standardization of some aspects of the DVI. Standardizing a warning interface, however, has both positive and negative potential. The possible drawbacks are significant, so that recommendations for standardizing should not be made lightly and without a strong empirical or analytical basis. Some of the concerns with standardizing the ACWS DVI include the following:

- A standard may constrain what industry can do, which may limit innovation
- Technology advancements may suggest new and better approaches, not compatible with the standard
- The standard may ultimately be inconsistent with aspects of future in-vehicle environments (e.g., new types of displays)
- Each manufacturer may have a different suite of warning systems and system features, and a single approach may not be optimal for all manufacturers or all drivers
- A standard may oppose manufacturers' interests in product differentiation and conflict with the given esthetic approach of a given vehicle. Another approach of equal effectiveness to the standard might be reasonable, yet not allowed

However, despite these concerns, there may be good reasons for promoting some common features for the ACWS DVI interface. Driver response to signals that are unfamiliar may be delayed or confused. Safety may be compromised if the user experiences negative transfer between one system and another. A driver who is accustomed to a particular interface in one vehicle may be confused by, react slowly to, or react inappropriately to a warning from a vehicle with a distinctly different ACWS DVI. In fact, the very same signal could have explicitly different meanings in two vehicles. Furthermore, if there is a range of possible displays, it may be difficult to keep crash alerts perceptually distinct from other non-emergency displays. The DVI must also convey the status of the warning system to the driver in order to achieve its function. A driver should be aware of whether a given vehicle has a particular type of warning system (e.g., FCW), have an accurate mental model of how that system operates, and understand whether the system is currently operational. For example, a particular FCW system may only work when the vehicle speed exceeds some threshold, or a LDW may not be functioning because lane markings are inadequate or because there is some sensor or electronic failure in the system. These are complex messages to convey and inconsistency among manufacturers in whether and how such messages are conveyed may lead to driver confusion.

In summary, ACWS are increasingly common in vehicles and their effectiveness will depend in part on the properties of the DVI. Common methods to objectively evaluate the DVI for a given ACWS function could provide benefits, but there are significant challenges in accomplishing this. Furthermore, there may be some aspects of the DVI for which there may be benefits to common treatment among different vehicles, even assuming an effective CWIM methodology.

1.2 Objectives

The goal of the CWIM project was to provide the basis for broadly accepted procedures for the objective assessment of the DVI for ACWS. It also had the goal of determining what ACWS warning modalities and features may be promising, and where variability of displays may lead to comprehension problems. The focus of this initial work was on two ACWS functions: FCW and LDW.

As part of this effort, the project also included a comparison of driver response to alternative DVI modes for LDW and FCW applications. These included modes beyond the traditional visual and auditory displays, such as haptic cues and active vehicle control response (e.g., countersteer). These comparisons were intended to provide useful new findings regarding driver response as well as serving as a basis for the assessment of evaluation methods.

1.3 Project Activities

The project involved a combination of empirical research and analytic activities. Figure 1 is a "project roadmap" that shows the various project activities and their interrelations. The initial efforts of the project were analytical (left side of figure). This work examined research literature, crash analyses, current practice for interface design and evaluation, and expert/stakeholder feedback. This defined needs, options, and preliminary suggestions for use in the subsequent project activities.

The various empirical efforts are shown in the second column of the figure. These include three experiments (blue boxes in Figure 1) that compared various crash warning interface modes and examined the methods used to evaluate them. There were two experiments (green boxes) that addressed various implications of DVI variation among vehicles. Finally, there was a review of the methods and findings of research under the Advanced Crash Avoidance Technologies (ACAT) program (purple box). While this was not new empirical work undertaken as part of the CWIM project, it was a source of recent empirical NHTSA research that might bear on CWIM issues.

The subsequent activities shown in Figure 1 integrated findings and derived implications from the various lines of research. Ultimately this was used to derive the final products of the project, which included recommendations for DVI evaluation methods, warning modes, and interface design principles.

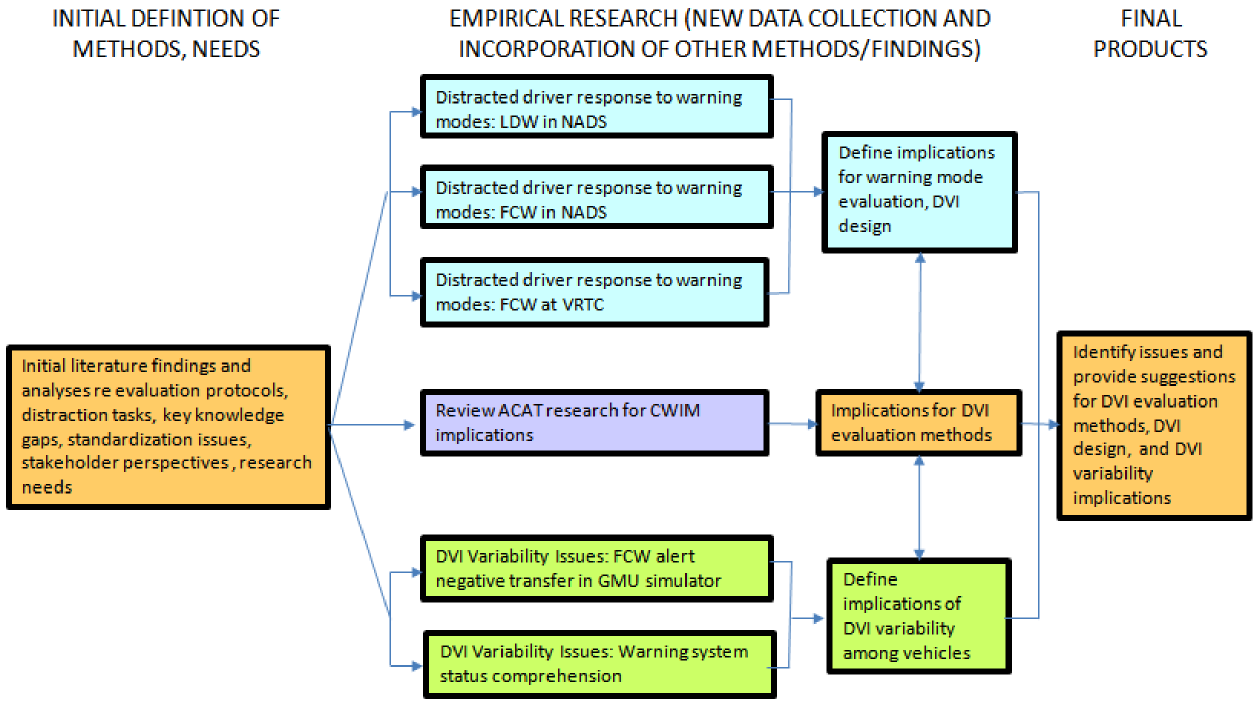

Figure 1. Roadmap of CWIM project activities

Further detail on the various project activities follows. Section 2 summarizes analytical activities and Section 3 summarizes the empirical research.

2 Summary of Analytical Project Activities

2.1 Document review and driving scenarios

The project included an initial literature review that addressed ACWS interface design for passenger vehicles, FCW systems, and LDW systems. The review effort was initiated under a predecessor project and completed as part of the present project. Particular attention was given to recent major review papers in these areas as well as key guidance documents. Over 300 primary sources were identified and subjected to initial scanning for prioritization. Approximately 150 sources were subsequently further reviewed, with particular focus on performance metrics and methods.

Based on the review of this literature, a taxonomy of driver responses to ACWS warnings was developed and provided a basis for consideration of potential evaluation measures. This analysis is further discussed in Section 2.2, below.

In parallel with this review, a search was made for information on the key driving scenarios associated with FCW- and LDW-relevant events. This is an important consideration for the development of any broadly applicable evaluation methodology. While a variety of information sources was found, analyses associated with the Integrated Vehicle-Based Safety Systems (IVBSS) program, the Crash Avoidance Metrics Partnership (CAMP), and the Advanced Crash Avoidance Technologies (ACAT) program were particularly relevant. The analyses of Najm and his colleagues (e.g., Najm, Smith, and Yanagisawa, 2007) were commonly referenced and provided a major source for at least the broader outlines of typical relevant crash events.

For rear-end crashes (the relevant case for FCW), the two most common scenarios were “stopped lead vehicle” (typically with subject vehicle traveling at 35-55 mph) and “decelerating lead vehicle” (typically with subject vehicle traveling at 45-60 mph). These two scenarios respectively accounted for 50.4% and 23.4% of applicable rear end crashes. Among the less frequent rear-end scenarios, one of particular interest is where the lead vehicle changes lanes, revealing a decelerating or stopped vehicle ahead of it. This is of interest because of its adaptability to test track applications, since it can employ a stationary surrogate “target” vehicle rather than an actual moving vehicle with a driver inside.

For crashes relevant to LDW, no scenario was as dominant as the top two for FCW. However, a majority of events are included by the scenarios of drifting off of a straight road to the right (typically with subject vehicle traveling at 25-55 mph), drifting off of a straight road to the left (typically with subject vehicle traveling at 30-60 mph), departing a road to the right while negotiating a left curve (typically with subject vehicle traveling at 30-55 mph), and loss of vehicle control while negotiating a curve (typically with subject vehicle traveling at 40-60 mph).

For practical purposes, only a limited number of event scenarios are likely to be employed in evaluation protocols for DVIs for a given warning function. This of course does not preclude researchers or system developers from encompassing more scenarios in their work. However, the analyses of crash scenarios reviewed in this task suggests that for purposes of a common evaluation basis, the predominance of the associated crash problem can be assessed with relatively few prototypical scenarios.

2.2 ACWS response taxonomy and potential measures

The literature review identified an extremely large number of measures and methods for assessing the effects of a crash warning. There are numerous aspects of the driver response to ACWS. For example, there are perceptual aspects (detecting the hazard), cognitive aspects, and vehicle control actions. For any of these aspects, there are many different measures that may be used to quantify or evaluate behavior. Therefore, this project developed a taxonomy that systematically organized and structured the dimensions of driver response to ACWS, to serve two primary purposes:

- It provides a way to comprehensively view all of the elements of driver response that relate to ACWS and to understand their relation to one another.
- It provides a structure for organizing and categorizing the many potential dependent measures that might be used in the study of ACWS, allowing for systematic comparison and selection.

In structuring the problem, it became evident that there are three broad evaluative domains for the driver response to ACWS. We have labeled these domains “crash avoidance,” “general driving performance,” and “driver/consumer acceptance.” “Crash avoidance” refers to responses that occur in response to an immediate threat situation. “General driving performance” refers to changes that occur in driving style as a result of experience with the ACWS, without regard to a specific threat situation. For example, as a result of a forward collision warning system, a driver might come to adopt longer vehicle following headways; or conversely, as a result of reliance on the system, may become less cautious in car following. There are many potential influences of ACWS experience that might affect a driver’s vehicle control strategies, attention allocation, hazard recognition, situation awareness, willingness to be distracted, and other safety-related behaviors and attitudes. The third evaluative domain, "driver/consumer acceptance," refers to emotional or subjective evaluative aspects of the ACWS. If the system is seen as excessively annoying, startling, intrusive, unreliable, or of low utility, it may not be accepted by the public, regardless of its performance in the crash avoidance domain.

A comprehensive evaluation of the DVI for ACWS should incorporate all three of these domains. However, the present project was focused primarily on the crash avoidance evaluative domain and less on consumer acceptance. Subsequent expansion of the taxonomy was therefore limited to this area.

In examining metrics used in the crash avoidance literature, it became evident that there are two very broad, more or less parallel sorts of measures. We have labeled these "driver cognition/awareness" and "driver behavior." "Driver cognition/awareness" metrics focus on the driver's knowledge of the situation. At what point does the driver become aware of a hazard? How accurately is the threat understood? Such questions require techniques for tapping the perceptual/cognitive aspects of processing ACWS warnings. In contrast, the "driver behavior" metrics are not concerned with cognitive measures but rather overt driving-related behaviors. These behaviors may include vehicle control actions, physiological or perceptual activity, or vehicle/hazard interaction. The taxonomy thus recognizes this two-part "behavioral domain": driver awareness measures versus overt, directly measurable driving actions.

A comprehensive ACWS driver response taxonomy was developed incorporating the three evaluative domains and the two behavioral domains. The intent was to help expose all of the aspects of ACWS DVI that may warrant inclusion in a set of base evaluative metrics. The general structure grew out of natural clusterings as well as broad evaluative aspects. The basic structure of the taxonomy is shown in Figure 2. The figure shows, but does not expand, the "general driving performance" and "driver/consumer acceptance" evaluative domains.

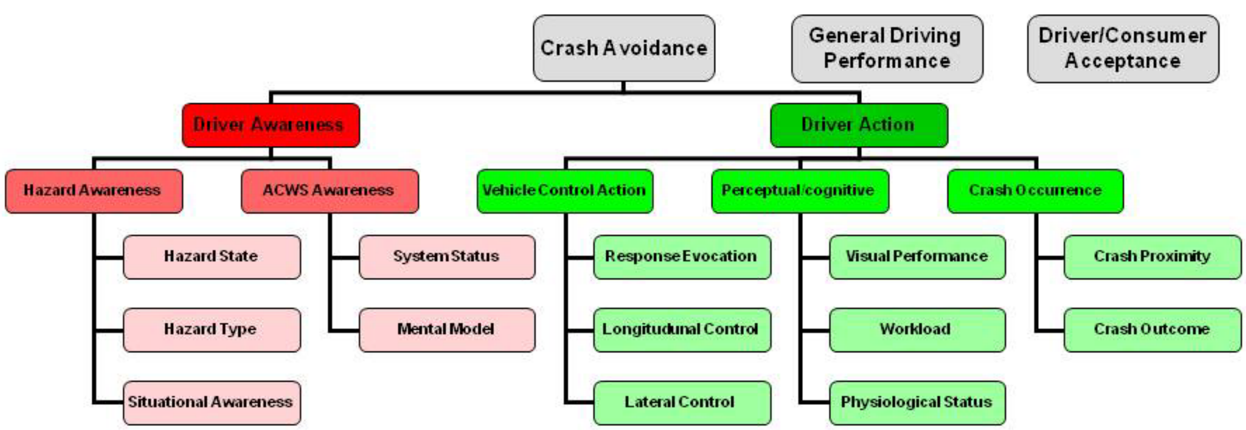

Figure 2. ACWS response taxonomy

The left side of the figure (red boxes) breaks out the "driver cognition/awareness" portion of the taxonomy. The right side of the figure (green boxes) breaks out the "driver behavior" portion of the taxonomy. The driver cognition/awareness measures are of two sorts: hazard awareness and ACWS awareness. Measures related to hazard awareness concern the presence of a hazard state, the type of hazard, the awareness of the context of the hazard event, and response options. ACWS awareness refers to the driver's awareness of aspects of the warning system itself. Little reference was found in the literature regarding such measures. However, various guidance for the design of crash warning systems clearly suggested the need to insure that aspects of the ACWS system are communicated to the driver. These include system status (on/off, optional settings) and the "mental model" of how the system operates and what its limitations are. Table 1 expands each of the boxes in the Figure 2 taxonomy with subcategories and illustrative examples of the type of factors that are associated with each.

The taxonomy shown in Figure 2 and Table 1 makes evident how many different types of behaviors and associated measures are potentially relevant for CWIM. Numerous dependent measures and research methodologies may be associated with any one of the categories in Table 1. While any one of the components of this taxonomy may be valuable for purposes of research into the driver response to ACWS, it would be impractical to try to develop commonly used CWIM for evaluation purposes that encompass all elements of this taxonomy. The structure provided here may serve as a basis for selecting and comparing alternative approaches for deriving an effective but practical set of CWIM. Using the taxonomy, the very large number of potential responses identified in the literature review were organized into 13 tables, corresponding to the 13 boxes in the five categories of the lower tier of Figure 2. For each table (e.g., Driver Vehicle Control Action: Response Evocation), for each entry there is a description of the general measure (e.g., deceleration avoidance response), a list of specific dependent variables for the response (e.g., accelerator release reaction time), an expanded definition of the measure beyond the descriptive term, stated advantages (as listed in cited literature), stated disadvantages, and comments from the cited document authors. The tables are extensive (about 40 pages) and are not reproduced here. Some of the measures presented in the tables are quite commonly used, while others are less common. Some require specialized equipment while others are easily implemented. Some are only now emerging as potentially viable measures (e.g., evoked potential) while others have long histories of use. The potential value of a particular measure may be very much context specific. Its usefulness may also be limited by constraints of resources, time, or practicality for a particular application. Thus it is neither simple nor necessarily reasonable to identify a given measure as "good" or "bad." These judgments ultimately need to be made based on the functional needs of the metric under development. It is also the case that in reviewing the literature, there was relatively little systematic critical comparison of alternative measures, and what existed was typically narrow in that the comparison was for a very specific application (e.g., test track evaluation of lane departure under a given scenario).

Table 1. Subtopics and example factors for component blocks of the ACWS response taxonomy

Driver Cognition/Awareness	**Vehicle Control/Action**
Awareness of hazard state • Hazard presence: signal meaning, conspicuity, discriminability • Hazard urgency/severity: perceived urgency to act, crash proximity, likely outcome events	Response evocation (definition of response and basis for response time) • Deceleration avoidance response: e.g., accelerator release, maximum brake force, deceleration rate • Steering avoidance response: e.g., steering response initiation, criterion lateral acceleration, change in heading • Inappropriate response: excessive or unnecessary braking, over-steering, improper path, startle, disorientation
Awareness of hazard type • External hazard: object, road geometry • Vehicle interaction hazard: conflicting paths, converging, hazardous proximity • Driver performance hazard: self-awareness of distraction, fatigue, speed selection	Longitudinal control • Vehicle following: time-to-collision, time/distance headway, car following coherence • Decelerating/stopping: maximum/mean deceleration, stopping distance • Speed control: mean/maximum speed, speed variance, relative speed
Situational awareness • Awareness of hazard event contributing factors: road surface, surrounding traffic actions, roadway features • Awareness of crash avoidance response limiting factors: situational affordance of driver maneuvers, e.g., steering left, steering right, braking	Lateral control • Lane position: lane exceedances, position variance, steering inputs • Dynamics: lateral acceleration, steering reversals, time to lane crossing, heading • Recovery time
Response options • Subjective decision questionnaires • Awareness of response options • Timed decision responses	Visual search • Target acquisition: target glance RT, time on target, number of target fixations • Allocation of visual attention: glance duration, eyes off road time, % glances
Awareness of ACWS system status • ACWS presence in vehicle • System status: on/off, operational • System parameters/options in effect	Physiological status and Workload • Alertness/awareness: evoked potential, blink rate, Perclos • Arousal: heart rate, eyelid closure • Subjective workload • Secondary tasks, e.g., peripheral detection • Physiological indices: galvanic skin response, heart rate variability

2.3 Stakeholder feedback

As part of the CWIM effort, stakeholder feedback was sought regarding the needs, issues, and concerns related to the development of a common evaluation method for ACWS DVIs. Three parallel activities were conducted to provide this feedback:

- A Federal Register Notice soliciting input from industry and the public on key issues related to CWIM.
- A meeting of representatives from various NHTSA offices and programs that have a direct interest in the CWIM effort and may be potential users of the CWIM protocols.
- Meetings with the automotive industry.

The Federal Register Notice (NHTSA Docket No. NHTSA–2007–0038) was published to elicit responses from interested stakeholders in industry and other members of the public. The Notice contained a detailed explanation of the CWIM project, and a listing of topics that the project team would like comments/feedback. Responses were received from a variety of manufacturers, researchers, and industry groups.

Members of the various relevant offices at NHTSA were invited to attend a session to discuss CWIM and possible uses to their respective departments. Each office sent at least one representative, and included Rulemaking, Research and Program Development (Behavioral Safety Research), Enforcement, and Vehicle Safety Research. Attendees were sent an overview of the CWIM project, and a list of discussion questions that included topics such as potential uses for their office. During the meeting, a presentation was given to describe the CWIM project, and then attendees were asked to discuss various topics and answer several questions in an informal group setting

The automotive industry is a key stakeholder and the project team sought additional information on their perspective, beyond responses to the Federal Register Notice. Follow up contacts were made with OEMs, Tier 1 suppliers, and industry groups that responded to the Federal Register Notice and follow up discussion meetings were held in Detroit. Additional OEMs and suppliers were also contacted about participation in these discussions. Meetings were held with the Automotive Alliance (attended by representatives from several companies) as well as with individual companies. The meetings included Westat and VTTI members of the project team and NHTSA staff.

Among the key points raised from stakeholder feedback were the following:

- Some members of industry had a strong reaction to any notion of standardizing the interface design or a relative comparison of DVIs across systems. Arguments ranged from statements of there not being a need to the process being intractable.
- There was also a concern by some in industry that a standardized evaluation tool could impede the natural evolution of products. Furthermore, there was a concern regarding cost, practicality, and realism in having a standardized evaluation protocol.
    - There was also a belief by some members of industry that testing was unnecessary because there has been no documentation of adverse safety effects.
- System integration poses a real challenge to evaluation of any single specific warning function. Likewise, separating DVI evaluation from broader system performance is a challenge.

- Alternatively, several offices within NHTSA expressed interest in utilizing CWIM to perhaps standardize some facets of the DVI or compare DVIs across systems. For example, the ACAT program does not have an evaluation tool for DVI, and could potentially utilize CWIM for that purpose.
    - It was noted that a threshold (pass/fail) score instead of a ranking of systems may yield greater cooperation from industry.
- Industry believed measuring consumer acceptance to be an important part of any evaluation tool.
- Not all industry members opposed CWIM. Several believed that a uniform evaluation metric, and even some form of DVI standardization, would be helpful.
- CWIM appears relevant to the activities of several NHTSA programs, including NCAP and ACAT.
- Some current DOT programs (e.g., ACAT) provide only limited consideration of driver-vehicle interface and/or driver behavior aspects. Therefore, the CWIM effort may complement and enhance those other programs.
- NHTSA staff generally suggested that simplicity in procedures and outputs is desirable
- There is a concern with the extent to which CWIM indices may be scenario dependent. How much would conclusions differ if other scenarios were chosen?

3 Overview of Empirical Studies

Five empirical experiments were conducted in association with this project. A sixth activity was not new data collection but a review of recent research methods employed by research teams working under NHTSA's ACAT program. Three of the five new experiments were formally executed under the contract that supported the project. The other two experiments were conducted by organizations under separate contracts, but were designed to be complementary to the goals of this contract. The project team worked collaboratively with these other organizations in planning the studies and interpreting the findings, although ultimate responsibility for final decisions and the actual conduct of the data collection and analyses was with the other organizations. Because of this interrelationship, the methods and findings of all five experiments are summarized in this report and used as a basis for conclusions and recommendations. Three of the experiments focused on the effects of alternative modes of alerting the driver. The other two addressed implications of DVI variation across vehicles. The five experiments were:

1. Driving simulator study of passive and active warning modes for LDW (led by the University of Iowa, under this contract)
2. Driving simulator study of warning modes for FCW (led by the University of Iowa, under a separate contract)
3. Test track study of warnings modes for FCW (led by NHTSA's Vehicle Research and Test Center (VRTC), under a separate contract)
4. Driving simulator study of potential negative transfer for FCW auditory signals (led by George Mason University, under this contract)
5. Laboratory study of ACWS status display comprehension (led by Westat under this contract).

The three experiments on warning modes were closely interrelated. The LDW experiment was conducted in the National Advanced Driving Simulator (NADS) facility. Participants drove a simulated two-lane rural highway and periodically engaged in a variety of distracting activities that directed driver vision away from the forward roadway. At three points in the drive, the simulator vehicle dynamics subsystem induced a lane shift (without motion cues) that caused the vehicle to be drifting out of its lane. Several different warning interfaces were compared with each other and with the control (no warning) condition. These included (1) acoustic alert; (2) tactile (steering wheel vibration) alert; (3) weak active countersteer; (4) stronger active countersteer; (5) no warning (control). The FCW simulator experiment was also conducted in the NADS facility with a similar procedure to the LDW experiment. The experiment included two FCW events, lead vehicle slowing and lead vehicle stopped (essentially very severe slowing). The interfaces tested included: (1) auditory/Head-Up Display (HUD); (2) brake pulse; (3) no warning (control). The FCW test track experiment involved a car-following procedure on a multi-lane test track. The FCW event was a “lead vehicle revealed” scenario in which the vehicle that the participant was trailing suddenly move out of its lane to reveal a stopped vehicle ahead. The warning modes tested included: (1); auditory (2) HUD; (3) seat belt pre-tensioner; (4) auditory and HUD: (5) seat belt pre-tensioner and auditory; (6) seat belt pre-tensioner and HUD; (7) seat belt pre-tensioner, auditory, and HUD; (8) no warning (control). In addition to assessment of the driver response to various warning modes, this set of studies also encompassed methodological interests. The experiments on warning modes used experimental methodologies that differed in a number of respects. This further allowed comparisons of the attributes of alternative methodological features.

The remaining two experiments dealt with DVI comprehension and the implications of DVI variability across vehicles. The FCW negative transfer study was a driving simulator experiment that examined what happens when drivers familiar with one auditory FCW warning encounter a different FCW auditory warning in another unfamiliar vehicle. The ACWS status display experiment presented participants with detailed reproductions of vehicle interiors under various system status conditions. At issue was how well people understood what crash warning systems were present in the vehicle and what their functional status was. The experiment also considered the effects of familiarity with the system (through the owner’s manual) and the extent that any beneficial effects of familiarity transferred to another vehicle.

In the following sections, we provide capsule descriptions of the procedures for each experiment and highlight a few key results. More extensive treatments of the methods and findings may be found in separate interim reports for the various experiments. The simulator study of LDW modes is reported in the project task report (University of Iowa and Westat, 2010). The simulator study of FCW negative transfer and the laboratory study of status display comprehension are reported in Robinson et al. (2010). The test track study of FCW is reported in Forkenbrock et al (2011). The simulator study of FCW will be detailed in an as-yet unpublished report under a separate NHTSA contract. We also include below a summary of findings from the review of ACAT test protocols. Then in Sections 4 and 5, we treat the combined findings of the experiments and their implications.

3.1 Driving simulator study of passive and active warning modes for LDW

The objective of this study was to determine how readily drivers are able to use LDW to improve lane recovery and crash avoidance, and in particular how this is related to warning modality and active warning strategies used by the LDW system. Active warnings (e.g., active countersteer) were of particular interest, both because they presumably have greater potential to promote rapid vehicle control responses and because their potential to induce inappropriate driver reactions is not well understood. Examples of inappropriate driver reactions include over-correction in steering, strong lateral acceleration, severe deceleration, and startle responses. The study also addressed driver acceptance issues. A system that is not well accepted by drivers may be disregarded or disabled and would therefore not be effective. Finally, through development of the experimental protocol, this study addressed issues surrounding the best approaches for evaluating the driver interface of LDW systems.

The experiment was conducted in the National Advanced Driving Simulator (NADS-1) at the University of Iowa. A two-lane bi-directional rural highway used in the study was representative of the most common roadway departure crash scenarios according to Najm et al. (2002). This road type with 3-meter (10-foot) lanes was selected. The roadway database was designed so that it had both long two-lane highway straight-aways as well as a variety of left and right curves. It consisted of a drive that was approximately 30 minutes in duration.

Choosing crash scenarios that are representative of real fatal and injury related crashes was an important goal of this study. The most common crashes and ones that are generally the most injurious and fatal were chosen to be examined in the study, these included:

- Vehicle drifts off road to the right;
- Vehicle drifts over the centerline, with on-coming traffic; and
- Vehicle fails to keep lane in a left curve entry.

Each participant in the study was exposed to these three scenarios while they were periodically distracted by a secondary task. Participants were also exposed to a false alarm scenario in which the LDW alert activated while they were passing through a construction zone.

This study compared driver responses to passive and active LDW warnings and to a control condition in which no warning was given about an impending lane departure. Passive LDW warnings included an acoustic alert and a tactile alert (steering wheel vibration). Active warnings included a weak active countersteer and a stronger active countersteer in the direction of the initial travel lane. The form and magnitude of all alerts tested were within the range of alerts seen on production vehicles now being sold in the United States or elsewhere, and on pre-production vehicles tested by staff at NHTSA's VRTC. Both countersteers had the same response profile, in which a gentle torque begins to be applied as the vehicle approaches the lane line, and then quickly ramps up to full force when the vehicle is departed between 0.5 and 1.5 feet out of lane. Once the vehicle has departed more than 1.5 feet from its lane, the torque ceases. When the torque is applied, the force is reflected with steering wheel motion, providing an additional haptic cue. The strong countersteer had a peak torque about three times as strong as the weak torque. Data were obtained from 90 participants (18 participants in each of five different LDW system groups, including one control group that did not experience any LDW).

Participants were instructed to perform a variety of secondary tasks while driving including a visual/manual "bug task" which distracted them from the forward roadway long enough that a lane departure could occur unnoticed. To achieve this, a simulated insect task ("bug task") was used as the primary distracter. This task required that the participant turn and reach into the back seat to trace the path of the insect on a touch screen display. To ensure that a LDW was obtained, a bias in the steering was triggered that nudged the car to the desired side of the lane during a distraction event. This was initiated based on driver engagement with the bug task. To mask the drift, a compensation technique was used to remove the portion of the vehicle dynamics associated with the drift from the motion cues in the simulator.

Although specific findings varied somewhat across the range of dependent measures included in this study, the general outcome was that all four warning conditions were superior to the baseline control condition and that frequently the "weak torque active LDW" conditioned performed best (although not always statistically significantly so). For example, the weak torque warning was significantly better than all or all but one other treatment for measures of maximum lane exceedance, severity of initial steering angle, total amount of time spent out of lane, and number of inappropriate behaviors elicited. Figure 3 shows an example for the metric of total area of lane exceedance exposure. In this case, all four warning conditions had statistically significantly less area outside of their lane than those who did not receive a LDW. Although the weak torque condition appeared to have the lowest degree of lane exceedance exposure, the differences among the four warning conditions were not statistically significant.

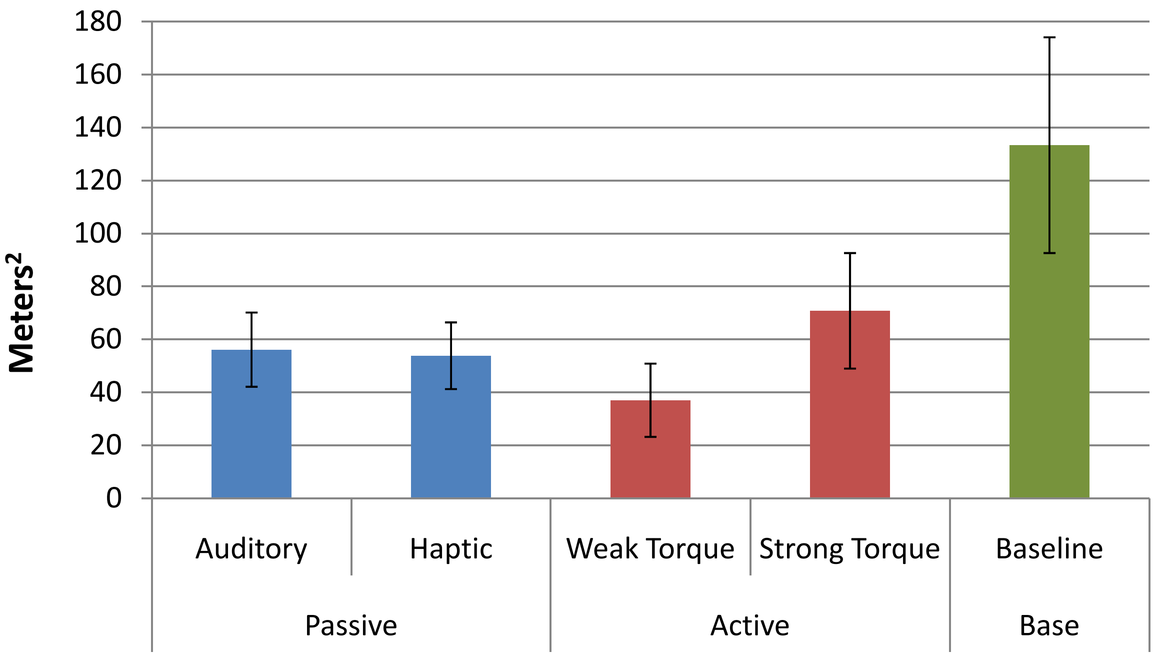

Figure 3. Area of lane exceedance as a function of LDW warning condition

Overall, the impact of the warning modes on the number of inappropriate behaviors observed from driver showed that the highest number of inappropriate behaviors was observed from those participants who did not have a LDW. There were significantly fewer inappropriate behaviors for participants who experienced either of the active systems. The number of subjects who fully departed their lane, or ran off road, was significantly less for the strong torque warning than for the auditory warning. For the false alarm scenario, in which there was no actual lane deviation, the strong torque differed from the other warning conditions in that drivers with this system responded with greater, though unnecessary, vehicle control actions.

Participants were not asked to directly compare different warnings in this study. However, those participants who experienced the weak torque rated that warning as less effective in capturing their attention as compared to other participants' ratings of other LDW warnings. The group experiencing the auditory warning found it more effective at capturing attention as compared to other participants' ratings of either of the active warnings. Furthermore, the passive systems were viewed by the participants as being more helpful than the active systems. Participants who experienced a passive warning felt that the system was more easily interpreted than those who experienced an active warning. Participants who experienced a passive warning also felt that the system was more reliable than those who experienced an active warning.

Conclusions from this study were that:

- The protocol was effective at getting the driver's attention off of the road long enough to initiate a lane departure, but refinements to the task and/or participant motivation could improve it.
- The weak torque active LDW was overall the most effective for the planned lane departures at minimizing the extent of lane departures; it provided enough force to help correct a lane deviation and alert the driver without promoting steering overcorrection or other inappropriate behaviors. All LDWs, however, reduced the risk to the driver as measured by area of exposure and maximum lane exceedance.
- Drivers considered the passive warnings to be more effective in capturing attention and more helpful than the active systems.
- There is a mismatch between driver perception of effectiveness and effectiveness as measured by the performance measures.
- When a false LDW alert was issued, drivers in the strong torque LDW condition displayed the most inappropriate vehicle control actions.

3.2 Driving simulator study of passive and active warning modes for FCW

This experiment compared driver responses to two different FCW systems (passive versus active) on two different crash scenarios and a false alarm event. The passive FCW driver interface incorporated a head-up display (HUD) and an auditory alert. The active FCW used a brake pulse to alert the driver by exerting momentary activation of the brakes. Specifications for these warnings were developed in consultation with NHTSA and VRTC. The two crash scenarios used in the study were a decelerating lead vehicle and a stopped lead vehicle. The experiment was conducted in the National Advanced Driving Simulator (NADS-1) at the University of Iowa. The roadway environment was similar to that used in the LDW study discussed in Section 1.1.

To support this research on the effectiveness of FCW system warnings, it was necessary to use a distraction task that would reliably and repeatedly insure that the driver's eyes are off road for several seconds prior to the forward collision events. Because drivers are able to use peripheral vision to monitor the roadway, it was essential that the driver's gaze be directed well away from the forward view. To achieve this, the same simulated bug task described in Section 3.1 was used here.

Thirty-two participants experienced one of the two FCW systems; 16 other participants in a third group (baseline) did not have any FCW system. The independent variables were warning condition (baseline, auditory/visual, brake pulse), gender, and scenario event (braking lead vehicle, stopped lead vehicle, false alarm). Warning condition was a within-subjects variable and scenario event was a between-subjects variable. For each forward collision event, measures of initial vehicle control and inappropriate responses were recorded. Following the drive the driver's acceptance of the FCW system was assessed.

An objective of the study was to determine which of several potential dependent measures would be most sensitive for assessing driver response to FCW alerts. An objective of the study was to determine which of several potential dependent measures would be most sensitive for assessing driver response to FCW alerts. There were no statistical differences (at the $p<0.05$ level) in response time among the conditions, although there was a trend across the several reaction time measures used. When looking at initial response to the event, drivers in the baseline condition took longer to release the accelerator than drivers in the warning conditions relative to their initial engagement in the distraction task. There were also trends towards faster performance for responses relative to the the time the alert was issued. On average drivers responded by releasing the accelerator and applying the brakes 375 ms sooner with a warning than without. This result is shown in Figure 4.

When applying the brakes, there were significant differences in both the level of braking and the maximum decleration achieved by the driver. Peak brake pedal force was less forceful for drivers with the brake pulse than for drivers in the baseline and auditory/visual warning conditions. Also, these drivers achieved a maximum brake pressure that was 36 percent less than was achieved in the other conditions. Drivers in the brake pulse condition achieved a peak deceleration level that was 15 percent less than for drivers in the other two conditions. These differences in braking response did not translate into differences in collisons. While the reasons for these differences are not clear, it is possible that the brake pulse resulted in less panicked responses or that the brief brake pulse made drivers feel that less urgent braking was required.

The collision outcomes by scenario and alert condition are shown in Table 3. Overall, there were no differences between groups in the number of participants who avoided collisions for the stopped lead vehicle (Fisher's Exact = 0.6400) or braking lead vehicle (Fisher's Exact = 1.0) scenario.

Table 2 provides the dependent measures that were analyzed for this analysis as well as the level of significance for the main effect of FCW condition. The highlighted rows show where there were statistically significant differences. Only Peak Brake Force and Peak Deceleration differed significantly between the three FCW groups.

There were no statistical differences (at the $p<0.05$ level) in response time among the conditions, although there was a trend across the several reaction time measures used. When looking at initial response to the event, drivers in the baseline condition took longer to release the accelerator than drivers in the warning conditions relative to their initial engagement in the distraction task. There were also trends towards faster performance for responses relative to the the time the alert was issued. On average drivers responded by releasing the accelerator and applying the brakes 375 ms sooner with a warning than without. This result is shown in Figure 4.

When applying the brakes, there were significant differences in both the level of braking and the maximum deceleration achieved by the driver. Peak brake pedal force was less forceful for drivers with the brake pulse than for drivers in the baseline and auditory/visual warning conditions. Also, these drivers achieved a maximum brake pressure that was 36 percent less than was achieved in the other conditions. Drivers in the brake pulse condition achieved a peak deceleration level that was 15 percent less than for drivers in the other two conditions. These differences in braking response did not translate into differences in collisons. While the reasons for these differences are not clear, it is possible that the brake pulse resulted in less panicked responses or that the brief brake pulse made drivers feel that less urgent braking was required. The collision outcomes by scenario and alert condition are shown in Table 3. Overall, there were no differences between groups in the number of participants who avoided collisions for the stopped lead vehicle (Fisher's Exact = 0.6400) or braking lead vehicle (Fisher's Exact = 1.0) scenario.

Table 2. Driving simulator study dependent measures

Measure	**Description**	**Condition Effect**	**Mean Values**
Time to Accelerator Release	Time from task engagement (ex: first button press) to a partial throttle release. Partial release is defined as dropping below 85% of the value at task engagement.	$F(2,27) = 2.93$ $p = 0.0705$	Base = 4.39 A/V = 3.13 BP = 3.32
Accelerator Reaction Time	Time from when the FCW alert was issued (or would have been issued in the baseline condition) until the driver releases the accelerator measured in seconds	$F(2,27) = 2.16$ $p=0.1347$	Base = 1.19 A/V = 0.81 BP = 0.80
Time to Brake Press	Time from task engagement to a brake pedal depress measured in seconds.	$F(2,30) = 2.46$ $p=0.1024$	Base = 1.30 A/V = 0.94 BP = 0.91
Peak Brake Force	The maximum applied brake force during the avoidance response measured in foot-pounds	$F(2,30) = 5.08$ $p=0.0126$	Base = 120.6 A/V = 128.0 BP = 78.9
Peak Deceleration	The peak deceleration during the avoidance response measured in meters per second squared	$F(2,30) = 4.75$ $p=0.0162$	Base = -8.10 A/V = -8.67 BP = -7.06
Minimum Distance	The minimum bumper to bumper distance between the driver's vehicle and the lead vehicle measured in meters.	$F(2,30) = 2.51$ $p=0.0985$	Base = 4.63 A/V = 5.53 BP = 7.82
Adjusted Minimum Time-To-Collision	The minimum amount of time the driver had to avoid colliding during a potential crash event, as a function of the relative speeds of the subject and target vehicles, and the distance between them. In the case of no collision, this is positive and is measured purely as the minimum time to contact. In the case of a collision, AMTTC represents how much earlier (in negative seconds) a participant would have needed to start braking to have avoided the collision, based on their collision speed and deceleration rate.	$F(2,30) = 2.88$ $p=0.0718$	Base = 0.11 A/V = 1.37 BP = 1.05
Startle	The peak rate of deceleration measured in meters per second cubed.	$F(2,16) = 2.13$ $p=0.1514$	Base = 11.35 A/V = 13.01 BP = 20.28

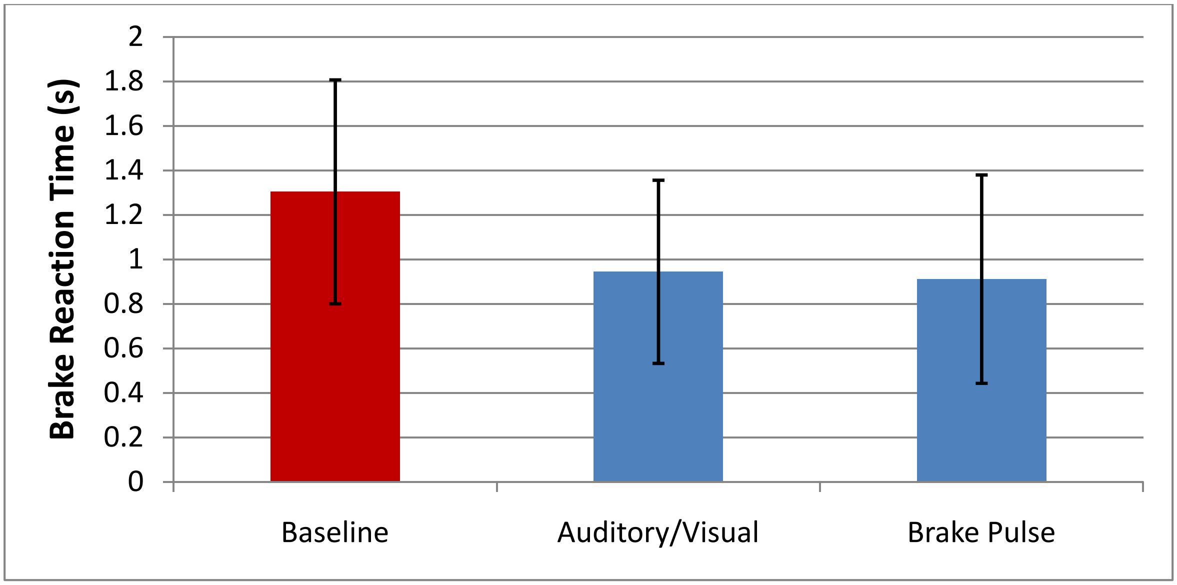

Figure 4. Elapsed time between initiation of FCW alert and brake pedal pressed (with standard error bars)

Table 3. Summary of collision outcomes by scenario and alert condition

	Braking lead vehicle			**Stopped lead vehicle**		
	Baseline	Auditory/ Visual	Brake Pulse	Baseline	Auditory/ Visual	Brake Pulse
No Collision	7	9	8	5	5	6
Collision	1	0	0	5	6	5
Total	8	9	8	10	11	11

Participants in each FCW group with alerts reported that they easily understood why the alert was presented, that the system successfully caught their attention, and that the alert was easy to see-and-hear or feel. The passive auditory/visual alert was rated significantly easier to interpret than the active brake pulse.

Several conclusions and lessons learned were drawn from the results of this study. From the standpoint of the effectiveness of the protocol for assessing the effect of active and passive FCW systems:

- As compared to the LDW study, shorter glances away from the road are needed for experimenters to successfully initiate a forward crash scenario event and the bug task generally seems to keep the drivers attention sufficiently long enough for the braking event to be triggered while the driver was inattentive to the road. However, there were still many cases in which the driver initiated a response prior to the activation of the alert (36 in total, spread nearly evenly across warning conditions). Drivers performing the bug task looked away from the forward view but were reluctant to look away for long periods of time, so they sometimes glanced toward the forward view; 38% of the FCW event data had to be removed because the driver had begun responding prior to reaching the alert threshold, indicating that drivers were not reliably and continuously distracted from the forward view. These findings suggest that the bug test may benefit from further refinement.

- Mounting the touch screen display further away from the driver in this study (as compared to its position in the LDW study) required a longer reach and seemed to help get drivers to commit to the task and make fewer glances back toward the road.
- Less data processing was needed to clean the data for the FCW study than for the LDW study. The procedure produced fewer outliers and there was less ambiguity concerning which data was applicable for the analysis.
- The data revealed some potential differences in responding between the two warning systems, but these differences were not statistically significant. A larger sample size might have permitted differences to be resolved more clearly.

3.3 Test track study of warnings modes for FCW

This experiment, conducted on a test track at NHTSA's Vehicle Research Test Center, compared driver responses to FCW systems that used either a HUD visual alert, an auditory beeping alert, a seatbelt tensioning device, or a some combination of two or all three of these alerts. For the full technical report detailing this study, see Forkenbrock et al. (2011). Each of 64 participants was randomly assigned to one of eight groups. Participants in the first group experienced no FCW alert while participants in the other seven groups experienced one of seven different possible combinations of FCW alerts. A key objective of the study was to develop a protocol for producing unexpected FCW events and for recording driver behavior metrics in a test track environment for use in testing the effectiveness of FCW driver interfaces. A second objective was to compare the effectiveness of a small set of FCW alerts using the protocol that was developed.

Adult participants were recruited from the general public. Each participant experienced only one FCW event and had no exposure to the FCW system prior to experiencing the event. Each participant was asked to follow a lead vehicle while attempting to maintain a constant headway. Feedback on current headway was provided to the driver on a visual display. The participant was also asked to perform a secondary task which involved diverting their attention away from the forward roadway toward a visual display inside the vehicle near the back of the front passenger seat. After performing this task several times and driving back and forth across a straight test track, the distraction task was performed for a final (i.e., fourth) time. During the final distraction task, while the participant was looking away from the roadway, the lead vehicle was abruptly steered out of the travel lane, revealing a stationary vehicle (a realistic-looking, full-size, balloon car) in the immediate path of the participant's vehicle. At a nominal time-to-collision of 2.1 seconds from the stationary vehicle, the FCW alert was presented to the driver.

All eight participants in the baseline group with no FCW alert collided with the balloon car. Similarly, all eight participants who received only the HUD alert collided with the balloon car and 7 of 8 participants who received only the beeping alert collided with the balloon car. Among the various FCW alert combinations tested, it was apparent that FCW systems which included the seatbelt pre-tensioner as an alert were more effective in helping drivers avoid a collision than other FCW driver interfaces that did not include this alert. Approximately half of the participants who received the seatbelt tensioning alert (alone or in combination with other alerts) avoided colliding with the balloon car. These results are shown in Table 4.

Table 4. Collision avoidance summary

FCW Alert Condition	Number of Participants	
	Collided	Avoided
No alert	8	0
HUD alone	8	0
Beep alone	7	1
Belt alone	5	3
Beep + HUD	7	1
Belt + HUD	3	5
Belt + Beep	5	3
Belt + Beep + HUD	4	4

The results of this study showed that the seatbelt pre-tensioner was effective at causing the driver to disengage from the secondary task (ending their visual commitment to the secondary task) and directing the driver's eyes back on the forward roadway in time to respond to the stationary vehicle in their travel lane. The timing of the protocol was such that many participants collided with the balloon car even though they had initiated some evasive maneuver. In addition to the outcome variable (collision/avoid) several other dependent measures were recorded in this study, such as timing variables (e.g., time from FCW activation until the driver's eyes were back on forward roadway), brake application timing and force, steering responses, speed of participant's vehicle at time of collision, etc.

The testing protocol developed for the study had several key features that were important for enabling researchers to successfully stage the unexpected FCW event. Some of these features were:

- A long, straight test track with turn-around loops on either end. This permitted researchers enough time to deploy the balloon car on the track out of the participant's direct view.
- Number recall distraction task – This task was designed to reliably keep the driver's eyes off of the road at a very precise location in the drive and to hold the driver's attention for a period long enough to mask the lead vehicle's maneuver around the stationary balloon car. The timing of the participant's engagement with this task was critical for creating the unexpected FCW event.
- Headway tracking – Participants were required to maintain a fixed distance between their vehicle and the lead vehicle. Headway feedback was provided by a visual display in the participant's vehicle.
- Participants' incentives and expectations – Participants were paid a variable amount based on their performance on the headway task and the number recall task. This encouraged them to perform the tasks as instructed. Participants' expectations were controlled by having them drive the same route several times and perform the same secondary tasks several times without incident prior to experiencing the FCW event.

- Screens – With the participant's vehicle in the correct position behind the lead vehicle, the participant was screened from seeing the balloon car ahead. A dark screen was installed inside the cabin of the lead vehicle to prevent participants from seeing through the vehicle to the forward roadway. Prior to deployment, the balloon car needed to be stored next to the test track but out of sight of the participant. This was accomplished by storing the balloon car in the back of a box truck that was parked near the track.

3.4 Driving simulator study of potential negative transfer for FCW auditory signals

This experiment addressed whether driver response to a FCW warning suffered when the participant switched from a familiar vehicle with one acoustic alert to a different vehicle with a different acoustic alert. A substantial decrement in response times after the vehicle change would suggest that there is a lack of transfer from one warning system to the other. Negative transfer, specifically, can be said to exist if drivers' existing understanding of how a warning system works actually interferes with responding to a novel, switched alert. If the difference is substantial, this would support the idea of having the same, or at least quite similar, FCW alerts in all vehicles. The experiment was conducted in the George Mason University driving simulator. To create a reasonable context in which participants encountered a different FCW system from which they were accustomed, the cover story for the experiment was one of testing how drivers handled various driving environments, tasks, and distractions when switching between vehicles. During the simulator drive, participants periodically engaged in a visually distracting task. Occasionally a forward event occurred (e.g., sudden slowing of a lead vehicle) that required an emergency avoidance response. Participants became familiar with a given warning system over the course of two driving sessions in the simulator. On the third day, superficial changes were made to the appearance of the simulator vehicle and the participant was informed that a different vehicle model was now being simulated. For half the participants, the FCW acoustic warning remained unchanged and for the other half, the warning was different. The key comparison was in response times to the warnings among drivers with or without a change in the FCW warning.

Two acoustic alerts were used in the experiment, counterbalanced across participants as the initial alert experienced in the familiarization stage. These were termed "light" and "heavy" warnings because they were drawn from previous studies employing light or heavy vehicles in the Integrated Vehicle-Based Safety Systems (IVBSS) program. The IVBSS light vehicle warning was faster (shorter inter-pulse interval and shorter pulse duration) than the IVBSS heavy truck warning. The IVBSS heavy vehicle warning incorporated a fundamental frequency of 600 Hz with one harmonic at 1800 Hz within a single burst, whereas the IVBSS light vehicle warning consisted of a fundamental frequency of 1500 Hz with five harmonics for a total of six frequencies. Table 5 summarizes the characteristics of both alerts.

Table 5. Negative transfer study FCW alert characteristics

Alert Characteristic	IVBSS Light Vehicle FCW Alert (A_1)	IVBSS Heavy Vehicle FCW Alert (A_2)
Tone	Abstract	Abstract
Frequency modulation	None	Two-tone
Frequency	1500 Hz, 4500 Hz, 7500 Hz, 10500 Hz, 16500 Hz, 19500 Hz	600 Hz and 1800 Hz
Pulse duration	50 ms	320 ms
Burst duration	700 ms	320 ms
Bursts per second	10	2
Interpulse interval	30 ms	0 ms
Onset ramp	5 ms	N/A
Offset ramp	20 ms	N/A
Number of bursts	2	3
Pulses per burst	7	4 (beginning with 1800 Hz then pulses of both frequencies
Warning duration	1300 ms	1300 ms

Figure 5 shows the primary finding from the experiment. As can be seen, brake reaction times (RT) in the two control conditions generally got faster with each subsequent exposure. Participants in the treatment conditions (in which the sound was later switched) also responded more quickly with repeated exposure to the same alert. Once the alert was switched to the alternative version, however, the pattern of faster responding was no longer apparent. Participants in the heavy-to-light warning sound condition displayed statistically significant and particularly dramatic increases in brake RT. Both the effect of familiarity through exposure (session to session) and the effect of a shift in the warning sound (at least in one direction) were quite substantial (roughly on the order of a half second), but the reason for the asymmetry in the shift conditions and the relationship of this to acoustic signal properties are unknown. While this study suggests that there may be substantial decrements in ACWS response for drivers who switch from a vehicle with a familiar warning to one with an unfamiliar warning, additional research would be required to determine the circumstances under which negative transfer is obtained and whether certain sound features (e.g., temporal pattern, primary frequency, tonal quality) may be adequate to maintain transfer.

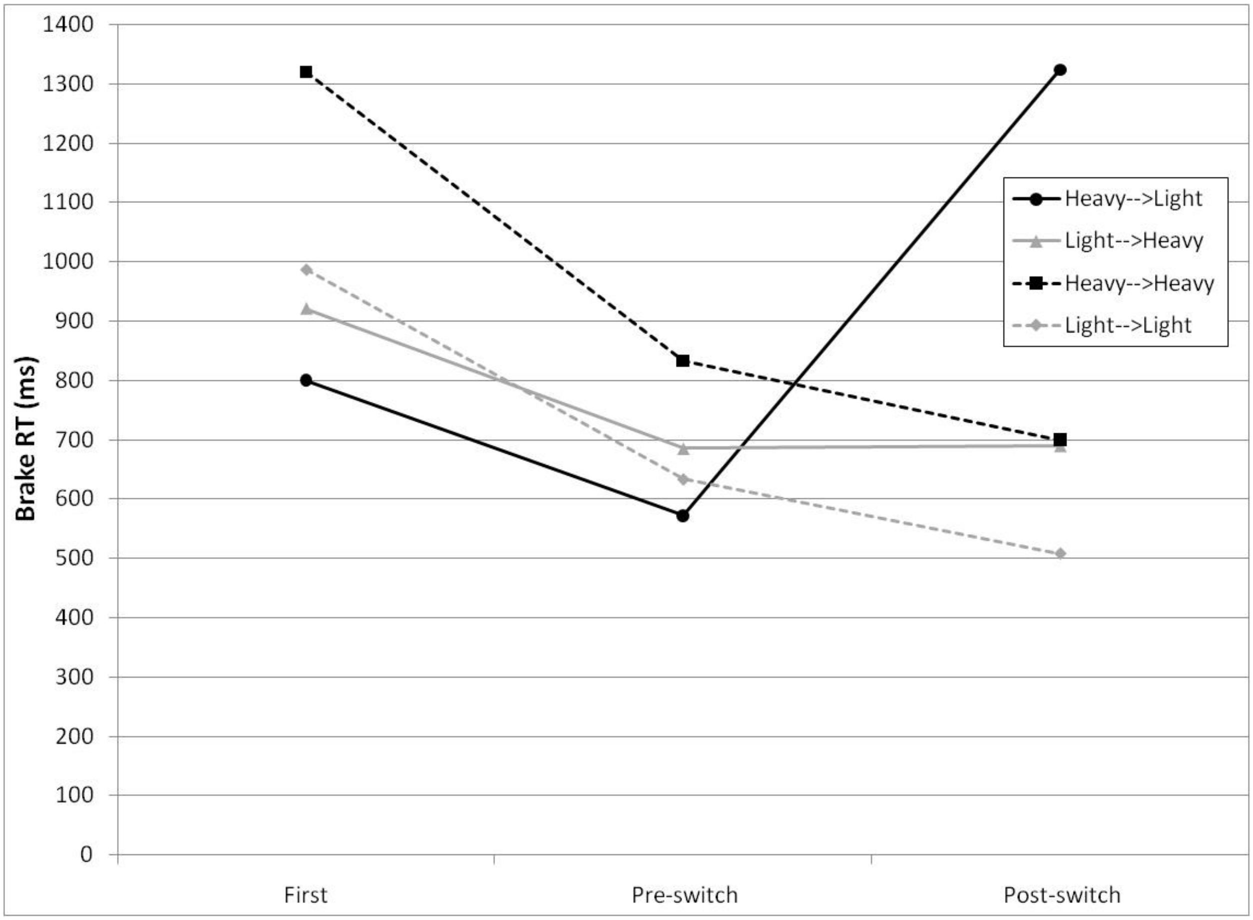

Figure 5. Brake reaction time as a function of warning condition and exposure time

3.5 Laboratory study of ACWS status display comprehension

This experiment addressed how well people comprehended status displays in vehicles with quite different display strategies and whether familiarity with one vehicle's system was helpful or interfered with the understanding of another vehicle's ACWS status. Guidance for improving the design of status displays could be helpful. However, given the extreme differences in the interfaces among various vehicles, a lack of transfer (or even negative transfer) from one familiar vehicle to another unfamiliar vehicle would still be problematic. Unlike the other empirical experiments in this project, this experiment did not deal with the immediate driver response to an imminent crash warning. Rather, it dealt with the driver's understanding of the status of the ACWS: Is a given warning function (e.g., FCW) present in this vehicle or not?; Is the function presently active ("on" or "off")?; Is the function currently fully operational? A driver who misunderstands these issues may fail to recognize a particular warning or may adopt a driving style that is based on a false assumption about driver support. The experiment visually simulated the vehicle interior that a driver sees (dashboard, steering wheel, and console displays). Three vehicles were selected to demonstrate a range of systems, based on the number and types of ACWS functions and their interface approaches (e.g., interface layout, use of icons, text, acronyms). For a variety of display scenarios, information was collected on what the participants understood, how confident they were in their understanding, how long it took to answer questions about the ACWS (processing time), and where they looked for particular information.

Prior to the experiment, some participants read owner's manual sections relevant to the vehicle they would see during the study, others read owner's manual sections for a different vehicle, and some read nothing. Reading the manual sections for the "same" vehicle allowed for an assessment of the effect of some degree of familiarization with the system, in contrast to a totally naïve user. Reading the manual sections for a "different" vehicle permitted examination of whether there are advantages to having general familiarity with other vehicle's systems (positive transfer) or whether vehicle-specific experience is required for better comprehension. The experiment allowed an assessment of viewer comprehension as a function of interface design and system familiarity (manual condition).

Each participant viewed the interior of one of three vehicles: 2010 Infiniti FX 35, 2010 Buick Lucerne, or 2010 Volvo S80. These particular vehicles were selected in part because they used very different display strategies from each other (e.g., icons, text, acronyms). Also, the vehicles ranged in the number of safety systems of interest, with one vehicle having the fewest (two systems) and another vehicle having the most (four systems). The subjects viewed the vehicle interiors at near full-size on a high-resolution (Dell UltraSharp 3008WPF) widescreen flat panel monitor (30-inch diagonal display area and a native resolution of 2560 x 1600 pixels). High-resolution photographs of the vehicle display areas were edited using Adobe Photoshop in order to clearly show all features and displays of interest, and saved as a high quality JPEG file. Symbols and messages of interest that could not be photographed in the stationary vehicle (e.g., certain malfunction indicators) were recreated using image editing software. Figure 6 shows example vehicle interior photos.

The experiment found that there were definite problems with both comprehension and transfer of knowledge from vehicle to vehicle. For pre-start, start-up, and en route situations, people had difficulty determining whether a particular safety system was present in the vehicle and whether it was operational. Furthermore, their decision times were quite long. Familiarity with the system through manufacturer-provided materials helped somewhat, but comprehension issues remained. Furthermore, there was little advantage (positive transfer) from being familiar with one vehicle's systems when the participant was tested in a different vehicle. Figure 7 provides a broad summary of the comprehension data (comprehension of system presence or status), averaged across scenarios and safety systems. Although differences among vehicles are apparent, we emphasize strongly that this study was not designed or intended to provide a systematic evaluation of the particular driver interfaces. Figure 7 shows generally low comprehension rates, modest overall advantages of familiarization (same manual) and no general positive transfer effect (different manual versus no manual).

Figure 6. Example vehicle interior photos: Volvo S80 before startup (top), Infiniti FX 35 at startup (center), Buick Lucerne en route (bottom)

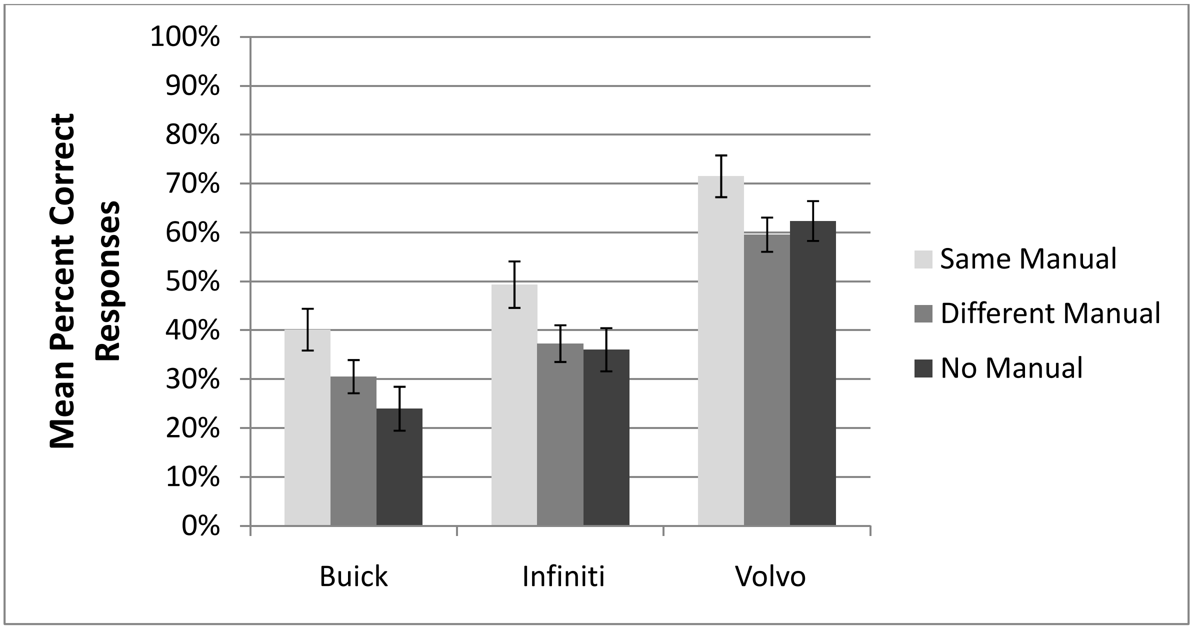

Figure 7. Mean percent correct responses for all systems (with standard error bars)

The implication of the findings is that as vehicles become differentiated in terms of the safety systems they include, the operational aspects of those systems, and the strategies for interface design, drivers may have difficulty understanding what is in the vehicle and what the status is (on or off, working properly, settings, currently functioning). There may be advantages to some level of consistency between vehicles for certain aspects of the interface, such as terminology, icons, acronyms, color coding, or location. Furthermore, since the improvements in comprehension as a result of reading manufacturer-provided information were rather limited for all three systems, enhancing the content, medium, or perceived value of these materials may help to improve driver awareness and familiarity with ACWS features.

3.6 Summary of ACAT test protocols

Research conducted under the Advanced Crash Avoidance Technologies (ACAT) program was reviewed for the application of the methodologies to CWIM issues. ACAT, funded by the NHTSA, is an effort to develop a basic methodological framework and computer-based simulation model to estimate the effectiveness and potential safety benefits of various crash countermeasure systems. The program includes a variety of parallel efforts, conducted by different manufacturer-led research teams, to address distinct application areas (e.g., LDW, FCW, backing warning). The program developed a standardized Safety Impact Methodology (SIM) – a computational tool that provides a framework for estimating safety benefits based on the results of objective tests of full vehicle systems. Data to support the SIM tool are drawn from objective tests designed to characterize and assess countermeasure system performance under a representative set of crash scenarios addressing a given safety problem. In general, performance-based tests characterized a system's ability to respond to obstacles (performance envelopes) under a range of situations, as well as false alarm performance assessing the extent to which countermeasures are likely to issue unhelpful alerts, warnings, or interventions (false system activations). Some ACAT models also included driver-in-the-loop performance tests using naive participants to gauge driver interactions and performance in response to the system in order to evaluate the effectiveness of a given countermeasure for breaking the chain-of-events leading to a crash. Thus, ACAT tests are intended to capture and represent different aspects or dimensions of system performance (e.g., system response envelopes, sensitivity to different types of hazards and in-path obstacles, etc) with data supporting computer-based models to estimate safety benefits and effectiveness assessments of advanced in-vehicle crash avoidance systems.

Some of the ACAT work included consideration of the human/machine interface (HMI), including the crash warning DVI. Although it represents just one of several inter-related components or sub-systems, experience has shown that the DVI must support effective crash avoidance by drivers; the DVI encompasses the displays and controls through which the driver and system interact. A poorly designed or implemented DVI may negate or substantially reduce system benefits if it confuses, annoys, distracts or otherwise impedes the driver's ability to quickly process and appropriately respond to system alerts, warnings, or control interventions. Currently, no uniform means exists for evaluating or quantifying the safety-related performance of crash warning DVI applications. Although work conducted under ACAT has advanced the state-of-the-art for estimating system effectiveness, it does not directly address the DVI issue.

This task of the CWIM project examined the methodologies used in the various ACAT performance evaluations and considered their implications for CWIM objectives. Two summary tables follow. Table 6 describes the general testing methods and approaches used by each ACAT team (inclusion of focused DVI testing, use of drive-in-the-loop tests, assessment methods, test scenarios, samples, evaluation measures and metrics, etc.). Table 7 identifies some of the key issues for assessment protocols and draws initial implications or suggestions from the ACAT experience.

Table 6. Summary table of ACAT objective tests & relevance to DVI assessments

Application Area (Driver Assistance System)	**Project Team**	**Focused DVI Evaluation?**	**Driver-in-the-Loop Tests?**	**Assessment Methods & Procedure**	**DVI Modes Assessed**	**Scenarios**	**Evaluation Measures & Metrics**	**Pass/Fail Criterion?**	**Notes**
Side Obstacle Warning (SOW), Lane Departure Warning (LDW)	Nissan and UMTRI	Yes Includes "Human Factors/HMI Testing"	Yes	Simulator & Survey 2 hour session, 16 subjects	Audible and blinking visual warning (properties unspecified)	Lane changes evoked in "natural ways." Specifies 2 scenarios (following lead that changes lanes, lane closure requiring merge). Host vehicle speed of 45mph	Lists "raw" vehicle simulator data. Primary measure: Probability that lane change is attempted, and latency of maneuver to abort following warning activation	None Specified	▪ Task 3 Interim Report for ACAT-2 Development of Objective Tests (October, 2009) includes a proposed plan for the evaluation. Study not run; no data available. ▪ Focus of HMI testing is to determine how drivers will respond to SOW warning preceding or during lane change. ▪ No treatment of LDW
(Forward Collision) Pre-Collision System: Providing Warning, Brake Assist, and Pre-Collision Brake (Automatic Braking)	Toyota	Yes, Limited Data captured to map driver reaction to PCS warning only	Yes	Simulator 133 drivers (64 valid usable cases), three age groups (20-30, 40-50, 60+)	Audible warning (unspecified)	Two scenarios: Lead Vehicle Stopped and Lead Deceleration (at 0.7 g). Host vehicle speed of 45 mph, Time gap to lead controlled at 1.8 sec, warning issued at TTC of 2.0 sec. Vehicle in adjacent lane used to discourage steering avoidance. Distraction task used to elicit off-road glance	Response to warning and deceleration time-history : Brake RT, Braking Level (max decel), and Speed reduction	None Specified	▪ Discussion does not focus on DVI assessment, but uses data from simulator as input to SIM model. ▪ No baseline or control group used in Simulator study (all conditions included warnings). Baseline data gathered via existing NHTSA Event Data Recorder data. ▪ Also conducted test track study with trained driver to evaluate brake system performance.

Application Area (Driver Assistance System)	**Project Team**	**Focused DVI Evaluation?**	**Driver-in-the-Loop Tests?**	**Assessment Methods & Procedure**	**DVI Modes Assessed**	**Scenarios**	**Evaluation Measures & Metrics**	**Pass/Fail Criterion?**	**Notes**
(Lane Keeping) Driver Alert Control (warns of reduced alertness, vigilance) Lane Departure Warning (LDW) Emergency Lane Assist (ELA; provides active steering intervention)	Volvo, Ford, and UMTRI	Yes Includes human factors testing with naïve subjects (primarily HMI testing for LDW systems).	Yes Model parameters for fatigued and distracted drivers derived in large part from data captured in driver-in-loop testing. Primarily limited to LDW systems	Simulator Three simulator studies: 1) sleep deprived drivers; 2) distracted drivers; and 3) assess effects of driver interference during ELA intervention Also included test track and public road evaluations.	LDW warning modes: steering torque, rumble strip sound, steering wheel vibration, and Head-Up Display (conditions created using combinations); haptic belt and warning.	Sleep deprived study: 3 hour drive, night-time, 60-70 mph, 2-lanes each direction separated by median, 23hr sleep deprived drivers; randomly induced "yaw deviation" lane drifts. Distracted driver study: Same as above except - secondary task, 20 minute daytime drive. N=16 (Test track) LDW HMI Acceptance & Usability testing had 23 employee participants, haptic belt vs. audio warning. Participants deliberately activated the warning system. Balloon car target. (Public road) Data collection to asses system performance & availability, >2000km in several countries.	Captured driver response to imminent lane departure events. Steering-based reaction time (peak steering rate), head turn reaction time (for distracted drivers),	None Specified	▪ Report does not comprehensively cover results of DVI mode tests across modes. Number subjects run in studies is not clear ▪ Found mean steering response time was faster for sleep deprived compared to distracted driver s; 0.66 vs. 0.99 sec, respectively. (I believe this result may be due to differences in urgency from scenario – day vs limited sight distance at night, and driver situational awareness) ▪ Steering data may not be sensitive, many cases with no clear steering response, therefore may be unreliable ▪ Driver compliance, adaptation, and unintended consequences issues not investigated ("outside scope").

Application Area (Driver Assistance System)	Project Team	Focused DVI Evaluation?	Driver-in-the-Loop Tests?	Assessment Methods & Procedure	DVI Modes Assessed	Scenarios	Evaluation Measures & Metrics	Pass/Fail Criterion?	Notes
(Forward Collision) Advanced Collision Mitigation Braking System (CMBS). Collision detection system combining audio-visual warnings, auto-braking, and seatbelt pretensioning	Honda and Dynamic Research Inc.	Limited Some testing with naïve drivers is captured to "evaluate driver-vehicle system response and to measure the driver's and vehicle's response characteristics to system warnings and interventions."Focus was not to evaluate DVI. Evaluated display luminance.	Yes Intended goal to "measure driver response and behavior, with and without ACAT"	Simulator (included test track study, but with trained driver) Simulator tests included naïve drivers with and without ACAT under 4 Technology Related Crash Types across 12 scenarios. Used distraction task to divert eyes from forward roadway. 12 drivers, ages 25 to 51, 57% males	CMBS provides staged levels: 1) Visual and audio warning, 2) Light tactile signals (light seatbelt tug and braking at 0.2g) 3) Strong tactile signals (strong seatbelt retraction, and braking at 0.6g)	Simulator study used 12 scenarios across four technology relevant collision types: Head-on, Intersecting path, Pedestrian, and Rear-end. Cruise control was engaged to maintain speed profile. Runs included 8 events (2 critical and 6 null)	Measured driver response (i.e., glance, brake and steer time delays and amplitudes) to the ACAT warnings	None Specified	▪ Limited data presented on relative effectiveness of ACAT in DIL tests ▪ Crash scenarios are based on real crashed, and detailed in an appendix to the project report ▪ Sound measurements taken to assess relative noticeability of audible warnings (engine off, idle, and at 30 mph) ▪ Attempted to standardize the simulator conflict scenarios ▪ Simulator study repeatedly exposed same drivers to crashes with and without ACAT (within-subjects) ▪ Test track study with trained drivers to calibrate SIM tool; used guided soft targets and 11 conflict scenarios (same as simulator)

Backing Crash Countermeasures (Park Aid, Rear Vision, Backing Warning, Auto-Braking)	GM and VTTI	Yes, limited Testing characterized driver interactions with and responsiveness to countermeasure classes, but not alternative implementations	Yes	On-road. Five parking lot scenarios requiring 40 naïve drivers who executed backing maneuvers , some trials included rear in-path obstacles.	Visual (Rear Vision, Park Aid), Auditory, Haptic (brake pulse), and Automatic Braking	Included pedestrian and fixed object scenarios mapped to real-world crashes. 1. Intermediate Static Child Ped 2. Near incurring child ped 3. Intermediate incurring child ped 4. Near static vehicle 5. Intermediate static pole Details related to scenarios are documented in the report, including: prior experience, age range, target travel speeds, etc.	Obstacle detection & avoidance, driver search and reliance on Rear Vision, system activations, responsive to warning, Response latency to warnings, response to interventions.	None Specified	▪ Special considerations required for staging, included amount of previous exposure to countermeasures. ▪ Obstacles always presented after experiment supposedly "ended" to create surprise situation ▪ False alarm testing also performed to map system performance under real-world environments

Table 7. Implications for the development of CWIM-based DVI assessment protocols derived from ACAT objective tests

ISSUES	IMPLICATIONS FROM ACAT DEVELOPMENT OF OBJECTIVE TESTS
Evaluation of Integrated or Stand-alone Advanced Crash Warning Systems (ACWS)	▪ ACAT focused on benefits assessments of individual Advanced Crash Avoidance Technologies, and structured individual projects to address specific technology applications (e.g., Forward Crash, Lane Departure, Backing, etc.). The implication is that CWIM's initial focus and emphasis should be on stand-alone, not integrated, systems; this simplifies the problems and challenges, and makes the task manageable. However, it is important to recognize that a larger system integration issue exists and will become increasingly relevant as a broader range of systems are introduced into the light vehicle fleet
Integrated Countermeasures Within a Single ACWS Function (Progressive Countermeasures and Combinations)	▪ Although individual ACAT teams focused on specific functional applications (e.g., Forward Collisions, Backing Crashes), many of the systems evaluated within the ACAT framework included sets of integrated countermeasure features within a single application. This approach included progressive warnings, or combination of warnings, and active interventions (e.g., audible/visual forward collision warning, seat belt tug, automatic braking intervention). This presents a challenge in terms of measuring and attributing driver responses to individual system elements, particularly when the time-line of features is closely coupled. The issue is similar to the evaluation of integrated systems that address two or more functional applications. It should not be assumed that ACWS systems will provide for the opportunity to activate and de-activate individual features. CWIM evaluation protocols should allow for the evaluation of the overall Interface (whether single or combination of DVI's), and provide for measures to be captured with sufficient resolution to allow responses to be mapped to individual warnings, or intervention combinations. This could involve examining driver response trajectories at fine-grain intervals both with the composite ACWS system and with component pieces in isolation. Data collection systems with relatively high sampling rates may facilitate this type of fine-grain analysis.
Evaluation Approach & Method	▪ ACAT approaches and methods were geared towards the evaluation of commercial or near-production systems, as opposed to early product designs; emphasis was devoted to overall system performance and not the DVI aspects. CWIM evaluation protocols should parallel ACAT and target the evaluation of production-ready Driver-Vehicle Interfaces operating within the context of an overall system. ▪ ACAT evaluation approaches generally involved use of high-fidelity, controlled environments (test track, simulator, parking lots, etc.) to present drivers with representative and realistic driving environments and tasks. The use of simulation to mimic or occasion real-world situations was commonly used. Most, but not all, "driver-in-the loop" assessments designed to characterize driver responsiveness to system outputs (warnings and control interventions) took place in a simulator. One notable exception was work by GM/VTTI team who performed all of their "driver-in-the-loop" testing in actual settings. CWIM assessment protocols should seek to replicate important real-world scenario characteristics and tasks under realistic settings; evaluations should provide the opportunity to make use of real-world settings and environments, where feasible.

Scenario Selection and Development	▪ The ACAT program devoted significant effort to defining the crash problem, with results generally yielding a wide array of crash types, situations and conditions. As a result, scenarios developed under ACAT were selected to be representative of the real-world crash problem, and, to a large degree, also targeted to address Technology Relevant Crash Types – crash scenarios where it was believed the ACAT could intervene or otherwise act to mitigate and/or avoid the crash. CWIM should build on the initial foundation of crash scenarios defined under ACAT since they represent real-world crash problems and will serve to advance a set of evaluation scenarios useful for system and DVI assessments. ▪ As a whole, ACAT identified and generated a large number of crash test scenarios. In many cases, efforts were undertaken by ACAT teams to reduce, or otherwise prioritize, the number of potential evaluation scenarios by consolidating and/or selecting specific individual scenarios for testing based on crash frequency and/or severity. The number of ACAT evaluation scenarios ranged from one to twelve. CWIM will necessarily need to define a manageable set of assessment scenarios that capture and prescribe key details of the driving environment. Initial efforts will likely need to start with a limited number of representative scenarios; this will require developing or defining a decision strategy for down-selecting among available scenarios ▪ The rational for the selection of scenarios should be clear and consistent; avoid selection on the basis of sensor capabilities, or other factors that provide an advantage for one particular concept solution over another. ▪ ACAT teams tended to make a large number of simplifying assumptions, such as limited traffic scenarios, weather, and speed conditions, in order to produce workable testing scenarios in light of missing or limited data, and/or limited available resources. CWIM scenarios must be sufficiently prescriptive to enable replication of important parameter values, including vehicle dynamics (speeds, headways, etc), road geometry, traffic conditions, environmental characteristics (day/night, dry/wet, etc), event timing, and driver states and activities, among other aspects. Initial CWIM scenarios will likely need to define a set of simplifying assumptions – these should be clearly documented.
Long Term Effects & Behavioral Adaptations	▪ ACAT evaluations typically relied on novice users and presented drivers with limited duration (single session) exposures to advanced features and functions, often under idealized conditions (e.g., limited false or nuisance activations). This approach fails to capture a range of system usage experience and exposure levels which potentially limits generalizability and may not necessarily reflect performance characteristics of experienced users. Ideally, CWIM should attempt to study behavior over a longer time-course, as driver interaction with systems may change with familiarity and exposure, particularly in regard to false alarms. Protocols should allow for a range of users to include experienced users in order to identify potential behavioral adaptations resulting from extended system exposure. In some cases this may require giving participant extensive system training; however, for novel systems it should be recognized that even training will likely fail to replicate the experience of users who experience the technology on a regular basis over the long term. Realistically, it may only be possible to test using a defined sample of novice users (refer to sample demographics section)

Modality Comparisons	▪ ACAT studies tended to assess a variety of DVI interfaces within a single application, often making comparisons among alternatives and inferring effectiveness on the basis of performance and subjective preferences. For technologies with multiple warning modalities or interfaces (e.g., haptic, audio, visual) and interventions (warning vs. active control), it would be beneficial to examine these in both isolation and combinations to ensure no unintended consequences. Further, it would be beneficial to ensure that warning modalities that occur together not interfere with one another across countermeasure technologies (e.g., confusing warning tones).
User-Adjustable Settings	▪ ACAT studies did not discuss whether any countermeasure settings were user-adjustable, or what level these were set on if they were. It is reasonable for CWIM testing to focus on default settings for user-adjustable systems, as these are likely what most users will remain on.
Scenario Parameters	
Instructions to Participants	▪ ACAT studies did not generally report detailed participant instructions or protocols, and therefore limited guidance is provided on the basis of the available documentation.
Sample Size	▪ ACAT evaluations included substantial variation regards the number of participant used, ranging from 8 to 35 drivers. The Toyota ACAT study had approximately 8 participants per condition, while Honda-DRI assessments used 12 participants each of whom experienced all scenarios. GM/VTTI used 8-35 participants. As a general rule of thumb, CWIM should have a sufficiently large sample size to provide a basis for reliable statistical analyses (at least 8-10 participants per condition); a power analysis should be conducted to determine the requisite number of participants necessary for a target effect size.
Sample Demographics	▪ ACAT studies tended to draw samples using novice (inexperienced system users) or employees of the institution conducting the studies. Characteristics of the sample should be generally representative of the driving population, and the crash problem in particular; several approaches are possible. Novice users may help identify common problems associated with initial learning or interactions with the HMI (and can sometimes be used represent a "worst "case situation), while experienced users may provide insight into potential issues that emerge over time. CWIM protocols should ensure that the sample is generally representative of the driving population; ideally, testing should include both novice and experienced users, since experience with a system may result in qualitative changes in behavior. Realistically, initial assessments may need to make use of convenience samples comprised of novice system users with some defined level of instruction, training, or practice provided.
Operation Issues	
False and Nuisance Alarms	▪ It appears that none of the ACAT studies specifically address false alarms, although the Nissan/UMTRI proposed research alludes to mixing presence/absence of conflict with presence/absence of SOW warning. CWIM should focus on both driver acceptance of false alarms and any associated performance consequences (e.g., learned neglect of the alarm). This should be undertaken in conjunction with longer-term system use, as driver reactions to false alarms may change over time and experience. At a minimum, information should be compiled to characterize the False Alarm rate of the system.
Starting and Triggering Conditions	▪ No universal or standardized method of triggering system warning and events emerged across ACAT studies. Strategies included manually triggering events during distraction episodes (Honda-DRI & Volvo-Ford-UMTRI) and automated triggers such as TTC (Toyota). CWIM may benefit from standardized triggering protocols to increase consistency and aid comparison across studies.

Driver Expectancy	▪ As detailed protocols were not included in the ACAT reports, it is unclear what drivers were told to expect. In some cases, participants experienced multiple crash scenarios, which likely raised expectancy. For "surprising" incidences like unexpected conflict vehicle movement, it would be valuable for CWIM to limit exposure and to use a ruse to minimize driver expectancy of potential emergency events.
Distraction Tasks (Lapse in Attention)	▪ Attentional lapses in ACAT were induced by a variety of visual or visuo-cognitive distracter tasks. Honda/DRI asked participants to gaze at a secondary task light located on the passenger's door when an audible cue occurred. This distraction preceded the critical event by approximately 2s. The Toyota study had participants locate a convenience store using a 7" LCD map display mounted low on the center stack, near the shift lever; the authors noted that this task was difficult and somewhat less natural than alternatives such as making a cell phone call. In the Volvo/Ford/UMTRI study, experimenters distracted alert participants 40 times during a 20-minute drive, during which yaw deviation was introduced on 16 trials. The secondary task involved reading aloud a series of 6 digits presented on a screen located low on the center console near the passenger seat; display time was 0.3s per digit with 0.2s blank between digits. No standard practices or rationale emerged for distraction/secondary task selection. CWIM testing could benefit from standardized secondary tasks sampled from the distracted driving literature. In addition, naturalistic studies such as Klauer et al. (2006) could provide insight on the types of distracting tasks that drivers engage in regularly.
Acceptance Criteria	▪ ACAT driver-in-the-loop studies were intended to generate data for use in modeling, and therefore did not necessarily define or identify pass/fail criteria. In contrast, CWIM analysis protocols should enable analysts to isolate and localize problems with the DVI in order to remedy areas that need improvement. This requires the development of specific performance criteria.

4 Conclusions Regarding Warning Modes

Three of the empirical experiments associated with this project provided comparisons of alternative ACWS DVIs which differed in terms of the modality used to convey the warning. Traditionally, in-vehicle warnings have been presented in acoustic and/or visual modes. Various sorts of haptic warnings have been the subject of recent research and have begun to appear in production vehicles. One objective of this project was to compare various modalities in terms of how they influence the driver response to FCW or LDW alerts. In particular, the more recent and innovative haptic and active systems bear comparison with more traditional modes. The experiments in this project provided informative comparisons. It should be made clear that, consistent with the general CWIM effort, the interest was with DVI typical of current or near-production systems, not purely experimental possibilities. Therefore the haptic and active systems were based on current domestic or foreign vehicle examples.

In any of these modes – visual, auditory, haptic – there are diverse methods of display. For example, visual displays may include HUD or locations other than the instrument cluster; auditory displays might include speech or might include location cues; haptic signals may be located in the seat pan, steering wheel, accelerator pedal, or seat belt system. Active warnings might include braking or steering and may range from mild to aggressive in their action. Details of the display, such as intensity, temporal patterns, locations, and specific message content, vary substantially.

The CWIM project was not designed to comprehensively compare this very broad range of display and mode alternatives, so the limited set of comparisons made for this project could not encompass all of the possibilities. The intent was to employ prototypical examples of each mode to provide a general comparison of driver response to each. Therefore, the conclusions provided in this section emphasize general principles and relative effectiveness of general concepts rather than detailed design specifications.

4.1 Warning modes for LDW alerts

LDW alerts in various modes were compared in the driving simulator experiment described in Section 3.1. The experiment included five warning conditions, including two "active" interventions that provided countersteer during a lane drift. The five conditions were: (1) passive – acoustic alert; (2) passive – tactile alert; (3) active – weak intervention; (4) active – stronger intervention; and (5) control – no LDW. All four shared the use of an icon to indicate the operational status of the system (i.e. off, on, or tracking). Although it is recognized that commercial systems might combine warnings in various modes (e.g., an acoustic signal might accompany an active countersteer), the intent in this experiment was to examine the effect of each mode as an independent DVI.

The experiment captured a wide range of driver performance variables, and detailed analysis and discussion of the full range of data from this experiment may be found in the project interim report (University of Iowa and Westat, 2010) on this experiment. This section focuses on the overall conclusions comparing driver response to the various warning modes.

- Initial steering response: The speed and magnitude of the initial steering response was best for the active steering mode with weak torque. Neither of the passive (auditory or haptic) modes nor the strong torque condition improved initial steering relative to the baseline control condition. Figure 8 shows example findings for steering response time.

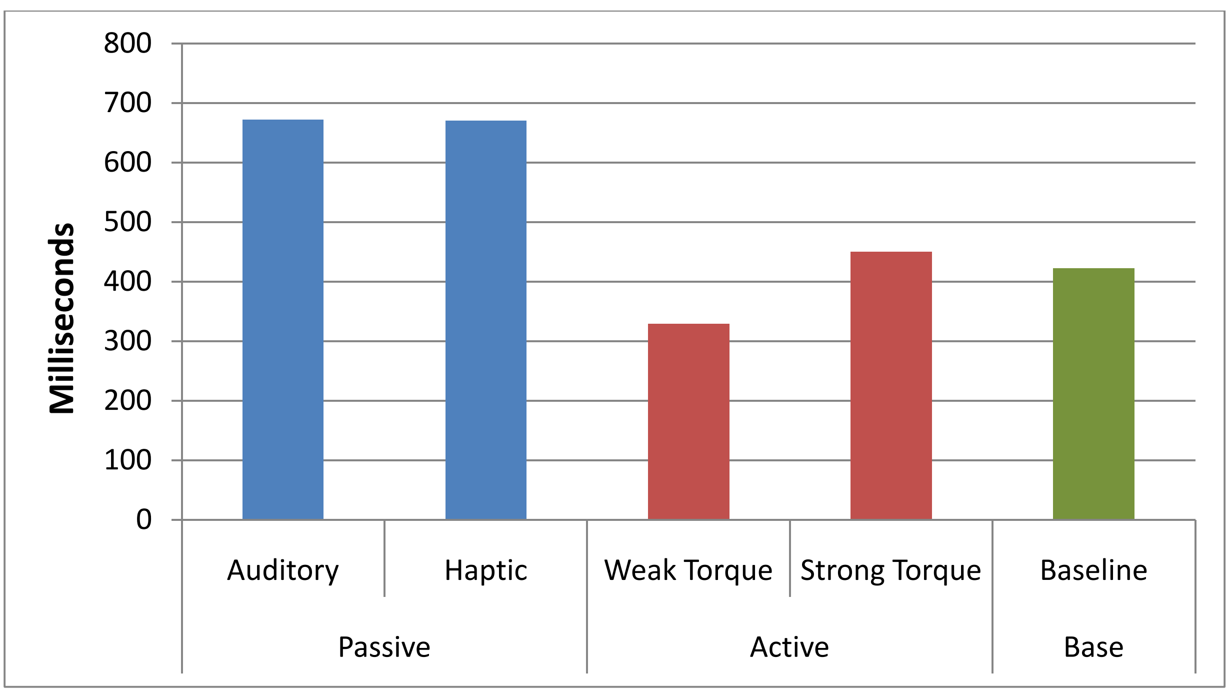

Figure 8. LDW steering response time findings

- Lane position: The weak torque active warning was the most effective system in terms of the various measures of lane position: standard deviation of lane position, maximum lane exceedance, duration of lane exceedance, and lane exceedance exposure (area). The relative effectiveness of the other warning modes depended upon the particular measure, but in general they were superior to the baseline control condition. Figure 9 shows an example of the findings, in this case for the maximum lane exceedance measure. There was an interaction of the system type with the particular event scenario (drift left, drift right, depart curve) but the general picture shown in the figure is representative. The magnitude of the effect may be seen in that the mean maximum lane exceedance for the weak active system is slightly more than half that of the auditory warning, and only about 40% that of the control condition, which exceeded 1.6 meters. Thus all warning modes appear to be somewhat effective in limiting the magnitude of lane exceedance, but the weak active mode is notably effective.
- Inappropriate behaviors: Various driver actions or failures are of interest, such as crossing lanes, running off the road, collisions with other vehicles, and inappropriate braking. The active warning systems were superior in limiting the number of these outcomes, as shown in Figure 10. For these measures, the weak torque active version was not superior to the strong torque.
- User acceptance: User acceptance findings contrasted with the driving performance measures. Participants rated the active systems less effective in catching their attention than the passive systems and active systems were seen as less helpful, interpretable, and reliable. These issues were generally more pronounced for the weak active system than for the strong active system.
- Response to false alarm: The simulator drive included a false alarm scenario, in which a lane departure alert was generated by incompletely removed old lane markings in a work zone. The vehicle was not actually moving out of its proper lane. The primary finding from the false alarm event was that the strong active system led to notably greater peak steering rate, steering acceleration, steering jerk, and startle response. Figure 11 shows this for the peak steering measure. Thus there may be a concern that driver actions in response to the strong active system could cause conflicts under false alarm conditions.

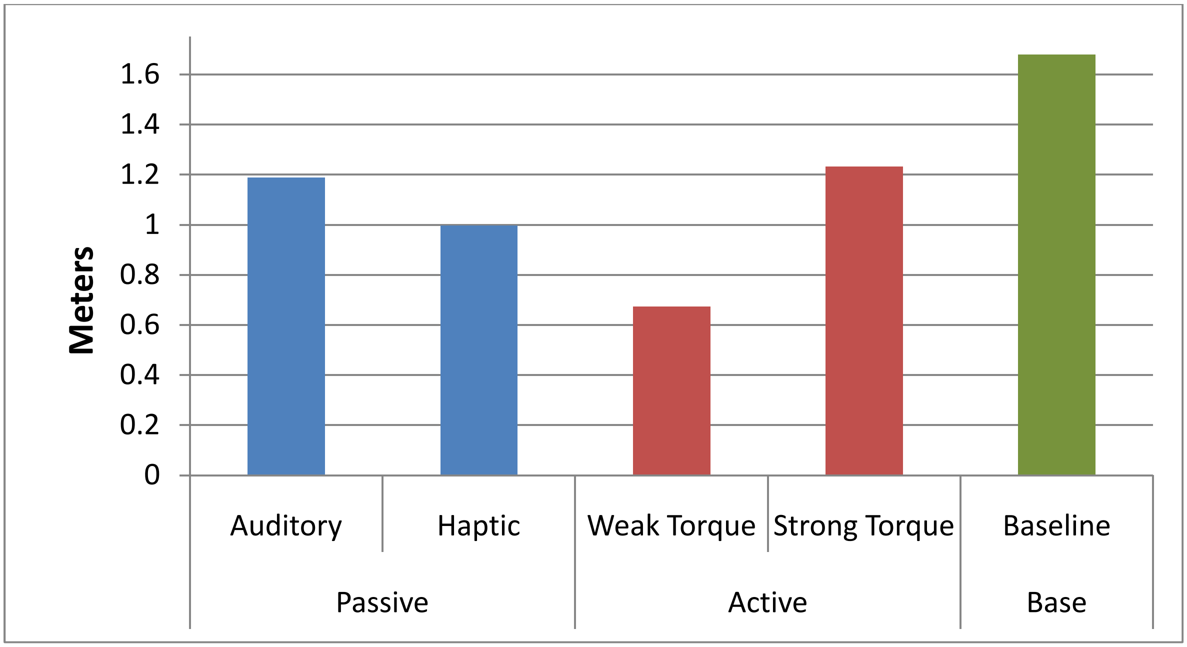

Figure 9. LDW maximum lane exceedance

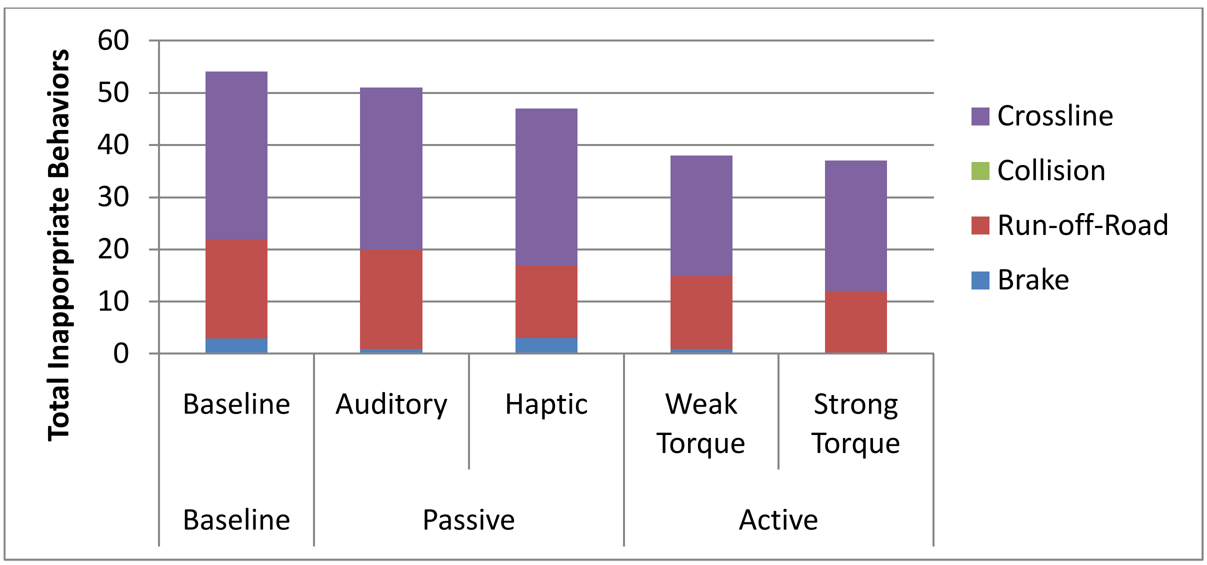

Figure 10. Inappropriate behaviors

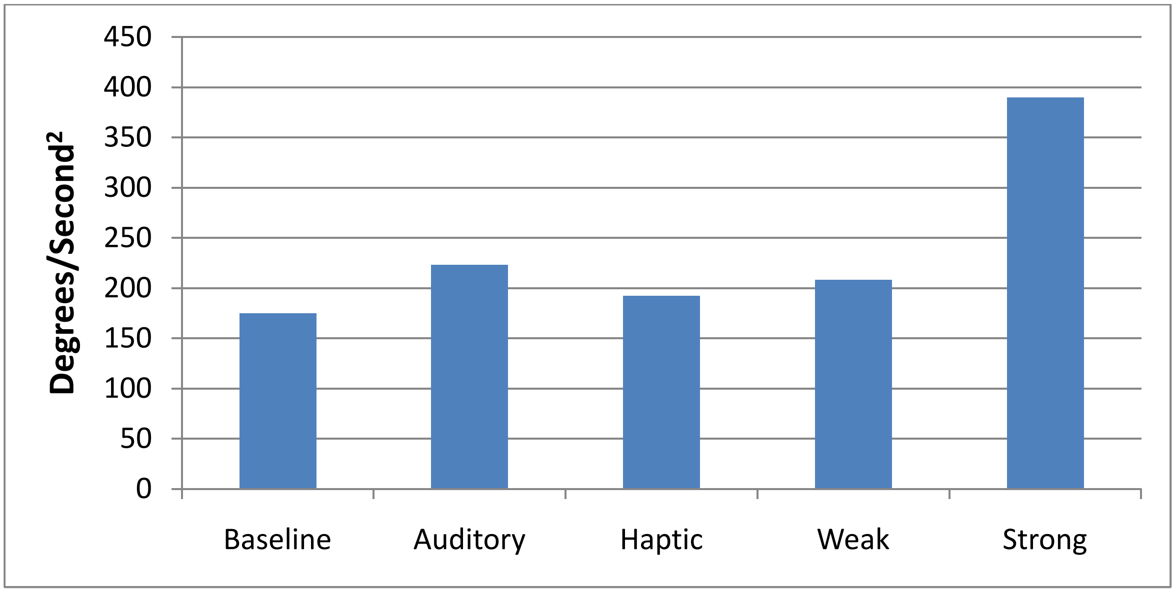

Figure 11. Peak steering acceleration in a false alarm situation

In overall conclusion, all of the modes tested had some beneficial influence, but the weak torque active system was generally superior in terms of reducing lane deviations. The haptic mode was generally at least as effective as the acoustic signal and therefore is a reasonable candidate for commercial application. User acceptance measures were not consistent with driving performance measures, with greater acceptance of the passive modes. This may suggest that active systems may be more acceptable if they also include an acoustic or haptic component, but such DVIs were not included in the study. The findings suggest that comprehensive LDW evaluation protocols should be able to accommodate active as well as passive systems and should include metrics related to user acceptance and negative behaviors (including false alarms).

The findings comparing the weak torque and strong torque active conditions are somewhat surprising, in that the weaker system appeared to provide better results across a range of measures. Subsequent research to better define the features of effective active LDW systems is suggested.

4.2 Warning modes for FCW alerts

FCW alerts in various modes were compared in two experiments: a driving simulator experiment (described in Section 3.2) and a test track experiment (described in Section 3.3). The two studies used somewhat different sets of DVIs. The driving simulator experiment investigated three DVIs (including the no warning control condition), and tested across two event scenarios (braking lead vehicle, stopped lead vehicle). The test track study investigated eight DVIs (including the no warning control condition) with a single event scenario (lead vehicle changes lanes to reveal a stopped vehicle ahead). As shown in Table 8 the two experiments shared in common only one DVI (the visual HUD and auditory beep) in addition to the no alert control group. While the two experiments are complementary, they were not designed to be directly comparable.

Table 8. FCW alert characteristics

FCW Alert Modality	Test Track Exp (VRTC)	Simulator (NADS-1)
No alert (control)	X	X
HUD	X	
Auditory beep	X	
Seat belt pre-tensioner	X	
HUD and beep	X	X
HUD and seat belt pre-tensioner	X	
Beep and seat belt pre-tensioner	X	
HUD, beep, and seat belt	X	
Brake pulse, active partial control		X

The two FCW modality experiments captured a wide range of driver performance variables, and detailed analysis and discussion of the full range of data will be found in separate project reports (Forkenbrock et al., 2011 for the test track study; the simulator study report will be provided under an independent NHTSA contract). For purposes of this section, we focus on the overall conclusions comparing driver response to the various warning modes.

The general finding of the VRTC test track study was that across a broad range of performance measures, systems that included the seat belt pre-tensioner were superior to the other DVIs. This was true for measures of redirecting visual attention, driver response initiation (e.g., throttle release, braking or steering initiation), and crash-related measures such as crash occurrence and time to collision when the participants first returned to a forward-facing viewing position after secondary task completion. The DVIs with the auditory component (but without the seat belt pre-tensioner) were generally superior to the visual (HUD) only system, which had little advantage relative to the control condition. The lack of a visual alert effect could be a result of the distraction task used in the study; any distraction task that diverts visual attention from the forward view is likely to degrade responding to a visual display in the vicinity of the driver's forward field of view. Figure 12 illustrates these points with the mean throttle release times as an example. The four experimental conditions that included the seat belt pre-tensioner (the four green columns on the left side of the figure) had faster release times than the two systems with the auditory component (right side of figure), while the control and HUD-only conditions were slowest. Table 9 summarizes crash outcomes, showing the percent of participants who avoided collision with the stationary balloon vehicle. No control participants avoided the crash and only 2 of 22 participants with the non-seat belt pre-tensioner systems avoided the crash. However, more than half of the participants with the seat belt pre-tensioner systems successfully avoided a collision.

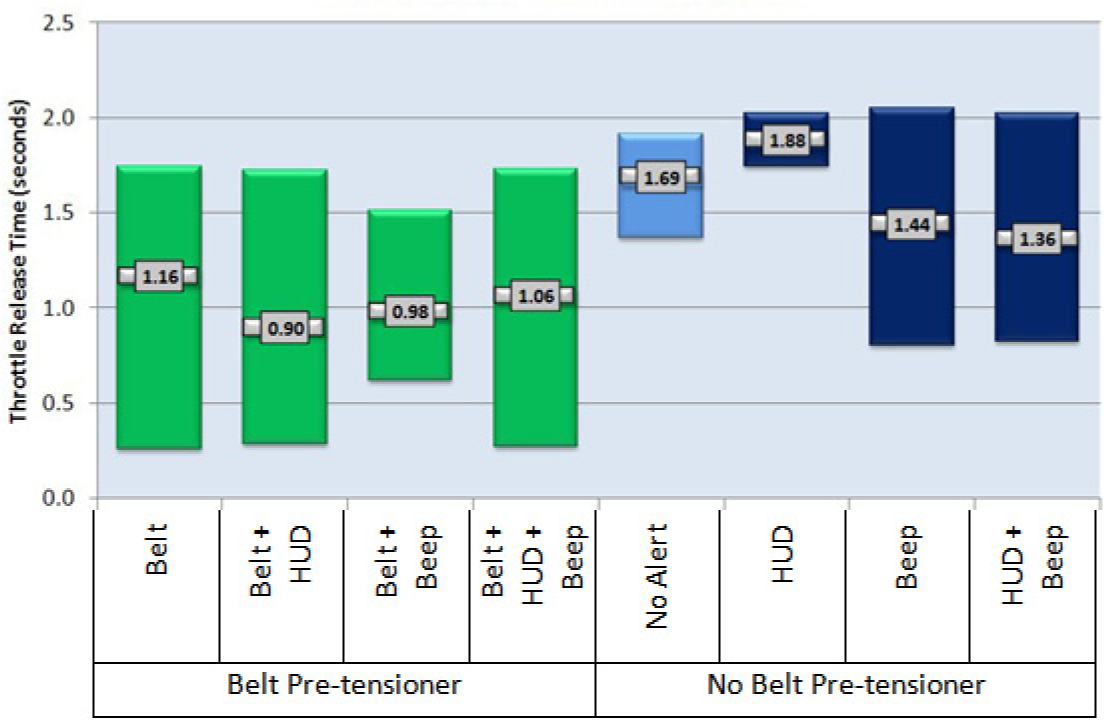

Figure 12. Throttle release times for FCW systems

Table 9. Crash avoidance outcomes for FCW systems

	Systems with seat belt pre-tensioner	Systems without seat belt pre-tensioner	No warning control
Crash avoided	17	2	0
Crashed	15	22	8
% crash avoided	53%	8%	0%

The findings suggest that the seat belt pre-tensioner is an effective warning (in addition to its crash injury mitigation effects). While the direction of some findings suggests there may be some advantage to adding an additional (auditory or visual) warning component to the seat belt pre-tensioner component (for example, see Figure 12 above), this is uncertain and not as robust as the seat belt pre-tensioner effect itself. The reason for the strong performance of the seat belt pre-tensioner systems is not evident. It may be inherent in the physical stimulus, it may be because it uses a sensory channel not used by other displays, it may be a novelty effect, or it may be because of the immediate association with the “crash” message. It is also not known if the effectiveness may be related to driver postural position associated with turning to engage in the distraction task. Since this experiment did not include other haptic alerts (e.g., vibrating seats), it is not known how the seat belt pre-tensioner compares to such alternatives. Clearly, however, seat belt pre-tensioner systems merit further consideration. DVI evaluation methods would benefit from the capability to simulate such systems as well as other haptic alerts.

The driving simulator FCW study included a haptic warning cue, though one quite different from the seat belt pre-tensioner. In this case, it was a brake pulse stimulus, which in addition to providing a movement sensation also provided some (though quite mild) automatic active partial control, since the pulsing slightly slowed the vehicle. This slowing, however, was quite minimal and best should not be viewed as a crash mitigation measure in itself.

The auditory/visual and the brake pulse DVIs each led to similar decreases in driver response time (throttle release and brake activation), relative to the control condition. Figure 13 shows this for brake response time data. The magnitude of the effect is comparable to that of the auditory/visual DVI in the test track study. Although the response times are similar, the auditory/visual and brake pulse systems led to somewhat different braking behavior. The peak brake force was less for the brake pulse condition and the peak deceleration was lower. Figure 14 shows this for the peak deceleration data that compares the three warning conditions. As might be expected, somewhat shorter minimum times to collision were found with the brake pulse system than with the auditory/visual warning. (Minimum time to collision was calculated as a function of instantaneous relative speed (between the subject vehicle and lead vehicle) and distance between the two vehicles.) However, for both warning systems the minimum distance or time to collision was substantially longer than in the control condition and there was no difference between the two systems in terms of collision avoidance. Figure 15 shows the substantial effect for the minimum time to collision measure. The indication is that while both DVIs speeded the reaction to the potential crash situation, the brake pulse system produced a more gradual and controlled braking profile. If this finding is generalizable, such as system may have additional advantages in preventing secondary crashes with a following vehicle.

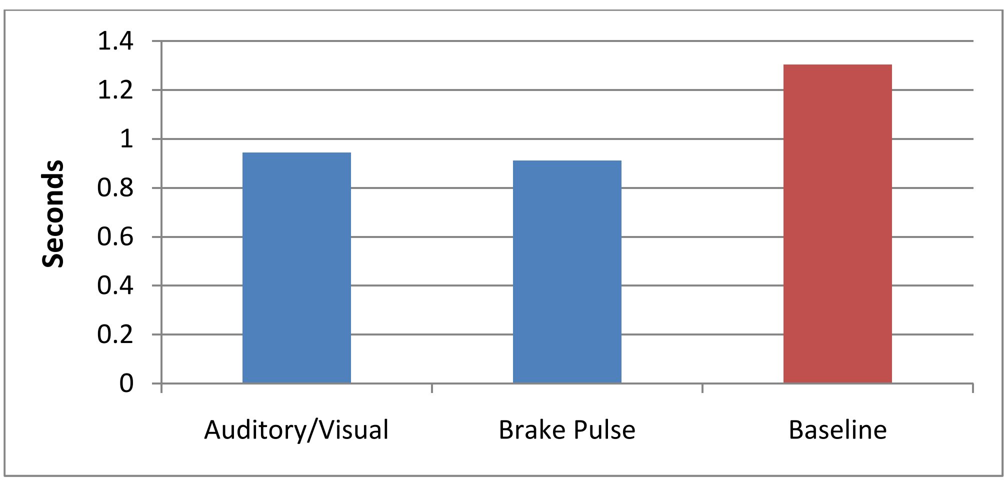

Figure 13. Brake response times for FCW systems

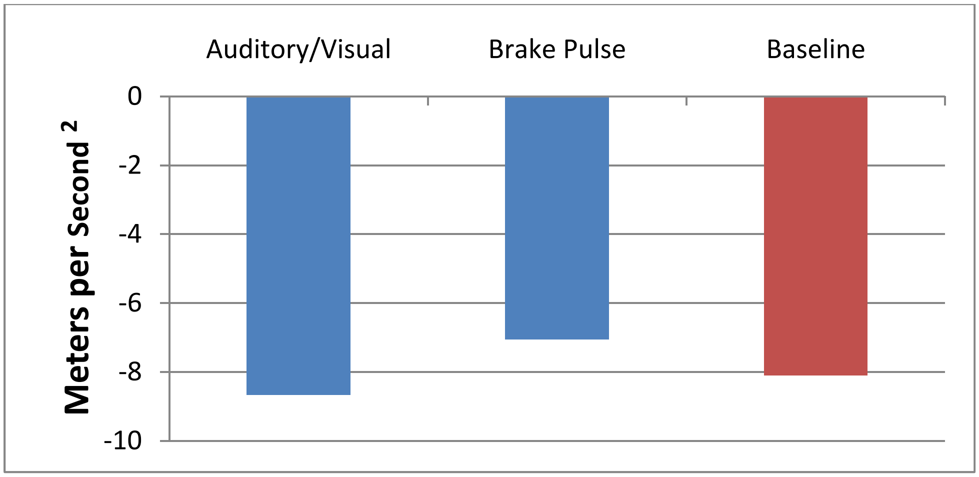

Figure 14. Peak deceleration while braking for FCW systems

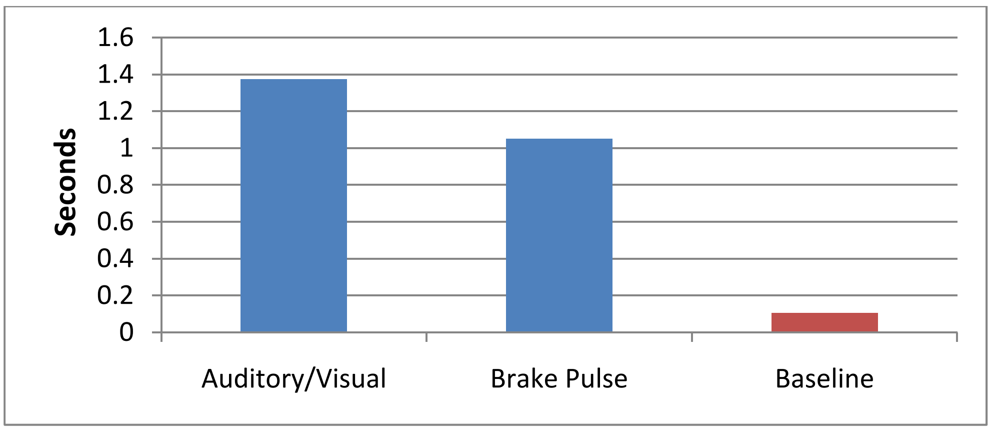

Figure 15. Minimum time to collision for FCW systems

Taking the findings of the two FCW experiments together, the results suggest that there may be advantages to innovative DVIs such as brake pulse or seat belt pre-tensioning. Visual-only DVIs may have only limited effectiveness. DVIs with an auditory signal improve driver response relative to the no-warning control, but may not prove as effective as the seat belt pre-tensioner or brake pulse. Certainly more exploration of these alternatives is merited. Such modalities may have additional advantages as more safety-related systems and communications technologies become present in vehicles, because they do not share the overloaded visual and auditory modes. This is an important consideration for system integration.

5 Conclusions Regarding DVI Consistency

As noted in Section 1.1.3, ACWS DVI variability among different vehicles might contribute to a lack of knowledge transfer between vehicles. Two experiments in this project provided a preliminary look at this issue. One experiment dealt with the effect of variability in the acoustic warning for FCW while the other dealt with system status displays for ACWS functions more generally.

5.1 Considerations for FCW alerts

The experiment on negative transfer for FCW auditory signals found a substantial performance decrement when participants familiar with one warning sound experienced a different warning sound in "another vehicle" (actually a modified interior of the same simulator cab, described as a different vehicle to the participants). Since this study was confined to auditory alerts, it is not clear how severe this transfer issue may be if the alerts are in different modalities (e.g., a shift from a familiar auditory alert to a tactile alert). Based on the auditory data alone, a shift in one direction (from the "heavy vehicle" sound to the "light vehicle" sound) led to a substantial increase in driver brake reaction time, of about 700 ms. Although the shift in the other direction was not statistically significant, the fact that the control group participants continued to improve their reaction times (by about 130 ms) may indicate that negative transfer may have played a role in this direction of shift as well. Therefore it appears that a novel DVI can meaningfully lengthen response time among drivers who are familiar with the FCW function and normally respond quite quickly. The substantial asymmetry in the data with respect to the direction of switch (light-to-heavy versus heavy-to-light) suggests that there are features of some sounds that make them more resistant to this problem, but what those features are is unknown.

The effects of increasing familiarity with a particular warning sound also are quite substantial. Across successive simulator sessions, from the first exposure in Session 1 to the first exposure in Session 3 (for the no-switch control groups), the response time to the signal decreased by about 400-500 ms. Therefore, presumably if people come to recognize a familiar sound from general experience as a driver or passenger, faster responding may be expected.

While the transfer and familiarity effects in this study were substantial, this single study leaves several questions unanswered. It is not clear if the negative effects would be similar if there was a sparser schedule of FCW events and more time between driver experiences. It is also unclear why the results of the shift in sound were asymmetrical and what degree of negative transfer is more typical. The two warning sounds differed in terms of frequency modulation, primary frequency, pulse duration, onset and offset ramps, pulses per burst, number of bursts, and the intervals between pulses and between bursts.

While the findings suggest that acoustic signal variability may impede transfer between vehicles, this does not necessarily mean that signals should be identical in all respects. It may be that some feature of the signal, such as the fundamental frequency or the temporal pattern, may be enough to support a transfer of the advantages of familiarity. In fact, some parameters, such as temporal pattern or intensity sweeps, may even be transferable among different display modalities (e.g., visual, voice, or haptic displays). It will be important to replicate and extend these findings before providing any recommendations.

5.2 Considerations for ACWS status displays

The experiment on ACWS status displays found problems in the accurate and timely comprehension of safety-relevant status displays. People did not understand what functions were present in the vehicle and whether they were operational. Getting some familiarity with the vehicle through manufacturer-provided materials helped, but problems remained. There were no general advantages to learning about one vehicle's interface and then experiencing another. Participants who read materials for a different vehicle did no better overall than participants who read nothing. Different manufactures use widely different status display approaches and user comprehension for any particular message may vary quite widely among vehicles. Conclusions from this study include the following:

- Current status displays, as represented by the examples used in this experiment, suffer problems in user comprehension. Improved design based on human factors principles may help alleviate this. However, the magnitude and extent of the problems across products suggests that the complexity of the display issue resists resolution by design guidance alone.
- There is a great deal of difference in how the displays are treated from vehicle to vehicle, in terms of terminology, icons, coding, location, and operational aspects. Similar functions go by different names and acronyms may be idiosyncratic. It is difficult to map similar meaning displays from one vehicle to another. While the Society of Automotive Engineers (SAE) has developed a standard that specifies symbols for use on vehicle controls, indicators, and tell-tales (SAE J2402, 2010), these standards do not currently address symbols for ACWS features. ACWS status information should be located where people expect to see it. It is not clear whether these expectations are related to other aspects of the driver-vehicle interface, so that the resolution might have to be empirically determined/performance based for the particular vehicle model.
- Familiarity with a particular system, even if limited to reading certain manufacturer-provided materials, improves comprehension for that system. However, this degree of familiarity has only limited benefits in improved comprehension. There is a need for effective quick-overview materials that convey what safety systems are in the vehicle, how status is indicated, and how they operate. Visual demonstrations would be appropriate and could be provided through web sites or other digital means. Additional strategies for improving comprehension are desirable.
- The benefits of familiarity with one vehicle's display show no consistent positive transfer to another vehicle's display. Thus it appears that the diversity among systems limits the ability of people to make use of their experience when they encounter an unfamiliar vehicle.
- If there were more commonality among displays, converging on good design strategies, drivers could become familiar with the display through exposure to common features they experience across the vehicles, as both drivers and passengers.
- Display limitations for difficult-to-code concepts can be at least partially overcome by learning, if such display features are used consistently across vehicles. Examples from current vehicle displays include the "low tire pressure" icon and the distinction between front and rear window defoggers. Over time and exposure to multiple vehicles, these displays come to be comprehended by drivers, even if their inherent understandability is limited.

6 Conclusions Regarding Evaluation Methodology and Protocol

An objective of the CWIM program is to promote common methods for the evaluation of the DVI for particular ACWS functions. This section presents conclusions regarding CWIM methodology, based on the full range of analytic and empirical work conducted under this project.

In discussing specific methodology issues, Section 6.2 provides a number of suggestions for a CWIM protocol. These suggestions are preliminary and may serve as “straw man” arguments for purposes of further development and consensus building. There may be needs for a stronger empirical basis or consensus. The rationales for such suggestions are provided, and include consideration of the need to balance scientific and technical criteria with practical issues for a protocol that is used as a shared tool among various users (e.g., cost, time, need for unique facilities, repeatability of findings, and simplicity/interpretability of outputs or scores).

6.1 Objective and scope of the CWIM application

In developing recommendations, it is important to keep in mind the intended role of CWIM testing. The methods are intended to assess the driver interface for a particular warning function in commercial or near-production systems. The methods are not put forth for purposes of product development or early design considerations. The intent is very specifically to have a common method for evaluating the driver interface of a commercial system.

Evaluating the driver interface is not the same as quantifying or rating the performance of the safety system itself. The DVI is only one component of the system. In fact, another Department of Transportation program, ACAT, has the specific objective of developing a basic methodological framework and simulation technique for estimating the quantitative safety benefits of particular crash countermeasure systems. Similarly, NHTSA’s New Car Assessment Program (NCAP) has developed detailed test procedures to assure that vehicles’ ACWS sensing and warning algorithms perform to minimum detection and timing requirements in different test track scenarios (e.g., NHTSA, 2010). The CWIM project is not specifically concerned with how well a system addresses a crash situation, but more narrowly with how well the DVI conveys the relevant crash-imminent information to the driver. How quickly and accurately does the driver perceive the threat and respond to the warning display and does the interface elicit appropriate actions? The CWIM recommendations are focused on this goal.

A crash avoidance system may have both immediate effects in terms of driver response to warnings and a more general and longer term influence on driver behavior and performance. While these longer term effects are important safety considerations and should be addressed, they are not the target of the current CWIM effort. An example of a longer term influence is driver adaptation, in which drivers frequently maintain a shorter car-following distance once they have experience with a FCW or are more likely to engage in distracting technology use if they are supported by an LDW system. Such potential effects of ACWS are beyond the scope of this work. The focus here is on the driver’s immediate response to a warning display in a potential crash situation.

Related to the concern above, the focus is also on the response to a particular warning function display, not on broader aspects of safety system performance. For example, driver response may be influenced by the frequency with which false alarms or nuisance alarms occur. This is important for assessing a system, but is not part of the DVI evaluation. Likewise, the effectiveness of a specific warning may depend on how well the particular function is integrated into the broader system of functions and information displays within the vehicle. This is again an important concern, but beyond the goal of the present project.

The CWIM recommendations also must be tempered by practical considerations. It would not be feasible for a standard evaluation procedure to experimentally manipulate all of the many factors that might interact with ACWS DVI performance. For example, these might include the number of event scenarios included, roadway types, driver impairment, weather conditions, types of distraction, and so forth. Some narrowing to a common set of conditions that will be practical for ACWS DVI assessment is required.

In order to achieve reproducible results among groups using a common general method, a number of key factors must be defined and controlled. The sections that follow discuss ten specific issues:

- Driving scenario (Section 6.2.1): Characteristics of the driving situation and details of the potential collision event scenarios
- Participants (Section 6.2.2): Test driver (participant) population characteristics
- Distracting the driver (Section 6.2.3): Need for distraction and criteria for the distraction task
- Warning system context (Section 6.2.4): Evaluation of single warning functions that are designed as part of a warning system
- Familiarity with the technology (Section 6.2.5): Control of the level of familiarization that research participants have with the general technology and specific product
- Participant expectancy (Section 6.2.6): Factors influencing what participants expect from the situation and understand their task to be, such as the presumed purpose of the study, exposure to multiple near-crash events, incentive structures
- Accommodating user settings and options (Section 6.2.7): How to test products that allow the user to modify performance features
- Comparison conditions/benchmarks (Section 6.2.8): Comparison groups or performance levels against which the DVI is measured
- Treatment of data (Section 6.2.9): Data quality and analysis, definition and treatment of bad trials
- General test method (Section 6.2.10): Simulator and test track methods

6.2 Specific methodological issues and recommendations for addressing them

6.2.1 Driving scenario

The driving scenarios under which the DVI is evaluated need to be specific to the warning system application. There are two aspects to the driving scenario. One is the general characteristics of the roadway, such as number and width of lanes, speed limits, presence of traffic, type of setting (e.g., urban, rural), environmental conditions, and so forth. The other aspect is the dynamics of the potential crash event. As discussed in Section 2.1, for the FCW and LDW functions, relatively few potential crash event scenarios account for a high proportion of the relevant crashes.

The general criteria recommended to define the event scenarios are the following:

- The relative frequency with which the scenario occurs among crashes that are relevant to the warning function

- Scenarios that have been selected and developed in major related programs, such as ACAT or CAMP. These tend to draw on the same sources and incorporate the criterion above. It would be advantageous to be able to relate CWIM findings to those of other evaluative programs, based on ACWS performance in common scenarios.
- The ability to safely and realistically simulate the potential crash event. The limitations of certain driving simulators or test tracks may preclude the practical use of certain scenarios. For example, "lead vehicle braking" is a dominant scenario in rear end crashes but may be difficult to safely accomplish in a practical manner on a test track. The CAMP program use this scenario with a towed mock vehicle designed for safe impact. However, this is a very demanding methodology and may be impractical as a standard approach. The FCW experiment conducted at the VRTC facility was able to devise a more practical assessment procedure by adapting a "lead vehicle cut-out" scenario in which the lane change of the lead vehicle revealed a stationary inflated faux balloon vehicle ahead. While this scenario was a realistic one, it does not occur in crash statistics as frequently as the "lead vehicle braking" scenario. Ideally, research should be conducted to determine the extent to which similar results emerge for different scenarios, so that the most practical but generalizable procedures can be used.
- The number of different scenarios to include depends on the ability to include multiple events during the experimental session while maintaining an element of surprise for the crash event(s) that will test the ACWS DVI. Both the NADS-1 and GMU simulator experiments reported a difference in findings between different scenarios, so it is advantageous to include more than one scenario.
- Multiple scenarios are also valuable in methods where there are multiple warning events, so that the occurrence of a potential crash event is less predictable. For instance, after a warning event in which a vehicle cuts in and rapidly decelerates in front of the subject vehicle, the participant is likely to be on guard for similar events.

In order for studies to remain comparable, the general characteristics of the roadway should be specified. In general, the driving situation should not be more complex than required to test the particular ACWS; for example, a rural two-lane road may be preferable to a complex urban setting. However, the application will determine when additional complexity is required (e.g., LDW or blind spot warning scenarios require multilane roads). Daytime and clear weather conditions are typical of the majority of relevant crashes (although potentially this may not be true for certain other warning functions, beyond FCW and LDW, that may ultimately require a CWIM protocol). The roadway characteristics of the warning mode experiments conducted in the NADS-1 simulator may be a good starting point, as they appeared successful in the FCW and LDW studies. Considerable piloting went into the selection of lane width and shoulder characteristics, and other details, for optimizing LDW testing. Obviously, test track methodologies have less flexibility in adopting common features, such as lane width, number of lanes, or sight distance.

6.2.2 Participants

In order for applications of a common evaluation protocol to remain comparable, the characteristics of the sample of research participants should be similar. For example, it will not be acceptable for a manufacturer to use a convenience sample of its employees while some other organization recruits participants from the general population and another organization samples from a customer base.

One might try to define the participant population with various strategies in mind. The sample could be designed to reflect:

- The general driving population
- A higher-risk sub-population (older drivers, younger drivers, crash-involved drivers)
- The most typical drivers
- Consumers likely to purchase the vehicle in which the system will be installed

A generally appropriate participant sample could be a relatively homogeneous and stable portion of the typical driving public which specifically excludes special groups based on diminished capabilities or risky actions or populations defined by consumer attributes. These other groups may be of particular interest in product design or in highway safety research and certainly merit attention, but a more homogeneous and typical group can provide a stable basis for comparison over a span of times and test sites. The goal of CWIM is to compare DVI "A" with DVI "B" in a stable repeatable manner, not to identify performance differences among user groups. The following sample characteristics may help to achieve this goal:

- Equal numbers of male and female drivers evenly distributed throughout the 25-59 age range. The amount and type of vehicle travel and the crash histories of drivers remain fairly stable over this age range. These are also the peak travel ages, accounting for the large majority of vehicle miles traveled.
- Participants should be fluent in English.
- Participants should hold a valid driver's license and insurance.
- Participants should meet minimum criteria for driving experience and driving exposure (e.g., at least three years of licensure and a minimum of 5,000 annual driving miles).
- Participants should not have excessive histories of risky driving (e.g., only drivers who have no license suspensions or serious driving-related convictions, few recent moving vehicle citations, few crashes). The intent is not to provide a population of especially "safe" drivers but to screen out those that are likely to be especially risky or aggressive drivers.
- The population should exclude those with driving-relevant health issues. These might include eye disease and vision problems, hearing problems, diseases that may compromise motor or cognitive functioning (e.g., Parkinson's disease), heavy alcohol use, and the use of drugs or medications that might impair alertness.
- The study should exclude drivers who are familiar with crash warning systems similar to the one to be tested.

The NADS-1 simulator studies of warning modes used even more restrictive age and mileage restrictions than suggested here (35-55 years old, minimum of 10,000 driving miles per year), without apparent recruiting difficulty. However, we feel that crash data and driver performance research support sampling a wider age range. Furthermore, in less rural areas than Iowa (where the NADS-1 studies were conducted), many regular drivers may fail to meet a 10,000 annual mileage criterion. Even 5,000 miles annually corresponds to roughly 100 miles per week. Therefore, the 25-59 year age range and 5,000 mile criterion allow for ease of recruiting while maintaining the desired homogeneity of the test population.

6.2.3 Distracting the driver

ACWS are intended to support the driver in recognizing emerging hazardous situations. To the degree that the driver is alert and scanning appropriately, a warning should be unnecessary, and if it does occur, it is likely to be redundant with what the driver already knows. The primary advantage of systems such as LDW and FCW is in alerting the driver who is distracted or otherwise incapable (fatigue, impairment) of detecting and responding to the potential crash situation. There are emerging warning applications, such as a variety of those foreseen under the IntelliDrive program, where ACWS may provide a useful alert even if the driver is not distracted. For example, the hazard may not be directly visible to the driver, due to intervening obstacles (e.g., other traffic), limited sight distance (e.g., curves), or poor discriminability (e.g., icy road surface). However for most current applications, including the LDW and FCW functions that are the focus in this work, it may be assumed that a competent and alert driver will not experience any benefit from the warning system. The warning is intended to support the driver whose momentary attention is not directed at the appropriate location on the roadway. Therefore any evaluation of the DVI must be based on drivers whose attention is distracted. The means of distracting the participant will be a key part of any common CWIM methodology. In normal driving for most people, relatively long glances away from the road are quite rare and are difficult to predict. Therefore the experimental method must have some means of inducing appropriate visual distraction at known times.

Distraction is a complex issue and there are many forms of distraction. Likewise, there is a very wide variety of tasks that have been used experimentally to distract the driver. The very different methods used in the various ACAT projects described in Section 3.6 illustrate this. To some extent, features of the ideal distraction task may be specific to the particular warning function. Several different distraction tasks were used in the empirical efforts of this project. Currently, there are various ongoing efforts to define preferred distraction methods, including major NHTSA programs in driver distraction and connected vehicles. Given this, we hesitate to recommend any particular task at this point. However, we provide a set of criteria for distraction tasks and discuss some of the issues related to the tasks we have used. Features of an ideal distraction procedure were established in earlier stages of this project, in order to aid in the planning of the research studies. These initial criteria, supplemented by subsequent experience with the distraction tasks employed, were used to generate the following set of criteria for an ideal method. In practice, it might be difficult to meet all of these criteria in a given study method, but they represent desirable goals.

- The distraction task should be visual/motor, not auditory/verbal or other solely cognitive distraction. This is based on the assumption that visual distraction is a worst case and is supported by the findings of the 100-car study (e.g., Klauer, Dingus, Neale, Sudweeks, and Ramsey, 2006; Klauer, Guo, Sudweeks, and Dingus, 2010), a naturalistic study of driver behavior using highly instrumented vehicles. This study found that crash and near-crash events were predominantly associated with episodes of driver-related inattention to the forward roadway and that crash/near-crash risk was greatly elevated when the driver was engaged in tasks that required multiple looks, multiple button presses, or long glances.
- The task should reliably draw visual attention away from the critical part of the visual field (based on the warning function) at a specified time. Alternatively (or additionally),

if eye tracking is done in real time, the task could be triggered when the driver's eyes are known to be averted from the road.

- The task should maintain visual attention for a period sufficiently long so that the potential crash scenario unfolds unnoticed and the warning system activates prior to the driver directly detecting the threat. It should be noted that this time requirement will vary based upon the implementation of the crash avoidance event.
- Ideally there would be a means of confirming that the driver's eyes were off the road. This would be done most directly with eye tracking capabilities, although that would add burdens of both system design and data reduction. Any eye monitoring should be unobtrusive so as not to influence the driver's normal behavior. It may also be possible to draw inferences about point of gaze from manual interactions with the task. However, in practice participants adopt a variety of strategies for inserting quick glances to the roadway ahead while physically interacting with the distraction task interface. An ideal distraction task should maximize visual commitment away from any visual cues of a developing threat.
- The driver's point of gaze should not permit the effective use of ambient (peripheral) vision to detect the potential crash event or allow the driver to maintain proper performance. For example, for the LDW application, even if focal vision is off the roadway, peripheral view of surrounding roadside features and delineation could allow the driver to maintain lane position. However, this consideration must be weighed against how representative the distraction is of real crash situations.
- Specifically for the LDW application, the task should require a body turn or other postural shift that would make experimenter-induced changes in vehicle position or heading more convincing to the participant as a self-initiated error.
- The distraction task should resemble some actual real-life distracter. It should not have the feel of an arbitrary, "laboratory" type of test that might detract from the immersive realism of the simulation.
- The distraction task should not artificially modify how the driver normally allocates attention to primary and distracting activities. Behavior should be as close to natural as possible.
- The task should not be subject to large practice effects so that the actual degree of driver distraction varies over the course of the experimental session.
- The distracter task should not in itself heighten the alertness or suspicions of the participant or signal the impending a potential crash event. For this reason, the distracter should initially and periodically occur without being accompanied by any significant events and various other distracting activities should also be included in the procedure. For example, the NADS-1 study of LDW included a CD changing task and a touch-screen interaction trivia game as additional distraction tasks, but the lane drift incidents were only programmed to occur during the primary "bug" catching task.

While these criteria define an ideal distraction task, it may be difficult to optimize all of them and there may have to be tradeoffs between them. This can be illustrated by comparing the two experiments on warning modalities for FCW systems. The driving simulator experiment used a distraction task that was designed to simulate the situation where a bee or other insect was in the vehicle, distracting the driver. The task was intended to be visually compelling and required continuous visual monitoring in order to "catch" the bug. Nonetheless, participants found it difficult to keep their eyes off the road for long and there was moderately high data loss because of early responding. In contrast, the test track experiment used a task that could only be successfully completed if the participant continually monitored a number sequence. The task had a short, fixed duration, so that the participant knew there would be an opportunity to look back to the roadway. The procedure included a payment incentive component, where a substantial portion of the total payment they could earn came from rewards for successfully completing the number recall task and there was a financial penalty for incorrect responding. The test track procedure resulted in very few instances of participants looking away from the display and toward the road during the distraction task.

Another trade-off is that most naturally occurring distracting activities in the real world do not require particularly long glance times. Where task time is long, people are able to "chunk" the activity into discrete segments and permit a glance back to the roadway in between chunks. Thus the tasks often used in research studies have an artificial feel to them. More realistic tasks will have more variable glance times and relative few long glances.

6.2.4 Warning system context

As noted above, there are issues regarding how to evaluate a particular crash warning interface with respect to the system context in which it occurs. The intended purpose of a CWIM method is to provide an objective, repeatable means of evaluating alternative interfaces for some specific warning function, such as FCW or LDW. Yet any particular ACWS functions within the context of the particular vehicle that it is designed to support. The warning will occur within the context of other safety functions, displays, and communications within the vehicle, may occur as part of a progressive warning strategy or be related in some way to a parallel safety-relevant system (e.g., intelligent cruise control), and may be accompanied by automatic vehicle control actions. The argument has been raised (Section 2.3) that a particular warning DVI may be "penalized" if it is tested out of this context. At the same time, there is no basis for comparing alternative DVIs if each occurs in a unique context.

The resolution to this issue goes back to the CWIM objectives discussed in Section 6.1. The purpose of the CWIM evaluation is not to quantify the effectiveness of the safety system in crash avoidance, but more specifically the ability of the DVI to convey the appropriate information and induce the appropriate driver response. Therefore a particular DVI should be evaluated on a stand-alone basis within the framework of a standard vehicle and driving context. The empirical studies of LDW and FCW warning modes conducted within this project were successful in adapting actual or simulated commercial systems to a common context and discriminating important differences in driver response among them.

The ACAT program also serves as a model for this approach. The ACAT assessment of benefits focused on individual, stand-alone systems (e.g., FCW, LDW, backing), not integrated systems. This simplified the problems related to context and made the evaluation task manageable. The same strategy appears reasonable for CWIM. The larger issue of integration and complementary features will become more important over time and this issue will ultimately need to be addressed.

An imminent crash warning display should be tested on a stand-alone basis even if it may occur in a particular vehicle within the context of earlier informational messages or lower level alerts. Obviously, performance might be better if these related messages were present. However, the earlier alerts might not always be sufficient; if that were the case, the imminent crash warning would not be required. Therefore the CWIM protocol should test the worst-case situation where the driver has not taken account of other messages and is responding only to the imminent crash warning itself. If resources allow, it may be of interest to include within-context testing as well. However, the primary context for evaluating the effectiveness of the warning display should be as a stand-alone presentation.

6.2.5 Familiarity with the technology

The response of a driver to a warning is likely to depend to some degree on the driver's familiarity with the warning system. At one extreme, a person may not even realize that the technology for a particular warning capability exists today. Or, they may not realize that the particular function is present in the vehicle they are driving. Or, they may understand that it is or may be present but have no idea what it looks, sounds, or feels like. They may or may not have familiarity with other commercial products that fulfill a similar function. At the other extreme, they may be highly experienced with the specific system present in the vehicle they are driving. Therefore, the question arises as to what degree of familiarity participants should have under the CWIM procedures.

The findings of the driving simulator experiment on potential negative transfer for FCW auditory signals (Section 3.4) clearly showed that familiarity was an important factor in driver response. Response times became progressively shorter with continued experience with a specific FCW, and the magnitude of this effect was substantial. Furthermore, when an unfamiliar warning sound (in an ostensibly "new" vehicle) replaced the one they were familiar with, response times suffered substantially in one direction of change. The experiment on ACWS status display comprehension (Section 3.5) indicated that drivers often did not know whether a particular safety function was present in a given vehicle and whether it was currently functional, but that familiarity acquired from reading the owner's manual increased comprehension. These experiments indicate that familiarity with the warning system may matter. The implication is that CWIM procedures should specify and control the degree of participant familiarity, prior to data collection. Unfortunately, the studies do not indicate what the ideal degree of pre-exposure should be and it may not be so much an empirical question as one of testing philosophy.

One perspective is that a totally naïve driver represents the "worst case" and therefore should be the basis for the evaluation. Another suggests that this is neither a representative nor fair basis for testing a particular DVI. According to this view, drivers may be assumed to at least be aware that a warning function is present in their vehicle; furthermore, they will only be totally naïve to the look, sound, or feel of the display once, and after that, all future driving will be done with some awareness of the system. Some types of ACWS warnings may be expected to occur with some frequency (e.g., lane departure or blind spot warnings) while others could be quite rare, so drivers may have less familiarity with the ACWS DVI in their own vehicles. We see both of these approaches reflected in the experiments on FCW and LDW warning modes. The studies conducted in the NADS-1 driving simulator provided participants with pre-familiarization about vehicle features, including the FCW or LDW system. Then in the initial training phase of the simulator drive, participants were exposed to various warnings by intentionally activating the system under experimenter instructions (i.e., the driver intentionally drifted out of lane or rapidly approached a lead vehicle, to trigger the warning). In this way, participants had had some knowledge of what the system did and what the DVI was like, although they were not experienced users of the system. In contrast, the VRTC test track study of FCW provided no specific pre-exposure to the presence of the warning systems of the DVI characteristics.

If pre-exposure to the DVI is provided, as in the NADS-1 experiments, it is critical that this not be done in such a way that the participant's attention becomes focused on the crash warnings. In the NADS-1 experiments, the FCW or LDW system was presented as one of a number of systems and task activities present in the vehicle and presented in advance for familiarization to the driver. The FCW negative transfer study conducted in the George Mason University driving simulator similarly embedded the FCW function within a broader range of features. Based on the NADS and GMU experience, this appeared to be successful in that participants did not demonstrate awareness that the focus of the experiment was on warning systems.

While the most appropriate degree of system familiarization remains open to debate, our recommendation at this point is to provide a limited and controlled degree of pre-exposure, as in the NADS-1 experiments. The argument is that the majority of real-world exposure to FCW or LDW alerts will be for drivers who have at least some awareness of the system and its nature. A totally naïve user is a case of interest, but a very limited case. Ideally, a study might incorporate both naïve and familiarized participants. However, since this would essentially double the effort, we hesitate to require it.

6.2.6 Participant expectancy

The nature of the participant's driving task and the expectancies engendered by the procedures are a critical concern. The intent is to impose the potential crash situation on drivers who are driving in their normal manner and are not anticipating the probable occurrence of an emergency event. The instructions defining the purpose of the experiment *from the participant's perspective* are critical. Participants should not have any indication that the researcher's interest is specifically with crash warning systems. Therefore the instructions to participants, as well as associated materials such as recruiting flyers, screening scripts, and consent forms, should not promote this perception. As much as possible, the procedure should foster the feeling that drivers can simply behave in their normal manner.

Based on the procedures used in the empirical studies of this project, we suggest a general procedure in which participants are told they are going to experience a new prototype vehicle that has a variety of innovative design features. These might include advanced electronics, but also other things such as (putatively) new seat designs, seatbelt features, climate control systems, infotainment features, or suspension features. What is important is that the ACWS under evaluation be presented as only one feature of many in a vehicle with a variety of innovations not specifically oriented to warnings or safety. The participant's stated task is just to drive normally over the course of an extended drive and they may be asked various questions about the system after the drive. They are also made aware that there will be various tasks they may be asked to engage in during the course of the drive, since the intent is for them to experience the vehicle under a range of driving conditions. Before starting the drive, participants are given an introduction to the various vehicle features and tasks. This allows some familiarity with the ACWS alert without focusing participant attention specifically on crash warnings. The initial phases of the drive are benign, with no potential crash alerts associated with the initial exposures to the distraction tasks. If the procedure includes multiple occurrences of relevant potential collision events, several distinct scenarios should be used so that the circumstances do not become predictive.

As much as possible, driving should be allowed to be normal and unconstrained. However, it may be necessary to constrain some aspects of driving to ensure successful and consistent potential crash events. For example, participants may be asked to always travel in the right lane, or to try to closely adhere to the speed limit. While constraints such as these probably have minimal influence on most driving behaviors and visual search behaviors, test track studies may need to impose more severe constraints than simulator research. For example, the test track FCW experiment required drivers to follow a lead vehicle in a closely coupled manner and drivers were awarded incentive bonuses for maintaining a specific range of headway. These constraints may come about both because it is more difficult to control the situation dynamics in the real world as opposed to the virtual world of the simulator and because of safety concerns. The test track provides a trade-off of actual, as opposed to virtual, driving with the need to constrain driving more than in the simulator. In either sort of test environment, however, the goal should be to minimally influence the baseline driving task upon which the potential crash scenario is imposed.

6.2.7 Accommodating user settings and options

Manufacturers may design systems that allow the user to select or program various aspects of the system response, or systems that adapt to the characteristics or performance of the driver. Thus there may be user-controlled or dynamic variance in DVI characteristics such as display intensity, display type, triggering criteria, or timing of displays. If a display attribute is adjustable in some dimension, what setting should be used for CWIM testing? There was little in available literature relevant to this question and the empirical studies conducted within this project did not address this issue. Should the procedure use the most conservative setting, the least conservative, a mid-point, a default setting, a setting selected by the research participant, or some combination of these? The recommendation here is to use the default or mid-point setting, because it carries the implication that this is the "normal" option and any deviation from this is the user's responsibility.

6.2.8 Comparison conditions/benchmarks

If CWIM methods are meant to evaluate the effectiveness of an ACWS DVI, there is the question of "effective compared to what?" Is the comparison made to some benchmark value, control condition, or "standard" interface? Is the evaluation meant to be taken in absolute or relative terms? Is the assessment quantitative score or a pass/fail decision through comparison with some criterion?

As noted in Section 2.3, NHTSA and industry stakeholders expressed the opinion that acceptance may be best for a threshold-based pass/fail method and that simplicity in output is desirable.

The recommendation here is that CWIM evaluations of a DVI include two benchmark conditions within the same study. One of the benchmarks is a "no warning" control condition. The other is a fully-specified "basic" DVI. These two benchmarks would define thresholds for three levels of performance: (a) no benefit (i.e., not better than control); (b) basic effect (i.e., not better than basic DVI); and (c) superior (i.e., better than basic DVI). Ultimately, it would be most desirable to define some absolute performance levels for a particular dependent measure, based on a sufficiently large study to define this threshold empirically. In that case, subsequent product evaluations would not need to include the control condition(s), but simply compare the DVI performance against the established criterion level. However, this is not feasible until some metric is agreed upon, a threshold is established through adequately large empirical efforts, and the measure is shown to be highly reproducible across different evaluation sites. Until absolute metrics have been adequately demonstrated, the performance of a given DVI must be made on a relative basis, compared to a benchmark condition included in the same evaluation study. The comparison with the no warning control condition is desirable because poorly designed DVIs may have no appreciable beneficial effect and in some cases may even prove worse than no warning at all. Furthermore, the control condition may provide a confirmation of the appropriate urgency of the potential crash scenario and the distraction procedures. For example, if all participants in the control condition responded quickly and easily avoided a crash or close call, this would indicate that the experimental procedures were not effective in developing the event scenario.

The comparison with a basic standard DVI is useful because merely showing an improvement relative to a no-warning control condition is a very minimal basis for evaluating a DVI. If a simple and common type of warning, such as a typical acoustic alert, is shown to have some beneficial effect, the CWIM procedure should determine whether a given DVI is similar to, worse than, or superior to this basic display. The comparison signal should be an exemplar of typical vehicle warnings, but should not be identifiable as uniquely the display of any specific OEM's product. An additional advantage of having these benchmark conditions is that it will permit "calibrating" comparisons across testing locations or testing times.

6.2.9 Treatment of data

Two aspects of data treatment are highlighted here: data quality control and analytic considerations. In terms of data quality, it is recognized that the experimental procedures for generating the incident scenario may not always work properly. On any given trial, the dynamics of the situation may not develop properly, the participant may fail to follow instructions properly, or the distraction task may not be effective. These sorts of failures were seen with varying degrees of frequency in all four empirical experiments (NADS-1 experiments on LDW and FCW, GMU simulator experiment on FCW, and VRTC test track experiment on FCW). The experimenter must be able to identify and exclude improper trials from the subsequent analysis. The criteria should be clearly specified to avoid subjective selection of data. Determining when the participant was not distracted from a direct view of the evolving potential crash situation is a particular concern. If eye tracking data are collected, they may provide a useful means of verifying that the participant is not directly viewing the event at the time of the warning. Another strategy is to eliminate from consideration any cases where the response time (relative to warning onset) is so brief that it may be assumed that the participant was responding to events prior to the warning. A criterion of 200 ms was used as a threshold in the FCW simulator experiments of this project; other researchers have used different values. For the LDW experiment, the criterion was based on relative lateral velocity; if the point at which the lateral drift ceased and changed direction (indicating a steering correction) occurred prior to the warning, the trial was excluded.

An issue in data analysis concerns the focus on traditional analytic methods that compare measures of central tendency. While these remain appropriate, they may not be sufficient. Figure 16 illustrates this concern with data from the LDW experiment.

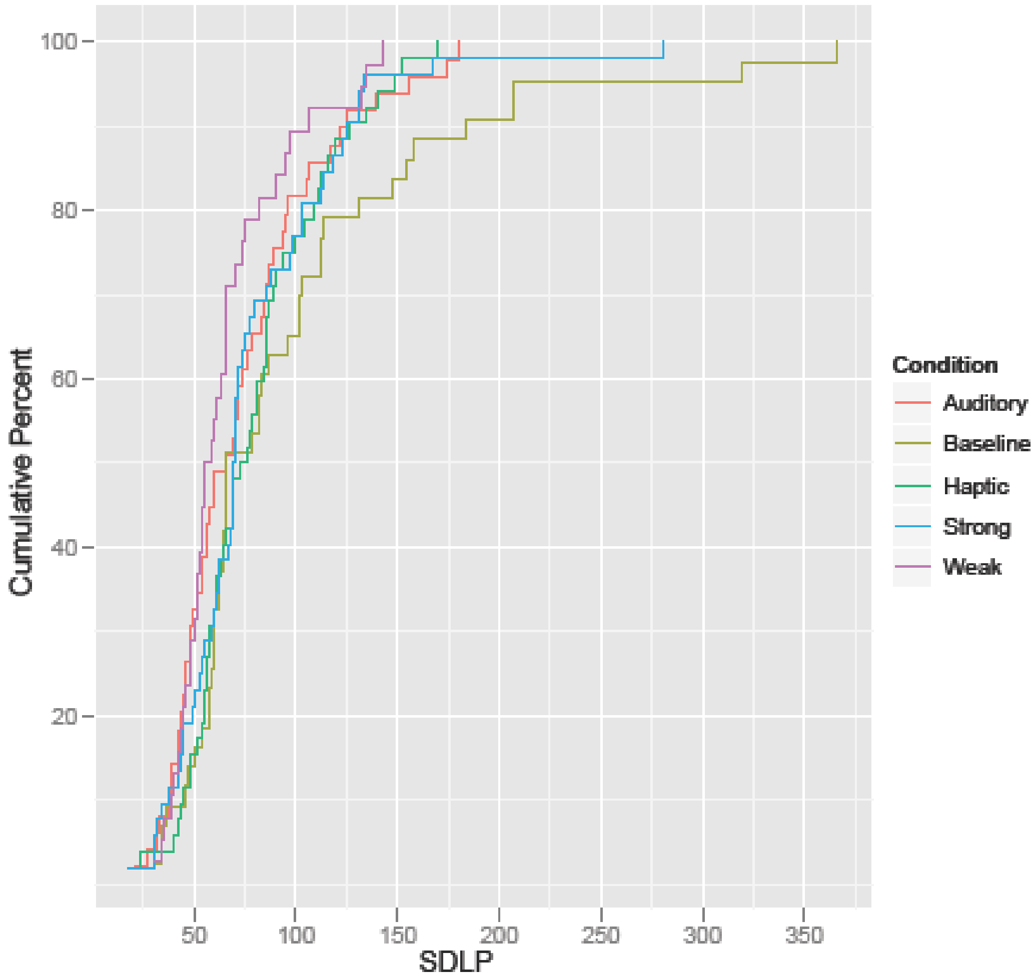

Figure 16. Cumulative percent plots of lane position standard deviation

The figure shows the cumulative percent distributions of the standard deviation of lane position, under each of the five warning conditions in the experiment. There is relatively little difference among warning treatments at the 50th percentile. At the 90th percentile and higher, however, differences are very pronounced. The advantages of the "weak active" system are clear and the problems of the no warning control condition are very evident. The more pronounced differences in the tails of the distributions suggest that the warning condition might have relatively small effects on routine situations but large effects in extreme situations. The figure also indicates how there may be crossover points where the relative performance of different systems may depend on the situation. A focus on central tendency (50th percentile) might obscure important differences. Another way to look at the issue is that the relative merit of a system may not be reflected in the bulk of cases, where all of the alternatives perform adequately, but in the elimination of the rare extreme cases that may be most associated with crashes. Thus it may be reasonable to focus on the frequency of events that meet some threshold, such as very short times-to-collision or very severe lateral accelerations. However, it is recognized that measures of central tendency tend to be more stable than estimates based on extremes, and the binary (nominal) events (e.g., exceeding a threshold) do not provide the statistical power of continuous measures. There will be implications for statistical power and associated requirements for the size of experiments. This issue merits further consideration, but the immediate point here is that considerations beyond central tendency will be useful.

6.2.10 General test method

In order to have a highly repeatable and meaningful measurement system, CWIM methods should include high-fidelity, tightly controlled test environments in which actual driving occurs. This implies the use of either driving simulator or test track methods. Other methods may have merit for the development of prototype systems or for safety and consumer acceptance research. For example, actual on-road driving with the ACWS may be valuable. However, it does not provide the control needed for a formal assessment tool to compare DVIs.

Both driving simulators and test tracks have advantages and disadvantages and both are potentially useful for CWIM testing. Research should be done to establish the comparability of various test track and simulator procedures for particular ACWS functions. If it is ultimately desirable to have a single methodology (simulator or test track) for a particular ACWS function, direct comparison of the alternatives would be desirable. Lacking such a basis at this point, a high-fidelity simulator environment may be most advantageous, for the following reasons:

- The availability of reasonably sophisticated simulators is increasing, with many research centers, universities, manufacturers, and private firms operating them.
- The costs of simulators continue to drop, so that more potential end users of CWIM will be able to have this capability.
- Simulators allow tight control of driving scenarios, which will be important for reliability (replicable findings).
- If there are a limited number of test scenarios, it may not be necessary for every user organization to individually program these scenarios, reducing the time and effort associated with simulation.
- Simulators largely eliminate safety concerns and many of the privacy issues associated with test track, instrumented vehicle, and naturalistic driving methods. They permit inclusion of even very severe hazard events. They allow relatively high travel speeds without safety concerns.
- Some desired scenarios simply may not be replicable on a test track, or may be impractical to implement.
- Simulators provide an ability to link the occurrence of hazard events to naturally occurring driver states or actions. For example, if the driver is looking away from the forward view, a safety-critical event can be made to occur.
- Unlike a test track environment, the simulator is not subject to environmental variation or cancelled sessions due to weather. However, the simulator cab environment may not recreate all potentially relevant conditions (e.g., bright sunlight and glare), which may have implications for evaluation of visual displays.
- While there may be some issue of whether risk-related driving performance is entirely natural in a simulation setting, the relative nature of CWIM values mitigates this concern.

One concern in simulator research is the problem of simulator sickness. However, this problem is more pronounced for older participants, and the suggested participant selection criteria (Section 6.2.4) do not include this group. The simulator experiments conducted under this project did not suffer significant problems with simulator sickness.

The argument for the use of simulators is based on the assumption that adequate validity can be achieved with reasonably inexpensive simulators and that similar results can be obtained when the same test protocol is used at different sites with different simulators. These are empirical questions that need to be addressed. Minimum criteria for driving simulator displays and performance will need to be derived.

7 Key Research Needs

This project has explored a broad variety of issues and a number of research needs are evident. This section highlights some of the more prominent research needs in the areas of ACWS modality, ACWS features, and CWIM assessment methods.

7.1 Key research needs for ACWS modality

- Greater attention needs to be given to modes other than visual and auditory. Various sorts of haptic and active systems appear promising for particular applications (e.g., brake pulse and seat belt pre-tensioning for FCW; active steering torque for LDW).
- There is some potential for inappropriate responding with innovative modes and these need to be included in future studies. Stronger active warnings do not necessarily result in more rapid or effective responding; a more detailed study of how drivers react to vehicle-initiated control actions would be beneficial.
- Seat belt pre-tensioning appears promising and should be systematically explored. Direct comparison with other forms of haptic warning (e.g., seat vibration, brake pulse, steering wheel vibration) would be desirable and should encompass a range of scenarios.
- Driver subjective response to the DVI is not necessarily parallel to driver performance data. Initial findings suggest that drivers did not perceive active LDW warnings as helpful or informative. User acceptance issues should be studied to understand this issue and address it.
- The present CWIM research evaluated both individual warnings and select multimodal warnings. Because of the potential for multimodal warnings to overcome the limitations of any individual mode, additional multimodal warnings using different combinations of alerts and modes should be studied.

7.2 Key research needs related to DVI variability

- Further research is needed to understand what warning signal features contribute to negative transfer problems. The FCW acoustic warning study found very large detrimental effects in one direction of transfer, but only moderate effects in the other direction. Research should determine what parameters of an auditory signal are most important to maximize transfer from vehicle to vehicle. Cross-modality transfer should be included.
- Very little research has been done on the effectiveness of in-vehicle status displays for conveying information about warning systems to drivers. The initial research in this project found that drivers frequently do not understand that safety systems are present, how they operate, and their current operating status. Research is needed to improve aspects such as terminology, icons, acronyms, color coding, and perhaps location. Features of effective status display interfaces need to be determined. Since manufacturer-

provided communications appear to have limitations for informing drivers about the nature of the safety systems and associated displays, research on more effective materials and strategies could be valuable.

7.3 Key research needs for CWIM assessment methods

- Distracting tasks are required as a component of CWIM methods and there is little consensus on what this task should be. Research should be conducted to directly compare the effectiveness of alternatives on multiple dimensions (e.g., diversion of glance, workload, influence on participant's expectancy, interference with driving task, stability of performance over time).
- Research should be conducted with multiple methods and multiple dependent variables to objectively determine what set of metrics and procedures will optimize measurement sensitivity, reliability, validity, practicality, and efficiency.
- Coordinated parallel experiments should be conducted among a range of simulators and test tracks in order to determine the ability to derive similar findings and to define minimum requirements for test facilities.
- Research should better define the effects of familiarization with the general type of warning function and with the specific DVI. The most desirable degrees of familiarity for CWIM applications, and the means of establishing them, should be determined.

8 Promoting the Acceptance and Use of Project Recommendations

To promote the acceptance of the recommendations in this report, several steps are required:

- Follow-on research to demonstrate the validity, reliability (repeatable measures), and practicality of CWIM methods.
- Government and industry input and continued involvement in review and decision making
- Coordination with relevant standards groups, such as SAE
- Work with consumers, manufacturers, and automotive dealers to understand how the outputs of a common CWIM method may be tailored for consumer use.
- Follow-on research to the initial findings related to the implications of variability among DVIs, to better determine whether there is a need for some common elements and what these might be.

References

Bareket, Z., Barnes, M., Blankespoor, A., Blower, D., Gordon, T., Isaksson-Hellman, I., & Theander, H. (2010). *Advanced Crash Avoidance Technologies Program (ACAT) Safety Impact Methodology – Final Report.* National Highway Traffic Safety Administration, Washington, DC.

Forkenbrock, G.J., Snyder, A.A., O'Harra, B.C., Hoover, R.L., Vasko, S.M., & Smith, L.A., (2011). *A Test Track Protocol for Assessing Forward Collision Warning Driver-Vehicle Interface Effectiveness*. Washington, DC: National Highway Traffic Safety Admininstration.

Hankey, J., Deering, R., Angel, L., Perez, M., Llaneras, R., Neurauter, M., Antin, J., & Green, C. (2010). *Advanced Crash Avoidance Technologies (ACAT) Program – Final Report of the GM-VTTI Backing Crash Countermeasures Project.* National Highway Traffic Safety Administration, Washington, DC.

Klauer, S., Dingus, T., Neale, V., Sudweeks, J., & Ramsey, D. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Data.* DOT HS 810 594. National Highway Traffic Safety Administration, Washington, DC.

Klauer, S., Guo, F., Sudweeks, J., & Dingus, T. (2010). *An Analysis of Driver Inattention Using a Case-Crossover Approach on 100-Car Data: Final Report.* DOT HS 811 334. National Highway Traffic Safety Administration, Washington, DC.

Najm, W., Smith, J., & Yanagisaw (2007). *Pre-Crash Scenario Typology for Crash Avoidance Research.* DOT HS 810 767. National Highway Traffic Safety Administration, Washington, DC.

Najm, W., Koopmann, J., Boyle, L., & Smith, D. (2002). *Development of test scenarios for Off-Roadway Crash Countermeasures based on Crash Statistics.* DOT HS 809 505. National Highway Traffic Safety Administration, Washington, DC.

NHTSA (2010). *Forward collision warning system confirmation test* (Docket NHTSA-2006-26555, 0128). Washington, DC: National Highway Traffic Safety Administration.

Nissan Technical Center North America, Inc. (2009). *Advanced Crash Avoidance Technologies Program (ACAT) Task 3 Interim Report for ACAT-2 Development of Objective Tests.* National Highway Traffic Safety Administration, Washington, DC.

Robinson, E., Lerner, N., Jenness, J., Singer, J., Huey, R., Baldwin, C., Kidd, D., Roberts, D., & Monk, C. (2010). *Crash Warning Interface Metrics Task 3 Report: Empirical Studies on Standardization Needs.* Interim report under Contract DTNH22-05-D-01002, Task Order 15. National Highway Traffic Safety Administration, Washington, DC.

SAE International (2010). *Surface vehicle standard J2402: Road vehicles – Symbols for controls, indicators, and tell-tales.* Society of Automotive Engineers.

Toyota Motor Company (2009). Advanced Crash Avoidance Technologies (ACAT) Final Report. National Highway Traffic Safety Administration, Washington, DC.

Van Auken, R.M., Zellner, J.W., Chiang, D.P., Kelly, J., Silberling, J.Y., Dai, R., & Sugimoto, Y. (2009). *Advanced Crash Avoidance Technologies (ACAT) Program – Final Report of the Honda-DRI Team, Volume I: Executive Summary and Technical Report.* Report under Contract DTNH22-06-H-00070. National Highway Traffic Safety Administration, Washington, DC.

University of Iowa and Westat (2010). *Task 9 (Report and Lessons Learned from NADS Study) Letter Report.* Interim report under Contract DTNH22-05-D-01002, Task Order 15. National Highway Traffic Safety Administration, Washington, DC.

DOT HS 811 470a
August 2011

U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

NHTSA
www.nhtsa.gov