

SECTION 3.

RESEARCH UNDERTAKEN OR PUBLISHED BY THE OUTDOOR ADVERTISING INDUSTRY

Over a period of many years, the outdoor advertising industry has commissioned a number of research studies from universities and private consulting organizations. To a large extent these studies, their methods and results, are not released to the public. Occasionally, or upon request, the OAAA will release the report of a commissioned study. In addition, internet research occasionally identifies excerpts of such work or information provided by manufacturers or sellers of space on billboards oriented to potential clients. Finally, patent searches occasionally identify new technologies of relevance in the field.

The on-premise sign industry, through its representative organizations such as the International Sign Association (ISA) and the United States Sign Council (USSC), has also sponsored research, some of which is available to the public for a fee through the organizations' web sites. The USSC website currently lists 15 documents available for purchase by the general public. Examples of such studies include those by Garvey, Thompson-Kuhn & Pietrucha, (1995), Garvey (1996), and Kuhn (1999). In addition, the ISA publishes a periodical called *Signline*, which reports on new developments, and often highlights legal challenges to on-premise signage.

Perception Research Services (1983), Young (1984).

A series of studies conducted by Perception Research Services (1983), and separately reported by its President (Young, 1984) was intended to “observe the attention-getting ability of outdoor boards from the perspective of the individual in an automobile (Young, 1984, p. 19). This work measured the eyegaze behavior of 200 licensed drivers who viewed a 27 minute video of a drive through three metropolitan areas to “observe the stopping power of outdoor” (p. 19). Although insufficient detail was presented in the published reports to independently review the research, the results are illuminating. First, the author suggests that recall scores (based on questioning of the participants immediately after the simulated drive) “grossly (understates) the true impact of outdoor advertising ... that outdoor is generating approximately two and one-half times as much attention as recall scores would ever indicate” (Young, p. 20). Second, the research found that “outdoor advertising located near highway signage tends to generate *greater* attention. We hypothesize that the highway signage tends to wake up the driver; his state of alertness increases and his attention to advertising and signage in the immediate area tends to get enhanced” (Young, p. 21). Finally, the research found that outdoor advertising attracts attention regardless of whether the displayed message is of interest/relevance to the driver or not. These findings, and particularly the last, obviously intended for an audience within the billboard industry, provide a useful comparison to the findings of several of the studies discussed in Section 2 of this report. In particular,

Young's finding that billboards attract a driver's attention whether or not the message is of interest or relevance, is quite similar to the findings of Crundall, et al. (1999), and Theeuwes, et al. (1998, 1999), both of whom showed that drivers do not, and cannot, ignore such irrelevant stimulation, even during the performance of a high priority task. Interestingly, Young's findings run directly counter to arguments routinely made by industry representatives in discussions with regulators – that there is no adverse safety consequence of billboards because, when a driver is engaged in a demanding task, he simply ignores the advertisement. An updated version of this report was issued in 2000, but has not been made public.

In addition to Perception Research Services, there are an unknown number of organizations that offer testing and assessment services to the billboard industry, or provide technologies to assist in such testing. Numerous technologies have been developed to perform such analysis, including simulator studies (PreTesting Company, Undated) billboard-mounted eye-tracking devices (Skeen, 2007), and others.

We are aware of only two billboard industry sponsored research studies that have addressed DBBs empirically. These studies have been comprehensively reviewed previously by Wachtel (2007), and the full details of those reviews are not repeated here. The interested reader can examine the full reviews at: http://www.sha.state.md.us/UpdatesForPropertyOwners/oots/outdoorSigns/FINALREPO_RT10-18-GJA-JW.pdf. Below, we have summarized the concerns that were discussed in the earlier reviews, as well as the comments of other independent peer reviewers. Overall, the reviewers have found serious weaknesses in both studies; weaknesses that call their findings into question. Conversely, in one of the two studies, data was collected but not fully analyzed or reported that should have led the researchers to conclude that there were, indeed, adverse safety consequences of roadside digital advertising signs.

Tantala & Tantala (2007)

This study was performed for the Foundation of Outdoor Advertising Research and Education (FOARE), an arm of the Outdoor Advertising Association of America (OAAA). The authors performed a post-hoc accident analysis study in which they reviewed statistical summaries of traffic collision reports, the originals of which had been prepared by investigating police officers. There are serious, inherent weaknesses in the use of this technique; such weaknesses have been understood and well documented for many years (see, for example, Wachtel and Netherton, 1980; Klauer, et al., 2006b, Speirs, et al., 2008). The use of this approach to relate crashes to driver distraction from DBBs, however, raises additional concerns. These issues are discussed below.

Limitations of Post-Hoc Accident Analysis.

Any post-hoc accident study, in which researchers review statistical summaries of traffic collision reports (TCRs) is limited, not only by the detail and accuracy of the original reports, but also by the inherent simplifications imposed by the coding system used to summarize the data in the first place. When a third party excerpts this summary

data for inclusion in a statistical data base, as is the case here, the level of detail and specificity that may have originally been present is further compromised. When such summary data are used to relate crashes to driver distraction that may or may not have been caused by the location and operation of DBBs, the interpretation of crash data is subject to further limitations, discussed below.

In addition to the general methodological concerns discussed above, there are several other important limitations to the viability of post-hoc accident analyses. These include:

- It has long been known that the majority of traffic collisions are never reported to, nor investigated by, the police. However, it was not until the conduct of the 100-car study (see, for example, Klauer, et al. 2006b) that researchers developed a “real world” understanding of the magnitude of this issue. The study documented 69 crashes that occurred to participants while driving their instrumented cars. Of these, 57, or 83%, were not reported to the police. If this statistic is applicable to the driving population at large in the U.S., then the fact that less than 20% of all crashes are reported to the authorities suggests that post-hoc crash studies are underreporting crashes by a factor of five. We believe that this problem is likely to be exacerbated with distraction accidents, for reasons to be discussed below.
- Unless a reported crash involves major property damage, serious injuries, or fatalities, any police investigation is likely to be cursory. In most States, only a serious crash requires a specialized investigative team to examine the precursors to the accident (evidence such as skid marks, debris fields, etc.) and to prepare a supplemental report. For the vast majority of police investigated accidents, no in-depth investigation is performed.
- As a result of the typical limited investigation, the crash location is generally identified as the point of rest (POR) of the involved vehicle(s) after the impact rather than the upstream location where the driver or drivers initially lost control or failed to pay attention. For a study of driver distraction or inattention, what matters is the location where the inattention or distraction occurred. The POR of the involved vehicle(s) is meaningless. In fact, since the POR may be a considerable distance downstream from the “distraction location,” not only will the TCR (and its statistical summary) fail to provide the relevant information needed, but this summary data may lead to an artificial understatement of the relationship between the source of the distraction and the accident, should one exist. This is because more crashes will be coded as having occurred at a roadway location that is not related to the source of the distraction.
- Drivers who are involved in crashes as a result of their inattention or distraction are unlikely to willingly report their pre-crash behavior to an investigating officer (Wallace, 2003b, Speirs, et al. 2008), due to concerns about legal liability, insurance surcharges, or points on a driver’s license.

Indeed, the driver may not even be aware of having been distracted or inattentive.

- As a result of a driver's inability or unwillingness to recognize distraction as a potential factor, an investigating officer is likely to check a box on the TCR such as "failure to yield right-of-way," or "following too closely" for expedience.

For these reasons, it is likely that the traffic collision summaries evaluated in this study represent a substantial underreporting of the true total number of crashes that occurred on the road sections studied within the analysis period. Further, it is likely that the classification scheme (using vehicle point of rest as the accident location) artificially reduces any true correlation between DBB distraction and driver errors that result in loss of control, and, at the same time, artificially increases correlations shown to be unrelated to DBBs.

Erroneous Underlying Assumptions.

The roadway sections for which data (accident report summary statistics) were collected for this study rest on two basic underlying assumptions made by the authors. The first assumption rests, in turn, on their determination of the distance from which a DBB could be seen by an approaching driver. The second rests on the researchers' decision to exclude from their data analysis those crashes that resulted from what they called "data bias" or "intersection bias." We believe that these determinations, and the assumptions based upon them, were seriously flawed. Each will be discussed in turn.

Assumptions about the Visibility Distance to DBBs.

The authors, justifiably, intended to analyze those crashes that occurred in the vicinity of DBBs, i.e. those roadway sections in which an approaching driver could first *see*, and subsequently *read* the message on such billboards. In other words, the crashes of interest would be those that were initiated (i.e. where a driver first lost control or first failed to attend to the driving task) during the time and within the road section that a DBB was within the visibility or legibility range of an approaching driver. We would want to compare such crashes to those that occurred on comparable roadway sections where no DBBs were visible or legible.

It is imperative, therefore, that the researchers identify, in advance of data collection, those roadway sections which were, and those which were not, within the visibility and legibility ranges of the seven DBBs that they studied. To support their determination of these locations, the authors provide the reader with five different terms, none of which are clearly defined in the report. These terms are: "visible range from route," "viewer reaction zone," "viewer reaction distance (VRD)," "viewer reaction distance zone", and "viewer reaction time (VRT)." The only one of these terms that is given a definition is this tautological and confusing description of VRD: "... Viewer Reaction Distance (is) how far from a billboard that the driver is potentially within the 'influence' of the

billboard” (p. 45, 79). In other words, viewer reaction distance is the distance in which the viewer can react to the DBB. Instead of providing a meaningful or operational definition of this key term, the authors explain that “reasonable values for VRD were previously determined in previous studies, and are a function of the driver’s speed.” But no such previous studies are cited, and no other basis for the VRD formula is offered. Regardless of the basis for the determination of VRD, however, the researchers’ statement that it is a function of speed is simply wrong. Clearly, the *distance* at which a driver can first see, and then read, any sign (DBBs included) is independent of speed; it is only viewer reaction *time* that would be affected by speed. This is a critical error, because this false assumption led the authors to identify those road sections upstream of each DBB for which they would collect and review the accident summary data. If these roadway sections were inappropriately truncated, and we will show below that this was the case, potential billboard-related crashes would be missed, and the identified correlation coefficients would be artificially and incorrectly reduced.

But the consequences of this error are even greater because of other mistakes made by the authors.

They report that, at 65 MPH, the VRD is approximately 0.2 miles with a VRT of 10 seconds (p. 79). But calculation demonstrates that, at 65 mph (95 ft/sec), 0.2 miles is traversed in 11 seconds, not 10. In addition, if the actual speed limit was 60 mph (88 ft/sec) and not 65 mph (see below), 0.2 mi requires 12 sec to travel. Thus, reviewing only those crashes that occurred within a 10 second VRT window would exclude an unknown number of crashes that might have occurred when a DBB was visible to an approaching driver. Further, the accuracy of the authors’ selected VRD is further reduced because they made no allowance for the fact that billboards on the opposite side of the roadway from the driver’s direction of travel (what they termed “left readers”) have a longer viewing time than those on the near side, and by their commingling of VRD with their measurement of “distance to the nearest billboard” (pp. 45-46) - a term which they do not define.

But their error in relating VRD to speed exacerbates this problem. Although Table 2-3 (p. 15), “Visible Range of Billboards Along Interstate Routes;” is never discussed in the report, a review of its content sheds light on the issue. The table shows the “visible range,” in miles and feet, for each of the seven DBBs in the study. Although never defined, visible range appears to represent the maximum distance at which each of the seven DBBs studied could be seen by an approaching driver; these distances range from a low of 0.28 to a high of 2.15 mi upstream of the sign. Translating these distances to time at 65 mi/hr, the DBB with the shortest visible range (#4) would be within an approaching driver’s visual range for 15.6 seconds, and the billboard with the longest visible range (#5) would be visible for 118.9 seconds, nearly two minutes. Thus, the researchers’ decision to review only those crashes within 10 seconds upstream of any billboard is insufficient even to assess the potential influence of billboard #4, the one with the shortest visible range - no less any of the other six, all of which were visible for greater distances, in one case more than ten times the limit chosen for data collection.

In summary, the consequences for compromising the validity of the data of this study are potentially high because the researchers' erroneous assumptions, even in light of their own documented sight distances, led them to exclude all crashes that might have been initiated in roadway segments further upstream from each of the billboards than they chose to study, but well within the visibility range of those billboards.

In addition to issues of sight distance, it should be obvious that every visible DBB along the route will have a different VRD and VRT depending upon numerous other factors, for example, sign location, elevation, angle toward the driver's eye, brightness, size of characters, roadway geometry, etc. None of these factors are addressed in the report.

If we look at legibility distance rather than visibility distance, the problem with the researchers' assumptions is similarly problematic. To take one example, if we assume (based on accepted industry practice) that 1 in of character height on a sign permits a legibility distance of 40 ft, and that a 14 ft tall billboard face (as were all seven in this study) with a character height of 75% of the available height or 10 ft 6 in (a reasonable assumption based on scaling the DBB images in Figures 2-4 and 2-8 of the report), then the legibility distance of such a sign would be 5040 ft (0.95 mi), nearly *five times* the VRD assumed by the authors.

So, if even the *legibility distance* of some of the DBBs studied is greater than the *visibility distance* accepted for analysis by the authors, there is a serious problem with the data that forms the basis of their conclusions. Further, given the size, brightness, and frequently changing imagery on DBBs, it is reasonable to assume that crashes initiated within a given sign's visibility distance must be considered, well beyond the legibility distance. In short, it is reasonable to assume that the gaze of an approaching driver might be attracted to, and that such a driver might be capable of reading, a DBB at far greater distances and for far longer periods of time, than the authors chose to evaluate in this study. It is reasonable to conclude, therefore, that the crash data accepted for inclusion in this study, based on the researchers' artificially constrained assumptions of VRD, has resulted in a substantial understatement of the true number of crashes that have occurred within the visibility and legibility range of the DBBs studied.

Finally, because Viewer Reaction Zone is never satisfactorily defined, the results reported in Tables 4-1 to 4-4 cannot be verified. Similarly, because the Visible Range is not defined, the results reported in Figures 4-2 to 4-9 must also be questioned.

"Data Bias" And "Intersection Bias"

One of the more troubling decisions made by Tantalala and Tantalala was to exclude from analysis any reported crashes that were attributed in the accident summaries to what they called "data bias." The reader cannot know exactly which such biases were excluded, because they are never clearly defined and because the descriptions of them change throughout the report. Indeed, as shown below, some of the stated biases are listed in certain report tables but not others. Their "data biases" included:

- Deer hits (sometimes called animal related)⁶

⁶ Discussed in Tables 4-5, 4-6, pp. 45, 49, 77

- Driving under the influence of drugs or alcohol⁷
- Adverse weather⁸
- Speeding⁹
- Senior related¹⁰

While it might be argued that deer hits, speeding, and DUI-related crashes were appropriately excluded from the data analysis, it is understood that most crashes have multiple causes, and it is possible that driver distraction may have played a role in some or all such crashes as secondary factors even if it had not been identified as the primary cause in the original TCR. On the other hand, it is recognized that adverse weather conditions place higher perceptual and cognitive demands upon the driver, the very kinds of increased workload for which researchers, traffic safety experts, and regulatory authorities express the greatest concern about the potential distracting effects of DBBs. In addition, older drivers (as well as young, novice drivers) may be at higher risk for distraction-related crashes, particularly when driving demands are high (see, for example, Chan, et al., 2008, Speirs, et al., 2008, Fisher, 2009). Thus, the exclusion of such “data bias” from their analysis raises additional questions about the basis for the researchers’ underlying assumptions. The authors’ supporting statement that: “A more fair and unbiased comparison of accident data would exclude accidents from known causes” (p. 63) is neither explained nor justified.

But it is their decision to exclude accidents in the vicinity of interchanges, called “interchange bias” (pp. 49, 77), that is particularly troubling. In their own words, the authors excluded interchange-related crashes because interchanges are where “drivers undertake additional tasks such as changing lanes, accelerating/decelerating, negotiating directions, and attention to others undertaking these additional tasks” (p. 78). It seems obvious that such driver demands associated with intersections are the very types of challenges that are of concern to the traffic safety community, and because interchange areas are among the prime locations for high visibility billboards, their elimination from this study is a cause for concern. If there is one issue about which all of the research about billboard distraction and all of the published guidelines and regulations agree, it is that billboards, and particularly DBBs, should not be located near interchanges, precisely for the reasons that Tantala and Tantala excluded such accidents from their analysis. Indeed, the Farby, et al., (2001) study for FHWA specifically noted that intersections and interchanges were highly demanding road locations, and that such locations should be included in any study of electronic billboards. Thus, the authors’ decision to ignore all such data is of concern.

Although the decision to exclude crashes in the vicinity of interchanges is problematic for the “temporal” (before-and-after) study that the authors conducted, it is more harmful in that section of the report that deals with “spatial” factors. One concern is that the reader cannot know which accidents were excluded due to “interchange bias” because the

⁷ Discussed in Tables 4-5, 4-6, pp. 45, 49, 77

⁸ Discussed in Table 4-5, pp. 49, 77 (“snowfall” and “icy roads” on pp. 49, 77)

⁹ Discussed in Table 4-6

¹⁰ Discussed in Table 4-6 (age 65 and above)

authors describe this exclusion zone in two conflicting ways within the same sentence. They state, in part, that they excluded “those accidents and billboards on interchanges (entrances/exits) within one mile (1/4 mile on each side of an interchange)” (p. 78). Regardless of whether they actually excluded accidents within ½, 1, or 2 miles from interchanges, any resulting findings are confounded by the fact that at least three of the seven billboards chosen for study (#3, Figure 2-8; #4, Figure 2-10; #7, Figure 2-16) appear, from photographs, to be in close proximity to interchanges. Thus, given that some percentage of accidents in the vicinity of these DBBs was excluded due to the signs’ proximity to the nearby interchanges, this artificially lowers the true number of crashes that may have been contributed by driver distraction due to these DBBs. As a result, the data for “bias adjusted” crashes in Tables 4-7 through 4-10, and in Figures 4-11 through 4-17 must be questioned.

Figure 1 below, taken from the ClearChannelOutdoor website, shows the researchers’ Billboard Number 3 and its proximity to an I-90 interchange. As discussed above, Billboards 4 and 7 are also close to interchanges. This leads to the rhetorical question – if accidents in the vicinity of interchanges are excluded due to “interchange bias,” and if DBBs are very close to interchanges, how can one capture and analyze accidents that occur close to the billboard? (Note that the authors provide no information about the proximity of any of the DBBs studied to the nearest interchange).



Figure 1. Proximity of DBB #3 to an interchange. This same DBB is shown in Figure 2-8, p. 16, of the Tantala study. It is also Site # 22 from the Lee, et al (2007) study, discussed in detail below.

(Source: <http://www.clearchanneloutdoor.com/products/digital/don/cleveland/index.htm>)

Decades of research into driver distraction has shown that alert, experienced drivers can tolerate some distraction when their task demands are not high, but that all drivers have upper limits on their cognitive capacities, and that there are certain road, traffic, and environmental conditions that may increase cognitive demands to the extent that additional sources of distraction should be avoided. Thus, the exclusion from analysis of some of the very types of crashes that might be *expected* to occur in the vicinity of DBBs is troubling, and, as with the decision to artificially truncate the data collection in road

sections upstream of DBBs, results in a likely substantial understatement of the actual crash statistics that took place in road sections where drivers were able to observe these DBBs. Taken together, the choice of crash types to exclude is a serious weakness of this study, given that some of the very kinds of crashes excluded are those that would be of direct relevance to the potential for distraction caused by billboards.

In summary, the authors' decision to exclude from study crashes that may have been affected by certain "biases" is critically flawed because it overlooks a basic understanding of traffic crashes – that they are frequently multi-causal – and it is precisely when such multiple factors are at play – adverse weather, older drivers, complex interchanges, speeding - that cognitive demands on the driver are increased and that irrelevant distraction cannot be tolerated. In other words, one should not exclude such factors because they cause "bias" – these are exactly the factors that interact to increase the likelihood of a crash when other factors such as inattention or distraction are present, and they must be investigated.

Inappropriate Statistical Methods, Assumptions, Analyses.

A key concern, raised by peer reviewers, about the findings of this study is that because of the limited before-and-after data collection periods (24 months) the sample sizes obtained are too small to conduct a meaningful statistical analysis. In addition to this concern, however, there remain others about the appropriateness of the research methods used and the results reported.

The analysis performed in this study is based on what the authors call "commonly accepted scenarios relating accident density to billboard density, to 'viewer reaction distance,' and to billboard proximity (how far the accident is from the nearest billboard)." But none of these terms is defined, no references to prior research are provided, and the conceptual drawing used to explain these assumptions in Figure 4-1 (p. 46) provides nothing more than a visual illustration of the authors' narrative statement. Thus, it is not possible for a reader to form an independent opinion of what was actually done, what assumptions were made, and how the data was collected and analyzed.

There are numerous examples of the erroneous use of statistics, both in terms of assumptions made, errors in application, and misuse of findings.

For example, the researchers define annual average daily traffic (AADT) as "the total volume of traffic in both directions of a highway or a road for one year divided by 365 days" (p. 33). But in their calculation of accident rates at "digital-billboard locations" (a term that they do not define), they fail to account for the fact that the seven DBBs studied were single-sided (i.e. they faced only one direction of travel). Thus, the authors have overstated the actual AADT by a factor of two, and the actual accident rate is therefore twice as high as reported.

In a section of the report titled "Accident Density and Billboard Density," it is clear that the researchers have inappropriately commingled DBBs with traditional billboards along

the route. By including all billboards in their metric for billboard density, they nullify both their ability to compare digital with conventional billboards, as well as their opportunity to compare digital billboards with the absence of billboards (an expressly stated objective of the study). This weakness is exacerbated because of their failure to control for the roadside environment (geometry, interchanges, presence of other objects in the roadside environment that might attract a driver's attention, etc.) in areas where billboards were present from those where they were not. For these reasons their statement that: "If a noticeable correlation between billboards and accidents exists, then one would expect a significantly larger number of accidents in areas with relatively high billboards densities" (p. 78) is unsupported.

As part of their statistical treatment of the data, the authors invent meaningless terms such as "noticeable correlation" (p. 78). Further, despite their correct understanding that correlation does not imply causation, they suggest otherwise on several occasions (see, for example, pp. 2, 98). Further, they inappropriately suggest that no correlation less than 1.00 is indicative of any relationship. For example, they state: "Statistically, a correlation coefficient of 0.7 or smaller is considered to indicate 'weak' correlation, at best, and does not indicate much difference from correlation coefficients of zero" (p. 81). Quite to the contrary, results from traffic safety research in the real world would typically consider correlation coefficients of 0.7 to be quite high.

The researchers undertook both a "spatial analysis," discussed above, and a "temporal analysis" to examine the incidence of crashes at locations where billboards had undergone conversion from traditional (fixed) to digital display. Data was collected for 18 and 24 months prior to, and after, the conversion. Although this before-and-after analysis is a necessary component of such an analysis, it is not sufficient. There is, in fact, an essential weakness to the temporal analysis performed in this study. That is that the researchers failed to compare the data at the billboard conversion sites to data at comparable locations at which there were either no billboards present, or billboards that were present but not converted to digital. It is possible that crash rates remained essentially the same in road sections featuring converted billboards (as these authors reported), but actually decreased in sections that included non-converted billboards, or at non-billboard locations, during the same before-and-after study period. This very result has been found in an earlier study of a single digital billboard near Boston (Massachusetts Outdoor Advertising Board, 1976), and led directly to the order that the sign be removed.

This failure of the temporal analysis underlies the authors' inability to answer the question that they posed early in the report: "... what is the statistical relationship between digital billboards and traffic safety?" (p. 4). This question is the one that should have guided this research. However, the next sentence, also posed in the form of a question, asks: "Are accidents more, less, or equally likely to occur near digital billboards compared to conventional billboards?" Unfortunately, it was this second question that guided the research, not the first. In other words, this study was not designed to investigate the potential impact on crashes of digital billboards compared to the *absence* of billboards; rather, it made the unjustified and unstated assumption that conventional billboards were the acceptable baseline for comparison with DBBs. As a result of this

assumption, the research methodology did not include true comparison sites where billboards were absent, and thus any assessment of the contribution to crashes from DBBs against a true baseline were impossible.

The announcement of the availability of this report on the website of the OAAA stated that this study “offers conclusive evidence that traffic accidents are no more likely to happen in the presence of digital billboards than in their absence.” Clearly, since the researchers made no comparisons between crashes in the presence and absence of DBBs, this claim is unsupportable.

Oversights, Omissions, and Evidence of Bias.

As discussed above, the metrics that the authors used to define the roadway sections for which accident report summaries were analyzed were called "viewer reaction distance" and "viewer reaction time". Obviously, each of these values is determined, in part, on the posted speed limit or on prevailing speeds. The authors claim that they used speed limit as their determinant, and that the posted limit was 65 MPH in all cases (p. 79). But this is incorrect. Figure 2 below clearly shows the posted Speed Limit to be 60 MPH. Although the reader cannot know whether this speed was in effect at all of the studied sites, it was clearly the case for DBB #4. The significance of this error would differ for each site, depending upon the sight distance for drivers approaching the billboard in question. At 60 MPH, a driver approaching a DBB will be able to see and read the billboard for a longer period of time than would be the case at 65 MPH, thus requiring data to be collected and analyzed for a longer roadway section upstream of the billboard, and far longer than any section that the authors chose to use. In other words, possible driver distraction from a DBB might well have occurred earlier than the authors reported, and, as a result, possible distraction-related crashes were artificially excluded from the database.



Figure 2. Image showing DBB #4 adjacent to posted Speed Limit signs. This image shows the same DBB depicted in Figure 2-10, p. 17 of the Tantala study. Interchange signs can clearly be seen, as can an additional billboard in the driver's view. This is the same sign represented as Site No. 42 in the Lee, et al. report discussed below. (Source: <http://www.clearchanneloutdoor.com/products/digital/don/cleveland/index.htm>)

The authors fill their report with information irrelevant to the study, while ignoring information of interest. For example, on pages 23-27 and in Tables 3-1 and 3-2, they describe in detail the total number of miles of interstate highways in the state and county, the terminus of each roadway, and the base and surface type of all pavements. On pages 29-31, they provide cursory information about the location of each of the studied billboards – again providing data on road surface and previous state work projects, and repeating, verbatim, information already presented on pages 10-11. However, no information is given about relevant concerns such as horizontal and vertical curvature, merges or lane drops, presence of official signage, proximity of DBBs to interchanges, multiple billboards within a driver's line of sight simultaneously, or intersection characteristics such as entrances, exits, gores, etc., either for the system as a whole or within the vicinity of the studied DBBs.

Bias is evident throughout the report. For example, the authors' state that their numbering system for the billboards studied was "arbitrary" (p. 10), but a review of the website of the billboard owner, ClearChannelOutdoor, shows that this information was supplied by them. Several figures and tables in the report are taken directly from the ClearChannel

website, and a ClearChannel executive was quoted as saying that his company had “hired” the researchers to perform the study (Slobodzian, 2007).

It is typical in a research study such as this for the authors to identify prior research and other sources that have informed their assumptions, methods, and conclusions. However, despite listing 17 references, none are actually cited in the text. In addition, references made within the report of prior research are not accompanied by citations; thus it is not possible for the reader to verify the basis of the authors’ claims.

Author Response.

One of the two authors of the paper, in a letter sent to the Director of Right-of-Way for the Texas Department of Transportation (Tantala, 2007) responded to the previous Wachtel (2007) review and took issue with a number of statements made in that review. This section discusses the Tantala response, and our conclusions based on a review of the response and a resultant re-review of the paper and our comments to it.

The Tantala letter takes issue with two major criticisms that were included in the Wachtel report (and are discussed in detail above). First, Tantala argues that the Wachtel criticism of the report’s exclusion of accident analyses beyond the VRD (approximately 0.2 miles upstream of the DBBs at 65 mi/hr) “misrepresents our analysis, because we did not exclude larger ranges. In fact, our analysis compiled statistics for a wide range of vicinities” (p. 1). A review of the Tantala letter and a re-review of the original report reinforce our original criticism. The key phrase in the Tantala letter is: “...we examined accident data and statistics...” While that may be true, any such data and statistics were not analyzed, and no supportable conclusions could be drawn from them. Indeed, the Tantala letter refers the reader to two report Tables (2-3 and 4-11) and two Figures (4-24 and 4-25) in support of his arguments. Our re-review of Table 2-3 (p. 11) confirms that this table merely identifies the “visible range” for each of the seven DBBs. Table 4-11 (p. 84) shows “correlation coefficients of various comparisons,” and the one of relevance here, accident density vs. VRD, simply reaffirms our criticism. Finally, the two cited figures (pp. 90, 91) present nothing more than summary statistics (raw accident counts) without analysis.

The second point made in the Wachtel review with which Tantala takes issue is that “the review opines that our analyses should not exclude ‘bias’ factors because accidents are often multi-causal and those are the very factors that increase the likelihood of accidents” (p. 2). Tantala expresses his agreement with Wachtel’s opinion, and states “we did include this in part of our study. In fact, we performed an analysis that included all data collected and compiled by the State of Ohio.... This robust, comprehensive and all-inclusive data-set includes the very multi-causal accidents that the review references” (p. 2).” But the Tantala letter provides no link or reference to any pages, tables, or figures in the report where a reader might find these all-inclusive analyses (those in which the stated biases were included in the analyses). Indeed, our re-review of the paper, undertaken as a result of the Tantala letter, finds no such analyses. In fact, only Table 4-5 (p. 54) addresses the all-inclusive vs. bias-adjusted accidents, and it merely presents the

summary statistics of raw accident counts and accident rates with no accompanying analysis. In contrast, after stating: “A more fair and unbiased comparison of accident data would exclude accidents from known causes” (p. 63), the report presents a series of four tables (4-7 through 4-10) and seven figures (4-11 through 4-17) that present “the number of accidents with statistical bias events excluded within the visible range” (p. 63). If there was any comparable presentation of the all-inclusive data within the report, this reviewer could not find it.

In summary, Tantalus’s letter defending the study against Wachtel’s criticisms does nothing to challenge the points made in the review and, as a result, reinforces the original concerns raised by Wachtel.

Lee, McElheny, & Gibbons (2007).

As is the case for the Tantalus and Tantalus study discussed above, this study was performed for the Foundation for Outdoor Advertising Research and Education (FOARE), an arm of the Outdoor Advertising Association of America (OAAA). It, too, has been previously reviewed (Wachtel, 2007), and the complete report can be accessed at:

<http://www.sha.state.md.us/UpdatesForPropertyOwners/ooots/outdoorSigns/FINALREPO RT10-18-GJA-JW.pdf>. Below we will review the major reported findings of the Lee, et al., study, and discuss our principal concerns about the efficacy of this work.

The approach to this study was completely different from that of Tantalus and Tantalus, although the two studies used the same DBBs. In this study, an instrumented car was driven along a prescribed route by a volunteer sample of drivers, and some of their driving behaviors and eye glances were recorded as they passed previously identified and defined locations.

Study Overview.

In the main study, 36 participants drove an instrumented vehicle along a pre-determined 50-mile route on surface streets and interstate highways in the Cleveland, Ohio area. During the drive, the participants passed a number of DBBs, conventional billboards, “comparison” and “baseline” sites. In the final 8 sec of their approach to each of these sites or “events,” the direction of their eye glances was recorded, along with their lane keeping and speed maintenance performance. A subset of 12 participants also drove a similar, but shortened, route at night.

Methodological Concerns.

Eye Glance Recording.

Eye movement recording and analysis is a time-proven method for determining where drivers are looking as they drive. Until recently, however, it has not been possible to obtain precise eye glance data (with a precision of 1 deg or better) without the use of

highly intrusive, head mounted equipment. The trade-off is to use recording equipment that is mounted on the dashboard or other interior vehicle structure, but the weakness of this less intrusive system is that eye glance information can then be obtained only for more gross directions of gaze. In other words, while it is possible to record the general direction in which a person is looking, it is not possible to know with confidence the exact object (no less an image within that object) being viewed, or the distance from the eye at which that object is located. Because this study employed such vehicle-mounted equipment, the researchers could report only on the general direction of gaze and could not identify if, or when, a participant was looking at a specific object (such as a DBB) in the visual field.

Eye movement recording equipment must be calibrated separately for each participant, and this calibration should be performed both before and after each participant's drive. This is because eyeglance recording equipment can "drift" over time, vehicle vibration during the drive could have changed the mounting position of one or more cameras, or the driver could have adjusted the seat or otherwise shifted his or her position while driving. Unfortunately, the authors calibrated the equipment only after each participant had driven the route, and thus could not know whether the eye glances that they captured were accurate and reliable.

Lack of control over site variables

The authors conducted their on-road studies on "interstate, downtown, and residential road segments" (p. 27). Given that all five DBBs (study sites) were on interstate highways, the decision to include some of the control sites (baseline, conventional billboards, comparison sites) on roads other than interstates confounded the data collection and made meaningful comparisons across sites impossible. When conducting field research, the goal must be to reduce, wherever possible, extraneous sources of variability. In this study, the decision to include study sites (DBBs) on interstates and some control sites (the reader is not told which or how many) on surface streets leads to additional uncontrolled sources of variability. Some of the significant differences between these two types of roadways, any or all of which may have affected the data, are: traffic speeds and flow; illumination levels; sight distances; access control; at grade vs. grade separated intersections; presence or absence of traffic signals; and divided vs. undivided traffic.

Even for the five DBBs that were the principal focus of this research, the authors seem to have made no attempt to identify, no less control, extraneous variables such as traffic speeds and volume, horizontal and vertical curvature, or other roadway and traffic characteristics that might have interacted with the variables of interest. Further, the distance between adjacent study sites was often very short. For example, using the Haversine formula, we calculated the distance between Site 37, a DBB, and Site 36, a baseline site, as less than 1.2km. Other studied sites might have been even closer to one another. Thus it is likely that the visibility ranges for adjacent sites overlapped, confounding eye gaze and vehicle performance measurements and comparisons.

The researchers selected some study sites on the right side of the road and some on the left, then recorded and analyzed whether drivers glanced in the direction of these sites as they approached and passed them. In some cases they found examples of participants looking in the direction opposite to the site being studied. When such behavior occurred in the presence of billboard sites, they interpreted this to mean that the billboard did not draw the driver's attention. But there is no evidence to suggest that they sought to identify or control for the possible presence of billboards or other attention-getting targets that may have existed opposite from their study sites or otherwise within the driver's field of view simultaneously. In other words, when they selected a study site on the right, there is no indication that they made sure that there was nothing on the left that might capture the driver's attention. If, in fact, they did not identify and control for such opposing sites, then the eye glance data that they captured are suspect. Since they do not report any efforts to evaluate and control for such conditions, one must assume that they did not do so. In short, it is entirely possible that glances to the left when a billboard was on the right (or conversely) were made because there was a competing, perhaps more compelling, site across the road from the study site that was neither controlled nor evaluated. Figure 1, for example, shows the DBB that served as Site # 22 on the right side of the road¹¹. But the figure also shows a large billboard on the left side of the road that appears in the center of the image. If the researchers captured eye glances straight ahead or to the left at this location, they might have been due to the participant looking at this uncontrolled billboard. A similar concern exists for uncontrolled sites that might exist on the same side of the road as a site of interest and within a driver's field of view as he or she approached that site. Given the lack of precision of the eye gaze data obtained, there was no way for the researchers to know whether a particular participant was looking at the study site or an unidentified site visible simultaneously for which they did not control.

Although the five DBBs studied were all of the same size, the reader is given little information about other important characteristics of these signs; characteristics that could have had a direct impact on their attention-getting qualities, such as their height, angle to the drivers' line of sight, and proximity to the road. Further, the reader is told little about roadway geometry, prevailing traffic speeds and volume, etc. Any of these factors may have affected the comparability of sites. Even though all five DBBs were 14' high and 48' wide, they were mounted at very different heights relative to the road surface. Further, there was no consistency of sizing of conventional billboards or signs on the comparison sites. Indeed, the researchers state that conventional billboards included a "few" that were of other sizes, including "standard poster, junior paint, and 10'6" x 36' bulletins" (p. 21). Since the size of a billboard or other sign, and thus the size of the characters that can be displayed on it, likely has a direct relationship to the distance from which it can be seen and read, this failure to control for sign size and other characteristics relative to a sign's visibility and legibility range is an important oversight. In our opinion, without any effort to control these basic site and sign characteristics, it is difficult for the researchers to defend any interpretations they may have made from their data in comparing driver responses to DBBs against responses in other locations.

¹¹ Note – this figure was taken from the ClearChannelOutdoor website – it was not shown in either of the two studies discussed herein although we have confirmed that it is the study location cited in the reports.

Confounding of data collection sites.

The researchers selected four types of “events” or “sites” at which to collect data. For the main (daytime) portion of this study, there were 5 DBB locations, which we have called study sites, and three other types of locations, which we have called control sites. The latter included 15 “conventional billboards,” 12 “baseline sites,” and 12 “comparison sites.” Because the report provides no images or drawings of any of the 44 locations, and because the descriptions and definitions of the site characteristics, particularly for the baseline and comparison sites, are vague and inconsistent, it is not possible for the reader to determine just how these site types compared to one another. For example, at one point, the authors state that baseline sites contained no signs of any kind (p. 6). At another, the reader is told that some baseline sites (the authors do not state how many) in fact, did contain signs. A more serious concern, however, is with the multiple, conflicting definitions and descriptions of the comparison sites. The reader is first told that comparison sites are “similar to items you might encounter in everyday driving” (p. 8). On page 21, these sites are described as “areas with visual elements other than billboards.” Later on the same page the reader is told that some of these sites included on-premise signs, variable message signs, and “digital components.” Finally, Table 2 (p. 22) describes one comparison site as a “tri-vision billboard” and three others as “on premise LED billboard(s).” To the average motorist, and from the perspective of driver distraction potential, the distinction between an on-premise and an off-premise digital sign display is meaningless. One must conclude that at least some of the comparison sites may have been just as visually compelling and distracting, if not more so, than the DBB sites that were the principle focus of the study. Clearly, this intentional confounding of study and control sites (the researchers selected each of the sites to study) would artificially reduce any adverse findings from DBBs by showing them to be no worse than existing sources of distraction present at the comparison sites.

As expected, the study’s findings bear out this concern in that, for many measures, the DBB and comparison sites elicited similar results, and these results differed, often significantly, from those obtained at conventional billboard or baseline sites. The problem for the researchers is how to treat these findings given their *a priori* inappropriate site selection decisions; the problem for the reader is how to interpret them. In our opinion the approach adopted by the researchers is seriously flawed. It takes the clear evidence found in this study that roadside digital advertisements (whether on- or off-premise) are associated with adverse driver performance, and manipulates this evidence to suggest that there is no problem with digital billboards because drivers are equally distracted by other “comparison” sites. In short, the authors’ false assumption that their chosen comparison sites were appropriate control locations against which to compare the effects of DBBs enables them to slant their findings to suggest that, because driver performance in the presence of digital billboards is similar to their performance in the presence of these equally distracting “comparison” sites, there is no cause for concern about the safety of DBBs. We believe that the data suggests otherwise, as discussed below.

The choice of an 8-second data recording interval.

The researchers chose a time period of 8-sec in advance (upstream) of each site during which to record driver performance and eye glances. This data recording period ended when the instrumented vehicle passed each event. The assumption that 8 sec was a reasonable data capture interval, and the researchers' ability to define and measure this interval, raises several methodological concerns.

At 65 mi/hr, the presumed speed on the freeways studied, a vehicle travels approximately 95 ft/sec. Thus, during an 8-second interval, a vehicle will travel 760 ft. The accepted practice for highway signs is that 1 in of letter height can be read from approximately 40 ft. So, for a billboard with 24 in high characters, the sign can be read from approximately 960 ft. Indeed, several of the billboards used in this study likely included characters much larger than 24 in and thus could be read at even greater distances (given clear sight lines upon approach). Figure 3, enlarged from Figure 2-4 (p. 13) of the Tantala and Tantala study, depicts characters approximately 84 in high (the DBB face is 14 ft tall). These characters are theoretically legible (no less visible) from a distance of 3,360 ft. At 65 mph, this sign could be read for approximately 35 sec, more than four times the data collection interval used in this study. In addition, because of the brightness, contrast, and image quality of digital billboards, and the fact that (in Cleveland) their messages change every 8-seconds, it is apparent that driver attention to the billboard may be initially attracted at far greater distances than those at which the message can actually be read. As a result, the choice of an 8-sec data recording interval is likely to result in a substantial understatement of the distracting effects of digital billboards compared to other roadside sites including more traditional billboards and on-premise signs.



Figure 3. An enlargement of the DBB that served in both the Tantala & Tantala and Lee, et al. studies. Scaled measurement shows the numerals to be approximately 84 in. high.

The authors state that they chose an 8-sec data collection period because the “digital billboards were programmed to change messages instantaneously once every 8 seconds; an event length of 8 seconds thus made it highly likely that a message change would be captured during the event” (p. 21). This argument is flawed for several reasons. First, as described above, the sight distance and legibility distance, coupled with the size of the signs studied and their character height, demonstrates that digital billboards can be seen and read far earlier than 8 sec in advance of the sign, thus suggesting that the data recording interval should have been much longer. Second, had the researchers selected *any* data recording interval longer than 8 sec, it, too, would have permitted them to capture a message change during each driver’s approach to the event. Finally, despite their understanding of the potential importance of a driver observing a message change during his or her approach to the DBB, the researchers never actually reviewed or analyzed any data related to this message change, and therefore had no way to evaluate any possible driver response to it.

Some signs are located perpendicular to the driver’s direction of travel. Others, such as some two-sided billboards and many on-premise signs, may be located at other angles, including parallel to the driver’s direction of travel (such as when mounted on a building façade). In addition, the lateral distance of each sign from the driver’s line of sight varies greatly as a result of factors such as: lateral distance from the road edge, and the number and width of lanes, medians, and shoulders. If the same 8-sec point for passing a sign was applied regardless of sign angle and lateral distance, then some signs would be visible to drivers for less time than others, thus rendering the 8-sec recording interval inconsistent across the studied sites.

In summary, the researchers’ choice of an 8-sec data recording interval was inappropriate for several reasons, and resulted in unequal exposure to signs of interest across sites. A more appropriate way to determine the data collection interval would have been to identify the point at which a billboard or other sign of interest fell outside a predetermined angle of view from the driver’s line of sight along the road axis, and to define the data recording interval upstream from that point. This would have assured a more equitable, and comparable, identification of sight distance and would not have had the effect of artificially reducing the available glance times and control measurements made for the signs of interest in this study.

Measurement of nighttime luminance levels.

The authors measured the luminance levels of different sites at night. They took these measurements from the participant-driver’s eye position, a decision which masked and minimized the actual brightness differences between the DBBs and the other sites. A more appropriate comparison would have been from measurements taken directly in front of each of the signs of interest (as recommended in, for example, TERS, 2002; NYDOT 2008a) so that the authors could be sure that they were comparing sign against sign without the contribution of the general ambient environment. Several other weaknesses affected this measurement approach. First, taking measurements from the driver’s position would have yielded non-comparable readings even if every sign had the same luminance, merely because the signs were positioned at different angles to the driver, and

were located at different horizontal and vertical distances from the driver's eye. Second, the authors do not state whether some of the (non-DBB) sites measured at night were those on surface streets and whether there were fixed luminaires within the range of the luminance meter at such sites. The presence of fixed lighting would also have reduced the actual luminance differences between DBBs and other sign sites. Third, since the DBB displays changed every 8 sec, the luminance levels on these signs changed accordingly. Thus, unless the researchers measured each DBB with the identical display (highly unlikely), they would have no way to compare the light output of the different DBBs. They would not know, for example, whether measured differences between DBBs were due to actual sign output, different brightness settings, or differences between displayed messages. Despite these limitations in measurement strategy, however, and despite the fact that the digital billboards were automatically dimmed at night, the authors recorded nighttime luminance levels at the driver's eye position that were, on average, 10 times greater for the DBBs than for baseline sites, approximately 3 times brighter than sites with conventional billboards, and approximately 2.5 times brighter than comparison sites. The authors' state: "this probably explains some of the driver performance findings in the presence of the digital billboards" (p. 68).

Inappropriate and Inconsistent Statistical Treatment.

Eye glance recording and long duration eye glances.

One of the greatest weaknesses of this study is the authors' failure to follow their own recommendations as expressed in their review of the work by Wierwille (1993), Horrey and Wickens (2006), and the "100 car study," (Dingus, et al., 2006). This error is compounded by their questionable decision to analyze and present only selected data that they collected, choosing not to report their own findings that might have undermined their conclusions. These actions require some explanation.

The authors collected and recorded four types of eye glance behavior at each of the four types of sites: glance frequency, glance duration, average duration per glance, and total eyes-off-road time. Of these four measures, those that deal with the duration of eye glances off the road are of the greatest relevance because long duration eye glances at distracting stimuli have been implicated as predictive of crash risk in several prior studies, including those by Wierwille (1993), Smiley, et al., (2005), Horrey and Wickens (2006), and Klauer, et al., (2006a). Lee and her colleagues are clearly aware of this work, as they state as early as the study abstract: "Various researchers have proposed that glance lengths of 1.6 seconds, 2.0 seconds, and longer may pose a safety hazard" (p. 6). The authors follow this statement with an overview of their own results, in which they claim to have found no pattern of longer glances to the digital billboard sites: "An examination of longer individual glances showed no differences in distribution of longer glances between the four event types" (p. 6); and: "An analysis of glances lasting longer than 1.6 seconds showed no obvious differences in the distribution of these longer glances across event types" (p. 9). These two statements are misleading, and wrong, as discussed below.

In their introductory description of eyeglance results (p. 52) the authors list the seven questions that they sought to answer with the eyeglance data collected. The seventh question was: “Are longer glances (longer than 1.6 s) associated more with any of the event types?” This is, of course, a key question, because of recent research that identifies such “longer glances” as being associated with a higher crash risk. After listing the seven questions, Lee and her colleagues present a summary and analysis of their findings relative to each. For six of the seven questions, they performed an analysis of variance (ANOVA) to analyze the data, and they report their tests of statistical significance in both graphical and narrative form (see Figures 17-22, pp. 53-58). It is only for the key Question 7, the one that addresses longer glance durations that the authors apparently performed no such analysis and offered no test of statistical significance (see Figure 23, p. 59). The reader might ask why, but the authors provide no explanation. After restating Wierwille’s recommendation that 1.6s be used as a criterion representing a long glance away from the roadway, and after again explaining that their approach in analyzing this data followed that recommended by Horrey and Wickens, “who suggest analyzing the tails of the distributions whenever eyeglance analysis is performed” (p. 59), Lee and her colleagues failed to perform this analysis. Instead, it appears that they performed nothing more than a visual inspection of the data presented in their Figure 23 (p. 59), the figure that depicts the distribution of glance durations for the four different event types. Perhaps as a result of performing only this visual inspection, they state: “As shown in Figure 23, the distributions of glance duration were similar across all event types, and there was no obvious pattern of longer glances being associated with any of the event types” (p. 59). This statement is wrong, as discussed below.

This failure to report key findings is even more surprising because of the results that the researchers obtained in response to their Questions 5 and 6. These two questions asked whether the “mean single glance time” varied according to the type of event. Question 5 asked this question for events on the left side of the road; Question 6 addressed events on the right side of the road. In both cases, the Lee and her colleagues found that digital billboards and comparison events had statistically longer mean single glance times than did baseline or conventional billboard events ($F_{3,73} = 3.59, p = 0.0176$ left, and $(F_{3,77} = 3.73, 0.0147$ right), and that the DBB and comparison sites did not statistically differ from one another. In addition, in an effort to “increase power and verify the above findings” (p. 60) the researchers aggregated the left and right eyeglance data. This combined analysis confirmed with statistical significance ($F_{3,91} = 4.98, p = 0.0030$) that “digital billboards and comparison sites did not differ from one another, but each differed from conventional billboards and baseline events” (p. 60).

These findings alone should have led the researchers to statistically evaluate the longest such glances, the tails of the distribution, as they said they would in posing Question 7, and as they did for every other question.. But they did not do so.

Figure 4, below, reproduces the authors’ Figure 23 (p. 59) together with its original caption.

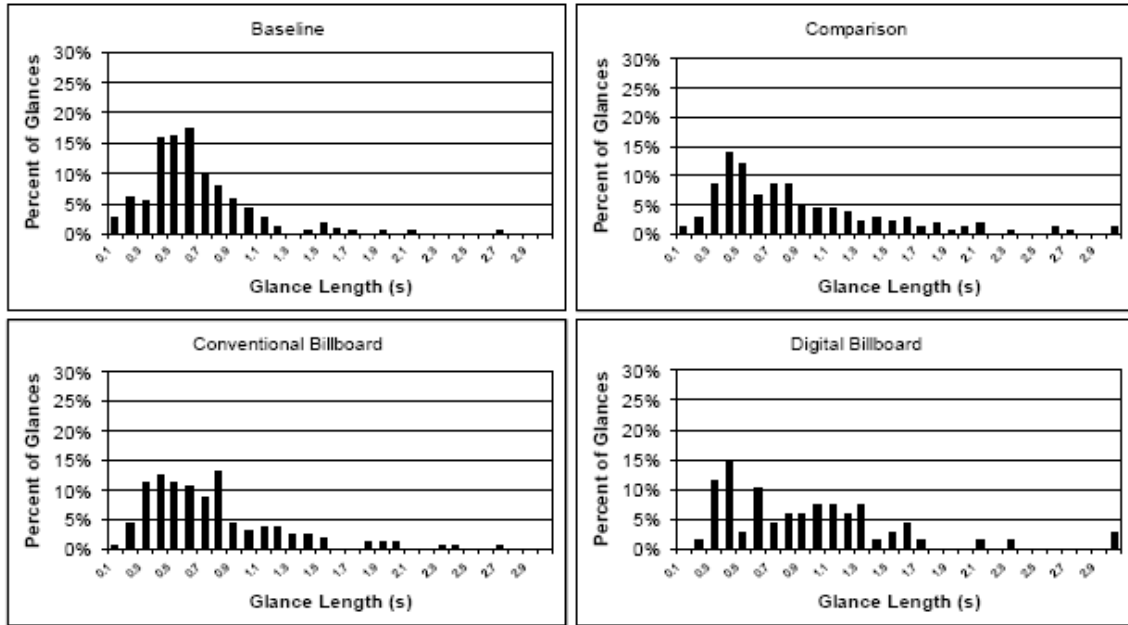


Figure 23. Tails analysis for the distribution of glance duration, (method described in Horrey and Wickens, 2007).

Figure 4. A reproduction, in original size, of the authors' Figure 23 (p. 59), together with its original caption.

The authors do not provide sufficient information about these measured glance durations to permit the reader to perform an independent analysis of their data. However, an inspection of enlargements of these four charts enables a non-statistical independent review of their findings. Using the tails analysis as recommended (but not performed) by the authors (following Horrey and Wickens), and using both 1.6 sec (the Wierwille criterion) and 2.0 sec (the 100-car study cut-point), we find the following:

Approximately 5.5% of baseline sites and 7.5% of conventional billboard sites captured glances of 1.6 sec or longer compared to 13% of DBBs and 16% of comparison sites.

Approximately 2% of baseline sites and 4.5% of conventional billboard sites captured glances of 2.0 sec or longer, compared to 7% of DBBs and 8% of comparison sites.

No glances longer than 3.0 sec were made to either the baseline or conventional sites, but glances of 3.0 sec or longer were made to both DBBs and comparison sites.

In summary, this visual inspection of the researchers' data suggests that long glances occur two-to-three times more often with DBBs and comparison sites than they do with baseline or conventional sites, and that the longest glances (3.0 sec or longer) occur *only* with these sites. These results suggest important differences for the longest glances, the

ones that highway safety experts are most concerned with. One must ask why the authors chose not to perform a statistical analysis of this data, particularly when they did so for every other set of eyegance data, and why they reported that their visual inspection of these data suggested that there was “no obvious pattern of longer glances being associated with any of the event types” (p. 59). The report offers no explanation.

Misleading and Inconsistent Reporting and Evidence of Bias.

Throughout the report, there are conflicting and inconsistent statements, and evidence of bias.

Was this a “naturalistic” study?

Although described by the authors as a “naturalistic study,” and modeled superficially upon the much larger, 100-car study performed at the same institution – (Dingus, et al, 2006; Klauer, et al. 2006a,b), this study exhibits few of the characteristics of a true naturalistic study (Hanowski, 2009).

Although they used an instrumented vehicle with on-board cameras, and although their participants drove the route without a researcher present in the vehicle, this study differs significantly from the 100 car study in several key ways. First, the four on-board cameras used to record views of the road and of the drivers’ glances were not unobtrusive as they were in the 100 car study. Rather, they were prominently located on the driver’s side A-pillar and adjacent to the rear view mirror. These camera locations are shown in Figures 8-10 of the report (pp. 32-33). Second, the duration of the present study was less than two hours per participant, whereas, in the 100 car study, participants kept their instrumented vehicles in their possession and used them daily for several months. Third, participants in the present study had to follow a prescribed route (to ensure that they would pass the DBBs and other events that were the subject of the study), using a set of printed instructions taped to the dashboard, whereas in the 100 car study, participants were free to drive when and where they chose in the course of performing their daily activities. In short, whereas the participants in the 100 car study may well have become acclimated to their test vehicles over time and ignored the fact that they were participating in a research study, the participants in the current study were fully aware that their performance and behavior was being monitored and recorded – thus their behavior could not reasonably be described as “naturalistic.”

Literature Review.

The authors’ approach to their literature review is illustrative of the bias shown throughout the report. There is a long history of published literature examining the relationship of roadside billboards to crashes and to driver behavior. Relevant studies dating as far back as 1934 have been identified and reviewed by others; and research continues to be conducted and reported to the present day. The authors chose to discuss only a small, highly selective subset of these studies. As will be seen below, it is clear that the studies reported, particularly the early work in this field, were selected because they were supportive of the authors’ position. When they cite studies that reported findings at odds with their position, the authors dismiss them as poorly done or irrelevant;

conversely, studies that report findings consonant with these authors' views are praised with descriptors such as "rigorous."

Their reporting about two early epidemiological studies is illustrative of their approach to the literature. The authors cite an article by Rykken (1951), a two-page interim progress report on a roadside study conducted in Minnesota. They quote from Rykken: "...no apparent relationship was found between accident occurrence and advertising sign type or location" (p. 12). What they fail to say, however, is that Rykken called his result "a very preliminary study of approximately 170 mi. of the 500 mi. study segment (p. 42). Significantly, Lee, et al. fail to cite the final report of the subject study (Minnesota Department of Highways, 1951) which concluded, in part: "An increase in the number of advertising signs per mile will be accompanied by a corresponding increase in accident rate" (p. 31), and "intersections at which four or more (advertising) signs were located had an average accident rate of approximately three times that for intersections having no such signs." This final report has been extensively cited and reviewed by previous researchers. Wachtel and Netherton (1980), in particular, discussed it at length. It is puzzling, therefore, why these authors cited the interim progress report and ignored the final document.

Lee and her colleagues followed the same approach in their review of a parallel study conducted in Michigan. They cite an interim study report by McMonagle (1951) that looked at only partial findings (p. 12), and ignored the study's final report (Michigan State Highway Department, 1952) which found that illuminated advertising signs showed "an appreciable association with accident locations" (p. 6).

In a confusing discussion about a study by Rusch (1951) which analyzed crash reports on Federal and State highways in Iowa, the authors fail to report on Rusch's own published results, and offer no evaluation of his actual study. Instead, they cite a brief review by Andreassen (1985) (ignoring all other published reviews of the Rusch work) which stated, in part: "the greatest number of inattention accidents occurred on the sections where business and advertising predominated as the roadside property usage, but this does not prove anything about the effect of advertising signs on accident occurrence" (p. 13). Given that Rusch's actual findings, despite methodological weaknesses that often affected these early field studies, demonstrated that the number of accidents was more than double in the study section (where 90 percent of the businesses and roadside advertising signs were located) than in either of the two control sections, given that "inattention" accidents predominated over both "business" and "other" accident categories in this study section, and given that the results were confirmed after statistical correction for mileage per segment, the researchers' treatment of this study is puzzling.

Obfuscation of Study Purpose and Intentional Confounding of Study Sites

The stated purpose of this study was to "assess the effects, if any, of digital billboards on driver behavior and performance" (p. 8), *not*, as suggested in the Abstract, to ascertain whether driving performance in the presence of digital billboards was similar to performance in the presence of other, primarily on-premise, digital signs. As discussed

above, the researchers clearly found that DBBs *did* have an adverse impact on driving performance, and the fact that this adverse impact was similar to the adverse impact from similarly distracting signs that might have been on- rather than off-premise does not diminish this finding nor make it acceptable. The authors admit that “there are measurable changes in driver performance in the presence of digital billboards” (p. 6), and, as demonstrated in the body of their report, these changes are adverse and statistically significant. It is inappropriate to suggest that such adverse impacts are deemed acceptable (or “safety neutral” in the authors’ coinage) merely because they “are on a par” with the adverse effects of other digital signs that happen to be other than billboards because they may be located on the premises of roadside businesses.

Baseline sites should have been, as stated in the abstract, “sites with no signs.” But, as described elsewhere in the report, an unidentified number of them *did* contain signs, thus diminishing their potential to serve as true control sites and, likely, minimizing the differences in glance behavior between DBBs and true baseline sites.

In direct conflict with a statement in the Abstract, and as discussed in detail above, longer individual glance patterns (greater than 1.6 and 2.0 seconds) *did* show differences (actually, rather dramatic differences) between the event types. In fact, per the authors’ own statements elsewhere in the report, and as shown by several other researchers, these differences at the tails of the distributions for glance duration may be critically important in assessing the true impact of digital billboards on driver performance and behavior. Similar misstatements are made throughout the Executive Summary, and will not be repeated here. However, the expressed “finding” that: “An analysis of glances lasting longer than 1.6 seconds indicated that these longer glances were distributed evenly across the digital billboards, conventional billboards, comparison events, and baseline events during the daytime” (p. 7) is clearly inaccurate. Critically, the data discussed in this “finding” was not analyzed by the researchers in accordance with their own data analysis recommendations, nor was such data even collected for the abbreviated nighttime study, when we would have expected such findings to be even more dramatic than they were in the daytime study.

The authors identified five DBBs for study. These are identified by latitude, longitude, route number, and side of road in Table 2 (p. 22), and shown graphically on a map in Figure 2 (p. 23). With this information, that reader can view images of these DBBs from either the Tantara report or from the website of ClearChannelOutdoor, at <http://www.clearchanneloutdoor.com/products/digital/don/cleveland/index.htm>. Examination of Figures 1 and 2 in our report may lead the reader to question the accuracy of the authors’ statement that: “The Cleveland digital billboards... were located off to the side of the roadway in straight-away sections of interstate with no interference from hills, curves, or intersections” (p. 19).

The authors provide voluminous data for irrelevant issues (e.g. 124,740 video frames analyzed, 96,228 data points collected, 8,678 eye glances identified, etc.) but offer no information useful to readers who might want to know what was actually studied. For example, there are no images of any of the billboards or other sites studied, there is no

indication of the precision with which eye gaze was captured, etc.). It appears as if the researchers intended to overwhelm the reader with useless information in an attempt to avoid questions about the real issues.

There are numerous statements throughout the report that, on the one hand, are irrelevant to the study, and, on the other, demonstrate a clear pro-billboard attitude. Some examples:

“The lead author of this report recently participated on an expert panel charged with providing recommendations for a minimal data set to be included on police accident reports; billboard were never raised as a possible distraction...” (p. 11).

“After a long gap in research, there were a few additional studies in the 1960’s through the 1980’s, none of which demonstrated that billboards were unsafe.” (p. 11)

“The national crash databases do not mention billboards in their list of driver distractions.” (p. 14)

Findings that DBBs are “Safety Neutral.”

The authors invented the term *safety neutral* (p. 10) to describe their conclusions about the impact of DBBs on driver distraction and performance. They state: “Although there are measurable changes in driver performance in the presence of digital billboards, in many cases these differences are on a par with those associated with everyday driving, such as the on-premise signs located at businesses” (p. 6). In other words, the authors say, because other roadside distractions such as their “comparison sites” (which, they note elsewhere, contained multiple signs, changeable message signs, and digital, flashing, and video displays) are also associated with difficulties in speed and lane maintenance and excessively long glances away from the forward roadway, DBBs should be considered *safety neutral* because their adverse effects on driver performance are similar to the effects from these other digital advertising signs..

The authors are able to reach this conclusion because of their intentional confounding of the DBB and comparison sites. The intentionality of this confound is demonstrated by the fact that the researchers had complete freedom to select the (50-mile long) study route and to choose the test sites anywhere along that route. That they chose “comparison sites” which often included digital signs, changeable message signs, and flashing and video signs, made it highly likely, even prior to data collection, that they would find similar results from these “control” sites and from the DBB sites, and that they would thus be unable to demonstrate whether the DBBs were more or less distracting to their participant drivers.

As expected, the researchers found quite similar driver performance and behaviors at these two types of sites, and these performance and behavior variables differed, in the critical area of eyeglance behaviors, from the two other types of sites studied (conventional billboards and baseline sites). The clear lesson, had the researchers chosen

to accept it, was that sites containing digital imagery with changing messages (whether on- or off-premise) were more demanding and more distracting than sites devoid of such sign characteristics. Yet, the authors took this obvious conclusion and twisted it in favor of their biases by reporting that DBBs were “safety neutral” because the adverse, and potentially unsafe, driver behaviors that they observed at such sites were generally similar to the behaviors that they observed at the comparison sites. This conclusion, accompanied by the authors’ contrived term “safety neutral” seems to reflect obvious bias, and flies in the face of efforts to promote highway safety by reducing, not increasing, the number of irrelevant, distracting, roadside stimuli.

Correlation and causation.

Throughout the report, the authors confuse the terms *correlation* and *causation*. Although it is clear that they understand the important differences between these two types of statistical analysis, they often slip into the erroneous mode of citing a study whose sole purpose was to measure correlation, and criticize that study because it failed to prove causation. These fallacious comments are in line with a long tradition in the outdoor advertising industry of suggesting that there can be no relationship between billboards and traffic safety because billboards have never been shown to *cause* accidents.

Nighttime data collection.

Digital billboards are of particular concern to traffic safety experts at night, due to their ability to achieve high brightness and contrast levels, their high resolution imagery, and their visually compelling message changes, all of which can act to capture the attention of the driver at the expense of other targets in the visual scene (such as official signs and signals, pavement markings, and other vehicles). Because of the recent emphasis on the tails of the distribution in research studies and the long-standing practice of road safety considerations for the 85th (or higher) percentile, it is increasingly recommended to researchers that they examine the “high risk” or “worst case” scenarios in their studies, particularly when time, budget, or logistical constraints limit the number of participants. We question, therefore, why Lee and her colleagues chose to perform only a limited night-time study, one which included, *by design*, too few participants to enable the researchers to analyze their data statistically. This decision is particularly troubling because, as might have been hypothesized, the researchers found indications of greater distraction by digital billboards vs. control sites at night. In fact, unlike the daytime study, they found that all four of their eyegance measures showed that DBBs and comparison sites were more distracting and attention-getting than the conventional billboard and baseline sites (pp. 64-66), and, they believed, at least some of these findings “would show statistical significance” in a larger study (p. 64).