Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques

William J. Horrey and **Christopher D. Wickens,** University of Illinois at Urbana-Champaign, Savoy, Illinois

Objective: The performance costs associated with cell phone use while driving were assessed meta-analytically using standardized measures of effect size along five dimensions. **Background:** There have been many studies on the impact of cell phone use on driving, showing some mixed findings. **Methods:** Twenty-three studies (contributing 47 analysis entries) met the appropriate conditions for the meta-analysis. The statistical results from each of these studies were converted into effect sizes and combined in the meta-analysis. **Results:** Overall, there were clear costs to driving performance when drivers were engaged in cell phone conversations. However, subsequent analyses indicated that these costs were borne primarily by reaction time tasks, with far smaller costs associated with tracking (lane-keeping) performance. Hands-free and handheld phones revealed similar patterns of results for both measures of performance. Conversation tasks tended to show greater costs than did information-processing tasks (e.g., word games). There was a similar pattern of results for passenger and remote (cell phone) conversations. Finally, there were some small differences between simulator and field studies, though both exhibited costs in performance for cell phone use. **Conclusion:** We suggest that (a) there are significant costs to driver reactions to external hazards or events associated with cell phone use, (b) hands-free cell phones do not eliminate or substantially reduce these costs, and (c) different research methodologies or performance measures may underestimate these costs. **Application:** Potential applications of this research include the assessment of performance costs attributable to different types of cell phones, cell phone conversations, experimental measures, or methodologies.

INTRODUCTION

Over the past 10 years, there has been accelerated use of cellular or mobile phones. In the United States alone, there are an estimated 167 million cell phone subscribers (Cellular Telecommunications Industry Association, 2004). This increase in usage has been coupled with an acceleration in the number of studies that have attempted to document the negative safety implications of their use while driving (e.g., Alm & Nilsson, 1994; McKnight & McKnight, 1993; Strayer & Johnston, 2001). Such studies have been carried out with a variety of methodologies: epidemiological (e.g., Redelmeier & Tibshirani, 1997; Young, 2001), simulator based (e.g., Strayer, Drews, & Johnston, 2003), and those conducted in basic laboratory settings, emphasizing low-fidelity simulations of the driver's information-processing demands (e.g., Strayer & Johnston, 2001). Finally, some legislative efforts aimed at banning or restricting the use of cell phones have been seen, based in part upon the research just cited.

At present, the impact of cell phone use on driving performance is complicated by conflicting findings. For example, Brookhuis, de Vries, and de Waard (1991) showed that in some conditions drivers exhibited *decreased* lane deviations while engaged in a cell phone task, whereas others have shown the opposite effect for tracking performance (e.g., Strayer & Johnston, 2001). Similarly, numerous studies have shown that people exhibit increases in response time on a variety of perceptual and cognitive tasks while they are engaged in

Address correspondence to William J. Horrey, Liberty Mutual Research Institute for Safety, 71 Frankland Rd., Hopkinton, MA 01748; william.horrey@libertymutual.com. ***HUMAN FACTORS,*** Vol. 48, No. 1, Spring 2006, pp. 196–205.

cell phone conversations (e.g., Alm & Nilsson, 1995; Consiglio, Driscoll, Witte, & Berg, 2003), whereas others have not shown any significant increases in response latency (e.g., Rakauskas, Gugerty, & Ward, 2004). Given these inconsistencies, we decided to integrate the collective wisdom of many of the empirically valid studies, through the technique of *meta-analysis* (Rosenthal, 1991).

Meta-analysis is a technique whereby the results of a number of studies addressing a hypothesis can be combined to provide a single estimate of the reliability and magnitude of the effect supporting (or refuting) that hypothesis. In the current case, the effect we measure is the degradation in driving performance when using a cell phone, compared with performance on a single-task (driving) control condition. An advantage of meta-analysis is that it allows one to combine data from separate experiments that may have differences in sample characteristics, experimental protocol, and dependent measures. In addition, meta-analysis allows the testing of more restricted hypotheses addressed by a subset of research reports. For example, one might wish to ask if the effect of cell phone use was the same on lane-keeping performance as on hazard response, or whether the effect was the same for both hands-free and handheld cellular phones. We refer to these as *moderating variables* – those that can modify the main effect of cell phone use.

In carrying out our meta-analysis, we identified five moderating variables that (a) we hypothesized might influence the costs of cell phone use on driving performance and (b) could be important in modeling the effect on the driver's attentional system. We acknowledge that this list is by no means exhaustive; however, it highlights a number of variables that have been documented in the literature and represents those variables that were of greatest interest to us.

1. Measures of driving performance. Prior research has suggested that continuous perceptual-motor measures of lane keeping depend on separate attentional resources and are differentially affected by concurrent task demand as compared with discrete measures of hazard response (Horrey & Wickens, 2004a). As such, we sought studies that have included measures of (a) lane position or tracking ability (e.g., mean error, variability) or (b) discrete response time to a target that is not related to the cell phone task (e.g., roadside pedestrian). Other measures, such as speed control, although relevant, are beyond the scope of our investigation.

2. Handheld versus hands-free phones. Some have argued that the primary source of interference for drivers using a cell phone is between the manual dexterity necessary to hold the phone and the manual steering activity – a source of interference that would be evident only in handheld phones. Others have argued that the primary source of interference is cognitive, related to the information-processing activities of listening and selecting vocal responses (e.g., Strayer & Johnston, 2001). Therefore, we consider the degradation of driving performance separately in studies using handheld phones versus those using hands-free phones.

3. Conversation versus information processing. A number of studies have employed realistic conversation tasks (e.g., Strayer & Johnston, 2001), which may "engage" the driver to varying degrees, depending on their level of interest. In contrast, other researchers employ tasks that simulate the demands of conversation on different aspects of information processing (e.g., Alm & Nilsson, 1994). We have therefore contrasted studies that have used the two types of tasks to simulate cell phone usage.

4. In-vehicle versus remote conversation. It has been argued by some that a major source of distraction in cell phone use is the inability of the nondriving speaker to be aware of the momentary demands on the driver and, hence, the inability to modulate conversation when the driving demands increase. This inability does not characterize the passenger in the vehicle, who may be carrying out the same conversation (e.g., Gugerty, Rando, Rakauskas, Brooks, & Olson, 2003). Hence we contrasted studies that used the two classes of conversation.

5. Simulator versus field studies. Given the frequent citation of cell phone costs in the context of simulator studies, we were interested in whether these findings were consistent with real-world in-vehicle field trials, which are more characteristic of the environments from which epidemiological accident data are drawn (e.g., Redelmeier & Tibshirani, 1997; Young, 2001). We note that there are differences in fidelity across many simulator studies. As such, the related findings should be interpreted with caution.

METHODS

Studies

As shown in Table 1, 23 experiments contributed to the current meta-analysis. Some studies had multiple conditions, which allowed us to increase the overall number (to $N = 47$). For example, some studies examined both handheld and hands-free cell phones – these conditions were included in the meta-analysis as separate entries. (Because some studies contribute multiple entries in certain aspects of the meta-analysis, the treatment of such nonindependent results as independent may draw some criticism. This may be the case for significance testing, however, not in dealing with estimates of effect size, as we are dealing with here. See Rosenthal, 1991, for a detailed discussion). All of the experiments were gathered from online databases (e.g., PSYC INFO, USDOT-TRIS Online), online resources (National Highway Traffic Safety Administration, 2000), or through backward referencing and included journal articles, conference proceedings, and technical reports. For database searches, we used key words such as *cell, cellular, mobile, phone, driving,* and *driver.* We did not set any restrictions on the year of publication. We were also able to use the reference sections in many papers to obtain further citations, especially in the case of review papers (e.g., Goodman et al., 1997).

From these sources, we amassed in excess of 50 papers; however, many were not included in the analyses for various reasons. For example, the paper needed to exhibit some features of relevance to the moderator variables described earlier. Also, meta-analyses require that all findings be presented in terms of single degree-of-freedom (*df*) main effects, most especially in cases where the raw data are unavailable. For example, if authors reported a significant omnibus analysis of variance (ANOVA) with more than two levels of an independent variable (i.e., $df > 1$) but failed to report any follow-up contrasts or pairwise comparisons (all of which are $df = 1$), readers could not be sure where the differences lie. Hence, we cannot incorporate these types of results in a meta-analysis. For these reasons, we urge authors to report a full account of their statistical results, including obtained F (or t) and p values. Other details, such as standard deviations or mean square error and effect sizes, would be desirable as well. Furthermore, we required that all the studies include a common comparison of cell phone use while driving against single-task baseline conditions (e.g., driving alone) and that the results be presented in terms of tracking performance, vehicle control, or response time to a non–cell phone event (e.g., roadside hazard). As such, studies that did not meet these criteria were dropped from this analysis (though several of these are summarized in Horrey & Wickens, 2004b). Also, because we were interested in the conversational aspects of cell phone use, we did not include studies that examined driving performance while the driver was dialing or manipulating the cell phone in some way.

For those studies that met the criteria described, we coded them along each of the five moderator variables. Specifically, we indicated whether the study measured tracking performance or response time (or both). For tracking, performance was typically assessed by absolute error or as an index of root mean square error (i.e., variability in tracking performance). Response time tasks, in contrast, involved a speeded response to some stimuli, whether a road hazard or an artificial stimulus presented in the traffic environment.

Second, we coded whether a particular study employed a hands-free or a handheld cell phone for the conversation task. Third, we categorized the type of phone task that participants performed – that is, whether the task involved a conversation (typically characterized by a free discussion of topics of interest or autobiographical information) or an information-processing task (e.g., mental arithmetic, word generation games, or the like). Next, we specified those studies that utilized remote conversations (i.e., over a cell phone or speaker) and those that incorporated in-vehicle (i.e., passenger) conversations and, finally, those that used simulators (whether low or high fidelity) versus those employing actual field trials.

Meta-Analysis

For the meta-analysis, we converted statistical results into effect sizes and combined these values (Rosenthal, 1991; Rosenthal & Dimatteo, 2001; Rosenthal, Rosnow, & Rubin, 2000). Effect sizes are advantageous because they focus on how large a particular effect is (as opposed to whether or not it differs from zero) and, when coupled with confidence intervals, they offer estimates for the upper

TABLE 1: List of Studies Contributing to the Meta-Analysis and Some of Their Attributes

Study	Phone Type		Task Type		Location		Measure		Study Type	
Alm & Nilsson (1994)	Hands-free		IP		Remote		Tracking		Sim	
Alm & Nilsson (1995)	Hands-free		IP		Remote		Tracking	RT	Sim	
Brookhuis et al. (1991)	Hands-free	Handheld	IP		Remote		Tracking	RT		Field
Consiglio et al. (2003)	Hands-free	Handheld		Conversation	Remote	In-vehicle		RT	Sim	
Cooper et al. (2003)	Hands-free		IP		Remote			RT		Field
Green, Hoekstra, & Williams (1993)		Handheld	IP	Conversation	Remote		Tracking			Field
Gugerty et al. (2003)	Hands-free		IP		Remote	In-vehicle		RT	Sim	
Hanowski, Kantowitz, & Tijerina (1995)		Handheld	IP	Conversation	Remote		Tracking		Sim	
Horswill & McKenna (1999)	Hands-free		IP		Remote			RT	Sim	
Irwin, Fitzgerald, & Berg (2000)		Handheld		Conversation	Remote			RT	Sim	
Kantowitz, Hanowski, & Tijerina (1996)		Handheld	IP		Remote			RT	Sim	
Laberge, Scialfa, White, & Caird (2004)	Hands-free		IP		Remote	In-vehicle	Tracking	RT	Sim	
Lamble, Kauranen, Laakso, & Summala (1999)	Hands-free	Handheld	IP			In-vehicle		RT		Field
Nilsson & Alm (1991)	Hands-free		IP		Remote		Tracking	RT	Sim	
Parkes & Hooijmeijer (2001)	Hands-free		IP		Remote		Tracking	RT	Sim	
Patten, Kircher, Östlund, & Nilsson (2004)	Hands-free	Handheld	IP		Remote			RT		Field
Rakauskas et al. (2004)	Hands-free			Conversation	Remote		Tracking	RT	Sim	
Spence & Read (2003)	Hands-free		IP		Remote		Tracking		Sim	
Strayer & Drews (2003)	Hands-free			Conversation	Remote			RT	Sim	
Strayer et al. (2003)	Hands-free		IP	Conversation	Remote			RT	Sim	
Strayer & Johnston (2001)	Hands-free	Handheld	IP	Conversation	Remote		Tracking	RT	Sim	
Tijerina, Kiger, Rockwell, & Tornow (1996)		Handheld	IP	Conversation	Remote		Tracking			Field
Waugh et al. (2000)		Handheld	IP		Remote	In-vehicle	Tracking			Field

Note. IP = information processing, Remote = conversation over cell phone, In-vehicle = conversation with passenger, RT = response time task.

and lower limits of the true effect size in the population.

We used reported test statistics to calculate the effect size for each study, based on the product moment correlation (*r*). In general, we employ *r* as a measure of effect size because it has a number of advantages over other measures (e.g., Cohen's *d*, Hedges's *g;* see Rosenthal, 1991, for details). The effect size can be calculated from *t* statistics or *F* statistics (with $df = 1$), as shown by Equation 1.

$$r_{\mathrm{ES}} = \sqrt{\frac{t^2}{t^2 + df}} = \sqrt{\frac{F}{F + df_{\mathrm{error}}}} \qquad (1)$$

In cases where the authors provided only *p* values, we converted these to their associated *z* score and used Equation 2 (from Rosenthal & Dimatteo, 2001),

$$r_{\mathrm{ES}} = \frac{Z}{\sqrt{N}}, \qquad (2)$$

in which *Z* is the standard normal deviate for the associated *p* value and *N* is the sample size. Whenever authors provided a range for the *p* value, we used the following values for *Z:* for $p < .05$, $Z = 1.645$; for $p < .01$, $Z = 2.362$; and for $p < .001$, $Z = 3.090$. In situations where the authors indicated there was no significant difference between the conditions of interest but failed to provide any statistical details (e.g., *F*, *t*, or *p* values), a conservative effect size of zero was assumed for the meta-analysis (see Rosenthal & Dimatteo, 2001).

Following the calculation of the effect size (r_{ES}) for each study, we coded the findings to denote costs or gains in driving performance with the concurrent cell phone task. In the first case, effect sizes were assigned positive values (i.e., predicted costs to performance). In cases where the pattern of results was opposite (i.e., gains in performance in dual-task situations), effect sizes were assigned negative values.

In order to combine the effect sizes from multiple studies, we first normalize our effect sizes by converting the r_{ES} scores to *z* scores using Fisher's *r*-to-*z* transformation (typically offered in tabular form in statistical textbooks; see Rosenthal & Dimatteo, 2001, for details). Next, the unweighted and weighted means of these transformed scores were calculated, in which the latter was weighed by the *df* of the study. The weighted and unweighted *z*-transformed means were then converted back into *r* values and reported (e.g., see Table 2). Finally, for both the unweighted and weighted means, we estimated the 95% confidence interval to determine whether the combined effect sizes differed significantly from zero (i.e., did not include zero in the interval), following Equation 3,

$$CI_{95\%} = \overline{Z}_r \pm t_{(.05)} S / \sqrt{k}, \qquad (3)$$

where $\overline{Z}_r$ is the mean of the transformed r_{ES} values, $t_{(.05)}$ is the appropriate *t* value for the .05 probability level, *S* is the standard deviation of the transformed r_{ES} values, and *k* is the number of studies included in the sample.

Although the weighted and unweighted mean results were highly correlated in our analyses (Pearson's $r = 0.94$), thus providing an equivalent picture, the former has the advantage of amplifying the greater impact of more reliable (higher *N*) studies. We do note some exceptions in the Results section.

After the combined effect sizes were obtained, we conducted a test of heterogeneity (Rosenthal, 1991) to determine whether or not the values contributed to the analysis are consistent (i.e., homogeneous) with one another or inconsistent (i.e., heterogeneous). Heterogeneous results may indicate the presence of moderator variables, such as those described previously, that warrant further investigation.

RESULTS

The results from the meta-analysis are shown in Table 2, including the unweighted and weighted combined effect sizes and corresponding tests of heterogeneity. When we examined all of the studies collectively (without factoring in moderator variables), we found that there is a significant cost of cell phone use on driving performance (see Row 1). We proceeded to break down the set of studies following from the moderator variables outlined previously. (When examining the interactive effects of multiple moderator variables, we did not analyze all possible combinations of variables; instead we focused on those combinations that were of greatest interest to us.)

As shown in Row 2, we broke down the overall set of studies into those that examined driving performance in terms of response time to a road

TABLE 2: Summary for the Meta-Analysis and Tests for Heterogeneity, Including Examination of Potential Moderating Variables

		Combined Effect Size (*r*)			
	No. of Entries	Unweighted (95% CI)	Weighted (95% CI)	Test of Heterogeneity	*p* Value
1. Overall	47	.39 (.27, .49)	.43 (.33, .52)	200.7	<.001
2. Measure					
a. RT	28	.53 (.41, .64)	.50 (.36, .60)	81.7	<.001
b. Tracking	19	*.13* (–.04, .29)	.23 (.16, .29)	49.9	<.001
3. Phone type					
a. Hands-free					
i. Overall	28	.41 (.28, .53)	.44 (.33, .54)	86.2	<.001
ii. RT	20	.51 (.37, .62)	.49 (.36, .61)	44.3	.001
iii. Tracking	8	*.13* (–.17, .43)	*0.25* (.0, .48)	26.7	<.001
b. Handheld					
i. Overall	19	.35 (.13, .53)	.40 (.18, .58)	82.5	<.001
ii. RT	8	.60 (.27, .80)	.51 (.13, .70)	37.1	<.001
iii. Tracking	11	*0.13* (–.12, .36)	*0.20* (–.04, .41)	22.5	.01
4. Task Type					
a. Conversation (RT)	10	.66 (.47, .79)	.66 (.49, .78)	21.1	.01
b. Info process (RT)	18	.45 (.29, .59)	.42 (.25, .57)	53.9	<.001
5. Location					
a. Remote (RT)	23	.51 (.36, .63)	0.48 (.33, .60)	78.0	<.001
b. In-vehicle (RT)	5	.63 (.36, .81)	.58 (.32, .76)	3.3	.51
6. Study type					
a. Simulator (RT)	19	.51 (.37, .64)	.42 (.27, .55)	42.2	.001
b. Field test (RT)	9	.57 (.29, .76)	.66 (.46, .80)	39.3	<.001

Note. Results in italics indicate nonsignificant findings (i.e., confidence interval includes zero).

event or discrete stimuli or in terms of lane keeping or tracking performance. For those studies examining response time (RT; Row 2a), the costs to driving performance were still significant (with a large effect size, which translated to approximately 130 ms in costs). However, for those studies that examined decrements in tracking performance or lane keeping (Row 2b), the effect size was substantially smaller and, for the unweighted mean, nonsignificantly different from zero.

Given this pattern of results, we elected to examine the other moderator variables (task type, location, study type) solely on the basis of their impact on RT tasks to determine the differential effects of these variables on this significant effect. However, we do include the lane-keeping (tracking) variable in the analysis of phone type to determine whether the manual aspects of holding a phone interfere more with tracking performance than does a hands-free phone.

An examination of the overall impact of phone type on overall driving performance (hands-free phone, Row 3a-i; handheld phone, Row 3b-i) showed significant costs. However, these costs were moderated by the measure of driving performance, converging with those findings shown in Row 2. For measures of reaction time, both hands-free (Row 3a-ii) and handheld (Row 3b-ii) phones showed significant costs. In contrast, tracking performance showed smaller and nonsignificant effect sizes, and there was no difference between hands-free and handheld phone types (Rows 3a-iii, 3b-iii).

We broke down the overall set of studies by those employing conversation tasks and those using information-processing tasks in conjunction with measures of RT (Row 4). In general, when a conversation task is employed there are higher costs to driving performance than for information-processing tasks. As shown in Table 2, the weighted combined effect size for the information-processing task (r = .42; Row 4b) lies outside of the 95% confidence interval for the conversation task (.49 to .78; Row 4a), and vice versa. However, both costs are still significant.

In the comparison shown in Row 5, we note that the distinction of in-vehicle or remote conversations does not appear to have a differential impact on the costs in RT performance (again, using the boundaries of the 95% confidence intervals). That is, the costs associated with a phone conversation versus a passenger conversation are roughly equivalent. Finally, the costs observed in simulator and field studies were similar for the unweighted means (Row 6); however, the weighted means showed a larger effect size for field studies. We speculate that, because of the lower number of studies contributing to the analysis, those studies with stronger weights (i.e., higher *N*s) may have driven this effect. Again, in both cases (i.e., field studies and simulators) the effects are significant.

A common criticism of meta-analyses is that there is a bias in the reporting of significant findings, such that studies that fail to show significant results are less likely to published and, as a result, less likely to be represented in a meta-analysis. (Such a criticism can also be applied to the typical literature review as well.) Although there is no definitive solution to this "file drawer problem" (Rosenthal, 1991), one can calculate the number of studies, showing a null result, that would be required before the overall p value reaches a nonsignificant ($p = .05$) level. (We do note however, that this technique has been criticized by some: e.g., Scargle, 2000). In Equation 4 (from Rosenthal, 1991),

$$X = \frac{K[K\overline{Z}^2 - 2.706]}{2.706} \tag{4}$$

X is the number of required studies (showing null effects), K is the current number of studies, and $\overline{Z}$ is the mean Z values obtained for the K studies. For those results shown in Row 2a of Table 2, following Equation 4, we would require approximately 1108 additional studies before the combined effect size would approach zero.

DISCUSSION AND CONCLUSIONS

From the current meta-analysis, we note several important findings. First, there are definite costs associated with cell phone use while driving; however, these costs appear to be manifested primarily in measures of response time to critical road hazards or stimuli. In contrast, the costs associated with lane-keeping or tracking performance are much smaller (and, for the unweighted means, nonsignificant). Horrey and Wickens (2004a) suggested that these tasks (e.g., lane keeping and hazard response) depend on separate resources (ambient and focal vision, respectively) and therefore may be differentially impacted by cell phone conversations. Furthermore, lane keeping may be a skill that is relatively automatic, requiring fewer overall resources to maintain performance (in addition to being supported by ambient vision). In contrast, responding to road events or stimuli may be less automated because drivers must not only detect critical objects but also select an appropriate course of action or response to them. The interference derived from cell phone conversations may manifest itself at these stages of processing (i.e., decision making, response selection). We note that although the magnitude of the reaction time effect was relatively small (an average delay of 130 ms), this represents a mean value, around which there was considerable variance. Accidents are often caused by "worst-case" performers under "worst-case" circumstances (Wickens, 2001), at the tail end of the distribution where reaction time delay can be expected to be considerably longer. Furthermore, we believe that delays in the response to road hazards are at least as compromising to highway safety as are increases in lane-keeping error.

Second, the meta-analysis suggests that costs in driving performance are equivalent across hands-free and handheld phones, suggesting that the larger part of these costs is attributable to the cognitive aspects of conversation and not to the manual aspects of holding the phone. We note, however, that this does not discount the possibility that the costs associated with handheld phones could be exacerbated in situations requiring significant amounts of manual steering inputs (e.g., intersection turns) and the manual aspects of keyboard entry ("dialing"), which are not considered here.

Third, conversation tasks, in general, showed greater costs in driving performance than did information-processing tasks. This may be attributable to the greater "engagement" associated with actual conversations. Although information-processing tasks involve perceptual resources and working memory, they do not share the same degree of engagement. We speculate that the costs of engagement may be more pronounced when the conversation is intense, though there was insufficient data along this dimension for the purposes of meta-analysis. Some researchers have defined an intense conversation as one that is difficult or challenging (e.g., problem-solving exercises; McKnight & McKnight, 1993). However, we differentiate here between phone tasks that may be considered difficult and those that may be emotionally loaded or heated (as we consider *intense* to be defined). Importantly however, information-processing tasks do produce substantial influence on performance and, so, should be able to effectively simulate many aspects of the demands of cell phone usage for research purposes.

From our analyses, in-vehicle (passenger) conversations were just as costly to driving performance as were remote (cell phone) conversations. This suggests that passengers, at least in those studies explored here, did not moderate their conversation in such a way as to alleviate the costs (as compared with remote conversers). These results must be interpreted with caution, however, given that relatively few studies directly examined the impact of passenger conversations.

Finally, in comparing the costs of cell phone use to driving performance in simulator studies versus field studies, we observed a trend for the costs to be greater in the field studies, at least when the weighted means analysis was used (r_{ES} = .66, field; .42, simulator). We do note here that both effect sizes are significant and in the same direction. As for the discussion of conversation type, we must insert some cautionary notes. First, the simulator studies included in our analyses represented a wide range of fidelities. In some cases, high-fidelity simulators were employed, whereas in other cases, desktop PCs were used. We have not, in our analysis, differentiated among the many potential levels of simulators. Second (and related to the first issue), we do not make any claims regarding the validity of simulators in the study of driving performance and behavior. We are merely noting that costs attributable to cell phone use can be observed in both settings.

Implications

The implications of the current research for the human factors of highway safety appear to be threefold: (a) We have demonstrated a significant cost (and estimated magnitude) to cell phone use on driver reactions to external hazards. Whether this time delay of 0.13 s, on average, warrants legislation or other actions should be driven by further safety analysis and modeling based on considerations of vehicle speed and hazard exposure. Such modeling should consider that the estimated mean delay will underestimate the delay of slower responding drivers and of those responding under worst-case situations (Wickens, 2001). (b) From a design standpoint, we have shown that hands-free cell phones do not eliminate, and may not substantially reduce, the cell phone interference costs. (c) Methodologically, we suggest that such costs may not emerge from measures of lane keeping and that their magnitude may be underestimated in simulator studies and with those using "simulated conversation demands" based on information-processing tasks, rather than on real conversation.

Although the technique of meta-analysis lacks the experimental control from a single experiment, it does afford investigators a useful tool for integrating the collective contributions of multiple studies. As noted by Hall and Rosenthal (1995), there is no single way to perform a meta-analysis. In our current analysis, we elected to focus on the impact of cell phone use on lane keeping (tracking) and event response time, as these relate to the driving tasks of vehicle control and hazard awareness. We do note that these two measures represent logical precursors to the less frequently observed loss-of-control and collision events, respectively. Other important measures, such as workload and collision frequency, may also provide the foundation for other meta-analyses. Furthermore, we focused on the conversational aspect of cell phone use. Others may tap the interference derived from the physical manipulation of a cell phone (e.g., dialing) on steering behavior.

ACKNOWLEDGMENTS

This research was sponsored by a grant from General Motors (GM TCS16231 WIC). John Lenneman was the scientific/technical monitor. Special thanks to Arthur Kramer, Charles Scialfa, Jeff Caird, Yueng-hsiang Huang, William Shaw, Barry Kantowitz, and three anonymous reviewers for their insightful comments and suggestions. The opinions and views expressed in this paper are the authors' own and may not reflect those of the sponsors.

REFERENCES

Alm, H., & Nilsson, L. (1994). Changes in driver behaviour as a function of handsfree mobile phones – A simulator study. *Accident Analysis and Prevention, 26,* 441–451.

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Analysis and Prevention, 27,* 707–715.

Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23,* 309–316.

Cellular Telecommunications Industry Association. (2004). *CTIA's world of wireless communications.* Retrieved August 10, 2004, from http://www.wow-com.com/. [Data indicating most recent numbers of cellular subscribers in the U.S.]

Consiglio, W., Driscoll, P., Witte, M., & Berg, W. P. (2003). Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. *Accident Analysis and Prevention, 35,* 495–500.

Cooper, P. J., Zheng, Y., Richard, C., Vavrik, J., Heinrichs, B., & Sigmund, G. (2003). The impact of hands-free message reception/response on driving task performance. *Accident Analysis and Prevention, 35,* 23–35.

Goodman, M., Bents, F. D., Tijerina, L., Wierwille, W., Lerner, N., & Benel, D. (1997). *An investigation of the safety implications of wireless communications in vehicles* (Rep. No. DOT HS 808-635). Washington, DC: National Highway Traffic Safety Administration.

Green, P., Hoekstra, E., & Williams, M. (1993). *Further on-the-road tests of driver interfaces: Examination of a route guidance system and a car phone* (Rep. No. UMTRI-93-35). Ann Arbor, MI: University of Michigan, Transportation Research Institute.

Gugerty, L., Rando, C., Rakauskas, M., Brooks, J., & Olson, H. (2003). Differences in remote versus in-person communications while performing a driving task. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 1855–1859). Santa Monica, CA: Human Factors and Ergonomics Society.

Hall, J. A., & Rosenthal, R. (1995). Interpreting and evaluating meta-analysis. *Evaluation and the Health Professions, 18,* 393–407.

Hanowski, R., Kantowitz, B., & Tijerina, L. (1995). *Heavy vehicle driver workload assessment Task 7b: In-cab text message system and cellular phone use by heavy vehicle drivers in a part-task driving simulator* (Rep. No. DOT HS 808 467). Washington, DC: National Highway Traffic Safety Administration.

Horrey, W. J., & Wickens, C. D. (2004a). Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors, 46,* 611–624.

Horrey, W. J., & Wickens, C. D. (2004b). *The impact of cell phone conversations on driving: A meta-analytic approach* (Tech. Rep. AHFD-04-2/GM-04-1). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Horswill, M. S., & McKenna, F. P. (1999). The effect of interference on dynamic risk-taking judgments. *British Journal of Psychology, 90,* 189–199.

Irwin, M., Fitzgerald, C., & Berg, W. P. (2000). Effect of the intensity of wireless telephone conversations on reaction time in a braking response. *Perceptual and Motor Skills, 90,* 1130–1134.

Kantowitz, B. H., Hanowski, R. J., & Tijerina, L. (1996). Simulator evaluation of heavy-vehicle driver workload: II. Complex tasks. In *Proceedings of the Human Factors Society 40th Annual Meeting* (pp. 877–881). Santa Monica, CA: Human Factors and Ergonomics Society.

Laberge, J., Scialfa, C., White, C., & Caird, J. (2004). The effect of passenger and cellular phone conversations on driver distraction. In *Proceedings of the 83rd Annual Meeting of the Transportation Research Board* (n.p.). Washington, DC: National Academy of Sciences.

Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: Safety implications for using mobile (cellular) telephones while driving. *Accident Analysis and Prevention, 31,* 617–623.

McKnight, A. J., & McKnight, A. S. (1993). The effect of cellular phone use upon driver attention. *Accident Analysis and Prevention, 25,* 259–265.

National Highway Traffic Safety Administration (2000). *Internet forum on the safety implications of driver distraction when using in-vehicle technologies.* Retrieved January 16, 2006, from http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/welcome.htm

Nilsson, L., & Alm, H. (1991). Elderly people and mobile telephone use: Effects on driver behavior. In *Proceedings of the Strategic Highway Research Program and Traffic Safety on Two Continents* (pp. 117–133). Gothenburg, Sweden: Swedish Road and Traffic Research Institute (VTI).

Parkes, A. M., & Hooijmeijer, V. (2001). Driver situation awareness and carphone use. In *Proceedings of the 1st Human-Centered Transportation Simulation Conference* (n.p.). Iowa City: University of Iowa.

Patten, C. J. D., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: Cognitive workload and attention resource allocation. *Accident Analysis and Prevention, 36,* 341–350.

Rakauskas, M., Gugerty, L., & Ward, N. J. (2004). Effects of cell phone conversations on driving performance with naturalistic conversations. *Journal of Safety Research, 35,* 435–464.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine, 336,* 453–458.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Rosenthal, R., & Dimatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59–82.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research.* New York: Cambridge University Press.

Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration, 14*, 91–106.

Spence, C., & Read, L. (2003). Speech shadowing while driving: On the difficulty of splitting attention between eye and ear. *Psychological Science, 14,* 251–256.

Strayer, D. L., & Drews, F. A. (2003). Effects of cell phone conversations on younger and older drivers. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 1860–1864). Santa Monica, CA: Human Factors and Ergonomics Society.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9,* 23–32.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. *Psychological Science, 12,* 462–466.

Tijerina, L., Kiger, S., Rockwell, T., & Tornow, C. (1996). *Heavy vehicle driver workload assessment Task 7a: In-cab text message system and cellular phone use by heavy vehicle drivers in a part-task driving simulator* (Rep. No. DOT HS 808 467). Washington, DC: National Highway Traffic Safety Administration.

Waugh, J. D., Glumm, M. M., Kilduff, P. W., Tauson, R. A., Smyth, C. C., & Pillalamarri, R. S. (2000). Cognitive workload while driving and talking on a cellular phone or to a passenger. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 6.276–6.279). Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. D. (2001). Attention to safety and the psychology of surprise. *Proceedings of the 2001 International Symposium on Aviation Psychology* (n.p.). Columbus, OH: The Ohio State University.

Young, R. A. (2001). Association between embedded cellular phone calls and vehicle crashes involving airbag deployment. In *Proceedings of the 1st International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Aspen, CO,* (pp. 390–400).

William J. Horrey is a research scientist at the Liberty Mutual Research Institute for Safety, Hopkinton, MA. He received his Ph.D. in psychology at the University of Illinois in 2005.

Christopher D. Wickens is a professor of psychology at the Institute of Aviation, Human Factors Division, University of Illinois. He received his Ph.D. in experimental psychology at the University of Michigan in 1974.

Date received: March 26, 2004
Date accepted: October 19, 2004