A Peer-Reviewed Critique of the Federal Highway Administration (FHWA) Report Titled: "Driver Visual Behavior in the Presence of Commercial Electronic Variable Message Signs (CEVMS)"

Prepared by Jerry Wachtel, President, The Veridian Group, Inc.

January 2015

Table of Contents

ABSTRACT.

On December 30, 2013, the Federal Highway Administration (FHWA) made available on its website, three interrelated documents concerning its recent research product about driver response to digital outdoor advertising: (1) a non-peer reviewed draft report, (2) the peer reviewers' comments to this report, and (3) a final report, described as peer reviewed, which was modified from the draft report, ostensibly to address the peer reviewers' comments. The present report, which was subjected to independent peer review, reviews these three FHWA documents, and concludes that the final report is seriously flawed due to confounding methodological issues, substantive factual discrepancies between the draft and final reports, failure to incorporate advances in the state of knowledge in the field from recent research, serious oversights in experimental procedures, and significant equipment constraints. In the opinion of the present author, the FHWA final report does not justify the conclusions as stated, and should not be accepted as an answer to the ongoing and important question of whether contemporary digital billboards contribute to driver inattention and distraction to the extent that traffic safety may be compromised. The present report calls on the FHWA authors to explain and justify their findings and conclusions, and the methods employed to achieve them; and it recommends that State and local governments, and private roadway operators, charged with regulating digital billboards within their jurisdictions, adopt a cautious and conservative approach to digital billboard control and regulation until such time as a definitive study is available.

FORWARD.

It is not an exaggeration to state that the origins of this Federal Highway Administration (FHWA) report date back more than 30 years. In the late 1970s, the predecessors to today's digital billboards were first coming into wide-scale commercial application. Until then, changeable message signs had been largely confined to "time and temperature" messages that most commonly appeared on banks. But with the growth of color capability, remote programming, and (crude) graphical images, such signs began to appear at shopping malls, automobile dealerships, airports, and performing arts centers, to name a few. The Highway Beautification Act (HBA), when drafted in 1965, and modified some years later, had not contemplated the roadside presence of advertising signs that could quickly and effectively change their message, imagery, color, or brightness. The Office of Right-of-Way (ROW) of the FHWA was charged with regulating off-premise signage under the HBA, but was in the dark about how this new technology should be treated. Accordingly ROW turned to FHWA's Office of Research (RES) and asked for help in identifying the safety and environmental concerns, if any, that such signs posed. As the Director of FHWA's Highway Aesthetics Laboratory within RES, the task fell to me and to my colleague, the late Dr. Ross D. Netherton, to develop and conduct the requested research study. We quickly found that, because this digital billboard technology was so new, there had been little research conducted about its potential impacts. The few exceptions varied greatly in their experimental and statistical rigor and, accordingly, in their findings. As a result, we were forced to go back to *all* of the research literature that dealt with roadside advertising from the perspectives of safety, aesthetics, and highway investment. Indeed, we found and addressed relevant research dating back to 1934.

Our report (Wachtel J., & Netherton, RD, 1980) consisted of a critical review of the available literature, an assessment of the psychological, physiological, and human factors considerations posed by this technology, and an analysis of environmental, zoning, and legal practice relevant to the issue.

In the end, we concluded that although there was no consistent correlation between such signs and traffic safety, the more recent, better controlled research studies had begun to demonstrate a concern for driver inattention and distraction that could be attributed to these newer signs, which we called *Commercial Electronic Variable Message Signs,* or CEVMS. We identified a specific list of CEVMS-related issues that had the potential to cause concern, and we recommended that specific field research be undertaken to close our knowledge gap. To determine the feasibility of such field research, I led a small FHWA team that, in conjunction with the Maryland State Department of Transportation, conducted an on-road pilot test of driver and vehicle response to a typical CEVMS of the time. Our study utilized FHWA's then state-of-the-art Traffic Evaluator System (TES), and we had the luxury of designing and displaying our own messages on the sign. This pilot study proved successful, and, as a result, RES moved forward with a competitive procurement for the conduct of a full-scale study. As the designated Contract Manager (today COTR) for the research, I prepared a Request for Proposals (RFP), chaired an FHWA-wide team that reviewed the submitted proposals, and ultimately identified the contractor of choice to conduct the research. The selected contractor was the late Prof. Helmut Zwahlen of Ohio University, one of the pioneers of in-vehicle, real-time driver eye tracking in the United States. Unfortunately, the funds for the project were cut just before the contract was to be signed, and the research was never performed.

It took 20 years for FHWA to return to the issue of CEVMS as a subject for its research, and the agency produced a report that it described as an “update” to our 1980 document (Farbry, et al, 2001). Although the technology of CEVMS had grown enormously by this time, and research was now being conducted into their safety consequences internationally, the new FHWA report again concluded that the research results were not consistent, and again recommended a series of research studies to answer the growing questions being raised about the safety of these signs. After a lapse of another six years, the agency initiated the first of what was contemplated to be a three-phase study (Molino, et al, 2009). I was brought on-board as a consultant to the study team, which was led by Dr. John Molino, of FHWA’s in-house contractor, SAIC (now Leidos). This study laid the groundwork for Phase II, the actual data collection (and the subject of the present report). The research team designed a study remarkably similar to the one that had been proposed in 1980, of course with the benefit of 30 years of improvement in data collection, recording and analysis technology. The FHWA Contracting Officer’s Technical Representative (COTR), Dr. Thomas Granda, described the logistical difficulties of the study by demonstrating that there were literally dozens of variables, and hundreds of combinations of sub-variables pertaining to CEVMS, *any one of which could have a measureable impact on driver response* (Granda, 2009). This list of variables was nearly identical to the one that had been defined in the 1980 report, including, but not limited to, those shown in Table 1.

Table 1

Billboard luminous intensity (brightness)
Billboard size
Billboard proximity to the travel lane
Length of message
Complexity of message
Size and font of message characters
Proximity to official traffic control devices
Proximity to roadway geometric design features (e.g. vertical and horizontal curves)
Proximity to other billboards
Complexity of the environment in which the billboard was located
Frequency of message change
Method by which message was changed
Traffic speed
Traffic density
Driver familiarity with the roadway and environment

Dr. Granda pointed out a simple truth - one that was obvious to researchers, but perhaps not fully appreciated by the stakeholders interested in the outcome of this research – that it was realistically impossible to undertake a study that accounted for all of these variables, either through manipulating them, eliminating them, holding them constant, or controlling for them statistically. In short, Dr. Granda pointed out, the proposed study could address only the most basic number and level of variables, and the remainder would remain uncontrolled. It was determined by the research team that, if the Phase II study found that the basic CEVMS variables (e.g. size, luminance, placement, message change interval, etc.) could be shown to differentially impact driver response and performance, then a follow-up study (Phase III) would be done in a laboratory setting (i.e. a driving simulator) in which levels of these variables could be manipulated to learn which were the cause for concern.

Due to a series of problems with the relatively unproven vehicle-mounted eye-tracking technology being employed, actual data collection was delayed and the study ran over-budget. Because of concerns related to these issues, Dr. Molino left the project, and was replaced as Principal Investigator by Dr. William Perez, also of (then) SAIC. Soon thereafter, Dr. Granda retired from Government service, and his position as COTR was assumed by Dr. Christopher Monk, then of FHWA.

Unfortunately, for reasons never made public by FHWA, the original design intent of the study, to hold key variables (such as sign size and height, message duration, etc.) constant while carefully controlling for others, was not followed (Gramatins, 2010, Monk, 2010). The consequences of this action, and other failures discussed below, have resulted in a study final report that sheds little, if any, new light on this important subject, and allows the reader to draw no conclusions about the potential safety impacts of these signs.

Since the report's issuance on the FHWA website on December 30, 2013, its availability has led to conflicting public statements by stakeholders, advocacy groups, and the popular press, to FHWA policy statements without scientific or research basis, and to growing legal challenges both for and against CEVMS throughout the country. The unfortunate irony of this is that State and local governments nationwide have waited anxiously for several years in the now lost expectation that the FHWA study would resolve the question of digital billboard safety for the benefit of all.

This brings us to the present report.

ACKNOWLEDGEMENTS.

This document is my own work. No person or organization has suggested or requested that I write it. It was done on my own time, and at my own expense. It is based on my reading and interpretation of the FHWA Report titled: "*Driver Visual Behavior in the Presence of Commercial Electronic Variable Message Signs,*" dated September 2012 (unnumbered) but issued by FHWA on December 30, 2013, as well as the draft version of this report, dated March 2011 (numbered FHWA-HEP-11-014), and a summary of the peer reviewers' comments to this draft report dated May 16, 2011. All of the above materials were made available on the FHWA website on December 30, 2013, and can be accessed at: http://www.fhwa.dot.gov/real_estate/practitioners/oac/)

This report is further based on several earlier FHWA reports on this topic, as cited throughout this report.

Finally, this report is based on personal discussions that I have held with key personnel involved in several phases of the FHWA work, including: Mr. Janis Gramatins, Drs. John Molino, Thomas Granda, and Christopher Monk, and others within and outside the agency.

Although this report is my own work, and I am fully responsible for its contents and for any errors that may be present, I believe that the subject of digital billboards, and role played by the FHWA report in influencing the ongoing debate about such signs, was too important to rely upon my opinions and conclusions alone. Therefore, I reached out to professional colleagues, worldwide, all of whom have specific expertise in the fields of research needed to properly and objectively review the FHWA report (e.g. eye movement recordings, luminance measurement, instrumented vehicle studies, statistics) and asked them to review my report as well as the three FHWA documents issued by the agency in December 2013. Each of these subject matter experts performed their review on their own time and at their own expense. I gave them no guidance other than providing them with a simple suggested outline for their comments, and I promised them nothing more than that I would take their comments into account as I prepared a revised, now final, report.

I am grateful to all of the reviewers who gave of their time and effort to undertake this critical and valuable process. In Table 2 I have listed, in alphabetical order, the names, affiliations, and countries of those reviewers who provided peer review. Two reviewers asked to remain anonymous because they have, have had, or may have, relationships with FHWA that they did not want to risk. One individual declined to participate because of what he considered to be a perceived conflict of interest given ongoing work between his organization and the FHWA. In addition, one of the reviewers on this list was retained separately, and was offered an honorarium, to perform a broader independent review.

Table 2 – Peer Reviewers of this Report

- Anonymous, Ph.D., United Kingdom
- Anonymous, Ph.D., USA
- Christer Ahlstrom, Ph.D., The Swedish National Road and Transport Research Institute (VTI), Linkoping, Sweden
- Frank Berra, Manager, Network and Land Use Policy, VicRoads, Kew, Victoria, Australia
- John D. Bullough, Ph.D., Senior Research Scientist, Lighting Research Center, Rensselaer Polytechnic Institute, Troy, New York
- Barry A.J. Clark, Ph.D., Director, Outdoor Lighting Improvement Section, Astronomical Society of Victoria, Inc., Australia
- Jessica Edquist, Ph.D., Monash University Accident Research Center, Melbourne, Victoria, Australia
- Donald L. Fisher, Ph.D., Professor and Department Head, Mechanical and Industrial Engineering, University of Massachusetts, Amherst, Massachusetts
- Christian B. Luginbuhl, Ph.D., U.S. Naval Observatory Flagstaff Station, Flagstaff, Arizona (Retired)
- Marieke Martens, Ph.D., Professor, Centre for Transport Studies, University of Twente; Department of Traffic Behaviour, TNO Human Factors, The Netherlands
- Richard F. Pain, Ph.D., Senior Technical Coordinator, Transportation Research Board of the National Academies (Retired), Washington, DC
- Christopher Patten, Ph.D., Senior Research Fellow, Swedish Technical Research Institute (VTI), Borlang, Sweden
- Bryan Reimer, Ph.D., Research Scientist, MIT Age Lab Associate Director, New England University Transportation Center, Massachusetts Institute of Technology, Cambridge, Massachusetts
- Alison Smiley, Ph.D., President, Human Factors North, and Adjunct Professor, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada

Several additional individuals were unable to perform the requested review due to the press of other work or because they were on leave or sabbatical from their positions.

BACKGROUND.

In 2009, nearly 30 years after a Federal Highway Administration (FHWA) report on CEVMS (Wachtel J. & Netherton, RD, 1980) recommended that the agency undertake a research study to examine the potential effects of such advertising signs on driver performance, such a study was begun by the FHWA's in-house contractor, Science Applications International Corporation (SAIC – now Leidos) (Molino, et al, 2009). In a report to the Transportation Research Board (TRB) Digital Billboards Subcommittee in January 2010, Gramatins, (2010) (the FHWA staff member who was the "customer" for the study), stated that the final report was expected to be issued three months later, by April 2010. One year later, at TRB's 2011 Annual Meeting, the FHWA COTR, who had agreed to present the final results of the study in a lectern session, was informed days prior to his scheduled talk that he could not do so, and offered instead essentially the same presentation that had been given a year earlier (Monk, 2010, 2011). As recently as May 2012, FHWA personnel publicly stated that the report was not yet available. Finally, on December 30, 2013, FHWA placed on its website (at http://www.fhwa.dot.gov/real_estate/practitioners/oac/) the final report (backdated September 2012), which the agency described as "peer reviewed," together with an unpublished draft report dated March 2011 (described as "non peer reviewed") and a document containing comments from three independent peer reviewers who had been retained to review the draft report.[1] Stakeholders and interested parties greeted the release of the final report with relief, and significant press coverage, including text quoted out of context, presentations by and for special interest groups, and press releases by advocacy groups, followed within days. After the release of the final report, FHWA was again invited to make its long-promised presentation to the TRB Digital Billboards Subcommittee at its January 2014 meeting, but declined to do so. As someone who has followed, and played a role in, the discussion and debate about the potential effects of digital billboards on driver performance for more than 30 years, I set out to review both the draft and final reports, as well as the peer review comments to the draft. This document is the result of that review.

As will become clear in the following pages, I identified numerous areas of the FHWA study that caused me concern. (By way of full disclosure, I was initially retained as a consultant by the contractor at the request of the FHWA COTR; however, my services were no longer requested after the departure of the study's initial Principal Investigator, Dr. John Molino). I have tried to identify those concerns as clearly as possible below. Where possible, I have provided references to the applicable page, figure, or table numbers in the FHWA report, so that the reader may quickly go to those sections of interest. Except where stated otherwise, these page

[1] Of concern is the fact that, nearly one year after the posting of these documents on the FHWA website, the "final report" has been given no official FHWA document number, whereas the non-peer reviewed draft report has been assigned the official FHWA report number FHWA-HEP-11-014. Thus, someone performing an online search for this report is likely to be led to the draft report, rather than the final document.

references are linked to the final (September 2012) report. It should be noted that the study was conducted in two different cities, with two different sets of participants. The two cities were Reading, Pennsylvania (Reading) and Richmond, Virginia (Richmond).

FHWA'S DECISION TO PUBLISH BOTH DRAFT AND FINAL REPORTS, AS WELL AS REVIEWERS' COMMENTS.

It is rather unusual for an organization to make available to the public both an unreviewed draft report and the final report itself, as well as the reviewers' comments to the draft. But FHWA took this action and made available all three documents on its website on December 30, 2013. This decision enabled any conscientious or interested person to review all of these documents, to compare the final report to the draft and evaluate the changes made, and to determine whether, and to what extent, the peer reviewers' comments were addressed in the final document. Although my report might well have been justified based on a review of the final FHWA report alone, it is stronger, more detailed, and more critical due to the availability of these multiple documents from FHWA.

As one of the peer reviewers to the present report stated: "It is not usual to include a discussion of changes that were made between the draft and final report. However, given the long wait, the great interest and the public nature of the work, it would be appropriate to address the differences with (a statement such as): 'Initially we did (a) but based on feedback from reviewers that (b), we did (c), and, at a minimum, to clearly explain the methods used in the final report."

This review raises several questions about FHWA's methodology, measurement approach, reference sources, and, ultimately, the agency's findings and conclusions. Some of the concerns raised herein may seem minor to the casual reader, but all of them contributed, in a non-trivial way, to significant weaknesses in the final report, and questions about the conclusions reached in that report. Because of the implications for policy at the Federal, State, and local levels due to the FHWA final report, I believe that this review and criticism is of importance to interested parties and cognizant officials involved with influencing or implementing such policies.

Different stakeholders in the field of roadside outdoor advertising have seized on the FHWA report in support of their own aims. My report has no agenda other than to shine a light on this long-awaited FHWA research study, to allow independent readers to review the FHWA documents and this report, and to reach their own conclusions about the validity of the agency's findings. Ultimately, I believe that, far from being the definitive research report that will enable State and local governments to establish meaningful regulations about roadside digital signs (on- and off-premise), the FHWA report provides little clarity about this contentious issue. As a result, I believe that public and private road authorities should look to the dozen or more research studies published, world-wide, in the past several years, and should err on the side of caution and traffic safety in addressing CEVMS and other commercial roadside signs.

RED FLAGS.

As originally identified in the first FHWA report on CEVMS (Wachtel J. &., Netherton, RD, 1980), and as repeated by Dr. Granda when he served as the COTR for the most recent FHWA project (Granda, 2009), certain characteristics of CEVMS, including, but not limited to, size, height above grade, proximity to the driver's lane of travel, and side of the road on which the sign is located, can each have an important effect on a driver's response to the sign. Of course, there are many other sign characteristics that are considered to be important contributors to potential driver distraction – characteristics such as sign luminance and dwell time (frequency of message change). But sign size, height and location characteristics have been deemed sufficiently important that they have been described thoroughly in nearly every scholarly study in this subject.

It is well accepted in the scientific research community that the state of knowledge progresses both from research that confirms its hypotheses and research that fails to do so. However, a key tenet in support of the ability to move research forward is the requirement that researchers report their experimental design and methods in sufficient detail that future researchers can attempt to reproduce their study in an effort to replicate their findings. But the FHWA researchers did not identify the roads driven or the signs (CEVMS and others) used in this study, thus precluding others to attempt to repeat the study. However, the decision to not identify the signs or roadways used brought with it additional adverse consequences - specifically that apparent errors made by FHWA in the identification of important CEVMS characteristics cannot be fully understood or interpreted, leaving readers without the ability to know just how widespread and significant these errors were.

I have begun the technical discussion in this report with what I have called "red flags," discrepancies between the draft and final reports that are so central to the conclusions reached by the authors that they call into question the adequacy and accuracy of the project as a whole. These red flags require no interpretation on the part of the reader to understand the serious, unresolved errors made by FHWA and their study and internal review process.

To fully understand the significance of these Red Flags, it is useful for the reader to know that during the 31 months (May 2011 – December 2013) between the receipt of the peer reviewers' comments to the draft report and the issuance of the final report, many changes were made to the report itself, in both major and minor ways. But one thing that was *not* done by FHWA was any retesting of participants or any reanalysis of the roads or the CEVMS that were studied. We know this because the final report makes clear that all Reading data was collected between September 18th and October 26th, 2009 (pg. 21), and all Richmond data was collected between November 20, 2009 and April 23, 2010 (pg. 43). Accordingly, a number of major discrepancies in the identification of certain key sign characteristics between the draft and final reports raise serous questions, not addressed by the researchers, about the applicability and validity of their data – specifically, the measured and analyzed eye gazes that were ostensibly made to these signs.

Both the draft and final reports contain inventories of the target billboards and control sites used in the two cities, albeit without sufficient specificity for a reader to actually identify any of these signs on the ground. The inventory for Reading appears in Table 2, pg. 21 of the draft report, and Table 2, pg. 17 of the final report. That for Richmond appears in Table 8, pg. 44 of the draft report, and Table 7, pg. 40 of the final report. All of the discrepancies discussed below were identified based strictly on the data in these tables. It is possible that additional discrepancies exist, but it was not possible to identify such discrepancies due to the lack of detail in the tables.

NUMBER OF TARGET BILLBOARDS STUDIED.

The actual number of billboards studied decreased dramatically, in both cities, from the draft report to the final, and the authors offer no explanation for this decrease. The comparison shows that there were a total of two fewer standard, and 12 fewer digital billboards included in the final report compared to the draft:

	DRAFT REPORT		FINAL REPORT	
	CEVMS	Standard	CEVMS	Standard
Reading	11	5	4	4
Richmond	9	5	4	4
Total	20	10	8	8

Since the study was conducted only once, the deletion of data for more than 50% of the CEVMS in the final report must represent a failure in either data collection or data analysis. Given that the study of driver eye glances to CEVMS was the principal purpose of the study, the elimination of more than half of these signs from the database raises serious concerns. It is noted that, as late as August 2010, FHWA personnel were reporting that there were 10-14 CEVMS studied in each of the two cities (5-7 per route; 2 routes per location) (Monk, 2010). The two (draft and final) reports provide insufficient information to identify which billboards were eliminated from consideration for the final report. We cannot, therefore, know whether there was a pattern to this process, or whether the smaller number of billboards studied for the final report resulted in some kind of bias.

APPROACH DISTANCE TO BILLBOARDS.

Between the draft and final reports, in both cities, there is a dramatic difference in the specified approach distance to the billboards studied. (In three of the four tables [both final reports and the draft report for Richmond], there is a column titled "Approach Length (ft.)." Only for the Reading draft report is the equivalent column labeled "Data Collection Zone Length (ft.)." This is more than a simple linguistic difference, as discussed later in this report where I express concerns with how these Data Collection Zones (DCZs) were established, and the implications of these decisions. (I will refer to these as DCZs, as do the FHWA authors).

CEVMS in Reading. There are discrepancies in the data provided by the authors in the DCZ column between the draft and final reports. In the draft report for Reading, 10 of the 11 CEVMS signs were described as having a 960 ft. DCZ; the 11th had a DCZ of nearly twice that length (1860 ft.). But in the final report for Reading, of the four CEVMS reportedly still studied, the authors report DCZs of 375, 853, 537, and 991 ft. There is not a single match, and three of the four described DCZs are considerably shorter than those reported in the draft report, which, with all else held constant, would clearly result in fewer eye glances to these signs.

Standard billboards in Reading. In the draft report, the 5 standard billboards had approach lengths of 960, 682, 960, 547, 960 ft. In the final, the four remaining signs had approach lengths of 644, 774, 833, 770. Again, not a single match, and generally shorter DCZs in the final report compared to the draft.

CEVMS in Richmond. The same inconsistency occurs in Richmond. In the draft report, the six CEVMS all had DCZs of 960 ft., but by the final report, the distances for the four remaining CEVMS were 696, 602, 297, and 321 ft. In this case, not only are the DCZs described in the final report shorter than those listed in the draft, but two of them are shorter by a factor of approximately three.

Standard billboards in Richmond. The differences between draft and final reports are again in conflict. In the draft, there were 5 standard billboards, with approach lengths of 889, 960, 863, 960, 960. In the final, the 4 standard billboards, were at approach lengths of 857, 651, 997, and 816. Again, the majority of these discrepancies are in the same direction, with the obvious consequence of fewer glances to billboards – i.e. shorter DCZs in the final report.

It is not possible to know whether the authors intentionally labeled three of their four charts as "Approach Length," and one as "Data Collection Zone." We may assume that their use of the term approach length conforms with standard practice in this field, i.e. that it refers to the earliest distance from which an approaching motorist could see any particular billboard. "Data Collection Zone," of course, is quite different. In the FHWA study, the DCZ is that section of roadway in advance (upstream) of a billboard that begins at a distance 960 ft. away (artificially constrained, as reported by the study authors, by the eye-tracker's visual field given its 2° field of view), and ends when the billboard disappears from the scene cameras' field of view. Thus the end point of the DCZ again creates an artificial constraint because, in all cases, the driver/participant's view of a billboard continues well after the billboard is no longer visible to the scene cameras. Thus, the Data Collection Zone can never be longer than the "Approach Length," and given the eye-tracker and scene camera limitations that bound it, it is almost always shorter, sometimes significantly so. Given our knowledge that the authors did not collect on-road data a second time after the draft report was reviewed, the dramatic differences reported for approach/DCZ length between the draft and final report is both puzzling and of concern, especially given the frequency, magnitude, and consistency in direction of the discrepancies. One possible explanation is that the roadway distance in which eye-glance data to billboards was collected was curtailed by the researchers in the laboratory, after reviewing the data that went into the draft report. But if this is the case, one still must ask why this decision was made, and why there was no effort to explain it in the final report. Nonetheless, if some collected data was simply purged between the draft and final reports, the implications for the appropriateness and validity of the findings as reported in the final report must be questioned. Without an explanation from the authors, this issue cannot be put to rest.

SIZE OF BILLBOARDS.

There are puzzling discrepancies in the reported size of the target billboards between the draft and final reports. These differences are important because the size of the billboard affects a number of relevant driver responses, including: the distance from which the sign can be seen, the nighttime luminance, the letter and character sizes that determine legibility distances, and the length and complexity of messages displayed which can affect reading and comprehension time.

To cite a few examples, and, remembering that, in all cases, there were more billboards of each type in the draft report than in the final:

- The final report for Reading shows three standard billboards of 14x48 ft. – yet there were only two standard billboards of this size in the draft report.

- The Reading final report shows one standard billboard measuring 10'6" x 22'9", yet there is no billboard of this size in the draft report.

    - In the case of this discussion, it must also be remembered that, because the studied billboards were not identified by FHWA, and because there were several billboards listed that were of the same size, there may have been additional cases of this same type of discrepancy that cannot be identified.

BILLBOARD SETBACK FROM ROAD.

Setback from the road is an important measure because it determines the length of time that the billboard will remain within the driver's forward field-of-view, as well as the distance to the billboard at which it will disappear from the driver's view. It is of greater importance in this study because of the limitations in the eye-tracker's resolution at distances greater than 960 ft., and because of the premature cut-off of eye-glance measurements to billboards at closer distances due to the inability of the scene camera array to capture more than a ± 40° segment of the driver's field of view. (This latter problem could have been solved with the addition of a fourth scene camera, or by using shorter focal length lenses on the scene cameras that were employed). The following paragraphs demonstrate this problem:

- In the Reading draft report, two different 14 x 48 ft. standard billboards are shown. The table shows these billboards set back from the road at 50 and 97 ft., respectively. However, in the final report, three standard billboards of this size as shown (one of which seems not to have existed in the draft report). These three are set back from the road by 10, 20, and 35 feet, respectively. The setback differences from draft to final are quite large, and the number of eye-gaze measurements made to these billboards would potentially be affected.

- Also in Reading, one of the CEVMS in the final report, which measures 10'6" x 22'9", is shown to be setback from the road by 12 ft. In the Reading draft report, however, there were 5 billboards of this size, and all of them were shown as setback from the road by at least 35 ft., ranging up to 128 ft. In other words, the draft Reading report shows the setback distance of this billboard roughly between 3 and 10 times farther from the road edge than does the final report.

- In Richmond, there are several similar cases. For example, the draft report shows two CEVMS of 14’0” x 28’0” each. The setbacks given are 56 and 119 ft. from the road. In the final Richmond report, the (presumably same) two digital signs are shown as having setbacks of 37 ft. each.

- A CEVMS measuring 11’0” x 23’0” is shown as having a setback of 35 ft. in the draft report, and 71 ft. in the final, more than twice the distance.

BILLBOARDS ON RIGHT OR LEFT SIDE OF ROAD.

Perhaps the greatest concern for a reader attempting to understand the findings of this study is that, between the draft and final reports, some target billboards appear to have crossed from one side of the road to the other. Three examples illustrate this concern:

- One of the standard 14’0” x 48’0” billboards in Reading is shown in the relevant table to be on the on the right side of the road in the final report; however, in the draft report, the only two standard billboards of this size are both on the left.

- In Richmond, the same 11’0” x 23’0” CEVMS discussed above is said to be on the right side of the freeway in the draft report, and on the left side in the final report. Additionally, one of the standard billboards, which measures 10’6” x 45’3”, shifts from the left to the right between the draft and final reports.

- There may be several more cases of these roadside switches from draft to final. However, because there are often several target signs of the same size listed in each report, and because the authors do not provide critical sign placement information (such as GPS latitude and longitude data), it is not possible for the reader to directly compare them.

SPECIFIC CONCERNS WITH THE FHWA REPORT.

As discussed in the sections that follow, the present report identifies several areas of concern with the FHWA study. Below, I have provided, wherever possible, references to the applicable page, figure, or table in the FHWA report so that the reader may quickly refer to the original material that led to the concerns expressed herein. Except where stated otherwise, these page references are to the final FHWA report.

CHARACTERIZATION OF LONG GLANCES TO BILLBOARDS.

In the Executive Summary (p. 3), the authors describe four long dwell times greater than 2,000 ms (2 sec.) each that were observed to billboards in the study. They state that their review of the data showed that these billboards “were not far from the forward view while participants’ gaze dwelled on them.” They conclude: “Therefore, the drivers still had access to information about what was in front of them through peripheral vision.”

Several of the peer reviewers to the present report expressed concern about the subjectivity of that statement. One asked: “What do they mean? How do they determine this? Are they calculating visual angles? If so, they need to state the visual angles for each glance.” Another reviewer said: “I don’t understand either quote.” Another opined that, since the authors did not define what they mean by “not far from the forward view,” the reader cannot assess the relevance of this statement.

Nonetheless, their conclusion is an empirical statement that requires testing; testing that they did not perform. Further, the statement is in direct conflict with research findings reached by Fisher and his students at the University of Massachusetts, Amherst in a series of studies spanning the past several years (e.g. Divekar GP, 2012, Chan, 2007). Their simulator-based studies have shown that looking at external distractors with peripheral vision available for the road ahead fails to provide the driver with the visual attentional resources necessary to anticipate and respond in a timely manner to hidden and emergent traffic hazards. Although publicly available, the FHWA authors did not cite the work of these researchers.

The 2-second criterion mentioned by the FHWA authors is based on work known as the “100 car study” (Klauer, 2006). As one of our peer reviewers noted, this work provides a useful, accepted definition of inattention/distraction, but one that is too limited for studying billboard distraction. The reviewer continues: “Glance duration is only one of three measures needed to characterize looking behavior away from the road and define distraction. If you only make one 2-second glance at a sign, there is a much lower risk than if you make 2, 4, or 6 glances at that sign. So, the frequency/number of glances is a partial measure of distraction. In particular a number of short glances, e.g. under 2-seconds, in fairly rapid succession may pose a risk similar to one glance of 2-seconds. Moving the eyes back and forth from the road means that the eyes may, in total, be away from the road for more than 2-seconds. The expectancy that peripheral vision will “fill in” for glances not too far off the roadway has not proven viable.

Combining the measures of glance duration and frequency provides a more complete picture of where attention is likely focused, i.e. distraction. However, to be operationally useful there needs to be some limit on the time over which these are measured. For example, two or three glances of 1-second each over a 2-minute period would not typically have the eyes off the road for enough time to create a heightened risk. However, what if the eyes were off the road, e.g. looking at a sign, for 2-seconds or more within a 6-second window? Would that not be a more meaningful measure of distraction? If your eyes are off the road for a cumulative 2 seconds during any 6-second period, then both duration and frequency are accounted for. This is the measure that was used in the Danish study (Herrstedt, 2013). It is an operational measure of distraction that is more comprehensive than any previously used measures yet can be used in both field and laboratory testing.

But there is a further concern with the use of the "2-second" glance criterion. In the FHWA report (and in other research), the researchers rely on the conclusion reported by Klauer and her colleagues that a 2-second or longer glance duration away from the forward roadway is generally considered distraction. But more recent research demonstrates, according to one of our peer reviewers, "that using 2-seconds as a criterion or threshold for distraction clearly is insufficient. Victor (2014), as well as the Danish study (Herrstedt, 2013) found that the length of glance duration defining distraction is highly situation dependent. On a clear open road with little traffic density, 2-seconds may not really be a distraction. With increasing vehicle density, especially shorter vehicle headways, and opportunities for other vehicles maneuvering in or out of the roadway/lane, glance durations under 2-seconds are, in fact, distraction with high risk consequences."

Finally, recent research (subsequent to the issuance of the FHWA report) by Victor and his colleagues, using a much larger naturalistic driving study cohort than the 100-car study, demonstrates that the majority of crash and near crash events involving distraction followed a distracted glance duration of less than 2-seconds (Victor, 2014).

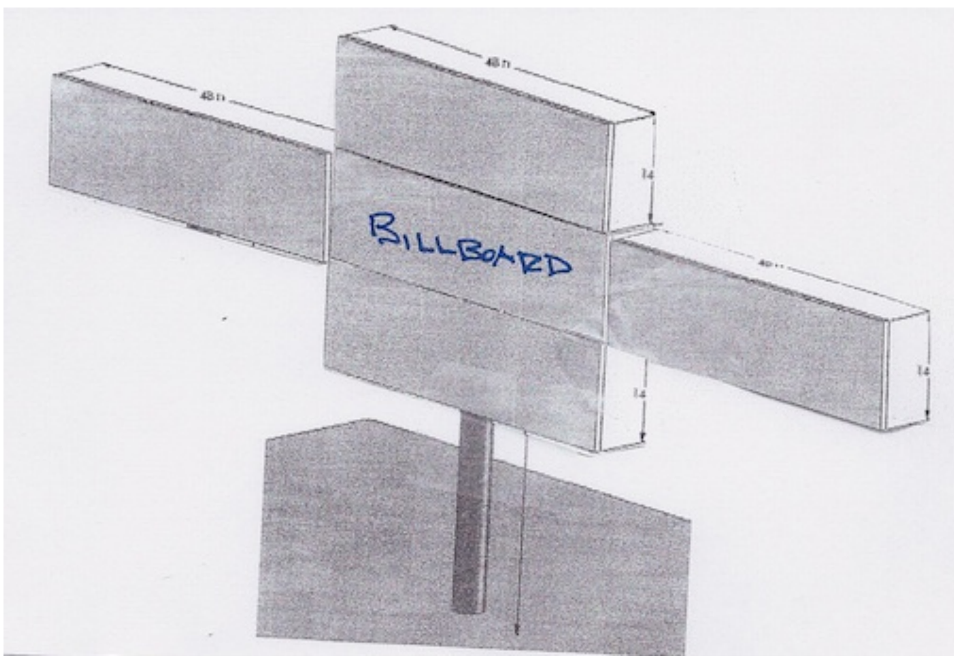
In short, one of our reviewers noted, "if you redefine distraction as a function of glance duration (not using a time criterion) relative to traffic density, the FHWA conclusions about the effect of billboards on distracted driver behavior will, in all likelihood, be significantly altered."

MEASUREMENT OF SIGN LUMINANCE.

When motorists express concern about the distracting effects of digital billboards, they typically seize on two operational characteristics of such signs: the length of time that each message remains on the screen before changing to the next message (called "dwell time" in the industry), and the luminance levels at which such signs often operate at night (Wachtel J., 2011). But, despite statements made by the COTR, after all data had been collected, that dwell time was studied (Monk, 2010), the FHWA study seems to have ignored this issue (except for noting it as part of billboard inventory). In addition, the study's treatment of sign luminance, discussed immediately below, is questionable.

On pp. 19-20, the authors describe their measurement methodology for determining the luminance (day and night) of both standard and digital billboards. The description states: "Measurements were taken by centering the billboard in the photometer's field of view with approximately the equivalent of the width of the billboard on each side and the equivalent of the billboard height above and below the sign." Although an illustration of the photometric measurement approach used was presented in the draft report [Figure 5, pg. 17], this figure appears to be erroneous in that it depicts the width, but not the height of the area that the authors say they measured. No equivalent illustration is provided in the final report. Without such an accompanying diagram, it is unclear exactly what was measured. However, taking the authors' description literally, the sketch shown below as Figure 1 provides my interpretation of the area included for measurement. If we assume the most common 14' x 48' billboard size, then the targeted measurement area would encompass an area 144 ft. wide by 42 ft. high.

FIGURE 1



If the above sketch accurately describes the measurement approach used by FHWA, then their approach would seem to be inappropriate, and different than other known or published approaches to the measurement of billboard luminance. Lighting experts agree that the appropriate way to measure the luminance of a billboard (day or night, digital or traditional) is to use a photometer with a narrow acceptance angle (1° and 1/3° are most often used) (Illinois Coalition for Responsible Outdoor Lighting, 2010) (Luginbuhl, 2010) (Bullough, 2011). The photometer is aimed at the billboard at a distance close enough that the sensor captures only a small, clearly defined section of the billboard presenting only a single color, and that becomes the luminance value of record. This process can be repeated to capture different colors of illumination and different LED output levels. To measure a billboard's maximum luminance level, it is generally recommended that the sign be set to display an all-white image. The method followed by the FHWA researchers, as interpreted from their narrative description, appears to have captured the luminance of the billboard plus the luminance of the area above, below, and to both sides of the billboard. This data was then used to report an "average" luminance reading. This is not billboard luminance, but rather billboard plus background luminance. If this is in fact what was done, we would expect that billboard luminances reported in the study for nighttime measurement would be far lower than billboard luminances captured using the widely accepted method discussed above. And indeed, this is what the FHWA data show. In Table 3 on pg. 27, the authors report billboard luminances averaging 2126 cd/m^2 and 56.0 cd/m^2 for digital billboards (day and night, respectively), and 2993 cd/m^2 and 17.8 cd/m^2 for traditional billboards (day and night). These readings, especially at night, are far below typical readings obtained by researchers in the field (Bullough, 2011; Luginbuhl, 2010), (ICROL, 2010), (Wachtel, 2014) and, indeed, bear little resemblance even to recommendations for nighttime luminance promulgated by the outdoor advertising industry itself (which are invariably higher than those recommended by lighting experts).

In addition, the authors report only mean luminance values and standard deviations. Since traffic safety and lighting experts, local officials, and even the lay public are primarily concerned with maximum luminance values at night, it is surprising that the authors failed to provide this information.

Further, given the hypotheses expressed by others, and the complaints regularly mentioned by the lay public in the media that it is the maximum luminance values of digital billboards that most effectively capture visual attention and contribute to distraction (and, potentially, veiling luminance or glare), it would have been appropriate for the authors to have recorded the billboard luminance value experienced by each study participant as he/she approached and passed each billboard. This would have added little complexity to the study and would have added substantially to a reader's ability to interpret its results. But this was apparently not done, with the result that the different driver participants were likely exposed to very different levels of billboard luminance (for the digital billboards) during their drives, a factor that could well have contributed to quite different eye glance patterns.

Finally, if our interpretation of the method followed by the FHWA researchers to record luminance values is incorrect, and they actually did record values for just the billboard without the backgrounds, then the only logical conclusion that can be drawn is that the billboards studied in these two cities are substantially less bright (produce less luminous intensity) than typical billboards nationwide.

The lighting specialists who served as peer reviewers of the present report differed in their opinions about the both appropriateness of FHWA's luminance measurement methods and the clarity with which FHWA explained their methods. They all agreed, however, as one put it, that: "the luminances they measured for the billboards are anomalously low compared to expectation and other work."

TARGET BILLBOARD CONTRAST WITH BACKGROUND.

The FHWA report discusses the authors' efforts to determine a billboard's contrast with its background, and this is, indeed, an important issue, especially at night, because higher contrast contributes to greater conspicuity and, hence, potentially more and longer eye glances. But, beyond FHWA's statements that they made such measurements, the authors appear to have done nothing with this information, and therefore the discussion is meaningless. As one of the peer reviewers to the present report stated: "The key variable apart from (luminance) range is the ratio of signal intensity (the billboard) to the background luminance, which is why I think (the authors) fell short in not considering something similar in connection with sign conspicuity." Another reviewer questioned why the authors did not perform a conspicuity measurement, for which methods are readily available. A third reviewer put it this way: "Absolute luminance tells one story. But if you don't know the contrast ratios involved you cannot predict the attention value of the sign. If, as in the FHWA study, you combine sign and background luminance, you lose the very parameter that is most prominent in determining the attention value of the sign."

Many years ago a research program studying official highway signs (i.e. not billboards) developed a model to quantify the attention value of a sign. Contrast ratios were the most significant contributor to attention value. The experiments, both field and laboratory, took into account both rural and urban (visually complex) environments (Pain, 1969).

VISUAL COMPLEXITY.

The authors provide a lengthy discussion about "visual complexity" (pg. 20) in which they criticize the work of Regan, et al (Regan, 2009), and of Horberry and Edquist (Horberry, 2009), as not providing a "systematic or quantitative way of classifying the level of clutter or visual complexity present in a visual scene." Instead, they recommend use of a method proposed by Rozenoltz, et al (Rosenholtz, 2007). Had the authors been more diligent in their literature review, they would have found ample documentation of such a classification scheme in the work by Regan, Horberry and Edquist (Edquist, et al, 2008, Edquist, 2010).

This discussion is relevant to the main theme of the study because it has been demonstrated that roadside billboards are a component of, and contribute to, visual clutter, and it has been shown that the presence of visual clutter can cause drivers to experience greater difficulty in identifying, and consequent delays in responding to, important road and traffic information (e.g. regulatory or warning signs, emerging traffic hazards) than would be the case with reduced clutter. The discussion of visual clutter in the final report, and the measurement strategy used to assess clutter, is substantially different from that described in the draft report, and the authors offer no information about the rationale for the change or the effect of the change on their findings. In addition, in the draft report, the authors state that they captured visual images for each DCZ and analyzed it to compute its visual complexity. Despite these extensive discussions, the final report is silent on whether or how the authors utilized this information, and why they followed such a different approach from that described in the draft report. As a result, possible contributions to the discussion of the impacts of roadside visual clutter on driver attention, and the role of billboards in contributing to such clutter, have been lost.

EYE-TRACKER PERFORMANCE.

The authors report that, in Reading, "if the eye tracker performance became unacceptable, then the researcher in the rear asked the participant to park in a safe location so that the eye tracker could be recalibrated." In Richmond, the situation was worse. The authors report that, because the route was somewhat longer in Richmond than in Reading, the eye-tracker data collection system "had problems dealing with the large files that resulted." They go on to state: "To mitigate this technical difficulty, participants were asked to pull over in a safe location during the middle of each data collection drive so that new data files could be initiated" (p. 43). In neither case do the authors state how many such occurrences there were, although it seems as if the overload problem occurred with every Richmond participant since all drove the same route. The authors are silent on the potential impact of these interruptions on participant performance or on the otherwise continuous data collection activity.

At a more basic level, it appears that the major impetus for the withholding of the draft report and the issuance of the final report nearly two years later was the fact that the eye glance measures (durations) presented in the draft report were clearly unreasonable, as pointed out by the FHWA peer reviewers. As one of our peer reviewers stated: "It is troubling to consider that those responsible for the report could get this so wrong." Our peer reviewers also pointed out that accuracy with the SmartEye system (the system used in the FHWA study and by several of our peer reviewers) is difficult to achieve, especially in the challenging environment of a moving vehicle. One reviewer said: "Around the straight ahead position, things work well, but once the driver's gaze drifts outside the line of the outer cameras, accuracy drops off rapidly."

SCENE CAMERA AND EYE-TRACKER FIELD OF VIEW.

The authors state that the three roof-mounted scene cameras captured an 80-degree wide field-of-view, which represented "the forward view area available to the driver through the left side of the windshield and *a portion of the right side of the windshield"* (italics added). They continue: "the area visible to the driver through the rightmost area of the windshield was not captured in the scene cameras." Of course, a typical driver has a field-of-view far wider than what was captured in the study's scene cameras. This is a serious concern because, as later described by the authors, the eye tracker did not record driver glances to the left or right of the scene camera limits, and thus the eye gaze data, central to the study, eliminated an unknown percentage of visual fixations, thus artificially understating both the number and duration of such fixations.

EYE-TRACKER CONSTRAINED FIELD OF VIEW.

The researchers' approach to eye-glance data reduction (p. 23 ff) is critical to the understanding of the results, and raises important concerns. The authors describe static "regions of interest" (ROIs), which include eight specified data collection regions; six within the scene camera view, and two that were outside (above and below) the view of the scene cameras but still accessible to the eye tracker. Critically, as discussed immediately above, the areas to the left and right of the scene camera field of view, and glances made in these areas, despite being accessible to the eye tracker, were ignored. This is important because, as discussed above, "the area visible to the driver through the rightmost area of the windshield was not captured by the scene cameras." There is no discussion in the report about the extent to which the scene cameras cut off the drivers' view through the right side of the windshield. In addition, the scene cameras did not capture the view through either the right or left side windows – areas where drivers would likely have to look to observe their side view mirrors, and where they might look at billboards as the instrumented vehicle approached them. Because the researchers did not analyze eye gazes to the right or the left of the scene camera boundaries, the eye gazes that *were* analyzed by the system therefore represented only a subset of all relevant eye-glances that were actually made by the participants. This issue can be better understood by examining the report's Figure 9 (pg. 23). The billboard shown in this image appears in the upper right segment of the screen. Had this frame grab image been taken one or two moments later, as the instrumented vehicle got closer to this billboard, the billboard would have appeared to move to the right, outside the recording limits of the scene camera, and the researchers would have dropped from analysis any such eye gazes made to this billboard during this time. (We assume that such gazes were captured by the eye-tracker since it is independent of the scene cameras and (presumably) operated continuously during each participant's drive; but that any such gazes were not analyzed). This deletion of critical data is central to the principal purpose of the study, because the billboard in this example (as well as an unknown number of others) would still be visible to the driver through the front windshield, and then through the right side window. But this valuable and relevant eye glance data that could have shown glances to the billboard was not analyzed. Worse, since the authors report their data as probabilities of gazes to billboards vs. to the road ahead (pg. 28 ff), had there been such unanalyzed gazes, the lack of analysis would have artificially reduced the reported probabilities of both the number and duration of views to billboards vs. views to the roadway ahead.

One of the peer reviewers of the present report referred to a study by Lamble, et al (Lamble, 1999) to shed further light on this issue. This was an on-road study that addressed the drivers' ability to detect the slowing of a vehicle ahead while they attended to displays within the vehicle and at various angles off the line of sight. Although data was not collected specifically for views toward the right side of the windshield, views were made to the rear view mirror, both side view mirrors, and the right-side window. The following table, adapted from the Lamble, et al data, shows the mean eccentricity from straight ahead to each of these targets, and the "time lost in detection, in terms of time to collision (TTC)", for each of these targets.

	Mean angle of eccentricity from forward view (degrees)	Time lost in detection (seconds) in terms of time to collision (TTC)
Left side view mirror	44	1.7
Interior rear view mirror	42	2.1
Right side view mirror	63	2.1
Right side window	90	2.8

Using these data, it is clear that the failure to analyze eye glances made to regions beyond the ± 40° cut-off of the scene cameras is a serious limitation to the FHWA study, especially given the fact that all on- and off-premise signs would be located within these angular regions as drivers approached them.

BILLBOARDS WITHIN VIEW OF THE ROAD AHEAD.

The authors state that, for their analysis, the top and bottom segments of the ROIs were combined since "this additional level of analysis was not needed in order to address the research questions." This resulted in three ROIs, defined by the authors as: "LSR – Left side of road," "RA – Road Ahead," and "RSR – Right side of road;" (remembering again that both the LSR and RSR views were artificially constrained by the scene cameras' limited horizontal field of view). Again using Figure 9 (pg. 23) as an example, it can be seen that the billboard appears in the RSR sector. Had this screenshot been captured a moment or two earlier, with the instrumented vehicle farther from the billboard, it would have appeared in the RA (Road Ahead) sector. Since, as discussed above, the authors report their findings of views to billboards as probabilities that are compared to the probabilities of gazing at the road ahead, the question becomes - how did they resolve the critical issue of coding a glance at a billboard that is at the boundary of two segments, or one that shifts from one segment to another? This is discussed below.

DYNAMIC ROIs.

The authors define "dynamic ROIs" on pg. 24. Dynamic ROIs include static objects (such as billboards) that appear to move within the video because of the movement of the instrumented vehicle through the scene, and actual dynamic objects (e.g. pedestrians or other vehicles) that move independently of the instrumented vehicle. They define four types of dynamic ROIs. Two make sense: target standard billboards, and target digital billboards. The other two, however, raise questions. One dynamic ROI is called "Other standard billboard," and is defined as "standard billboard(s) located in the DCZ, other than the target standard billboard or the target digital billboard." But an examination of the few roadside images included in the report suggests that there were many more non-target *on-premise* signs located within DCZs than there were non-target billboards, raising the question as to whether such on-premise signs were simply ignored in the data collection. (The average motorist is not attuned to the technical and legal differences between off- and on-premise signs. In fact, in many locations, digital signs that are considered on-premise display off-premise advertising).

The fourth dynamic ROI is defined as a "driving-related safety risk." The authors' define this as a car that is either actively turning or entering the roadway or one that appeared to be in a position to enter the roadway. This raises several questions – (1) was this dynamic ROI limited to cars (at the expense of other vehicles, pedestrians, or bicyclists) and, more importantly, (2) why did the researchers exclude from this category vehicular traffic in front of the instrumented vehicle that might represent a short headway (following distance) or that might suddenly slow or stop? Recent studies of driver distraction (whether from sources inside or outside the vehicle) have increasingly and appropriately used, as dependent measures, the driver's response to sudden braking by a lead vehicle, or recognition of and response to imminent or emerging hazards, whether in high fidelity simulator environments (Milloy, 2011) or in the real world (Herrstedt, 2013). Unfortunately, such realistic and commonly employed potential hazards were not included in this study. Further, the "driving-related safety risks" that *were* used in the FHWA study are highly subjective and do not appear to rise to the level of concern that would be representative of an immediate threat or hazard (Ayres, 2005).

One of our peer reviewers discussed this concern extensively. He said, in part:

> What is the likelihood of (a driver) missing a cue or event during a glance away from the roadway? In part that is determined by how close the driver is to an object to hit. For example, driving on a four lane road with cars spaced every quarter mile suggests that there is time to make a 2-second glance off the roadway without missing a time-critical cue with immediate safety consequences. On the other hand if there are cars next to you, in front, and behind you, perhaps 1-4 seconds away from you, the risk of glancing away increases dramatically.
>
> For a measure of glancing away to have meaning, it has to be placed in the context of the traffic situation. Traffic volume, typically measured as ADT or AADT, only tells us the number of vehicles per unit of time. A much finer measure of the traffic situation (becoming known as "traffic density") provides the context for putting glance behavior into a more meaningful context. Both the Dukic, et al (2013) and Smiley, et al (2004) studies successfully employed a measure of traffic density to place the risk of glance behavior to billboards into an appropriate real-world context.

EYE-GLANCE DATA REDUCTION/ANALYSIS METHODS.

The data reduction and analysis method used by the researchers changed significantly from the draft to the final report, presumably in response to the peer review comments made to the draft. But the authors describe the methods used in the final report with no reference to the draft report, no reference to the concerns of the peer reviewers, and no discussion of the dramatic and potentially significant changes that took place between the draft and final reports. Was the method used in the final report subject to a separate peer review? Have the revised findings presented in the final report been vetted by independent analysis? Have the original peer reviewers been given the opportunity to review the final report before its release to the public in order to verify that their concerns were properly addressed? Given that none of these questions are addressed in the final report, one must question FHWA's statement accompanying its release that it was peer reviewed.

CONTROL SECTIONS.

On pg. 14, the authors provide an initial description of their roadway "control" sections, which they define as "areas without off-premise advertising." As discussed above, however, at least some of these control sections (no information is provided to enable the reader to know how many) included prominent on-premise advertising. As becomes clear later in the FHWA report, some of these sections contained no signs of any kind (which would make them appropriate as control sections), but others contained on-premise signs, some of which were possibly digital (the report provides no description of signs in the control sections, and the photographs are incomplete and of insufficient detail to support any reader determination). As discussed above, since the typical driver is unlikely to distinguish between on- and off-premise advertising, roadway areas that include on-premise advertising (particularly if such signs are similar in size, location, etc., to off-premise signs) are inappropriate choices as control sections. If we assume that areas with bright (and perhaps changing) signs will attract a driver's gaze to a greater extent than areas with no signs (especially at night), then, by selecting control sections that included advertising signs, including illuminated advertising signs, the study creates the unreasonable consequence of artificially reducing the likelihood of capturing differences in eye glance patterns between treatment sections (those with digital or conventional billboards) and control sections. This concern is exacerbated because the researchers grouped all control sections together for analysis (i.e. they did not separate those with signs from those without signs). Even the Lee, et al study (Lee, 2007), despite its flaws as pointed out by the FHWA authors and this writer (Wachtel, 2007), recognized this peril: Lee and her colleagues categorized their road sections into four subsets: (a) digital billboards; (b) traditional billboards; (c) control sections (no advertising, but possibly some official signs) and (d) comparison sections. The "comparison" sections were road sections that included no billboards (either digital or traditional), but could have included other advertising signs, particularly on-premise and digital signs. In other words, the FHWA control sections were closer in function to Lee's comparison sections than they were to Lee's control sections. And Lee, et al found longer glances to digital billboards *and* comparison sections than to either traditional billboards or control sections. (Although Lee and her colleagues did not evaluate the statistical significance of these differences, an independent, post-hoc analysis of the Lee, et al data showed that these differences were significant) (Hurtz, 2011). As depicted in Figs. 7 and 8 of the FHWA final report (for Reading), and Figs. 29 and 30 (for Richmond) at least some of the control zones on the freeways seem to have had no visible signs of any kind (as is appropriate), whereas the control zones on the arterials had numerous on-premise signs (inappropriate). Further, in Reading, at least one control section also included two large overhead official signs, further reducing its suitability as a control zone.

One of the peer reviewers to the present report raised this question that may apply both to control sections and DCZs. "If 76% of arterial glances and 82% of freeway glances (in the CEVMS condition) were on the forward roadway, then what other objects were the participants looking at in the driving environment? Were there other billboards (or on-premise signs) in the environment that might not have been considered for analysis? If that is the case it is a major flaw in the design."

LACK OF REPRESENTATIVENESS OF BILLBOARD SIZE.

While there are many possible sizes of billboards (traditional and digital), the most common sizes, especially on freeways, are 14'x48' (672 sq. ft.), 20'x60' (1200 sq. ft.). In the final FHWA report, however, only two of the four CEVMS in Reading were of 14'x48', the other two measuring 10'6"x22'9" (239 sq. ft.) each. In contrast, no fewer than five 14x48' digital billboards in Reading were studied in the draft report. In Richmond, none of the four CEVMS discussed in the final report were of typical dimensions, averaging 384 sq. ft., with one of the CEVMSs measuring as small as 253 sq. ft. And it's not that such typically sized CEVMS don't exist in the cities studied. In Richmond, just one of the city's several billboard operators claims two such digital sign faces, and in Reading, seven. Why did the research team select billboards to study that were not representative of the most common sizes? In all cases in which this divergence from standard size existed in this study, the billboards chosen were smaller, by as much as half, than such standard sizes. More puzzling is the fact that, in the draft report, several more billboards were included than were reported in the final report. This "loss" of studied billboards is puzzling, and no explanation is provided. Why were so many billboards that had been included in the draft report removed from the final report?

LACK OF REPRESENTATIVENESS OF BILLBOARD LUMINANCE.

Our concerns about the study's luminance measurement methods are discussed elsewhere in this report. If, however, our understanding of the technique used is not correct, that is, if the researchers did actually measure luminance of billboards without their surroundings, then it must be concluded that the luminance values for the signs used in this study bear little resemblance to those of typical billboards (traditional and digital) nationwide. And it is luminance that draws the most criticism from the public, and that first captures the driver's eye. Recall that the FHWA study reported average luminance values for the two cities as shown in the following table. (This data is taken from the final report in Table 3, pp. 27 for

Reading, and Table 8, pp. 44 for Richmond):

LOCATION	TME OF DAY	TYPE OF BILLBOARD	AVERAGE MEASURED LUMINANCE (cd/m^2)
Reading	Daylight	CEVMS	2126
	Daylight	Standard Billboard	2993
	Night	CEVMS	56.00
	Night	Standard Billboard	17.80
Richmond	Daylight	CEVMS	2134
	Daylight	Standard Billboard	3063
	Night	CEVMS	56.44
	Night	Standard	8.00

Compare these luminance values to those reported elsewhere. Bullough and Skinner (Bullough, 2011) measured average daytime luminance of traditional billboards in New York State as 6,871 cd/m^2, with average readings at night of 123 cd/m^2. For digital signs, the same authors found average luminance values of 3990 cd/m^2 and 225 cd/m^2 for daylight and nighttime, respectively. The State of New York (Marocco, 2008) in promulgating regulations for CEVMS, proposed upper limits of 5,000 cd/m^2 for daytime use, and 280 cd/m^2 for nighttime use. And the government of Queensland, Australia (Douglas, 2002), in publishing its required method for measuring billboard luminance, set maximum nighttime upper limits of 300-500 cd/m^2 depending on the environmental zone in which the billboard was located. Other studies (Luginbuhl, 2010, ICROL, 2010, Wachtel, 2014) have found similar results and/or produced similar guidance or regulation. In short, other studies have reported daytime luminance values of digital billboards that are at least twice as high as those measured in this study, and nighttime values (the measures of greatest concern and greatest public complaint) that are five times or more higher than the luminance levels found in this study at night. Why do the FHWA's measured luminance values, particularly for CEVMS at night, differ so greatly from those found elsewhere? Why are they always lower than those measured elsewhere? Did they simply measure signs that were unusually dim, did they take their measurements during a stage of the digital display sequence that was lower than other displays in this sequence, or was their measurement methodology so different than that used by Universities, government agencies, and lighting specialists elsewhere? Since CEVMS luminance is so important to the question of attention-getting glance behavior, this substantive difference is vitally important to the understanding of the results of the FHWA study.

One of our peer reviewers, after seeing the discrepancy in luminance values between the FHWA and other studies, went back and reviewed his *own* luminance data, confirming that his measured values were correct. Another reviewer, commenting on this important discrepancy, suggested that it would be worthwhile to remeasure the signs used in the FHWA study to help determine the reason(s) why their luminance values are so low.

INSUFFICIENT LENGTH OF DATA COLLECTION ZONES.

It appears that there are serious issues with the definition and use of Data Collection Zones (DCZs) in the FHWA report. The relevant discussion appears on pg. 16 of the final report, and pg. 27-28 of the draft report.

- In the draft report, the authors say that they chose 960 ft. as the maximum distance for the DCZ because: (a) the MUTCD recommends 1 in. of letter height for 30 ft. of legibility distance, and (b) given an average letter height of 32" for a CEVMS, this resulted in a 960 ft. upper limit to their DCZ.

    - This reasoning is disingenuous because the authors provide no basis to assume a 32" letter height for a CEVMS. Indeed, there is no evidence that they made an attempt to measure billboard letter heights. A review of industry-supplied guidelines for outdoor advertising demonstrate that letter heights of 36" to 48" and even larger are frequently recommended (Signazon.com) (Meadow Outdoor Advertising) (Elliott Sign and Design).

- The actual reason for the selection of a 960 ft. upper limit appears to be that it coincides with the 2° limit of resolution of the eye tracker used. In other words, at distances greater than 960 ft., the glance target provided by the eye tracker covers an area larger than the billboards being studied - thus the researchers cannot be sure of where drivers were actually looking. As a result, they limited the maximum DCZ distance to accommodate the limitations of the eye tracking equipment, regardless of either the sight distance or legibility distance of the billboards studied. In short, even in cases where billboards could be seen and even read at distances greater than 960 ft., any such glances were not analyzed due to resolution limits of the eye tracker. Had the researchers utilized billboards of more standard sizes, as discussed above, or eye-tracking equipment capable of finer resolution, more reasonable DCZs, greater than 960 ft., could have been used. As one of our peer reviewers commented: "The potential for distraction is not only a function of letter height and legibility. With high contrast ratios and high luminance levels, any type of sign may be seen far before it is legible. An unaddressed question is whether glancing at a commercial sign before it can be read increase or decrease the likelihood that a driver will continue to glance at it until it becomes legible."

- A similar concern exists with respect to the *minimum* distance defined for the DCZ.

    - In the draft report, the authors state that the end of the DCZ was "marked by (the) billboard" (pg. 28). Although not adequately described, this seems to mean that the end of the DCZ was identified as the point where the instrumented vehicle passed the billboard. This is how this has been done in other studies, and is appropriate.

- However, the final report says something quite different. It defines the end of the DCZ as being "marked when the target billboard left the view of the scene camera" (p. 16). Of course, this point would occur considerably earlier than the point at which the vehicle actually passed the billboard.

- These differences in on-road location marking the end point of the DCZ are substantial, and have significant implications for the inclusion and exclusion of eye glances – not only those attributed to billboards, but for all eye glances made by participants in this study. If the definition used the final report is correct, this distance would vary depending on the size of the billboard, its setback from the road edge, and the side of the road on which it was located. Conversely, sign size, setback, or location would have no effect on the minimum distance if the definition provided in the draft report was used. This weakness is compounded by the fact that all three billboard parameters (size, setback, and location) appear to have mysteriously changed between the draft and final reports.

- Human peripheral vision extends to roughly 180° (roughly 90° on each side) with regard to looking straight ahead. Of course, if a person turns his/her head left or right, then the included angle of peripheral vision is extended accordingly. In other words, if I turn my head to look directly at an object (e.g. a billboard) that is 30° to my right, then my peripheral vision to the right extends to approximately 90°+30° = 120° (again referenced to the straight ahead position). In the draft report, the defined end of the DCZ occurs at the point when the target billboard was just about to leave the drivers' peripheral visual field, i.e. essentially 90° to the right or left of the instrumented vehicle as the vehicle came abreast of the billboard (assuming that the driver was looking straight ahead). However, using the definition in the final report, the DCZ ended when the billboard was no longer visible to the *scene camera*. The authors previously stated that the scene cameras provided a maximum view of 80° horizontal. If we assume that the center of the scene camera image was aligned with the heading of the instrumented vehicle (and this must be our assumption since: (a) the scene cameras were fixed to the roof of the vehicle, (b) it makes the most logical sense, and (c) the authors make no statement to the contrary), then the scene camera array covers an area of 40° left and 40° right of this heading. The authors previously acknowledged that the scene cameras did not provide the full field of view to the right side of the windshield, but they did not report the key information of how much of this view was eliminated (p. 13). Nonetheless, with the end of the DCZ marked by a 90° view to the left or right in the draft report (the extent of human peripheral vision with the instrumented vehicle in line with the billboard), compared to a 40° view left and right in the final report (the limit of the roof-mounted scene cameras' field of view, and also variable based on billboard size, offset, and side of road), it is easy to see that many glances to target billboards that would have been captured and analyzed in the draft report were not analyzed (although likely captured), in the final report.

- In short, the authors eliminated from data analysis any glances toward billboards that may have been made at distances greater than 960 ft. (in both draft and final reports), and any glances toward billboards that may have been made at distances closer than the point where the billboard exited the scene cameras' field of view 40° to the left or right of straight ahead. Even if a participant driver turned his/her head to the left or right to look directly at a billboard as the vehicle got closer to it, the authors would *not* have analyzed this eye glance data if that billboard was outside the ± 40° limit of the scene camera – this is because even though the driver's eyes and head were in motion, the scene cameras were fixed – they were mounted to the roof of the vehicle and were aimed only straight ahead and captured only ±40°.

- Several peer reviewers of the present report brought up the concern about the authors' choice of a 960 ft. maximum eye glance distance. One said: "There is a very good chance that drivers scanned the billboard/CEVMS in advance of the DCZ, and a long glance/dwell to the billboard/CEVMS will be completely missed." Another cited work by Smiley and her colleagues (Smiley, 2004) that demonstrated that billboards on an expressway achieved minimum legibility distances of 410 to 1476 ft., based on the results of one test subject. In terms of legibility time, Smiley et al wrote: "The expressway sign images were first legible about 20 seconds away, but the view was interrupted several times, reducing the available time to 18 seconds (at the speed limit)." If we assume a 65 mph (95.3 fps) speed limit, the legibility distances studied by Smiley, et al reached 1906 ft. (20 seconds) or 1715 ft. (18 seconds), nearly twice the distance captured by the FHWA study's 960 ft. cutoff.

- These two constraints, on both the leading and trailing edge of the defined DCZ, potentially had the effect of reducing both the recorded number and duration of glances to target billboards on both sides of the road. Accordingly, this warrants explanation or clarification from FHWA, particularly because much more of this critical data was presumably captured and made available for analysis for the draft report.

CODING AND ASSIGNMENT OF REGIONS OF INTEREST (ROIs).

A conflict seems to exist with regard to the coding of ROIs. The authors allude to this problem, but are silent about if or how it was addressed, and they provide no information to assist the reader with regard to understanding their coding process. The issue is this: As stated on pg. 24 of the final report, the eye tracking data reduction and analysis software (which was not used for the draft report) determined the gaze "intersection" (i.e. location) every 60 Hz, and automatically assigned each such gaze to an ROI. But, as the authors note, ROIs may overlap. Because the software allows the researcher rather than the software itself to specify the "priority" for each ROI, when this (presumably frequent) overlap occurs, whichever ROI was assigned the highest priority by the researcher will be "given" (i.e. assigned) any such overlapping glance at the expense of the "lower priority" but overlapping ROI. So, if I wanted to demonstrate that eye glances that overlap both a billboard and the road ahead are really to the road and not to the billboard, I merely assign higher priority to the RA (road ahead) segment. The authors cite this specific example on pg. 24, but provide no information about the process that they followed in such cases or how many such cases there were during the study. A look at Fig. 13, pg. 30 of the draft report shows a billboard at the intersection of what would be two ROIs (recall that the draft report used a different (manual) method for coding eye glances, and so ROIs did not exist as a paradigm until the final report). The authors do not discuss how the prioritization of ROIs would have handled this eye glance, or how many such instances actually occurred. While it is true that the authors also identified "dynamic ROIs," there is no explanation given of how such Dynamic ROIs were handled vis-a-vis the overlapping static ROIs.

One of the peer reviewers of the present report suggested that there should be no concern about allowing the researchers to assign priorities to ROIs "*as long as they were blind to the purpose of the experiment."* There is, however, no indication in the FHWA report that this was the case. Was it?

RELEVANT RECENT RESEARCH IGNORED.

The reference list for the final report shows that the authors used the 31 months between the draft and final reports to update their literature review (the final report includes citations dated as late as June 27, 2012). One wonders, however, why they ignored a number of available, peer-reviewed research studies of direct relevance to this project, such as: Backer-Grondahl, 2009, Dukic, 2013[2], Edquist, 2011 Gitelman, 2012; Edquist, 2008; Edquist, 2010; Milloy, 2011; Young, 2009, all of which were available during this time period or earlier. This is further troubling because they *did* include a number of billboard industry sponsored studies that received little if any peer review, and where the full studies were restricted from public access (Tantala, 2010, 2011).

DISCUSSION OF LITERATURE REVIEWED.

In their summary of the Lee, et al (Lee, 2007) study, the FHWA researchers state: "(the authors) did not show any significant effects of CEVMS on driver glance behavior." As the FHWA researchers were aware from prior research that they reviewed in preparation of this report, the study by Lee and her colleagues was paid for and overseen by the outdoor advertising industry. Earlier reviews of the Lee, et al report, (e.g. Wachtel, 2007) have shown that these authors did not, in fact, perform tests of significance for the eye glance duration data that they had collected (despite performing such significance tests for all six other measures studied) (Wachtel, 2009). Had they done so, they would have found significant differences (Hurtz, 2011; Placeholder1). Further, as Lee and her colleagues stated, and as reported in an earlier FHWA report (Molino, et al, 2009), the participant population for their nighttime study was too small to support tests of statistical significance, but that, had their sample size been larger, some of these findings "would show statistical significance" (p. 7).

Why did the FHWA authors accept the Lee, et al data at face value, despite evidence in their possession that important, relevant findings were ignored?

SALIENCE, ATTENTION CONSPICUITY, AND BOTTOM-UP PROCESSING.

The authors' discussion of "attention conspicuity," "salience," and "bottom-up processing" warrants clarification. On pg. 10 of the final report, the authors refer to a review by this author (Wachtel, 2009) of research done by Theeuwes. The FHWA authors state: "Wachtel leads one to consider CEVMS as stimuli in the environment where attention to them would be drawn in a bottom-up manner; that is, the salience of the billboards would make them stand out relative to other stimuli in the environment and drivers would reflexively look at these signs." They go on to state that the Theeuwes work used "simple letter stimulus arrays in a laboratory task," and continue: "Research using simple visual stimuli in a laboratory environment are (sic) very useful for testing different theories of perception, but often lack direct application to tasks such as driving." It is surprising, therefore, that, on the same page, the FHWA authors cite the work of Cole and Hughes (Cole, 1984) that reinforces the point made by Wachtel in his review of Theeuwes' work. The FHWA report states: "Standard and digital billboards are often salient stimuli in the driving environment, which may make them conspicuous. Cole and Hughes define attention conspicuity as the extent to which a stimulus is sufficiently prominent in the driving

[2] Although not published in print form until 2013, this study was made available electronically after approval for publication a year earlier.

environment to capture attention. Further, … attention conspicuity is a function of size, color, brightness, contrast relative to surroundings, and dynamic components such as movement and change. It is clear that under certain circumstances image salience or conspicuity can provide a good explanation of how humans orient their attention" (pg. 10).

Several of the peer reviewers of the present report weighed in on this issue and added their views that outdoor advertising, and particularly CEVMS, seeks driver attention through bottom-up processes by managing the visual stimuli with which drivers are presented.

ROAD TYPE AS AN INDEPENDENT VARIABLE.

The final report lists "road type" as an independent variable. The draft report did not.

- Why was an independent variable added after the study was completed and all data collected?

- What effect did this change have on the analysis of data or results?

FIXATION DIFFERENCES BETWEEN DRAFT AND FINAL REPORT.

On pg. 25 of the final report, the authors discuss their approach to measuring visual fixations. They do not mention the findings of the draft report, which were criticized by the independent peer reviewers for reporting gaze fixation durations (to billboards and road scenes alike) that were too brief to be reasonable.

- What was done after completion of the draft report to ensure that the results for gaze fixation as reported in the final report were valid?

- Given the major changes to the document between draft and final reports, did the authors submit for independent peer review the version of the report that was revised subsequent to receipt of peer review comments?

- If so, what were the results?

THE ROLE OF ADVERTISING CONTENT.

It is well understood that the content displayed on outdoor advertising can have a profound effect on driver distraction. Several reviewers noted the absence of any discussion of sign content in the FHWA report. One stated: "There may be characteristic differences in the products advertised by CEVMS vs. standard billboards – products that may appeal to one demographic more than another." One reviewer noted that other studies of distraction due to billboards made a concerted effort to match, to the extent possible, the attention getting nature of the billboard stimuli across the signs to which participants were exposed to eliminate this otherwise potentially confounding variable from the study.

While a study conducted on public roads, such as this FHWA study, does not lend itself to such controls, several reviewers wished to see illustrations of the billboard images which the participant drivers faced as they drove the instrumented vehicle. Given that the scene cameras continuously recorded the drivers' view of the road, the inclusion of this data would have been straightforward.

Another reviewer noted that some of the pictures provided of the signs used in the study showed images as well as text. Images may attract attention at distances beyond those required to resolve text, but the report is silent on this issue. The likelihood of earlier glances to imagery on a sign provides further evidence of the need for the leading edge of the DCZ to extend beyond 960 ft. from the sign.

UNANSWERED QUESTIONS.

In addition to the concerns and questions raised above, there are a number of issues in the final report, and in its obvious differences from the draft report, about which the authors are silent. For a reader to have confidence in the conclusions reached by the final report and in the appropriateness of substantive methodological changes made between draft and final reports, it is suggested that FHWA address the following questions.

ANALYSIS OF EYE GLANCE DATA.

The draft report makes clear (pp. 28-9) that the eye glance data was reduced manually, frame-by-frame. The final report describes a completely different, automated eye glance data reduction system (MAPPS, p. 23). Presumably, the change of eye glance analysis methods was made as a result of the peer review comments to the draft report, although the authors are silent on this issue. The following questions are suggested by this methodological change:

- When, why, and how was the decision made to scrap the system used for the draft report and replace it with another system?
- How was the accuracy/validity of the new (automated) system tested?
- How were the results determined to be valid given the failure of the earlier effort?
- Was the revised approach subject to peer review; if so, what comments were made, and how did the researchers respond?

REGIONS OF INTEREST (ROIs) FOR DATA ANALYSIS, AND BOUNDARIES BETWEEN ADJACENT REGIONS.

The scene camera views as segmented for analysis in the final report were substantially different than those used in the draft report.

- How and why were these changed?
- How was the revised system tested for accuracy and fidelity?
- Was the revised system subject to independent peer review; if so, what were the results?

- In the system described in the final report, the scene views from the three vehicle roof-mounted cameras were divided into six "static ROIs" as well as two additional areas ("inside vehicle" and "top"), which were beyond the view of the cameras, but where, the authors state, eye tracking was still possible. It is implied, but not stated, that the study's authors were the ones who chose/identified the static ROIs and their boundaries. Is this correct? If not, how were the static ROIs determined, and by whom?

GAZE DIRECTION PROBABILITIES.

The authors discuss the methodology they used to "analyze the probability of a participant gazing at driving related information" (which they describe as gazes at the ROIs identified as "road ahead, road ahead top, and driving-related tasks," which they also confusingly call "driving related risks" (p. 28). Their approach requires that they use only two possible outcome measures to classify a participant's gaze behavior. They state: "If the participant gazed toward the road ahead, road ahead top, or driving-related risks, then the value of 'RoadAhead' was set to 1" (which they deemed "success"). But, "If the participant gazed at any other object in the panoramic scene, then the value of "RoadAhead" was set to zero" (deemed "failure"). The authors are silent about gazes outside the limits of the scene camera views, even though, in some cases, eye tracking "was possible." In their discussion (second paragraph on pg. 28) the authors seem to conflate the terms "driving-related information" (which is presumably what they are interested in) and "road-ahead information" which is what their analysis captured. There is, of course, considerable driving-related information that is visually obtained by glances beyond the road-ahead view (especially given that the road-ahead view was artificially constrained in this study). Views to both side view mirrors, rear view mirror, and instrument panel, as well as to potential hazards such as passing or merging traffic, as well as emergent threats such as pedestrians, turning vehicles, or bicyclists, are all understood to be driving-related information, and all may require glances outside the view afforded by the scene cameras in this study. Given that the reported probabilities of gazing at the road ahead and at the specified ROIs (see tables 4 and 5, pg. 28-9) necessarily added to 100%, it must be understood that this represents 100% of only those gazes made by participants that were captured by the eye-tracker and analyzed by the researchers. In addition to these glances, there were an unknown number of instances, and an unknown percentage of time behind the wheel, when eye gazes were not recorded, simply because they fell outside the range of the limited field of view afforded by the scene cameras.

- We recommend that FHWA provide information about such unrecorded/unanalyzed eye glance events so that the reader can understand how often, and for how long, such potentially relevant eye gazes were made that were not recorded or analyzed.

BILLBOARDS (OFF-PREMISE) vs. ON-PREMISE SIGNS.

Throughout the report, the authors refer to this as a study of "billboard" or "outdoor advertising signs." FHWA programs and policies distinguish between "billboards" and "on-premise" signs. Billboards, which are considered off-premise signs, are designed, placed, and operated for different purposes than on-premise signs - signs that FHWA does not regulate. Yet it appears from some of the photographs in the final report that some of the signs referred to as billboards are, in fact, on-premise signs. As discussed above, the average driver is not familiar with the terms "on-premise" and "off-premise" or "billboard" and, if shown examples of signs of each type, might have a difficult time distinguishing them. In other words, given equal size, luminance, etc., a driver might not be able to distinguish whether he/she was looking at a billboard or an on-premise sign. This potential conflation of billboards with on-premise signs raises three questions:

- Can FHWA confirm that all of the signs referred to in the report as "target" billboards were, in fact, off-premise signs?

- When discussing the target billboards in each of the two cities, the authors identify (Table 2, pg. 17 and Table 7, pg. 40) "other standard billboards" that were, presumably, visible to drivers at the same time and, at least to some extent, in the same location, as the target billboards. But the authors never identify the presence of on-premise signs at these locations, even though such signs are clearly visible in several of the report photographs. Is it the authors' position that there were no on-premise signs located at the same general location as target billboards, or have any such on-premise signs not been accounted for in these Tables?

- Why, when identifying "control areas" (as distinguished from areas containing billboards), did the authors accept the inclusion of on-premise signs, including, perhaps, digital signs, when to the average motorist, such "control" areas (to the extent that they contained on-premise signs) would not be distinguishable from "treatment areas," those in which "target" billboards of interest were located?

TASK DEMANDS AND VISUAL SALIENCE.

The authors' discussion of eye gaze behavior in dynamic environments such as driving "suggests that task demands tend to override visual salience in determining attention allocation." The authors state: "When extended to driving, it would be expected that visual attention will be directed toward task relevant areas and objects (e.g. the roadway, other vehicles, speed limit signs, etc.), and other salient objects, such as billboards, will not necessarily capture attention." But the authors seem to ignore the fact that it is for this very reason that highway and traffic engineers have long recognized that there are times where they must capture the drivers' attention, to break task-driven visual attention from its common complacency, in order to communicate a timely or critical message. As a result, the

MUTCD contains numerous instances where unique colors (e.g. fluorescent yellow-green) are applied to specific signs, where high energy flashing signs, lights, and beacons are employed, and where specific pavement markings are installed – all because their visual salience is intended to command drivers' attention. Official Changeable Message Signs (CMS) are often set to flash a message of particular urgency, for the same reason. And the authors seem to ignore the fact that advertisers, seeking to capture drivers' attention, rely upon the visual capture techniques of high luminance, contrast, and frequent message change in an effort to accomplish this.

- Why has the FHWA report ignored these frequently employed examples (both by traffic officials and advertisers) of the use of visual salience to capture attention independent of task demands?

- Why have the FHWA authors ignored recent research showing that roadside advertising signs, including digital and video billboards, are able to capture the driver's attention at the expense of high levels of task demands, e.g. (Milloy, 2011), (Herrstedt, 2013)?

OTHER EXPERIMENTAL ISSUES.

CONTROL OF EXTRANEOUS VARIABLES.

Extraneous variables are those which are not relevant to a study's purpose but which may have an effect on the dependent variable (in this case, eye glances) and which therefore must be either eliminated or controlled. If such extraneous variables remain in the study and are not controlled they may have the effect of reducing the likelihood of finding significance in hypothesis testing, because they add to the error variance. Let us say that we want to study a driver's eye-glance response to CEVMS, and to compare that response to glances to traditional billboards and to roadway areas in which no billboards of any kind are present. The result would be a "clean" experimental design in which we examine driver glance behavior (the dependent variable) when the driver is exposed to each of the three independent variables (CEVMS, traditional billboards, and roadway areas without billboards). The FHWA authors claim that this is what they did. But, by their choice of words and by some of the photographs that accompany both the draft and final reports, the reader can see that, although they made sure that there were no other *billboards* present within the field of view at any of the three types of study sites, there were often other signs, typically on-premise signs, that were present. As discussed above, the average driver does not understand or appreciate the difference in purpose and function between billboards and on-premise signs, and may be just as likely to glance at an on-premise sign as at a billboard (CEVMS or traditional). Let's go back to our example, and discuss the situation that seems to have occurred in this study, an unknown number of times. If a target billboard happened to be located on or immediately adjacent to a property that included one or more on-premise signs, how might the researchers have recorded and coded eye-glances made by their participant drivers when approaching such an area? We can assume that any glances that were clearly centered on the target billboard would have been properly correctly coded as a glance to that billboard. But what about glances immediately before or after the billboard glance – what if such glances were made to one or more of the on-premise signs located quite near the billboard – how would such glances be coded? (See, for example, Figures 4 and 6, pg. 18, or Figure 23, page 45, of the final report). There are three possibilities: (1) They could have been coded as glances to the target billboard – of course, this would be erroneous and there is no indication that it was done; (2) They could not be coded at all; i.e. such glances could be discarded from the data set and not analyzed – this would also be an error, and there is no indication in the report that this was done; (3) They could be coded as glances, not to the billboard, but to one of the road-related ROIs (LSR, RA, RSR) described above. The authors are silent on this issue, and thus the reader cannot know whether, or how often, this may have occurred. But if this *did* occur, any such coded glances to an extraneous variable such as an on-premise sign would compromise the study results in two ways. First, such a glance would be coded as having been made to the "road ahead" (or road left or road right) and second, such coding would eliminate from analysis the situation that actually occurred; that the glance was actually made to a roadside advertising sign, simply one that was not a "billboard." Since the authors treated the probability of glancing at the road ahead to be a zero-sum game, with "success" and "failure" that must add to 100%, *any* target billboard site where other advertising signs were proximal to the billboard may have suffered this fate. We believe that the authors should clarify their coding procedures in this regard, and identify the frequency with which such situations occurred.

AGE RANGE OF PARTICIPANT DRIVERS.

It has been shown in studies of driver distraction that younger drivers (frequently identified as those age 25 and below), and older drivers (frequently identified as those age 65 or above [although researchers increasingly classify older drivers into the "young-old" – age 65-74, and the "old-old" – age 75 and above]) have more difficulty dealing with, and overcoming, distraction than the broad cohort of drivers between these age groups. But the FHWA researchers seem to have made no effort to recruit representatives of these two important age groups. There were no drivers included in either city who were above the age of 64; and although there were participants as young as age 18, the authors do not tell us how many. It is typical that authors of research papers that address issues of driver performance provide the mean, standard deviation, and range of ages of participants. Here, only the mean was provided. We recommend that complete data regarding participant ages be provided.

DROPOUT RATE OF POTENTIAL PARTICIPANTS.

The participant dropout rate was quite high. The authors attribute this to the unusability of the data of certain participants either because the eye-tracking equipment could not be calibrated to them or because of equipment failures. In Reading, 12 potential participants were excluded; in Richmond, the number was 17. This represents a loss of 24% (Reading) and 41% (Richmond) of all participants recruited. Although it is possible that the researchers intentionally and reasonably "overbooked" the number of participants in anticipation of some dropouts, there is no indication that this was done. Further, it appears that no effort was made to recruit additional participants to make up for those who were lost, and no discussion is provided to assist the reader in better understanding whether there were common characteristics among those participants who were eliminated such that the representativeness of the remaining participants might have been compromised. (Farbry, 2001) (Molino, 2009).

INTRUSIVENESS OF EYE TRACKING SYSTEM.

In the section titled "Experimental Approach," the authors describe the eye-tracking system as "non-intrusive." This language was added subsequent to the draft report. If, by non-intrusive, they mean that it was not physically attached to the driver's head as was the case for earlier generation eye tracking systems, we agree. But when the eye tracking system includes four prominent cameras mounted to the vehicle dashboard in front of the driver, and when the duration of the average drive for each participant was only about 20-30 minutes (so that the driver could not fully acclimate to the equipment to the point of ignoring its presence) it seems inappropriate to describe this as non-intrusive. This might be seen as a minor issue, except when, as discussed below, a reader must analyze the extent to which the results of this study might be generalizable to driving in general.

EXPERIMENTAL CONDITIONS INFLUENCE PARTICIPANTS' BEHAVIOR AND PERFORMANCE.

As discussed above, the authors describe the eye-tracking equipment as "non-intrusive." We disagree. Of equal concern, however, is the totality of the participants' experience, the question of whether the participants could be expected to perform "as they normally would," and the implications of this for the generalizability of the study. Human factors and human performance research in road safety is increasingly performed in one of two settings – either in "naturalistic" studies, or in studies conducted in advanced driving simulators. Although the discussion of these two approaches is beyond the scope of this review, it is useful to understand that these two methods, each with its own strengths and limitations, have proven to be more "generalizable" to real driving (i.e. have more applied validity to the real world) than most other forms of driving research. This FHWA study, conducted in an instrumented vehicle, provides certain key benefits in that it places participants in an actual vehicle in which they drive on actual roads under actual traffic and weather conditions, while viewing actual billboards and other signs. On the negative side, the vehicle being driven is instrumented in such a way that the participants know that they are being observed and recorded – and this fact has been shown to contribute to a likely change in participants' behavior from what might have been expected had they performed in more naturalistic setting. In addition, in instrumented vehicle studies, there is often an experimenter (researcher) in the vehicle with the participant, and this experimenter, typically sitting in the rear seat, is likely to be recording data in a computer or on a clipboard, monitoring equipment, and/or interacting with the participant. In this study, however, there were two experimenters in the vehicle at all times. Any one of these study conditions individually (cameras or other monitoring equipment; presence of an experimenter) could have an adverse effect on the "realism" of the participant's performance. But in this study the conditions were more unrealistic due to the presence of the camera equipment, two experimenters, and the limited amount of time (20-35 minutes) that each participant spent in the vehicle. In addition, the authors discuss several situations in which the eye-tracker had to be recalibrated, and others in which overloaded data files required the researcher to initiate new files. Each of these occurrences required the researchers to instruct the participant driver to pull off the road for some period of time, interrupting the continuity of the drive and increasing the interaction between researcher and participant. In addition to the Hawthorne Effect (Landsberger, 1958) which suggests that the very fact that participants are being observed is enough to modify their behavior, often more strongly than the experimental manipulation itself, it is likely that another well known study phenomenon known as "The Good Participant" (one of three types of participant roles described under the principle of "Demand Characteristics") (Whitley, 2002) was active in this study. This is the situation in which the participant tries to help or please the researcher by performing "well," and it can alter a participant's behavior to such an extent that true differences in performance that might be due to the independent variables are overridden or masked by participant behaviors stemming from the artifice of the situation in which the participant is asked to perform. As explained by one peer reviewer with extensive experience in the conduct of driving related research: "When being involved in an experimental study, most people want to drive as well as possible. This is illustrated by the fact that people often ask: 'How did I do? How well was I driving compared to other participants?' This means that people indeed might have ignored the billboards more often than they normally would have."

Another peer reviewer looked at the issue of experimental conditions from a different perspective. He said: "We are talking about fairly fresh drivers – they had not been driving for hours, the tests were conducted in good driving conditions (i.e. apparently none of the tests were conducted in inclement weather), the driver had been prepared with a map of the route *and* a GPS device providing turn by turn directions *and* a researcher in the front passenger seat to provide route guidance… is it at all surprising that no near misses or driver errors were observed?"

INSTRUCTIONS GIVEN TO TEST PARTICIPANTS.

On pg. 14 of the final report, the authors describe the instructions provided to drivers, which were: "to drive the routes as they normally would." In the draft report, however, the instructions provided to drivers were more specific and comprehensive, including that they should pay attention to other traffic, speed limits, etc."

- What were the actual instructions provided to drivers, and why did the description of this change between draft and final?

One peer reviewer of the present study questioned why the researchers provided such explicit instructions to the driver participants; specifically that their eye glance behavior was being studied. He considered this to be a serious flaw, in that it could have contributed to the participants' modifying their typical behavior to pay more visual attention to the road ahead, thus contributing to "The Good Participant" phenomenon.

POST-STUDY DEBRIEFING.

In their discussion of the post-study debriefing (pg. 22) the authors describe a process that seems to have differed from the draft report to the final. In the draft report, the participants "completed a driver feedback questionnaire." But the final report says nothing about this. In the draft report, the authors explain that the participants "were informed of the study's true purpose." Again, the final report is silent on this issue. Even though any differences in the debriefing as described in the final report vs. the draft report may have not been significant to the results of the study, the existence of such differences raises issues about the changes in the two versions of the report, and raises questions about what other changes might have been made that have not been reported.

As one of the peer reviewers of the present report put it, the post-study debriefing could be a gold mine of information. For example, this reviewer suggested that the subjective component of an advertisement is what gives it its 'value' in terms of personal interest from the driver. This *psychological effect* can be divided into valence, arousal, and motivational intensity – and the "success" (for the advertiser) is the aspect "most likely to create the extended dwell times (and eyes-off road episodes."

We recommend that FHWA address such questions as: Were the participants actually told the true purpose of the study? Were they given an opportunity to comment? What were their opinions? We further suggest that FHWA make public a copy of the "driver feedback questionnaire" that was used.

RATER (AND INTER-RATER) RELIABILITY.

The authors report that, during data collection (i.e. during the actual participant drives), the front-seat researcher observed and recorded driver behavior using "subjective measures." Human factors and experimental psychology tells us that human raters, particularly when judging along subjective scales, are susceptible to low reliability and validity without specialized training and practice. But the report is silent on key issues including: how many researchers were used for this process, how they were trained, how their ratings were reviewed and compared to those of other raters to measure inter-rater reliability, etc. If only one rater was used throughout this study, the question of rater bias arises. Finally, given that the front-seat researcher had other tasks to perform, there remains the important question of whether this researcher was actually able to observe and annotate driver behavior continuously. As one peer reviewer to the present report expressed concern about rater and inter-rater reliability – "what constitutes a researcher feeling 'slightly uncomfortable, but not to a significant degree'? How does one ensure the reliability of ratings when the rating criteria themselves are so subjective?" We recommend that FHWA clarify the entire issue of raters, and how reliability was assured.

CONCLUSIONS.

This review has raised several questions and identified a number of critical concerns that, taken together, suggest important deficiencies in the FHWA final report: *Driver Visual Behavior in the Presence of Commercial Electronic Variable Message Signs (CEVMS)*. Given the lack of information provided by the study's authors about key details of their research, the apparent internal conflicts in critical data provided, and the problems with the experimental equipment, a reader is unable to assess the validity of the findings as presented. In light of the harsh criticism of the draft study report provided by FHWA's retained independent peer reviewers, and the nearly three years that elapsed between the issuance of that draft and the final report, and further given the lack of information from the authors concerning important details of what was done, and how, to address and resolve those reviewer comments, the concerns of a reader of the final report are further heightened. When evaluated against the growing number of recent research studies, conducted world-wide, that increasingly demonstrate concerns for the adverse effects of billboard distraction on driver performance, particularly under conditions in which the driver must respond to suddenly appearing or developing traffic hazards, one must question the contribution of this study and the conclusions that can be drawn from it to this important field of research. As relevant new research (Edquist J. H., 2011), (Herrstedt, 2013), (Divekar G. P., 2012), (Belyusar, 2014) continues to be published, we urge the authors of this eagerly anticipated FHWA study to clearly document their methods and results in light of the peer reviewed comments directed at the draft report, and the concerns expressed herein.

As one of our peer reviewers said: "If FHWA can't appropriately address the issues raised in this report, it owes it to both sides of this debate to fund a replication of this effort with reasonable methods and a scientific advisory committee." In the meantime, other reviewers expressed the precautionary principle. One, heavily involved in road and traffic safety, said: "if there is a lack of scientific certainty and there is a question around safety – the response should be no. In the context of (outdoor advertising sign) permits, this is particularly important as permits for signs have a minimum life of a decade.

REFERENCES.

Ayres, T. K. (2005). Intersection Right-of-Way: What is an Immediate Hazard? *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting.* Santa Monica: HFES.

Backer-Grøndahl, A., & Sagberg, F. (2009). "Relative crash involvement risk associated with different sources of driver distraction." Presented at the: *First International Conference on Driver Distraction and Inattention.* Gothenburg, Sweden: Chalmers University.

Belyusar, D., Reimer, B., Shoup, A., Jokubaitis, B., Pugh, B., Mehler, B. and Coughlin, J.F. (2014). "A Preliminary Report on the Effects of Digital Billboards on Glance Behavior During Highway Driving." *Presented at the 93rd Annual Meeting of the Transportation Research Board.* Washington, DC: The National Academies.

Bullough, JD, & Skinner, NP. (2011). "Luminance Criteria and Measurement Considerations for Light-Emitting Diode Billboards." *Paper No. 11-0659 Presented at the 90th Annual Meeting of the Transportation Research Board*. Washington, DC: The National Academies.

Chan, E., Pradhan, A.K., Knodler, M.A., Jr., Pollatsek, A., and Fisher, D.L. (2007). "Empirical Evaluation on a Driving Simulator of the Effect of Distractions Inside and Outside the Vehicle on Drivers' Eye Behaviors. *Presented at the 86th Annual Meeting of the Transportation Research Board.* Washington, DC: The National Academies.

Chattington, M., Reed, N., Basacik, D. Flint, A., & Parkes, A. (2009). Investigating Driver Distraction: The Effects of Video and Static Advertising, *PPR409.* Transport Research Laboratory. United Kingdom: Crowthorne.

Clarion Associates/Clanton Associates. (2010). *Pittsburgh Electronic Sign Ordinance Review: Staff Review Draft.* Pittsburgh.

Cole, B. L., and P. K. Hughes. (1984). A Field Trial of Attention and Search Conspicuity. *Human Factors*, **26**, 299-313.

Divekar, G., Pradhan, AK, Pollatsek, A. & Fisher, DL. (2012). "Effect of External Distractions: Behavior and Vehicle Control of Novice and Experienced Drivers Evaluated." *Transportation Research Record: Journal of the Transportation Research Board. 2321,* 15-22.

Divekar, G., Pollatsek, A. & Fisher, DL. (2012). External-to-Vehicle Distractions: Dangerous Because Deceiving. *Fifth International Conference on Traffic and Transportation Psychology.* Groningen, The Netherlands: University of Groningen.

Douglas, J. (2002). *Guide to the Management of Roadside Advertising – Edition 1.0. TERS Product No. 80.500. Appendix D: Brightness/Luminance Levels.* Brisbane: Queensland Government, Department of Main Roads.

Dukic, T., Ahlstrom, C., Patten, C., Kettwich, C., & Kircher, K. (2013). "Effects of Electronic Billboards on Driver Distraction." *Traffic Injury Prevention* 14(5).

Edquist, J. &. (2008). Visual clutter in road environments - what it does,and what to do about it. *2008 Australasian Road Safety Research, Policing and Education Conference.* Adelaide.

Edquist, J. H. (2010). Effects of advertising billboards during simulated driving. *Applied Ergonomics*.

Edquist, J., Horberry, T., Hosking, S. & Johnston, I. (2011). "Advertising billboards impair change detection in road scenes." *2011 Australasian Road Safety Research, Education & Policing Conference.* Adelaide.

Elliott Sign and Design. (n.d.). "Letter Size to Visibility Chart." Retrieved 2014, from http://www.elliott-design.net/letter_size_chart.htm

Farbry, J, Wochinger, K., Shafer, T., Owens, N., & Nedzesky, A. (2001). *Research Review of Potential Safety Effects of Electronic Billboards on Driver Attention and Distraction: Final Report.* Washington: Federal Highway Administration.

Gitelman, V., Zaidel, D., & Doveh, E. (2012). "Influence of Billboards on Driving Behavior and Road Safety," Presented at: *Fifth International Conference on Traffic and Transportation Psychology.* Groningen, The Netherlands: University of Groningen.

Gramatins, J. (2010). The Effects of CEVMS on Driver Eye Glance Behavior: Phase II Update. *Presentation to the Transportation Research Board Subcommittee on Digital Billboards.* Washington, DC.

Granda, T. (2009). The Effect of CEVMS on Driver Eye Glance Behavior. *Presentation to the NAHBA Annual Conference* Branson, MO.

Herrstedt, L., Greibe, P, & Andersson, P. (2013). "Roadside Advertising Affects Driver Attention and Road Safety." *Proceedings of the 3rd International Conference on Driver Distraction and Inattention.* Gothenburg, Sweden: Chalmers University (No. 05-P).

Horberry, T. &. (2009). Distractions Outside the Vehicle. In M. L. Regan, *Driver Distraction: Theory, Effects, and Mitigation.* CRC Press, Taylor & Francis Group.
Hurtz, G. (2011). Billboard Data Statistical Analysis. Sacramento, CA: Unpublished.

Illinois Coalition for Responsible Outdoor Lighting (ICROL). (2010). *Digital Billboards: New Regulations for New Technology.* Retrieved 2014, from http://www.illinoislighting.org/billboards.html

Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics, 42*(6), 807-815.

Landsberger, H. (1958). *Hawthorne Revisited.* Ithaca: Cornell University.

Lee, S. E., McElheny, M.J., & Gibbons, R. (2007). Driving Performance and Digital Billboards. *Report prepared for Foundation for Outdoor Advertising Research and Education*. Virginia Tech Transportation Institute.

Luginbuhl, CB., Israel, H., Scowen, P., Polakis, T, and Polakis, J. (2010). *Digital LED Billboard Luminance Recommendations: How Bright is Bright Enough?* Unpublished Draft.

Marocco, M. (2008). *Criteria for Regulating Off-Premise Commercial Electronic Variable Message Signs (CEVMS) in New York State.* Albany: State of New York Department of Transportation.

Meadow Outdoor Advertising. (n.d.). *The Effect of Font Size on Readability.* Retrieved 2014, from http://www.meadowoutdoor.com/billboard-design/good-vs-bad/font-size

Milloy, SL; and Caird, JK. (2011). "External Driver Distractions: The Effects of Video Billboards and Wind Farms on Driver Performance." Published in: *Handbook of Driving Simulation for Engineering, Medicine and Psychology.* Edited by: D.L. Fisher, M. Rizzo, J.K. Caird, & J.D. Lee. Boca Raton: CRC Press.

Molino, JA., Wachtel, J., Farbry, JE, Hermosillo, MB., & Granda, TM. (2009). *The Effects of Commercial Electronic Variable Message Signs (CEVMS) on Driver Attention and Distraction: An Update.* Report No. FHWA-HRT-09-018. Washington: Federal Highway Administration.

Monk, C. (2011). Driver Visual Attention in the Presence of Commercial Electronic Variable Message Signs. *Presentation to the 90th Annual Meeting of the Transportation Research Board.* Washington , DC.

Monk, C. (2010). The Effects of CEVMS on Driver Eye Glance Behavior. *Keynote Presentation to the NAHBA Annual Conference.* Wisconsin: Madison:

Regan, MA., Young, KL, Lee, JD, Gordon, CP. (2009). Sources of Driver Distraction. In *Driver Distraction: Theory, Effects, and Mitigation.* M. A. Regan, J. D. Lee, and K. L. Young, Eds., CRC Press, Taylor & Francis Group.

Rosenholtz, R., Li, Y. &, Nakano, L. (2007). Measuring Visual Clutter. *J Vis*, **7**, 11-22.

Signazon.com. (n.d.). *Sign Letter Height Visibility Chart - How to Choose the Correct Letter Height.* Retrieved 2014, from : http://www.signazon.com/help-center/sign-letter-height-visibility-chart.aspx

Smiley, A., Smahel, T., & Eizenman, M. (2004). Impact of Video Advertising on Driver Fixation Patterns. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1899, 2004**,** 76-83.

Tantala, MW & Tantala, AM. (2010). A Study of the Relationship Between Digital Billboards and Traffic Safety in Henrico County and Richmond, Virginia. The Foundation for Outdoor Advertising Research and Education (FOARE).

Tantala, MW & Tantala, AM. (2010). An Examination of the Relationship Between Digital Billboards and Traffic Safety in Reading Pennsylvania Using Empirical Bayes Analysis. *Moving Toward Zero 2100. ITE Technical Conference and Exhibit.* Buena Vista, Florida: Institute of Traffic Engineers.

Wachtel, JA., & Netherton, RD. (1980). *Safety and Environmental Design Considerations in the Use of Commercial Electronic Variable-Message Signage.* Report No. FHWA/RD-80/051. Washington, DC: Federal Highway Administration.

Wachtel, J. (2007). *A Critical, Comprehensive Review of Two Studies Recently Released by the Outdoor Advertising Association of America .* Hanover: Maryland State Highway Administration.

Wachtel, J. (2011). Public Attitues Toward Roadside Advertising Signs: Forty Years of Surveys, Interviews and Unsolicitred Opinions. *90th Annual Meeting of the Transportation Research Board.* Washington: TRB.

Wachtel, J. (2014). *Report on Digital Sign Brightness: Prepared for the Nevada State Department of Transportaton, Washoe Couty, City of Reno and City of Sparks.* Reno.

Wachtel, J. (2009). *Safety Impacts of the Emerging Digital Display Technology for Outdoor Advertising Signs.* Report No: 20-7/256. Washington: National Cooperative Highway Research Program.

Whitley, BE. (2002). *Principles of Research in Behavioral Science, 2nd Edition.* Mayfield Pub. Co.

Young, MS, Mahfoud, JM, Stanton, N. Salmon, PM, Jenkins, DP & Walker, GH. (2009). "Conflicts of Interest: The implications of roadside advertising for driver attention." *Transportation Research Part F: Traffic Psychology and Behaviour, Vol. 12*(5), 381-388.