Accident Analysis and Prevention 129 (2019) 108–118

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Mobile phone conversation distraction: Understanding differences in impact between simulator and naturalistic driving studies

Kasun P. Wijayaratna[a,d,*], Mitchell L. Cunningham[b,c], Michael A. Regan[a], Sisi Jian[a], Sai Chand[a], Vinayak V. Dixit[a]

[a] Research Centre for Integrated Transport Innovation (rCITI), School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW, 2052 Australia

[b] School of Psychology, The University of Sydney, NSW 2006, Australia

[c] Australian Road Research Board (ARRB), 2-14 Mountain Street, Ultimo, NSW, 2007 Australia

[d] School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW, 2007, Australia

ARTICLE INFO

Keywords:
Mobile phone
Conversation
Driver distraction
Simulator studies
Naturalistic driving studies
Odds ratio

ABSTRACT

A current issue within the driver distraction community centres around different findings regarding the impact of mobile phone conversation on driving found in driving simulators versus instrumented vehicles employed in real-world naturalistic driving studies (NDSs). This paper compares and contrasts the two types of studies and aims to provide reasons for the differences in findings that have been documented. A comprehensive review of literature and consultations with human factors experts highlighted that simulator studies tend to show degradation in driving performance, suggestive of increased crash risk as a result of mobile phone conversation. Whilst NDSs, at times, present data suggesting that mobile phone conversation distraction actually reduces crash risk. This study identifies that these differences may be attributed to behavioural hypotheses associated with driver self-regulation, arousal from cognitive loading, task displacement and gaze concentration – all of which need to be explicitly tested in future driving studies.

Metric estimation and application was also revealed to be polarising results and the subsequent assessment of the crash risk. A common metric applied in this domain is the 'Odds Ratio', particularly prevalent in NDSs. This study presents a detailed investigation into the assumptions and application of the Odds Ratio which revealed the potential for over- and under-estimation of the metric depending on the core data and sampling assumptions. Furthermore, this research presents a comparative analysis of select driving simulator studies and an NDS considering only driving behaviour data as a means to consistently compare the findings of both methodologies. The findings from this investigation implores the need for greater consistency in the application of analysis methods and metrics across both simulator and NDSs. Improvements can yield a more robust platform to systematically compare and interpret data across both approaches, ultimately leading to enhanced planning and safety regarding mobile phone use while driving.

1. Introduction

Driver distraction is a significant road safety issue. Recent estimates in the US indicate that the phenomenon contributed to approximately 9% of crashes in 2017 which led to a fatality (National Highway Traffic Saftey Administration, 2017). In New South Wales, Australia, there has been a threefold increase in distracted driver crashes that resulted in injury or fatality since 2008 (Centre for Road Safety, 2017), leading to increasing safety concerns surrounding driver distraction.

Although defined in the literature in a number of ways, driver distraction is commonly conceptualised as one mechanism of driver inattention and has been defined as *"…the diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving"* (Regan et al., 2011). As such, this diversion of attention to a competing source can result in driver inattention (Regan and Strayer, 2014).

A particularly common, and risky, source of driver distraction is mobile phone use (Beanland et al., 2013; Dingus et al., 2016). This is perhaps unsurprising given not only the increasing range of functionality (text messaging, music, navigation, social media) accessible through the devices themselves, but also considering their ability to 'pair' or connect with in-vehicle systems (e.g., through Bluetooth) over

* Corresponding author at: School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW, 2007, Australia.
E-mail address: Kasun.Wijayaratna@uts.edu.au (K.P. Wijayaratna).

https://doi.org/10.1016/j.aap.2019.04.017
Received 20 November 2018; Received in revised form 5 April 2019; Accepted 22 April 2019
Available online 28 May 2019
0001-4575/

the past decade. Driver actions involved in using such functions (e.g., hold the phone, view device and content, type/press keys to write a text message) require attention to be diverted away from the primary task of operating the vehicle and, thus, can impair activities critical for safe driving (see, for review, Cunningham et al. (2017)).

The impact of a secondary activity, such as using a mobile phone, on activities critical for safe driving, can be quantified by measuring changes in driving performance or changes in the level of risk associated with driver engagement in that activity. Historically, there have been a number of methodologies used for this purpose including: driving simulator studies, on road testing, test track experiments and naturalistic driving studies (NDSs). This paper focusses on the contrasting results obtained from driving simulator studies and naturalistic driving studies (NDSs).

Driving simulator studies are conducted in controlled laboratory environments. They offer researchers the ability to assess the impact of driver engagement in a secondary activity on driving performance by measuring changes in driving performance metrics (going from the no-distraction to the distracted condition) such as reaction time, eye tracking, lane position, speed, inter-vehicle headway, and gap acceptance.

On the other hand, NDSs record the real-world driving behaviour of participants, including their engagement in secondary activities, over a specified period of time (generally 3–12 months). These studies have the ability to monitor driving behaviour and performance, as well as record crash and near-crash events, allowing researchers to gauge the likely safety risk (e.g., crash risk) associated with driver engagement in different secondary activities.

With these two methodologies in mind, the following sections provide a brief overview of the literature that have investigated changes in driving performance and safety risk associated with driver conversations using a mobile phone.

1.1. Mobile phone conversations and driving performance

A plethora of research has been undertaken to investigate the impact of mobile phone conversations on driving performance. For example, an early meta-analysis of this research by Horrey and Wickens (2006) concluded that conversing on a mobile phone was associated with longer response times to critical road hazards or stimuli. Furthermore, the authors note that, although the magnitude of this impairment in reaction time was seemingly small (an average delay of 0.130 s), it was the considerable variance in reaction times that was most concerning. The study also found similarly delayed reaction times across both conversations using handheld and hands-free mobile phones. In contrast, costs in lane-keeping or tracking performance were much smaller (Horrey and Wickens, 2006).

A more recent meta-analysis conducted by Caird et al. (2008) mirrored the above findings of Horrey and Wickens (2006). This study found that mobile phone conversations were associated with delayed reaction times to road events. Moreover, the mean reaction time was argued to be within the range of 0.14-0.33 s, depending on the cognitive demands of the conversation, with higher demand conversations being associated with a greater delay in reaction times. In addition, like Horrey and Wickens (2006), Caird et al. (2008) found that these delays in reaction time did not differ as a function of the mobile phone type (e.g., handheld vs. hands-free).

Caird et al. (2018) reported the outcomes of a comprehensive review of studies examining the impacts of mobile phone conversations on driving performance. Their meta-analysis, building on the work of Horrey and Wickens (2006) and Caird et al. (2008), covered 93 studies which consisted of 106 experiments. They found a moderate deterioration in reaction time, stimulus detection, and an increase in collisions when drivers engaged in mobile phone conversations (either handheld or hands-free) compared to baseline driving. However, no significant impact was found on driving performance metrics such as lateral and longitudinal variability, and little evidence that drivers compensated while conversing on a mobile phone by increasing headway or reducing speed.

Cognitive distraction, such as that deriving from a phone conversation, has been shown in a number of driving simulator studies to reduce a driver's ability to respond to critical cues and traffic signals in the roadway environment (e.g., Strayer et al. (2003); Strayer and Johnston (2001)). In addition, drivers that are cognitively distracted tend to have longer visual fixations and a denser gaze concentration (i.e. drivers' eyesight is directed at forward roadway) in the centre of the forward roadway, which may result in failures to perceive important information in peripheral vision (Recarte and Nunes, 2003; Regan et al., 2011).

1.2. Mobile phone conversations and safety risk

The link between driver engagement in mobile phone conversation and safety risk (i.e., the risk of being involved in a safety-critical event, such as a crash or near-crash) is inconclusive.

Naturalistic driving studies, which analyse data from real-world instrumented vehicles over an extended period of time, are typically used to estimate the risk of a safety-critical event (SCE) occurring (e.g., crash or near-crash) when the driver is engaged in a specific secondary task. For example, a number of NDSs have found that conversing on a mobile phone is not associated with any change in safety risk, suggesting that the activity poses no more of a safety risk than just driving (Simmons et al., 2016). Some NDSs have even yielded data suggesting that conversing on a hands-free mobile phone may actually decrease crash risk (Fitch et al., 2015; Hickman et al., 2010; Olson et al., 2009). For example, Victor et al. (2015) found that driving during a conversation using any kind of mobile phone was ten times safer than driving while not engaged in a phone conversation. On the other hand, Dingus et al. (2016) found that conversing on a handheld mobile phone was associated with a 2.2 times increase in crash risk. But more recently Dingus et al. (2019) presented another analysis of NDS data that indicated neither hands free or handheld mobile phone conversations had a significant impact on crash risk, highlighting the inconsistencies in findings within these studies.

Following on from the contrasting results above, a key research question was raised within the Caird et al. (2018) paper that closely aligns with the focus of the research presented in this paper - "Are the results on mobile phone conversation from naturalistic, epidemiological, and driving simulation studies convergent or divergent?". Caird et al. (2018) indicates, from a statistical perspective, that there appears to be a general convergence of results from simulated driving studies (e.g., delayed reaction time), however the safety risk associated with mobile phone conversations remains unclear and varies considerably across NDSs. This begs the question as to why the degraded driving performance associated with this secondary activity observed in simulated driving studies may not reliably translate into increased safety risk in real-world driving studies (i.e., NDSs).

The research presented in this paper suggests that the inconsistencies of the impact of mobile phone conversation between driving simulator studies and NDSs may be attributed to psychological phenomena and the diversity in methodological practices. Thus, the focus of this paper is to postulate what psychological and/or methodological factors may be contributing to these inconsistencies. With this in mind, the remainder of this paper is structured in the following manner.

First, Section 2 presents a summary of the methodology applied in this study. Second, Section 3 presents some psychological hypotheses that may account for the discrepancies observed. Third, Section 4 discusses the potential methodological intracacies involved in measuring safety risk using data from NDSs. Then, Section 5 presents a direct comparative analysis of driving performance data from simulator studies and NDSs offering a like for like comparison between the two approaches. Finally, the paper will conclude with some remarks that

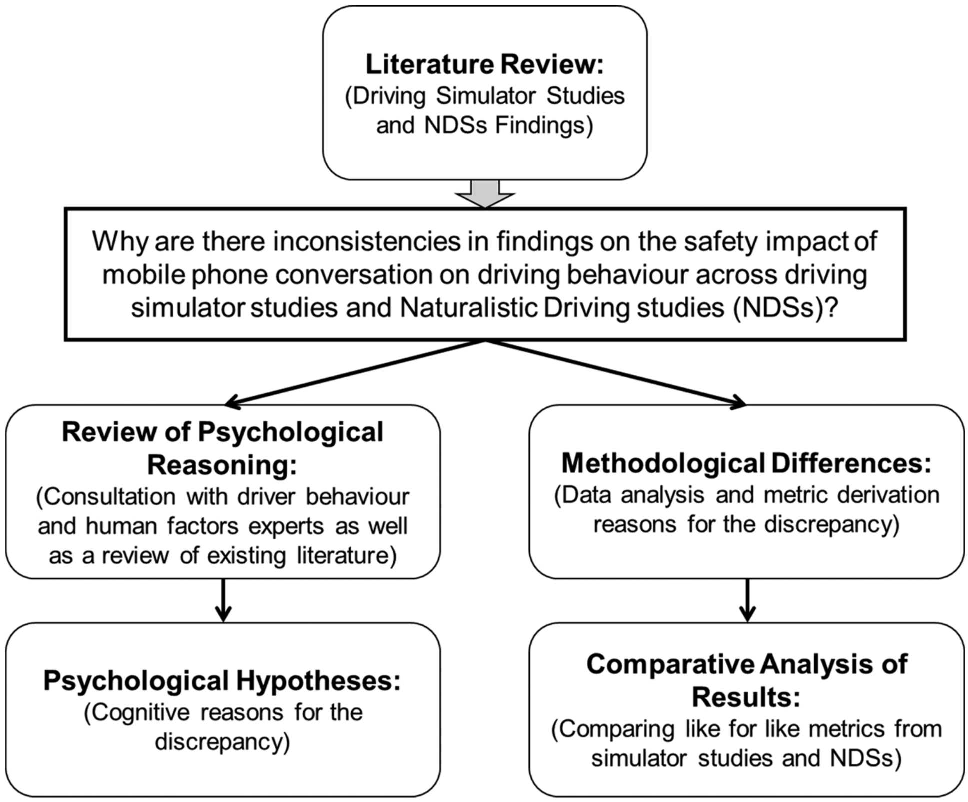

Fig. 1. Methodology flow chart.

may assist in reconciling the differences between the studies and highlights the value of both forms of studies in understanding the impact of mobile phone conversation on driving performance and safety.

2. Methodology

Fig. 1 presents a flow chart of the methodology used to conduct this study. Initially, a literature review was conducted to determine the discrepancies in findings related to driving behaviour phenomena across NDSs and driving simulator studies. As described in Section 1, the differences may be attributed to data collection methods, experimental design and data analysis procedures (Caird et al., 2018). In addition, the authors believe that these differences in the impact of mobile phone conversation on driving behaviour may be attributed to the gap in knowledge regarding cognitive processes associated with the driving task. Accordingly, the study is separated into two parts: the first part focusses on psychological hypotheses concerning driving behaviour in the context of mobile phone distraction, while the second part investigates the impact of methodological and analytic differences between NDSs and simulator studies.

Psychological factors were hypothesised in an effort to explain the existence of the discrepancy between the study methodologies. These were derived from targeted consultations with international distraction and human factors experts who assisted in exposing these factors, as well as an extensive review of existing psychological literature. (The experts, acknowledged at the end of this paper, were interviewed in small groups, by telephone. All had research expertise in driver distraction, and in researching driver distraction using driving simulators and/or instrumented vehicles in naturalistic driving studies. They were grouped according to the international times zones in which they resided, and each interview lasted for, on average, 1.5 h.) These factors propose a variety of behavioural explanations for why certain protective qualities of mobile phone conversation may exist whilst driving a vehicle. Details of the various hypotheses proposed are presented in Section 3 of the paper.

In general, driving simulator studies report differences in driving performance metrics with and without the presence of distractions and conduct standard statistical analyses to determine whether these differences are significant. On the other hand, NDSs derive metrics, such as the Odds Ratio, as an estimate of risk relative to a defined baseline scenario. It is essential to understand how these risk metrics are calculated, as differences in the way they are calculated may account for the degree of variation in results presented within NDSs. This is discussed in detail within Section 4.

The final analysis conducted in this study aimed to compare available results of driving simulator studies and NDSs in an effort to reconcile the discrepancy. Both NDSs and driving simulator studies collect driving performance data and available data were compared to understand the existence of measurement inconsistencies. The findings of the analysis are discussed in Section 5 of the paper.

The methodology highlights the presence of the discrepancy and attempts to provide reasoning for its presence. Upon gaining this understanding, it is envisaged that processes can be developed to obtain consistency across the assessment of the impact of mobile phone distraction on driving behaviour between NDSs and simulator studies.

3. Hypotheses for the discrepancy

Conversing on both hand-held and hands-free mobile phones can degrade driving performance in simulated environments. However, as noted, a majority of NDSs suggest that mobile phone conversations may not necessarily increase safety risk (e.g., Klauer et al. (2006)); and, in some cases, may even improve safety (e.g. Victor et al. (2015)). A number of psychological hypotheses regarding the existence of this discrepancy have been suggested by international experts in human factors and driving behaviour as well as being detailed within the literature. The following section of the paper presents a summary of the psychological hypotheses identified in the study.

3.1. Self-regulation

In real-world driving situations, drivers make decisions about using the mobile phone while driving and thus can exercise self-regulation. Self-regulation refers to how well a driver can change their behaviour to maintain adequate driving performance in the face of competing tasks (Young and Regan, 2007). For example, self-regulation can be *preparatory* in nature, which involves a driver preparing themselves *a priori* for the likely effects of potential distraction (e.g., turning off a cell phone before driving in case it rings during the drive). Self-regulation can also be *reactive*, in which case the driver strategically behaves in a way that compensates for the effects of an existing source of distraction (e.g., asking a passenger to stop talking when navigating a turn at a busy intersection).

However, within simulated environments, participants are essentially coerced into using mobile phones through instruction. It is feasible that these same sets of participants, when placed in real driving conditions, would not have used the mobile phone due to their awareness of the safety risk. Thus, this coercion through experimenter instruction may exaggerate the negative safety outcomes of simulator studies relative to NDSs. In naturalistic driving studies, drivers have some latitude to decide whether to engage in phone conversation, and therefore may choose not to do so in situations that they perceive to be risky.

There are several specific ways in which a driver may commonly attempt to self-regulate in response to distraction. For example, drivers may partake in behaviours perceived as being 'safer' such as reducing speed (Engström et al., 2005; Törnros and Bolling, 2005) or increasing headway to a lead vehicle to compensate for the engagement in secondary tasks (Hosking et al., 2009; Fitch et al., 2013). Another simple and common self-regulation strategy that drivers tend to use includes taking shorter glances away from the road when driving demand increases (e.g., keeping eyes on the road when the road is foggy) (Tivesten and Dozza, 2014; Tsimhoni et al., 2004).

Tivesten and Dozza (2015) used naturalistic driving data to determine how the intensity of the driving task moderated drivers' self-regulation in terms of their engagements with visual-manual (VM) phone tasks. Findings showed that drivers were more likely to engage with the VM task when the driving task was less intense (e.g., not when navigating sharp turns, roundabouts, etc). Through this self-regulation strategy of varying their engagement in VM phone tasks in response to changing roadway demands, drivers may reduce the level of interference between driving performance and interacting with the VM phone task.

3.2. Arousal

Mobile phone conversations may be arousing and help drivers to maintain alertness and counteract the effects of drowsiness and fatigue.

The differences observed between NDS studies and simulator studies may be attributed, at least in part, to the context in which distraction occurs. For example, a person who receives an emotional phone call (such as winning a prize or the passing of a loved one) whilst driving in the real world will most likely be distracted, in contrast to a person receiving a phone call in a driving simulator about a mundane topic, who may not be affected at all. Cognitively loading tasks, such as mobile phone conversation, have been shown to reduce crash risk associated with drowsiness and fatigue, which may be a factor behind the "protective" effect of such tasks found in several NDSs. There may be more scope for this effect to operate in the real world than in driving simulators. This proposition has been made for the studies of driving large vehicles carried out by Olson et al. (2009) and Hickman et al. (2010). However, Victor et al. (2015) contrasts the above findings in their study of light vehicle driving where drowsiness was rarely encountered in either baseline driving or immediately preceding SCEs. In a similar vein, Wiegand et al. (2008), in a study of fatigue in commercial motor vehicle drivers, found that the majority of the safety-critical incidents in their study occurred while the driver was alert.

3.3. Gaze concentration

The gaze concentration induced by cognitive load and the consequent focus on the forward central field of view may lead to a greater chance of detecting a sudden closing in of a vehicle ahead (Carsten and Merat, 2015). Cognitive loading tasks have been shown to induce gaze concentration to the forward roadway (Recarte and Nunes, 2000, 2003) and gaze concentration has been observed in a number of simulator studies (Victor et al., 2005; Engström et al., 2005; He et al., 2014; Boer et al., 2016; Li et al., 2018) and real-road studies (Victor et al., 2005; Yang et al., 2018). It can be viewed as an autonomic response to the cognitive load of the conversation. This gaze concentration may lead to a greater chance of responding automatically to looming cues created by rapidly decelerating lead vehicles, and in turn may increase the likelihood of a successful forward avoidance manoeuvre (Engström et al., 2017). This is more likely to occur in NDSs as opposed to simulator studies that present more hazards in the periphery (such as with peripheral detection tasks).

3.4. Task displacement

It is generally not possible to carry out two distracting activities simultaneously, and phone conversations may therefore act to suppress any possibility of engaging in other, more risky, activities, such as texting.

In a simulated environment, all distractions except for mobile phone use are controlled for between baseline and distracted conditions. Thus, in the distracted condition, the task of using the mobile phone becomes the only distraction and reduces performance. In the real world, these same participants may be distracted by a number of other activities. The interruption of a phone call during the performance of such activities may be displacing one distraction task with another, which is less risky. Victor et al. (2015) extend this task displacement to a proposition of glance displacement: while a conversation is being conducted on a mobile phone there are comparatively few glances away from safety-relevant scanning. Drivers still consult their mirrors and peripheral areas in the windscreen, and while doing so may still be capable of peripheral detection of the looming of the lead vehicle. However, from their data, the size of this effect looks to be rather small.

3.5. Controlled versus automatised performance

The apparent inconsistences between simulated driving studies and NDSs may be reconciled based on the distinction between automated and controlled task performance. Automatic performances occur without cognitive effort and are generally unconscious, and thought to be the product of repetition between, and the learning of, certain stimuli and their appropriate responses (Engström et al., 2017). By contrast, controlled performances are cognitive demanding, requiring attentional effort to deal with novel or unlearned tasks (Engström et al., 2017). The Cognitive Control Hypothesis, put forward by Engström et al. (2017), posits that cognitive loading tasks (e.g., a mobile phone conversation) will impact controlled driving performance, but not automatized driving performance. Therefore, it may be argued that, since simulator experiments often require the driver to perform tasks that rely on executive attention/cognitive control (e.g., responding to artificial stimuli, maximising performance on lane keeping), while real-world crash avoidance may be more governed by more automatised skills such as braking in response to a looming braking lead vehicle, mobile phone conversations have increased potential to impact on performance in the former, but not safety risk in the latter.

Self-regulation, arousal, gaze concentration, task displacement and the contextual differences between controlled and natural driving environments are all potential explanations for the differences in results observed between simulator and NDSs; and there may be others. However, as indicated by Caird et al. (2018), both these approaches have methodological differences which need to be investigated to further understand the variety of reporting that has been documented.

4. Assessment of methodological differences

Driving simulator and laboratory-based studies attempt to determine the safety impact of mobile phone conversation using driving performance metrics. Examples of these metrics include reaction time, headway and lane positioning and the studies report absolute and percentage changes to these metrics, going from the undistracted to the distracted driving condition, suggestive of an increase or decrease in safety.

In contrast, NDSs have access to documented crash statistics and use these to derive risk ratios, commonly the "Odds Ratio", which compares "crash conditions" to "baseline conditions" (typical driving, non-crash, scenarios) indicating the relative change in likelihood of crash events occurring. Odds Ratio determination involves categorisation of data as well as applying an estimation process to calculate the ratio which may change across the different studies. Furthermore, research has used the Odds Ratio as evidence to present both heightened crash risk (Dingus et al., 2016) and protective qualities (Fitch et al., 2013) with regards to mobile phone distraction while driving. Accordingly, it is imperative to understand the assumptions and mechanics of the Odds Ratio metric used in NDSs.

4.1. Odds ratio calculation method

Several available NDSs which investigated the impact of mobile phone conversation on driving behaviour have been reviewed (Dingus et al., 2016; Farmer et al., 2015; Fitch et al., 2015; Hickman et al., 2010; Klauer et al., 2014; Olson et al., 2009; Victor et al., 2015; Dingus et al., 2019) to understand the derivation of results and presentation of conclusions within these studies.

Klauer et al. (2006) presents a clear definition of the odds ratio in the context of safety studies. The odds ratio measures the frequency of event occurrence (e.g. number of times a driver is distracted by a mobile phone) relative to the frequency of event non-occurrence (e.g. number of times a driver is not distracted by a mobile phone). In other words, the odds of an event occurrence can be considered as the probability of occurrence divided by the probability of non-occurrence.

Consider the generalised event matrix presented in Table 1. Baseline conditions reflect non-crash events experienced during the driving period, while crash conditions reflect all crash and near-crash events that occur within the sample. Let the probability of a distraction occurring during baseline conditions be denoted by $\pi_1 = \frac{n_{11}}{n_{11}+n_{12}}$, and the probability of a distraction not being present be defined by $(1-\pi_1) = \frac{n_{12}}{n_{11}+n_{12}}$. Therefore, under baseline conditions, the odds of a distraction event occurring can be expressed as $\frac{\pi_1}{1-\pi_1} = \frac{n_{11}}{n_{12}}$. A similar set of definitions can be formed for crash conditions, assuming that the probability of distraction occurring is denoted by π_2. Thus, under crash conditions, the odds of a distraction event occurring can be expressed as $\frac{\pi_2}{1-\pi_2} = \frac{n_{21}}{n_{22}}$.

Consequently, the odds ratio is defined as the odds of distraction resulting in a crash event divided by the odds of a distraction not resulting in a crash event (baseline conditions).

Table 1
Generalised event matrix.

	Frequency of 'Distraction' events	Frequency of 'No Distraction' events	Total conditions
Baseline conditions	n_{11}	n_{12}	$n_{11}+n_{12}$
Crash conditions	n_{21}	n_{22}	$n_{21}+n_{22}$
Total events	$n_{11}+n_{21}$	$n_{12}+n_{22}$	

$$\textbf{Odds Ratio}\ (\boldsymbol{OR}) = \frac{\textbf{Odds of Distraction}\ (\textbf{crash conditions})}{\textbf{Odds of Distraction}\ (\textbf{baseline conditions})} = \frac{\frac{\boldsymbol{n_{21}}}{\boldsymbol{n_{22}}}}{\frac{\boldsymbol{n_{11}}}{\boldsymbol{n_{12}}}} = \frac{\boldsymbol{n_{21}} \times \boldsymbol{n_{12}}}{\boldsymbol{n_{22}} \times \boldsymbol{n_{11}}}$$

An odds ratio value less than 1 indicates that the distraction has protective qualities, while an odds ratio value in excess of 1 suggests that risk has increases as a result of the distraction. In addition to presenting the odds ratio, many studies present the 95% confidence interval of the odds ratio, which is calculated using an estimate of the standard error of the natural logarithm of the odds ratio (as shown below):

$$\textbf{Standard Error}\ (\boldsymbol{LN}\ (\boldsymbol{OR})) = \sqrt{\frac{1}{\boldsymbol{n_{11}}} + \frac{1}{\boldsymbol{n_{12}}} + \frac{1}{\boldsymbol{n_{21}}} + \frac{1}{\boldsymbol{n_{22}}}}$$

$$\textbf{Confidence Interval}\ (95\%) = \left(\boldsymbol{e}^{\boldsymbol{LN}(\boldsymbol{OR}) - 1.96\times\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{21}}+\frac{1}{n_{22}}}},\ \boldsymbol{e}^{\boldsymbol{LN}(\boldsymbol{OR}) + 1.96\times\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{21}}+\frac{1}{n_{22}}}}\right)$$

As an example, Dingus et al. (2016) present the odds ratio statistic for 'Cell talk (hand-held) = 2.2 (1.6–3.1)'. This indicates that the estimated value for the odds ratio is 2.2; however, for 95% confidence of the estimate, the odds ratio value may range between 1.6 and 3.1. This statistic provides an indication of the reliability of the estimate and is critical in evaluating the final conclusions of NDSs.

All NDSs reviewed applied the above methodology to determine the odds ratio in relation to different mobile phone uses, including making hand-held phone calls, hands-free phone calls, texting and browsing the phone. There are a number of aspects of the calculation of the odds ratio metric which can affect the interpretation of the impacts of mobile phone distractions on crash risk. Thus this could potentially be a source for the inconsistency associated with findings across NDSs as well as the discrepancy between driving simulator studies and NDSs. These aspects associated with the calculation are discussed in the following sub-section.

4.2. Factors contributing to Bias of odds ratios

Recently, Young (2017) presented a re-analysis of the SHRP-2 NDS data completed by Dingus et al. (2016), which clearly articulates the potential for bias in estimating the odds-ratio. Young (2017) posits that, by removing selection bias and confounding bias associated with driver behaviour errors from the sample, this resulted in a reduction of the odds ratio from 2.2 (as documented in Dingus et al. (2016)) to 0.72, a shift from an at risk scenario to a protective one. This clearly highlights the need to investigate the procedure for estimating the odds ratio and potentially calling for a set of standards that all driving behaviour research should follow to maintain consistency.

4.2.1. Identification and application of safety critical events

NDSs identify safety critical events (SCEs) to define crash conditions which are essential for the estimation of odds ratios and other crash risk estimates. SCEs aim to account for both crash and near crash events present within the data set (Simmons et al., 2016). However, there are issues in both the identification and application of SCEs to estimate crash risk. Identification of SCEs is an evolving field that aligns with advances in video image processing (Dozza and González, 2013). Accordingly, depending on the level of technology available in analysing NDS data there is the potential for variability of identification of SCEs across different NDSs resulting in inconsistencies between studies. Furthermore, a majority of SCEs identified in NDS data are associated with crashes related to following vehicles interacting with a lead vehicle. Thus, this will not capture the breadth of crashes present on a road and crash risks estimated are those related to a particular type of crash (lead vehicle interactions) (Knipling, 2017).

A key assumption of the Heinrich Triangle, a theory of accident prevention, is that 'crashes within the triangle have identical or highly similar causal factors regardless of the outcome severity' (Knipling, 2017). Road crashes are heterogeneous in nature and as such contradict this assumption. However, as NDSs generally capture SCEs, which are mainly non-crash dynamic events and minor crashes, the Heinrich Triangle assumption is explicit in the analysis of NDS data. Accordingly, crash risk estimates derived from existing NDSs may not reflect the crash risk associated with injury crashes or fatalities (Knipling, 2017), limiting the impact of the crash risks measured through the odds ratio and the conclusions derived from such studies.

4.2.2. Sampling procedures used to determine baseline conditions

Estimation of odds ratios within NDSs generally require a "case-control design" to investigate the impact of distractions on driving behaviour (Klauer et al., 2006). In the context of driving behaviour research, this metric measures the association between the presence of cases (defined in a number of studies as SCEs) and the controls, which are defined as baseline driving periods. The baseline driving periods consider the scenario where all non-driving secondary tasks are excluded. The intention is to remove all other sources of distraction - for example, eating and drinking or smoking while driving -and focus on a specific distraction. Even though this case-control approach allows for a degree of isolation of the focal distraction source, the method to derive the baseline frequencies can affect the calculation of the ratio.

Data reduction procedures are necessary to identify the presence of SCEs and also gain an understanding of the impact of the distraction. Though this is a necessary process, it may be eliminating data essential to the understanding of the problem; and if this approach is not standardised across NDSs, it can potentially result in the discrepancy observed across the studies as further evidenced with the study completed by Young (2017).

The method of collecting 'baseline frequency data' (conditions where there are no distractions and no crashes are recorded) is generally done by employing a sampling procedure. This procedure involves sampling baseline data events for each participant as a proportion of driving time relative to the total driving time of all participants in the study (Fitch et al., 2013). For example, in (Fitch et al., 2013), the total number of hours driven in the data set was calculated and then the number of baseline periods sampled from each driver is determined by multiplying the driver's percentage of driving time (relative to total hours in the data set) with the total number of baseline periods. Consider a single driver comprised 10% of the total driving time of the pool of participants and the total number of baseline periods obtained is 1000, then the number of baseline periods sampled from the driver is 10% of 1000, which would be 100 samples.

The issue concerning this method is that the total number of baseline periods is pre-defined based on a separate sampling procedure that is not clearly explained in the literature reviewed. In Fitch et al. (2013), for example, the study simply states 'The study's timeline facilitated the selection and reduction of 2000 baseline periods', without any clear description of the validity of using 2000 as a sample size. Variation in sample sizes of the baseline period can affect the estimation of the odds ratio, as shown in the example presented in Table 2. In this example, if the sample size of the baseline periods doubled (from 1000 to 2000), but the frequency of distraction events did not exactly double, the odds ratio would change. Table 2 shows that the odds ratio increases from a protective value of 0.91 to an enhanced crash risk value of 1.11, highlighting the concerns regarding the estimation procedure. Accordingly, it is imperative to, clearly define baseline conditions and also ensure the validity of the sample size used for baseline periods in order to obtain stability in the odd ratio calculation process.

Table 2
Example of the impact of underestimating baseline frequency.

Incorrect Value	**Frequency of 'Distraction' events**	**Frequency of 'No Distraction' events**
Baseline Conditions	220	1000
Crash Conditions	30	150
Total Events	250	1150
Odds Ratio $= \frac{30 \times 1000}{220 \times 150} = 0.91$		
Correct Value	**Frequency of 'Distraction' events**	**Frequency of 'No Distraction' events**
Baseline Conditions	360	2000
Crash Conditions	30	150
Total Events	390	2150
Odds Ratio $= \frac{30 \times 2000}{360 \times 150} = 1.11$		

4.2.3. Definition of 'baseline driving conditions'

The definition of 'baseline driving conditions' varies from study to study. In Fitch et al. (2013), base condition data are extracted 30 s prior to the start of mobile phone usage while Victor et al. (2015) considered base lines of 12 s prior to a crash event or form of distraction. This has been undertaken by the authors to ensure that the environmental context for baseline and mobile phone use periods are similar, which is a reasonable argument. However, there is doubt as to whether the "30 s prior to mobile phone usage" is appropriate to define base conditions (Simmons et al., 2016; Young, 2017). There is a possibility that participants are anticipating a phone call or distracted by an entirely different event that is not being recorded by the study, and 30 s prior to the event may not be an accurate reflection of base conditions. The 30 s value was selected without clear justification, and similar assumptions have been made within other NDSs when deriving the odds ratio, potentially acting as a source of error. Furthermore, baseline driving conditions are affected by the state of prevailing traffic. Driver behaviour during highly congested stressful peak periods differ from behaviour present during lower levels of road congestion (Hennessy and Wiesenthal, 1997, 1999). This change in behaviour results in variation between baseline data points and could potentially bias metrics derived from the data set. There may be the need to account for traffic characteristics through a standardisation approach which to date has not been documented in the literature. These are aspects that require further study to identify the true meaning of base conditions in the context of these NDS studies.

4.2.4. Independence of variables in the odds ratio

Further to this, it is evident that another key confounding factor of using the odds ratio is the assumption of independence between the categories of data defined to calculate the value of the odds ratio. For example, there is a separation of distracted (mobile phone use) and not distracted (not using mobile phone) data, when in reality it is possible that participants may be shifting from one form of distraction to another (task displacement (see Section 3.4): for example, texting or using the Visual Display Unit). These types of interactions and other cause-and-effect relationships within driving behaviour are generally not captured using the odds ratio approach.

4.2.5. Equivalence of 'Risk rate ratio' definition with that of Odds ratio

Another confounding factor is the use of the 'risk rate ratio' derived from a risk rate approach as a proxy for the odds ratio (Dingus et al., 2016). Unlike the case-control approach, the risk rate approach does not isolate the distraction and considers a baseline which contains the presence of other distractions (Fitch et al., 2013). The risk rate estimate can be computed by dividing the number of SCEs during mobile phone use by the total number of minutes spent using a mobile phone while driving. This rate is then compared to the risk rate of general driving. Unlike the odds ratio, the risk rate approach considers the possibility of distractions other than mobile phone use being present during the

Table 3
Confounding factors potentially affecting Odds Ratio calculation.

Confounding Factor	Potential Impacts
Identification and Application of Safety Critical Events (SCEs)	As SCEs are commonly used as a proxy for crashes in existing NDSs, crash risk estimates may not reflect the crash risk associated with injury crashes or fatalities reducing the impact of the conclusions.
Sampling of 'baseline driving conditions'	Underestimation of baseline frequency of 'no distraction' events resulting in an underestimation of the odds ratio.
Definition of 'baseline driving conditions'	Possibility that baseline conditions include distracted conditions and accordingly reducing the impact of the distraction and hence a lower odds ratio.
Independence of variables in the odds ratio	Inability to capture interactions and cause and effect relationships due to the categorisation of data resulting in an incomplete perspective of risk by only considering the odds ratio. This could potentially lead to an overestimation or underestimation of crash risk.
Equivalence of 'Risk rate ratio' definition with that of Odds ratio (Dingus et al., 2016)	Risk rate ratio has been used as a proxy for the odds ratio within Dingus et al. (2016) study, potentially causing biased results in comparison to other NDS studies.

driving activity. Fitch et al. (2013) computed the number of SCEs during non-mobile phone use divided by the total number of minutes driving without mobile-phone use. A regression model was then developed to investigate whether the risk rate for mobile phone use differed from the risk rate for the baseline conditions. Furthermore, a risk rate ratio can also be calculated serving as an approximation to the odds ratio. This approximation may lead to inconsistencies between NDSs (Fitch et al., 2013). It is not possible to determine whether this approach is overestimating, underestimating or even accurately assessing crash risk. However, the different methods affect the estimation of the odds ratio.

Table 3 presents a summary of the confounding factors which may potentially impact the calculation of the odds ratio and ultimately the assessment of safety in the context of the distraction.

4.3. Methodological factors affecting simulator studies

Finally, it is also important to discuss factors that may influence the results and conclusions of simulation studies. Across the studies that were reviewed (for example: Beede and Kass (2006); Strayer and Drew (2004)), the definition of the 'distracted state' varied. There were instances where distractions were defined by the action of picking up the phone while other studies focussed on identifying distracted states through measurements of deviations in awareness through eye tracking. These differences in definition can result in variability in the results of simulation/laboratory-based approaches.

However, upon reviewing both NDSs and simulation/laboratory studies, the most fundamental observation is the contrasting approach to measuring the safety implications of mobile phone conversation, which may be of more importance than any of the above factors.

All NDSs investigated safety from a crash risk perspective. Simulation studies assessed safety by observing changes to performance metrics such as speed, reaction time and lateral movements, as these studies do not have access to real world crash or near-crash event data. Thus, to appropriately compare the findings between these two forms of studies, it is essential to distinguish between them based on a 'performance metric' by extracting relevant data from NDSs and simulation studies alike.

5. Comparative analysis of driving simulator studies and NDSs

A demonstration comparative analysis of results from NDSs and simulation studies obtained from available literature is presented in the following section.

In general, simulation studies completed in the past have not collected or observed real world crash or near-crash data. Accordingly, it is not feasible to determine odds ratios for these studies. Thus, the reanalysis involved comparing performance metric data of NDSs and simulation studies.

Across the 11 NDSs investigated, the only set of performance metrics available for comparison was obtained from Fitch et al. (2013), which provided the following data:

- speed (km/hr)
- headway (s)
- time to collision (s)
- standard deviation of lane position (m)/ percentage deviation of lane position (%).

Simulation studies provided additional performance metrics such as reaction time, perception response time and following distance. However, these were not considered for comparison as they could not be extracted from the NDSs.

5.1. Results of the comparative analysis

Table 4 and Table 5 present a summary of the comparison data for the simulation approach and naturalistic driving approach, respectively. The tables present the raw mean and standard deviation results extracted directly from the literature for 'baseline' conditions, where no distractions are present, and 'distracted' conditions, where users are engaged in mobile phone activity. Consistency in the comparison is maintained as the NDS applied a case-control approach and baseline conditions were observed by removing the impacts of all other possible distractions (Fitch et al., 2013).

To understand the impact of the distractions, the percentage difference in means and the percentage difference in the coefficient of variation (CV) have also been calculated. The CV, also known as the relative standard deviation, is a standardised measure of dispersion of the data set and is calculated by dividing the sample standard deviation by the sample mean of a data set. The CV provides another measure of the degree of difference between baseline and mobile phone usage scenarios.

It is important to highlight that simulator-based studies focus on the variation in performance metric for a given driving manoeuvre; for example, a study will measure the variation in speed with and without the distraction of the mobile phone. In contrast, NDSs will focus on the impact of a particular mobile phone task and the subsequent impact on the performance metric. This emphasises the differences between the approaches.

The analysis of the performance data suggests some similarities between the two forms of studies. Though some simulator studies indicated an increase in speed with the use of mobile phones, in general, both sets of data suggest that mobile phone distractions result in a minor reduction in speed, generally between 2% to 5%. Visual-manual phone interactions (such as reading, browsing and texting) contained the greater reductions in speed relative to the auditory-vocal interactions (talking on the phone) which had lesser effect. Furthermore, mobile phone usage also results in drivers leaving greater headways, with mean differences between baseline and distracted conditions varying from 10% to 40%. These are both intuitive conclusions as drivers are likely to be more cautious whilst multitasking between

Table 4
Summary of Simulation Study Performance metrics.

Simulation Study Performance Metrics

Performance Metric	Study	Sample size	Baseline Mean	"Distracted" Mean	Baseline Standard Deviation	"Distracted" Standard Deviation	Percentage difference of Mean values	Baseline CV	Distraction CV	Percentage difference of CV
Speed (kph)	Beede and Kass (2006)	36	34.26	34.58	3.14	2.74	0.93%	0.09	0.08	−13.55%
	Haque and Washington (2015)	32	37.90	37.30	4.10	6.20	−1.58%	0.11	0.17	53.65%
	Horberry et al. (2006)	31	58.48	58.94	N/A	N/A	0.79%	N/A	N/A	N/A
	Strayer and Drew (2004) (Age: 18-25)	20	101.28	99.36	14.31	16.10	−1.90%	0.14	0.16	14.69%
	Strayer and Drew (2004) (Age: 65-74)	20	83.84	85.92	14.31	14.65	2.48%	0.17	0.17	−0.11%
	Strayer et al. (2006)	40	88.80	86.08	7.08	19.60	−3.06%	0.08	0.23	185.45%
	Rakauskas et al. (2004)	24	72.99	71.89	3.12	5.93	−1.51%	0.04	0.08	93.09%
Headway (s)	Ranney et al. (2004)	12	0.78	0.87	N/A	N/A	11.54%	N/A	N/A	N/A
Time to Collision (s)	Strayer et al. (2006)	40	8.50	8.1	1.90	1.17	−4.71%	0.22	0.14	−35.50%
Lane Position, Standard deviation (m)	Beede and Kass (2006)	36	1.50	1.34	0.34	0.43	−10.67%	0.23	0.32	41.57%
	Ranney et al. (2004)	12	0.53	0.45	N/A	N/A	−15.09%	N/A	N/A	N/A

manoeuvring the vehicle and using the mobile phone. A remarkable consistency was observed with the 'lane positioning variable' which indicates that, in both forms of studies, the lane position improved while drivers communicated using the phone (hand-held communication). The degree of improvement varied between 5% and 25%. This could be suggesting that the hypotheses of task displacement and forward gaze concentration allow drivers to focus better on the driving task at hand. Such a finding is supported by a past driving simulator study undertaken by Atchley and Chan in 2011 where young drivers attention to a monotonous road environment improved as a result of multi-tasking (Atchley and Chan, 2011).

However, there are also differences in trends between the two forms of studies. The single simulator study which recorded time to collision reported a 4.7% reduction in time with the presence of a mobile phone (Strayer et al., 2006). Even though Fitch et al. (2013) show a reduction in time to collision for three phone activities (dial, browse/read; talk/listen; end task), two of the activities (locate and answer; talk and listen) indicate an increase in time to collision. This could be an artefact of variations in time to collision measurements between studies, but could also be suggesting that some mobile phone usage activities are stimulating the driver and heightening his or her awareness, which in turn increases safety (consistent with the hypotheses of task displacement and arousal presented in Section 3).

The changes in the coefficient of variation highlight the differences in the magnitude of impact between the two forms of studies. Driving simulator studies reveal percentage differences of the coefficient of variation between baseline and distracted conditions that are in some cases more than 80%, highlighting the significance of the change in performance metric. However, within the NDS data, the percentage change in coefficient of variation remains less than 75% in all cases. Though still significantly different, there is less change within the NDS performance data analysed. These findings are consistent with the reasonable conclusions raised by a number of NDS studies (Caird et al., 2018; Simmons et al., 2016), suggesting that mobile phone usage in the real world does not pose as great a risk to the driver as what is proposed in some driving simulator studies.

6. Discussion and conclusions

The overarching aim of this study was to highlight several possible explanations for the apparent inconsistency in findings between naturalistic driving studies and experimental/simulator studies with regards to the impact of mobile phone conversations. A significant limitation to the study has been access to raw data sources from both NDSs and simulator studies. Though this paper has provided a demonstration of the possible approach to reconciling the inconsistency between NDSs and simulator studies, further detailed research must be conducted to confirm the approach and the observations presented in this paper. Overall, the research has provided a foundation of possible psychological and methodological reasons that can cause the identified discrepancies.

There may be a number of psychological factors that contribute to these inconsistencies. These include hypotheses concerning:

- Self-regulation: The possibility for a driver to self-regulate mobile phone use in the real-world to a greater extent than what may be allowed in highly controlled experimental settings.
- Arousal: Phone conversations being of higher significance and arousal in NDSs relative to a simulated environment leading to greater real-world compensatory driver behaviour.
- Gaze Concentration: Real-world phone conversations enhance enough cognitive loading to improve gaze concentration and as such longitudinal hazard detection in NDSs.
- Task displacement: Phone conversations may prevent drivers from engaging in other, more risky, activities in real world scenarios which cannot be captured in a simulation environment.

Table 5
Summary of Naturalistic Driving Study Performance metrics (Fitch et al., 2013).

Naturalistic Driving Study (NDS) Performance Metrics

Performance Metric	Mobile phone use task	Sample size	Baseline Mean	"Distracted" Mean	Baseline Standard Deviation	"Distracted" Standard Deviation	Percentage difference of Mean values	Baseline CV	Distraction CV	Percentage difference of CV
Speed (kph)	Locate/Answer	202	69.0	69.5	29.85	29.85	0.72%	0.43	0.43	−0.72%
	Dial	131	70.1	72.2	30.90	29.76	3.00%	0.44	0.41	−6.50%
	Talk/Listen	207	68.8	67.4	30.21	30.21	−2.03%	0.44	0.45	2.08%
	Browse/Read, Talk/Listen	56	74.4	69.6	31.43	32.18	−6.45%	0.42	0.46	9.44%
	End Task	179	69.6	68.2	30.77	30.77	−2.01%	0.44	0.45	2.05%
Headway (s)	Locate/Answer	72	0.5	0.6	0.00	0.00	20.00%	0.00	0.00	N/A
	Dial	42	0.5	0.7	0.00	0.65	40.00%	0.00	0.93	N/A
	Talk/Listen	65	0.6	0.6	0.00	0.00	0.00%	0.00	0.00	N/A
	Browse/Read, Talk/Listen	22	0.5	0.5	0.47	0.47	0.00%	0.94	0.94	0.00%
	End Task	58	0.5	0.5	0.00	0.00	0.00%	0.00	0.00	N/A
Time to Collision (s)	Locate/Answer	24	50.8	73.2	52.91	118.56	44.09%	1.04	1.62	55.50%
	Dial	12	53.4	50.9	63.74	44.34	−4.68%	1.19	0.87	−27.02%
	Talk/Listen	20	62.2	101.7	69.77	128.80	63.50%	1.12	1.27	12.91%
	Browse/Read, Talk/Listen	5	115.6	70.2	85.19	89.67	−39.27%	0.74	1.28	73.32%
	End Task	22	60.0	43.5	57.22	58.16	−27.50%	0.95	1.34	40.19%
Percentage Deviation of Lane Position (%)	Locate/Answer	123	40.7	33.3	N/A	N/A	−18.01%	N/A	N/A	N/A
	Dial	71	42.3	42.3	N/A	N/A	0.00%	N/A	N/A	N/A
	Talk/Listen	112	38.4	41.1	N/A	N/A	7.01%	N/A	N/A	N/A
	Browse/Read, Talk/Listen	41	25.6	31.7	N/A	N/A	23.67%	N/A	N/A	N/A
	End Task	104	34.6	40.4	N/A	N/A	16.64%	N/A	N/A	N/A

- Controlled versus Automatised performance: Cognitive loading tasks (e.g., a mobile phone conversation) will impact driving performance in controlled environments (driving simulator studies), but may not have an impact on automatized driving performance (NDSs) (Engström et al., 2017).

It is difficult to offer firm judgements on which of these theories may be more ‘plausible’ than others in accounting for changes in driving performance when cogntively loaded simply due to a lack of empirical research that has set out to test these theories against each other. However, one recent driving simulator study examined whether changes in lane keeping performance, when drivers are cognitively loaded, could be better accounted for by increased arousal or gaze concentration towards the forward roadway, two of the theories noted above (Li et al., 2018). While the study found that cognitive load was associated with improved lane keeping performance, the authors concluded that this effect was likely a result of *both* increased arousal and gaze concentration towards the forward roadway ‘working’ simultaneously. This prelimimnary research suggests that changes in driving performance due to cognitive load are likely to be complex, multifaceted, and due to multiple mechanisms (e.g., increased arousal *and* gaze concentration) being simultaneously active (Li et al., 2018).

From a methodological perspective, it is important to understand that the two approaches are inherently different and each provides important information. However, as a result of the differences, they are difficult to reconcile. NDSs collect real world data of actual day-to-day driving events whereas driving simulator studies present a controlled environment in which a researcher can monitor driving performance in the context of a distraction stimulus. Though the realism of NDSs present a rich data source for analysis there is a limitation in understanding crash risk assessments derived from these studies. Generally, crash types that are observed and analysed in NDSs are low severity, non-fatal and non-injury instances and thus crash risks are associated with these crash types and not serious or harmful crashes which are considered to have a greater community wide impact. A potential option to reduce this constraint is to conduct more extensive NDSs that capture high severity crashes from a larger fleet of vehicles. Furthermore, this limitation can be resolved within simulator studies as a variety of crash scenarios and distraction events can be tested. However, the weakness of simulator studies is that the stylised environments may result in participants displaying behaviour that they would not normally conduct in a real scenario. Though these differences exist and reconciliation is difficult it is important to obtain consistency in the conclusions concerning the impact of mobile phone conversation on driver safety.

In addition, inconsistency in the analysis of data between NDSs and driving simulator studies may help account for their discrepant findings. NDSs are designed to measure crash risk while simulator studies measure changes in driving performance. Once this inconsistency is resolved, the discrepancies in conclusions from these studies may reduce. It was possible, however, to compare speed, headway, time to collision and lane positioning variables across simulator studies and the NDS considered in the analysis. There were consistent trends across all NDS and driving simulator studies for these variables. However, the degree of difference between baseline and distracted conditions was greater in magnitude within simulator studies compared to NDSs, suggesting more significant risks for the driver when partaking in mobile phone use during simulator studies as well as small sample contributions. This is consistent with the relatively moderate performance results presented in most NDSs.

A fundamental difference between NDSs and driving simulator studies is the assessment of risk. Driving simulator studies do not investigate crash risk. They collect driving performance data with and without the presence of distractions and provide an assessment on the potential risks associated with the change in performance. In general, these studies do not collect crash data and, thus, it is not feasible to obtain a crash risk assessment of the impact of the phone conversation, especially based on the odds ratio measurement.

There are three future research directions, deriving from the research discussion, which will be of benefit for crash risk assessment using NDSs and driving simulators in the future:

- **Accessing raw NDS data to investigate the calculation of the Odds ratio**: As highlighted in the report, there are many confounds related to the estimation of the odds ratio. Accordingly, a beneficial study would be to conduct an in-depth investigation of the methods used to derive the odds ratios metric with a particular focus on understanding the sensitivity of the metric based on the different interpretations and measurements of baseline conditions in NDSs.
- **Simulator studies investigating crashes and near-crashes**: As noted within the paper, the greatest hurdle in comparing simulator studies and NDSs is the fact that simulator studies do not collect real-world crash data. Further research should be conducted into relating performance metrics that are generally extracted from simulator studies, which can also be collected in NDSs to understand relationships with crash risk.
- **Simulator studies which attempt to replicate crash and near-crash events observed in NDS studies**: Replication of crash events and near-crash events in simulator studies can provide the ideal base to compare the results of NDSs and simulator studies to identify if there are systematic differences between the two data collection methods.

It must be acknowledged that there are systematic differences between simulator and naturalistic driving studies which both contain strengths and weaknesses.

Simulator studies provide control, while naturalistic studies provide greater realism; and, as suggested by (Caird et al., 2018), both methodologies are key to understanding more about driving safety. However, as explained in the present paper, there must be standardisation and transparency surrounding both simulator studies and NDSs to ensure comparable and replicable research into the future. Specifically, standard guidelines for the estimation of the odd-ratio will reduce the potential for bias that can overestimate or underestimate crash risk. In addition, ensuring common performance metrics across studies offers comparability and consistency of evaluations between the two formats. Improved appraisal of crash risk and safety assessment in the context of mobile phone conversation will yield to improved and, most importantly, globally consistent legislation surrounding mobile phone use in vehicles potentially leading to a safer road environment for all users.

Acknowledgements

We are grateful to the following research colleagues for participating as experts in the groups convened to expose and discuss factors that might explain the differences in findings from simulator and NDS studies that were the focus of this study: Prof Oliver Carsten; Dr Johan Engstrom; Dr William Horrey; Dr Katja Kircher; Prof John D Lee; Adj Prof Nick Reed; Prof David Strayer; Adj. Prof Trent Victor; and Prof Richard Young.

References

Atchley, P., Chan, M., 2011. Potential benefits and costs of concurrent task engagement to maintain vigilance: a driving simulator investigation. Hum. Factors 53, 3–12.

Beanland, V., Fitzharris, M., Young, K.L., Lenné, M.G., 2013. Driver inattention and driver distraction in serious casualty crashes: data from the australian national crash in-depth study. Accid. Anal. Prev. 54, 99–107.

Beede, K.E., Kass, S.J., 2006. Engrossed in conversation: the impact of cell phones on simulated driving performance. Accid. Anal. Prev. 38, 415–421.

Boer, E.R., Spyridakos, P.D., Markkula, G.M., Merat, N., 2016. Cognitive driver distraction improves straight lane keeping: a cybernetic control theoretic explanation. IFACPapersOnLine 49, 627–632.

Caird, J.K., Simmons, S.M., Wiley, K., Johnston, K.A., Horrey, W.J., 2018. Does talking on a cell phone, with a passenger, or dialing affect driving performance? An updated

systematic review and meta-analysis of experimental studies. Hum. Factors 60, 101–133.
Caird, J.K., Willness, C.R., Steel, P., Scialfa, C., 2008. A meta-analysis of the effects of cell phones on driver performance. Accid. Anal. Prev. 40, 1282–1293.
Carsten, O., Merat, N., 2015. Protective or not? 4th International Conference on Driver Distraction & Inattention, 9 - 11 November.
Centre for Road Safety, 2017. Faituged and Distracted Driver Trauma Trends. Australia: Tranport for New South Wales, Sydney.
Cunningham, M.L., Regan, M.A., Imberger, K., 2017. Understanding driver distraction associated with specific behavioural interactions with in-vehicle and portable technologies. J. Australas. Coll. Road Saf. 28, 27.
Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M., Hankey, J., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proc. Natl. Acad. Sci. 113, 2636–2641.
Dingus, T.A., Owens, J.M., Guo, F., Fang, Y., Perez, M., McClafferty, J., Buchanan-King, M., Fitch, G.M., 2019. The prevalence of and crash risk associated with primarily cognitive secondary tasks. Saf. Sci.
Dozza, M., González, N.P., 2013. Recognising safety critical events: can automatic video processing improve naturalistic data analyses? Accid. Anal. Prev. 60, 298–304.
Engström, J., Johansson, E., Östlund, J., 2005. Effects of visual and cognitive load in real and simulated motorway driving. Transp. Res. Part F Traffic Psychol. Behav. 8, 97–120.
Engström, J., Markkula, G., Victor, T., Merat, N., 2017. Effects of cognitive load on driving performance: the cognitive control hypothesis. Hum. Factors 59, 734–764.
Farmer, C.M., Klauer, S.G., McClafferty, J.A., Guo, F., 2015. Relationship of near-crash/crash risk to time spent on a cell phone while driving. Traffic Inj. Prev. 16, 792–800.
Fitch, G.M., Hanowski, R.J., Guo, F., 2015. The risk of a safety-critical event associated with mobile device use in specific driving contexts. Traffic Inj. Prev. 16, 124–132.
Fitch, G.M., Soccolich, S.A., Guo, F., McClafferty, J., Fang, Y., Olson, R.L., Perez, M.A., Hanowski, R.J., Hankey, J.M., Dingus, T.A., 2013. The Impact of Hand-Held and Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk. Washington DC, USA. Available: http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2013/811757.pdf [Accessed 15/6/2017]. .
Haque, M.M., Washington, S., 2015. The impact of mobile phone distraction on the braking behaviour of young drivers: a hazard-based duration model. Transp. Res. Part C Emerg. Technol. 50, 13–27.
He, J., McCarley, J.S., Kramer, A.F., 2014. Lane keeping under cognitive load: performance changes and mechanisms. Hum. Factors 56, 414–426.
Hennessy, D.A., Wiesenthal, D.L., 1997. The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. Ergonomics 40, 348–361.
Hennessy, D.A., Wiesenthal, D.L., 1999. Traffic congestion, driver stress, and driver aggression. Aggressive Behav. 25, 409–423.
Hickman, J.S., Hanowski, R.J., Bocanegra, J., 2010. Distraction in Commercial Trucks and Buses: Assessing Prevalence and Risk in Conjunction With Crashes and Near-crashes. Washington DC, USA. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.173.3995 [Accessed 15/6/2019]. .
Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J., Brown, J., 2006. Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. Accid. Anal. Prev. 38, 185–191.
Horrey, W.J., Wickens, C.D., 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Hum. Factors 48, 196–205.
Hosking, S.G., Young, K.L., Regan, M.A., 2009. The effects of text messaging on young drivers. Hum. Factors 51, 582–592.
Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The Impact of Driver Inattention on near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data.
Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A., 2014. Distracted driving and risk of road crashes among novice and experienced drivers. N. Engl. J. Med. 370, 54–59.
Knipling, R.R., 2017. Crash heterogeneity: implications for naturalistic driving studies and for understanding crash risks. Transp. Res. Rec. 2663, 117–125.
Li, P., Markkula, G., Li, Y., Merat, N., 2018. Is Improved Lane Keeping During Cognitive Load Caused by Increased Physical Arousal or Gaze Concentration toward the Road Center? Accid. Anal. Prev. 117, 65–74.
National Highway Traffic Saftey Administration, 2017. 2016 Fatal Motor Vehicle Crashes: Overview. US Department of Transportation., Washington DC, USA.
Olson, R.L., Hanowski, R.J., Hickman, J.S., Bocanegra, J.L., 2009. Driver Distraction in Commercial Vehicle Operations. Washington DC, USA. Available: https://rosap.ntl.bts.gov/view/dot/17715 [Accessed 18/6/2017]. .
Rakauskas, M., Gugerty, L., Ward, N., 2004. Effects of cell phone conversations with naturalistic conversations. J. Safety Res. 35, 453–464.
Ranney, T., Watson, G.S., Mazzae, E.N., Papelis, Y.E., Ahmad, O., Wightman, J.R., 2004. Examination of the Distraction Effects of Wireless Phone Interfaces Using the National Advanced Driving Simulator-Preliminary Report on Freeway Pilot Study. Washington DC, USA. Available: https://trid.trb.org/view/1142075 [Accessed 18/6/2017]. .
Recarte, M.A., Nunes, L.M., 2000. Effects of verbal and spatial-imagery tasks on eye fixations while driving. J. Exp. Psychol. Appl. 6, 31.
Recarte, M.A., Nunes, L.M., 2003. Mental workload while driving: effects on visual search, discrimination, and decision making. J. Exp. Psychol. Appl. 9, 119.
Regan, M.A., Hallett, C., Gordon, C.P., 2011. Driver distraction and driver inattention: definition, relationship and taxonomy. Accid. Anal. Prev. 43, 1771–1781.
Regan, M.A., Strayer, D.L., 2014. Towards an understanding of driver inattention: taxonomy and theory. Ann. Adv. Automot. Med. 58, 5.
Simmons, S.M., Hicks, A., Caird, J.K., 2016. Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: a systematic review and meta-analysis. Accid. Anal. Prev. 87, 161–169.
Strayer, D.L., Drew, F.A., 2004. Profiles in driver distraction: effects of cell phone conversations on younger and older drivers. Hum. Factors J. Hum. Factors Ergon. Soc. 46, 640–649.
Strayer, D.L., Drews, F.A., Crouch, D.J., 2006. A comparison of the cell phone driver and the drunk driver. Hum. Factors 48, 381–391.
Strayer, D.L., Drews, F.A., Johnston, W.A., 2003. Cell phone-induced failures of visual attention during simulated driving. J. Exp. Psychol. Appl. 9, 23.
Strayer, D.L., Johnston, W.A., 2001. Driven to distraction: dual-task studies of simulated driving and conversing on a cellular telephone. Psychol. Sci. 12, 462–466.
Tivesten, E., Dozza, M., 2014. Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. Transp. Res. Part F: Traffic Psychol. Beha. 26, 258–272.
Tivesten, E., Dozza, M., 2015. Driving context influences drivers' decision to engage in visual–manual phone tasks: evidence from a naturalistic driving study. J. Safety Res. 53, 87–96.
Törnros, J.E., Bolling, A.K., 2005. Mobile phone use—effects of handheld and handsfree phones on driving performance. Accid. Anal. Prev. 37, 902–909.
Tsimhoni, O., Smith, D., Green, P., 2004. Address entry while driving: speech recognition versus a touch-screen keyboard. Hum. Factors 46, 600–610.
Victor, T., Dozza, M., Bärgman, J., Boda, C.-N., Engström, J., Flannagan, C., Lee, J.D., Markkula, G., 2015. Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk. Transportation Research Board, Washington, D.C., USA. Available: http://www.trb.org/Main/Blurbs/171327.aspx [Accessed 15/6/2017].
Victor, T.W., Harbluk, J.L., Engström, J.A., 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. Transp. Res. Part F Traffic Psychol. Behav. 8, 167–190.
Wiegand, D.M., Hanowski, R.J., Olson, R.L., Melvin, W., 2008. Fatigue Analyses: From 16 Months of Naturalistic Commercial Motor Vehicle Driving Data. Virginia Tech. Virginia Tech Transportation Institute, Virgina, USA. Available: https://vtechworks.lib.vt.edu/handle/10919/7411 [Accessed 26/3/2019].
Yang, S., Kuo, J., Lenné, M.G., 2018. Analysis of gaze behavior to measure cognitive distraction in real-world driving. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Young, K.L., Regan, M.A., 2007. Use of manual speed alerting and cruise control devices by car drivers. Saf. Sci. 45, 473–485.
Young, R.A., 2017. Talking on a wireless cellular device while driving: improving the validity of crash odds ratio estimates in the shrp 2 naturalistic driving study. Safety 3, 28.