How to select distracted driving countermeasures evaluation metrics: A systematic review

Meiyu (Melrose) Pan [a*] and Alyssa Ryan [b]

[a]Buildings and Transportation Science Division, Oak Ridge National Laboratory, 2360 Cherahala Blvd, Knoxville, TN 37932, USA

[b]Department of Civil and Architectural Engineering and Mechanics, The University of Arizona, 1209 E 2nd St., Tucson, AZ 85721, USA

Corresponding author email: panm@ornl.gov

How to select distracted driving countermeasures evaluation metrics: A systematic review

While there are numerous performance metrics that have been developed for the evaluation of distracted driving prevention programs, there is little information on how to select them depending on the requirements and/or objectives of the study. This paper describes a systematic literature review that was conducted on the metrics for evaluating distracted driving countermeasures in order to bridge this research gap. A summary of the evaluation metrics used for the existing distracted driving countermeasures was provided. Guidance for choosing an evaluation measure was provided by analyzing the metrics from the perspectives of functionality, spatial-temporal dimension, and equity. Three examples of distracted driving countermeasure evaluation processes were thoroughly reviewed in order to offer insight into metric selection and measurement. This paper contributes to the body of knowledge by discussing the implications for policy on how to enhance the thoroughness and accuracy of the evaluation of distracted driving countermeasures. Analysis from multiple angles, the development of data collection tools and direct behavior indicators, taking into account temporal dimensions, and equity considerations, are proposed.

Keywords: Distracted driving, Evaluation, Comprehensive, Metric selection, Equity

1. Introduction

Distracted driving is any activity that takes focus away from the core goal of driving. Distracted driving has caused extensive death and injury across the country and the greater globe. Over three thousand individuals were killed in car crashes in the United States in 2019 due to distracted driving (NHTSA 2021). To deal with distracted driving behaviors, many distracted driving countermeasures have been developed in various cities and counties (Hedlund 2011). Selecting a suitable set of evaluation metrics is essential to provide accurate and comprehensive evaluation results.

While there are numerous performance metrics that have been developed for the evaluation of distracted driving prevention programs, there is little information on how to select them depending on the requirements and/or objectives of the study. Because of the limitations in the collection of some data, depending on just one or two metrics may not always produce comprehensive and conclusive findings. For instance, one of the most direct ways to gauge how effective distracted driving countermeasures are is to examine crashes caused by distracted driving. Collecting data on crashes involving distracted driving is challenging since police officers cannot always tell whether their lack of attention while driving was the main cause of the crash (Sajid Hasan et al. 2022; (Raymond) Huang et al. 2020). Additionally, as more and more types of countermeasures are produced, particularly with the aid of cutting-edge technology, it is crucial to compare them to choose the ones that are most suited for a region or demographic group. Without the right set of metrics, these comparisons might result in many instances of nuanced or mixed outcomes and weak generalizability to other places or people (L. Arnold and Horrey 2022).

On the other hand, multiple distracted driving countermeasures, including marketing campaigns and smartphone applications, have been developed with the aim of improving usability, public acceptability, and an equitable impact across all demographics. However, there have not been adequate studies on developing metrics that effectively take into account these goals in addition to the reduction of crashes and/or distracted driving (Hill et al. 2020; Vegega et al. 2013).

To summarize the evaluation metrics applied for current distracted driving countermeasures and provide guiding information for evaluation metrics selection, this paper presents a comprehensive review of the current research. First, the evaluation metrics for various forms of countermeasures were examined. Several categories of evaluation metrics were compiled in terms of the countermeasure type, data collection method, and sample size. The evaluation metrics were then compared. The trade-offs of each metric type were then discussed. Following, the secondary evaluation dimensions were examined. Third, examples of how data collection and performance metrics selection were carried out in evaluation studies were discussed. Finally, the recommended performance metrics that could be taken into further consideration in future evaluation studies were stated.

The review paper's contribution to the existing body of knowledge stems from the useful policy implications it offers for a more thorough and accurate assessment of distracted driving countermeasures. The proposed guideline is critical to make it easier for distracted driving prevention program developers to choose from a wide range of performance metrics and data collection methods. Future policymakers can use this review to more easily identify and adopt distracted driving countermeasures depending on their individual needs and objectives.

This paper is organized into four sections. The following section describes the materials and analysis methods for conducting the systematic literature review. The study findings, policy implications, and limitations are then presented. The paper concludes with a study overview with primary research results.

2. Scope of the Review

The effectiveness of distracted driving countermeasures has been widely examined (L. Arnold and Horrey 2022; NHTSA 2020; Khan et al. 2022; Regan and Oviedo-Trespalacios 2022). For instance, the AAA Foundation for Traffic Safety has conducted two systematic literature reviews to assess the effectiveness of technological, legislative, and behavioral countermeasures in 2019 and 2022 (L. S. Arnold et al. 2019; L. Arnold and Horrey 2022). The National Highway Traffic Safety Administration has also released a number of guidelines that examine the effectiveness of various highway safety programs, such as those that prevent distracted driving (Richard et al. 2018; NHTSA 2020).

The comparison of the efficacy of various countermeasures is made possible when more countermeasures are implemented. Yet, if the performance metrics are not appropriately selected, the evaluation results may be biased and unable to adequately depict the true impact. According to some studies, the choice of performance indicators can significantly affect the outcomes of the evaluation. For instance, a review study assessing the effectiveness of technologically-based countermeasures revealed inconsistent findings. In some studies, only one or two behavioral metrics were employed, which made it difficult to compare the effectiveness of different interventions (L. S. Arnold et al. 2019). Furthermore, studies utilizing well-established criteria such as crash rate and self-report behavior have indicated that some countermeasures, particularly educational programs, have weak effectiveness (Fisher, Pollatsek, and Pradhan 2006; Groeger and Banks 2007). However, additional metrics regarding how individuals perceive or engage with these countermeasures may be helpful and provide supplementary meaningful insight, as the findings may show how the phrasing or engaging mechanism can be improved to have a greater impact on people's behavior.

In summary, choosing the right performance metrics is crucial for the evaluation. Nevertheless, there is currently little guideline for selecting performance indicators that will make it easier to compare various countermeasures in the future. As a result, the emphasis of this study is on synthesizing the performance metrics. In particular, the performance metrics are compared based on three major dimensions: functionality dimension, spatial-temporal dimension, and equity dimension. Best practices of metrics selection are discussed. This synthesis approach can help safety program developers comprehend how to choose the performance measures based on their own goals while taking a thorough evaluation technique into consideration.

3. Methodology and Materials

The methodologies widely utilized in systematic literature reviews were applied in this research to gather a broader and more balanced sample of empirical studies than those that traditional reviews usually depend on (Dibaj et al. 2021). In particular, a standard reporting procedure, PRISMA, was utilized for reporting the inclusion and exclusion criteria in this systematic literature review (Page et al. 2021) .

The literature search was conducted between July 2022 and April 2023. The literature search was conducted in four widely popular academic databases: Transport Research International Documentation (TRID) database, Taylor Francis database, Science Direct, and Web of Science. To ensure an efficient literature search process, different keywords and operations were used when identifying a set of relevant studies. This search process involved using three sets of keywords. This process is summarized in Table 1. The first set of keywords were "*distracted driving*" and they were restricted to document titles to weed out any records that are not relevant to this general subject. The second group of keywords were "*effectiveness or evaluation*" to restrict the paper's content to include an evaluation of the prevention and intervention programs against distracted driving. Based on the preliminary search results, people's behavior and/or their behavior impacts have primarily been used to evaluate the effectiveness of the distracted driving countermeasures. As a result, the third set of keywords were "*behavior, crash, loss, or attitude*." The "AND" operators were used to connect them in order to limit the records that contain these three types of elements. Using the "AND" operators also help contain important performance indicators as well as keep the number of search results within reason. In total, 271 records in total were identified using the search criteria.

Table 1. Keywords used for systematic search

To identify the most relevant studies from the literature search process, the present study followed a systematic literature selection procedure, which involved four key steps, namely identification, screening, eligibility, and inclusion of documents. A schematic representation of the review process is presented in Figure 1. During the screening phase, a rigorous assessment of the abstracts of the articles was conducted to identify those that were aligned with the research focus of evaluating countermeasures against distracted driving. Articles that were less related, such as those that focused on detection of distracted driving or brain activities related to distracted driving, were excluded from further analysis. In the eligibility phase, the full texts of the remaining articles were scrutinized. Records that focused on preventive measures or crash causes were also excluded, leading to 44 articles for the next step. In the inclusion step, eight additional project reports on the evaluation of distracted driving prevention programs, such as enforcement, were included. These reports, which were conducted by transportation agencies such as state Departments of Transportation agencies and the National Highway Traffic Safety Administration, were collected mainly using Google search. **Five** additional articles were included from Google Scholar to include the latest relevant research. Most of these studies involves using technologies such as mobile phones or driving simulators to assess individual distracted driving behavior. Overall, **57** articles were included in this literature synthesis.

Figure 1. Literature review process

4. Results

4.1 Evaluation Metrics Overview

Among the 57 studies identified from the literature, the relevant information, including authors, year, type of countermeasure, data collection tool, and evaluation metrics, was comprehensively summarized and synthesized. Appendix A provides a summary of all the literature used for this review. Based on the review, the evaluation metrics can be divided into three categories based on the findings of the literature review: direct behavior indicators, behavior impact indicators, and indirect behavior indicators. Table 2 provides a summary of the information extracted from the current research about performance metrics. The paragraphs below provide detailed descriptions of eight different metrics.

Table 2. Summary of the performance metrics

4.1.1 Direct Behavior Indicators

The behaviors of distracted driving that have been observed are referred to as direct behavior indicators. Observed distracted driving behaviors have frequently been utilized to assess the effectiveness of legislation or enforcement. According to the Governor's Highway Safety Association, there are three main categories for distracted driving laws (Li, Gkritza, and Albrecht 2014): (1) texting laws explicitly banning text messaging while driving, (2) handheld cellphone bans in which all handheld mobile electronic devices are prohibited from use while driving for drivers of all ages, but hands-free device use is allowed, and (3) all-cellphone-use bans that prohibit the use of all hands-free and handheld devices only for novice drivers. As a result, this metric is typically the number of cases of texting or using a phone observed over a specific period of time.

A common statistic used in numerous states, including Virginia, Connecticut, and California, is observed distracted driving behavior (Chaudhary et al. 2012; Bommer 2018). In the evaluation, the frequency of eating, drinking, using a cell phone for talking or texting, and grooming were recorded within a 45- to 60-minute time window. The selection of the study and control locations was made in advance to guarantee that the population size, demographics, roadway characteristics, vehicle characteristics, and traffic density were equivalent or similar (Kidd and Chaudhary 2019; Retting et al. 2017). This type of evaluation required a large sample size, which frequently included one to ten thousand observations spread across a considerable geographic area, such as a state, city, or town (Fournier, Berry, and Frisch 2016; Retting et al. 2017). The environment in which the observed studies were conducted was also controlled. Particularly, this observational study was carried out in areas with relatively slow traffic, such as at stop signs, traffic signals, or close to parking garages, to make accurate observations and to ensure both the safety of the observers and reliable observations (Bommer 2018; Joseph et al. 2016).

Driving behavior has also been examined extensively using driving simulators. These metrics were typically developed to assess the effectiveness of in-vehicle feedback in real-time or training programs. The sample size in these studies was often under fifty, which was smaller than that of the enforcement evaluation metric; however, much richer personal driving information is gathered in simulator studies. In addition to the type of distractions, these metrics take into account lane-changing behaviors, braking reaction time to hazards, braking aggressiveness, steering wheel angle, and speed (Chan et al. 2010; Atchley and Chan 2011; Prevedouros, Miah, and Nathanail 2020; Bamney 2022; Alessandro Calvi, Benedetto, and D'Amico 2018). Direct phone use behavior such as eye movement, texting frequency, and/or dialing frequency can also be detected using the monitoring systems on people's phones (Peng et al. 2023; Usami et al. 2023; Aljasim and Kashef 2022; Ma et al. 2022). Other research made use of glance data gathered by eye trackers or in-car cameras that recorded the driver's face (Krishnan et al. 2015; Donmez, Boyle, and Lee 2007; Zangi et al. 2022; Monzer et al. 2022; Xue et al. 2023). The metrics included the number of glances, the number of blinks, and the duration of each glance at a particular point on or off the road. Currently, many of these types of evaluations have examined the outcomes of interventions for a particular demographic, such as young, novice, or elderly drivers.

4.1.2 Behavior Impact Indicators

Behavior impact indicators are referred to the impact or outcome of distracted driving. Many distracted driving countermeasures have been evaluated using crashes or near misses, especially for legislation or enforcement (Osman, Ye, and Ishak 2017; Huisingh et al. 2019). One research study, for instance, investigated the impact of legislation and discovered that the states with the strictest hand-held cellphone and texting laws had the greatest reductions in the number of fatalities involving teenage drivers (Flaherty et al. 2020). Common metrics used for evaluation include the severity and crash odds ratio (i.e., the frequency of a crash event occurrence in comparison to the likelihood of an event not occurring) (Choi et al. 2020). Crash databases, such as the Fatality Analysis Reporting System (FARS), National Automotive Sampling Systems (NASS), and General Estimates System (GES), were utilized for analysis (Hossain et al. 2023). These systems used the police accident report (PAR) as its main data source to track whether driver distraction played a role in the crash (Mitran et al. 2019). Because identifying crash causation can be difficult and police crash reports differ between jurisdictions, estimating the impact of distraction from these crash databases can be biased (Abay 2015). To address this gap, the National Motor Vehicle Collision Causation Survey (NMVCCS) and Naturalistic Driving Study (NDS) database served as the data sources that can offer more comprehensive and detailed crash information to help with the cause analysis (Bakhit, Guo, and Ishak 2018; Singh 2010).

To understand the safety and economic effects following the implementation of distracted driving countermeasures, financial loss metrics have also been used (Karl and Nyce 2020). Data from auto insurers, such as the Insurance Research Council (IRC), were used for the evaluation. These metrics include claim frequency (i.e., total number of paid claims by all insurers), loss cost (i.e., total dollar amount of losses paid by insurers), and private passenger automobile physical damage Loss Ratio (i.e., losses incurred divided by premiums earned) (Karl and Nyce 2019).

4.1.3 Indirect Behavior Indicators

Indirect behavior indicators, which are gathered through surveys, pertain to people's attitudes, knowledge, or awareness about distracted driving behavior or countermeasures. These metrics have been primarily used to assess the impact of education programs or campaigns (Adeola, Omorogbe, and Johnson 2016; Hurwitz et al. 2016). The use of indirect metrics supplemented the use of direct behavior or behavior impact indicators. To determine the level of distraction of the participants while driving, studies employed both individual lane deviations obtained via driving simulator self-report behavior such as perceived mental workload obtained via survey methods (Prevedouros, Miah, and Nathanail 2020; Chen et al. 2022). The indirect behavior metrics that have been utilized in past studies include people's understanding of state laws against distracted driving, attitudes toward it, perceptions of the risks associated with it, perceptions of their own capacity for it, and other people's propensity to do so. Utilizing several different indicators, including both direct and indirect indicators, can help in determining the different levels of behavior change. This can range from awareness of the negative effects of distracted driving to comprehension of the purpose of a countermeasure, to actually reduce the frequency of driving distracted (Berlin et al. 2021; Rupp, Gentzler, and Smither 2016; Tontodonato and Drinkard 2020; Buczek et al. 2022). Studies further have examined how usable a countermeasure was considered by the public. (Krishnan et al. 2015). For instance, despite believing that smartphone-based countermeasures had the potential to improve safety, people were shown to be reluctant to adopt these types of countermeasures (Oviedo-Trespalacios et al. 2020). This result demonstrates the importance of taking a human-centered perspective while designing and evaluating various countermeasures.

The technique of surveying is important in the development of indirect metrics (Riaz et al. 2019; Donmez, Merrikhpour, and Nooshabadi 2021). Table 3 includes some representative examples of the survey questions that can be used to evaluate different aspects of behavior changes that occur following the implementation of a countermeasure.

Table 3. Representative Examples of Survey Evaluation Method

4.2 Metrics Comparisons

4.2.1 Functionality Dimension

According to the performed literature study, there are many advantages and disadvantages associated with the evaluation criteria for distracted driving countermeasures. A summary of these advantages and disadvantages is provided in Table 4.

Table 4. Summary of the advantages and disadvantages of the metrics

The number of distracted driving cases has been measured via on-site observations to evaluate the effectiveness of enforcement in many states throughout multiple years; thus, it is easy to compare statistics across states and years using this indicator (Bommer 2018). However, to gather the required amount of data, these methods are costly and call for more staff and resources.

A high sample size is typically not required if the countermeasure is a training program or in-vehicle reminders as this metric, which describes each individual's driving behavior or gaze behavior, can provide rich information in a simulated environment. Additionally, research has shown that visual-manual distractions are the main cause of crashes, which has sparked the development of visual attention metrics as an important and useful statistic (Gershon et al. 2019). However, the use of this kind of statistic has some drawbacks as well. Eye trackers are sometimes unreliable, challenging to operate, and costly to incorporate in driving simulator studies. The utilization of eye trackers can be even more difficult in naturalistic driving experiments, where dozens of drivers often take part in their own vehicles. For some driving simulators, the actual driving environment or the countermeasure itself might not be accurately simulated or represented (Karlsson 2005). It may be difficult to evaluate the impact of the distracted driving countermeasure accurately in several different conditions as it lacks elements of reality, such as severe weather conditions or the understanding that, no matter what they do, they can never be involved in a real crash (A. Calvi and D'Amico 2013; Meuleners and Fraser 2015).

The direct consequence of distracted driving is a higher likelihood of crashes and financial losses. As a result, crash rate and financial losses are the strongest indicators of the effectiveness of a countermeasure, especially when considering those large-scale preventative efforts, such as enforcement and campaigns. However, there are challenges with utilizing crash data in particular as a performance metric. This is because some crash data might not have been recorded as police officers often only considered distraction as a cause when the crashes were severe (García-Herrero et al. 2021). The Naturalistic Driving Study data could be used to address this problem as it contains richer and more accurate information, such as different types of distracted driving (Simmons, Hicks, and Caird 2016). A notable disadvantage of these data is the absence of sufficient information regarding crashes with cell phones as the primary contributing factor. This is due to the low prevalence of smartphone distraction when this data collection began in 2002 (Neale et al. 2002). As a result, the data from the Naturalistic Driving Study may not be helpful if the purpose of the study is to assess the behavior related to technology or cell phone use.

Self-reported attitudes or actions can be easily gathered via survey methods. These metrics can be useful for developing potentially cheaper behavioral interventions such as using social norms (Carter et al. 2014). Furthermore, survey methodologies could be used to collect a wider range of sociodemographic and socioeconomic characteristics. Thus, they serve as a useful addition to other more direct metrics. These surveys shed light on who might be performing secondary tasks and how frequently they are performing them; however, the information is constrained by self-reported behavior, low response rates, and social desirability of responses. Self-reported attitudes or behaviors, in particular, cannot be a reliable indicator of or a reflection of real behavior. For instance, inquiries such as how many days a person texted or e-mailed while driving in the previous month can barely produce reliable results as people may not want to record their undesirable behavior or may just not remember it (Rudisill and Zhu 2015).

4.2.2 Spatial-temporal Dimension

In some analyses, metrics were further divided into time and geographical regions. In these studies, the frequency of distracted driving incidents was broken down by the types of roads, the counties, the severity of the road conditions, and/or other situational elements (Bommer 2018; Nowosielski, Trick, and Toxopeus 2018). One study, for instance, discovered that local roadways saw the most use of electronic devices compared to highways and secondary roadways. On the basis of this conclusion, policy implications were recommended regarding where more enforcement could be used on local roads to improve safety (Bommer 2018). Some analyses compared states to examine if different regions might respond to the same countermeasure, such as a texting ban, differently (Retting et al. 2017). For example, there was a significant decline in handheld phone use only among male drivers in Connecticut while there was a significant decline in texting only among female drivers in Massachusetts. It was found that these comparisons can aid in the modification of legislation in various states to meet their particular driving, demographic and/or travel patterns.

The temporal impact of a countermeasure was seldom studied in the literature. In order to evaluate the temporal impact of a distracted driving training program, one study included both first-time participants and those who had completed the training the previous year (Buczek et al. 2022).

4.2.3 Equity Dimension

In recent years, equity dimensions have been taken into account in many safety program evaluations. Some analyses took into account whether a countermeasure affects all demographics equally, such as in terms of gender and age. According to one observational study evaluating the impact of a handheld phone ban, there was a statistically significant decline in handheld phone use among male drivers in Connecticut. They also found that there was a statistically significant decline in texting among female drivers in Massachusetts (Retting et al. 2017). Another study that analyzed the crash risk for both drivers and non-drivers revealed that comprehensive handheld bans were linked to decreased driver fatality rates compared to no ban, but not for non-driver fatalities (Zhu et al. 2021). Surveying methods can also be used to infer the relationship between attitudes and gender. For instance, one study utilized subjective norm (i.e., the belief that an important person or group of people such as family members will disapprove of distracted driving) as a countermeasure in a high school distracted driving workshop. However, it was found that only female students are significantly affected by the program (Riaz et al. 2019).

Age has been taken into account in numerous research studies that use self-reported behavior or glances as a metric. According to studies, teenage drivers may be less receptive to countermeasures and less aware of the risks (Donmez, Boyle, and Lee 2007; Hassani et al. 2017). Based on these characteristics, some programs were developed primarily targeting novice teen drivers (Durbin et al. 2014).

While all of these research studies used demographic data in their analyses, they did so as a supplemental analysis. Studies that specifically aimed to investigate how distracted driving countermeasures can affect people equitably are lacking. While many current countermeasures have developed targeted training programs for teenagers, additional equity-related research can help to facilitate the development of targeted and effective countermeasures for other demographic groups. In this case, datasets or collection tools with more individual information, including demographics, attitudes, or cultural background information, could provide a better understanding of the role equity plays.

5. Discussions

5.1 A Framework for Countermeasure Evaluation

A general framework is summarized for selecting the performance metrics for evaluating distracted driving countermeasures based on literature review. Three steps are comprised of the selection process:

1) Choosing performance indicators
2) Selecting evaluation subjects
3) Gathering data

The specific recommendations are outlined below. These recommendations aim to help safety program developers choose performance metrics throughout the evaluation process.

1) Choosing performance indicators
    a) Driving simulators should be considered as the tool for evaluation. Diverse evaluation metrics could be developed with the help of this tool.
    b) Multiple metrics for each dimension (i.e., functionality, space and time, and equity) should be employed in the evaluation.
2) Selecting evaluation subjects
    a) The potential demographic variables that could affect the countermeasure's effectiveness should be taken into account. Different experiment groups could be separated based on demographic traits.
    b) When evaluating the effectiveness of a combination of countermeasures, different experiment groups can be divided to test both single and combinations of countermeasures.
3) Gathering data
    a) The evaluation results of similar countermeasures that were implemented at different locations or contexts could be used as a reference in the evaluation process.
    b) The effectiveness sample size should be calculated for the control and experiment groups before collecting the data.

5.2 Examples of Metrics Selection and Evaluation

This section offers a detailed discussion of examples using the proposed framework to aid transportation safety professionals in identifying appropriate metrics and evaluation methods. The discussion focusses on three countermeasure types: enforcement, technology (i.e., a smartphone app), and training programs. Each of these countermeasures undergoes evaluation using the proposed framework for assessing distracted driving countermeasures. The goal of this section is to provide transportation safety professionals with a comprehensive evaluation guide that is tailored to their specific countermeasure type.

These three examples were selected due to the following reasons. First, these studies evaluated three representative types of countermeasures that have been widely used recently. Second, these studies had detailed documentation of the evaluation processes and the rationale for each step. Last but not least, these studies used more than two different kinds of metrics. The range of metrics used suggests that these research studies have thoroughly assessed the countermeasures. As a result, using these studies as a reference, transportation safety experts can better understand what and why metrics for various countermeasures were chosen. A summary of the metrics that were used in these studies is provided in Table 5. The paragraphs that follow emphasize the metrics that were used and how they were measured in these three steps in the selected studies.

Table 5. Summary of the example studies

Step 1: Choosing performance indicators

Study A focused on the evaluation of enforcement. Campaigns and enforcement against using a hand-held cell phone while driving have been established in the cities of Hartford, Connecticut, and Syracuse, New York. The goal of these countermeasures is to raise both awareness of and compliance with current regulations. The direct behavior indicator in this study was chosen as the evaluation's principal metric (i.e., the number of distracted driving cases two months before and after the enforcement and campaigns). Survey methods were used to measure several additional indicators, such as public awareness of enforcement, campaigns, and attitudes toward distracted driving. These attitude metrics were selected in light of research from a previous campaign that showed people did not perceive distracted driving as a risk even when they were aware of it (Chaudhary et al. 2012).

Study B focused on the evaluation of technology. Recent developments in emerging technology can provide more tailored services to reduce distracted driving (L. Arnold and Horrey 2022). The types of metrics used are typically determined by the sensors that are available. The purpose of this study in particular was to reduce the driving risk caused by voice-to-text applications when texting. Individual driving behaviors, glances, and self-reported behavior were all analyzed in this study. These metrics were chosen based on the measurable potential crash risk indicators (Yager and Texas A&M Transportation Institute 2013).

Study C focused on a training program. Only self-reported metrics were used in this study. The metrics took into account perceived risk to oneself and others from a range of distracted driving behaviors that were used in numerous survey research studies (Buczek et al. 2022).

Step 2: Selecting evaluation subjects

In Study A (enforcement evaluation), the heaviest traffic road segments were used as the criterion for choosing the observation sites, which were also geographically scattered among the areas. Based on a reasonable demographic similarity in terms of population size, density, and median income, comparison (control) areas were selected (Chaudhary et al. 2012).

Study B (smartphone app evaluation) used an active recruitment strategy. Although this method of recruiting can make target group analysis easier, no eligibility requirements besides a valid driver's license were imposed for the subjects in this study (Yager and Texas A&M Transportation Institute 2013).

Study C (training evaluation) had a different goal than the first two: it sought to ascertain the long-term impact of a high school student-based training program. As a result, the study recruited students from the same high school who were both first-time participants and those who had previously taken part in the program. For a less biased comparison of the participants' behavior change after the training, the demographics of the two groups of participants were kept to be similar (Buczek et al. 2022).

Step 3: Gathering data

In both Study A and Study C, changes in the indicators were quantitatively identified using the chi-square test. In Study A (enforcement evaluation), sixty minutes of observation were conducted as part of the enforcement evaluation at each site. A target of 1000 respondents to the survey was set. Due to the limits in the ability to control other confounding factors such as demographics at the same time, this extensive investigation was necessary in order to ensure that the findings regarding the significance of the behavior change were valid (Chaudhary et al. 2012).

Study C (training evaluation) followed the traditional sample size calculation procedure to select the minimal survey sample size (i.e., 157) based on the effect size, power, and significance level (Buczek et al. 2022).

A qualitative analytic approach was employed in Study C (smartphone app evaluation). Altogether 239 people were recruited. A 2009 Ford Explorer with sensors to record the performance indicators served as the data collection vehicle. A GPS receiver, a response button, and an eye-tracking system comprised the sensors (Yager and Texas A&M Transportation Institute 2013).

5.3 Summary of Findings and Policy Implications

A comprehensive summary and comparisons of the distracted driving countermeasure evaluation metrics were presented in this review paper. Direct behavior indicators, behavior impact indicators, and indirect behavior indicators made up the three categories of evaluation metrics. Eight different metrics were examined and discussed. The discussion focused on the functionality dimension (i.e., the tradeoffs of different countermeasures), equity dimension (i.e., impact on different demographics), and spatial-temporal dimension (i.e., impact at different locations and times). Examples of the metrics selection and measurements were provided.

Several policy implications were synthesized from the review. The following paragraphs describe the four identified policy implications in further detail.

Policy Implication 1: Multi-faceted analysis could be adopted in more evaluations

Multiple studies used a range of metrics to analyze the effects of the countermeasures from various perspectives. To guarantee that the evaluation of the countermeasures is thorough, this multifaceted analysis technique using more than one metric should preferably be used. Alternative direct behavioral indications and the viability of other data-gathering technologies can be explored. More research can be done to examine how various metrics interact and provide consistent results (e.g., observed driving behavior and self-reported behavior). The results offer suggestions for improving the choice of the appropriate collection of metrics used in the evaluation.

Policy Implication 2: Driving simulators could be utilized in more evaluations

Individual driving and gaze behaviors are among these countermeasures that are generally accurate because these experiments are frequently conducted in controlled environments. These metrics can capture drivers' actual behavior most of the time. Although many other safety analyses have already used driving simulators, the potential of utilizing them to assess the effectiveness of training programs or smartphone applications combatting distracted driving has yet to be fully investigated. It is possible that more quantitative crash indicators might be developed to accurately and comprehensively evaluate the impact of distracted driving countermeasures with the addition of more advanced capabilities to driving simulators and other behavior monitoring devices. Further, it is easier to study distracted driving directly, as the road user does not have the capacity to be harmed, as in a real-life roadway scenario.

Policy Implication 3: Temporal dimensions could also be considered

Temporal dimensions can reveal a countermeasure's long-term effects. Studies have shown that some countermeasures may not be effective long-term and may need to be modified depending on how people respond to them (Buczek et al. 2022). It is necessary to employ many evaluation cycles to track the behavior change (e.g., every 1~2 years). These assessments can also demonstrate if particular metrics are robust enough to represent how countermeasures' effects change over time.

Policy Implication 4: Equity considerations should be incorporated more in future evaluations

Many safety programs emphasize equity as a key concept to increase everyone's safety. Promoting socially equitable outcomes should be one of the primary goals of distracted driving prevention strategies. This indicates that it is not preferable to see if only a select set of drivers with similar characteristics is influenced by a countermeasure, as it may have a lower or higher impact on other groups. Understanding different people's acceptability and receptiveness is also important to develop human-centric technology, which more people will be willing to use and comply with. The concept of equity has not been fully taken into account, despite the fact that many recent studies have included demographic considerations in their evaluations. However, a challenge to overcome to include this type of analysis is the additional time and cost load associated with recruiting and analyzing a more diverse pool of subjects.

5.4 Research Trends

Recent research trends in the field of distracted driving countermeasures evaluation have shown a shift towards the use of advanced technologies such as driving simulators and smart-phone based technologies to develop more accurate evaluation metrics. These evaluation metrics aim to more accurately capture the distracted driving behavior.

One of the future trends identified in the systematic review is the increasing use of driving simulators for data collection and evaluation (L. Arnold and Horrey 2022). Driving simulators are powerful tools that can provide a controlled and safe environment for testing different scenarios and evaluating the effectiveness of countermeasures. In addition, driving simulators can collect a wide range of data including gaze patterns, driving behavior, physiological responses, and self-reported measures. The use of driving simulators can help researchers to obtain more nuanced data and identify the specific factors that contribute to distracted driving (Boot et al. 2015).

Another trend identified in the systematic review is the use of smart-phone based technologies for real-time monitoring and evaluation of distracted driving behavior (Dumitru et al. 2018; Oviedo-Trespalacios et al. 2019). Recent advances in image processing algorithms have made it possible to detect facial expressions and other physical indicators of distracted driving, such as texting or talking on the phone while driving. These technologies can be integrated with real-time monitoring systems to provide immediate feedback and warnings to drivers when they engage in distracting behaviors.

Furthermore, as more and more distracted driving countermeasures are being tested in different regions in the future, benchmark evaluation results should become more available (Sajid Hasan et al. 2022). This will allow for more comparisons to be conducted between different evaluation results, providing researchers with a better understanding of the effectiveness of countermeasures in different contexts. By examining demographic, geographic, and implementation details, researchers can identify the underlying reasons for these differences, providing valuable insights into how to improve the effectiveness of countermeasures. These trends have the potential to significantly improve our understanding of distracted driving behavior and help us develop more effective countermeasures to reduce the prevalence of this dangerous behavior.

5.5 Limitations

The publications that examined the causes of distracted driving, which might possibly have other metrics that indirectly reflect drivers' behavior, were excluded from the review. Professionals working in transportation safety may have a broader understanding of performance metrics to utilize in their evaluations from an additional review that incorporates this information. Furthermore, there are some technology-based countermeasures, such as digital campaigns, that were created via web tools or social media. Thus, browsing information such as page views and time was used for evaluation (Ehrlich, Costello, and Randall 2020). These metrics were not specifically examined in this paper due to the lack of adequate studies that have examined the associated countermeasures.

6. Conclusions

This paper serves as a comprehensive review of the literature, summarizing the evaluation metrics applied for the current distracted driving countermeasures and providing some guiding information for evaluation metrics selection. Eight metrics that were used for evaluating the distracted driving countermeasures were examined and discussed. Three examples of metric selection and data collection were presented. This paper contributes to the body of knowledge by discussing the implications for policy on how to enhance the thoroughness and accuracy of the evaluation of distracted driving countermeasures. The primary takeaways are summarized in the following bullet points.

- Different types of metrics and secondary dimensions should be employed in future evaluations to provide insight into the countermeasure impact from multiple angles. In particular, equity should be considered to understand the varying impact on different demographic groups.
- More advanced data collection tools focusing on individual behavior and performance should be explored. Direct behavior indicators should also be developed based on the data collection tools.
- The long-term impact of a countermeasure should be considered in future evaluations.

AUTHOR CONTRIBUTIONS

The authors confirm their contribution to the paper as follows: study conception and design: Meiyu (Melrose) Pan, Alyssa Ryan; data collection: Meiyu (Melrose) Pan; analysis and interpretation of results: Meiyu (Melrose) Pan, Alyssa Ryan; draft manuscript preparation: Meiyu (Melrose) Pan, Alyssa Ryan. All authors reviewed the results and approved the final version of the manuscript.

References

Abay, Kibrom A. 2015. “Investigating the Nature and Impact of Reporting Bias in Road Crash Data.” *Transportation Research Part A: Policy and Practice* 71 (January): 31–45. https://doi.org/10.1016/j.tra.2014.11.002.

Adeola, Ruth, Ashleigh Omorogbe, and Abree Johnson. 2016. “Get the Message: A Teen Distracted Driving Program.” *Journal of Trauma Nursing | JTN* 23 (6): 312–20. https://doi.org/10.1097/JTN.0000000000000240.

Aljasim, Mustafa, and Rasha Kashef. 2022. “E2DR: A Deep Learning Ensemble-Based Driver Distraction Detection with Recommendations Model.” *Sensors* 22 (5): 1858. https://doi.org/10.3390/s22051858.

Arnold, L. S., A. J. Benson, B. C. Tefft, D. Barragan, L. Jin, S. Kolek, and W. J. Horrey. 2019. "Effectiveness of Distracted Driving Countermeasures: A Review of the Literature," November. https://trid.trb.org/view/1669958.

Arnold, Lindsay, and William J. Horrey. 2022. "Effectiveness of Distracted Driving Countermeasures: An Expanded and Updated Review of the Scientific and Gray Literatures." *AAA Foundation for Traffic Safety*, March. https://aaafoundation.org/effectiveness-of-distracted-driving-countermeasures-an-expanded-and-updated-review-of-the-scientific-and-gray-literatures/.

Atchley, Paul, and Mark Chan. 2011. "Potential Benefits and Costs of Concurrent Task Engagement to Maintain Vigilance: A Driving Simulator Investigation." *Human Factors* 53 (1): 3–12. https://doi.org/10.1177/0018720810391215.

Bakhit, Peter R., BeiBei Guo, and Sherif Ishak. 2018. "Crash and Near-Crash Risk Assessment of Distracted Driving and Engagement in Secondary Tasks: A Naturalistic Driving Study." *Transportation Research Record* 2672 (38): 245–54. https://doi.org/10.1177/0361198118772703.

Bamney, Anshu. 2022. "Examining the Relationship Between Driver Distraction, Crash, and Near-Crash Risk Using Naturalistic Driving Data." Ph.D., United States -- Michigan: Michigan State University. https://www.proquest.com/docview/2707713887/abstract/58BF5550C1FE4620PQ/1.

Berlin, Helena, Courtney Coughenour, Jennifer Pharr, Timothy J. Bungum, Heidi Manlove, and Guogan Shan. 2021. "The Impact of an Educational Intervention on Distracted Driving Knowledge, Attitudes, and Behaviors Among College Students." *Journal of Community Health* 46 (6): 1236–43. https://doi.org/10.1007/s10900-021-01014-y.

Bommer, William H. 2018. "Observational Study of Distracted Driving Due to Electronic Device Use Among California Drivers for 2018 Summary Report Provided to the California Office of Traffic Safety." *The California Office of Traffic Safety, Fresno, CA*, 12.

Boot, Walter R., Neil Charness, Ainsley Mitchum, Nelson Roque, Cary Stothart, Kimberly Barajas, and Florida State University. Dept. of Psychology. 2015. "Driving Simulator Studies of the Effectiveness of Countermeasures to Prevent Wrong-Way Crashes." BDV30-977–10. https://rosap.ntl.bts.gov/view/dot/39272.

Buczek, Lindsay, Laura K. Gryder, Samantha Slinkard-Barnum, Kavita Batra, Cassandra Trummel, Allison G. McNickle, Douglas R. Fraser, Deborah A. Kuhls, and Paul J. Chestovich. 2022. "Evaluating Long-Term Outcomes of a High School-Based Impaired and Distracted Driving Prevention Program." *Healthcare* 10 (3): 474. https://doi.org/10.3390/healthcare10030474.

Calvi, A., and F. D'Amico. 2013. "A Study of the Effects of Road Tunnel on Driver Behavior and Road Safety Using Driving Simulator." *Advances in Transportation Studies* 30 (July). https://trid.trb.org/view/1277154.

Calvi, Alessandro, Andrea Benedetto, and Fabrizio D'Amico. 2018. "Investigating Driver Reaction Time and Speed during Mobile Phone Conversations with a Lead Vehicle in Front: A Driving Simulator Comprehensive Study." *Journal of Transportation Safety & Security* 10 (1–2): 5–24. https://doi.org/10.1080/19439962.2017.1310161.

Carter, Patrick M., C. Raymond Bingham, Jennifer S. Zakrajsek, Jean T. Shope, and Tina B. Sayer. 2014. "Social Norms and Risk Perception: Predictors of Distracted Driving Behavior Among Novice Adolescent Drivers." *Journal of Adolescent Health*, Driver Distraction: A Perennial but Preventable Public Health Threat to Adolescents, 54 (5, Supplement): S32–41. https://doi.org/10.1016/j.jadohealth.2014.01.008.

Chan, Elsa, Anuj K. Pradhan, Alexander Pollatsek, Michael A. Knodler, and Donald L. Fisher. 2010. "Are Driving Simulators Effective Tools for Evaluating Novice Drivers' Hazard Anticipation, Speed Management, and Attention Maintenance Skills?" *Transportation Research Part F: Traffic Psychology and Behaviour* 13 (5): 343–53. https://doi.org/10.1016/j.trf.2010.04.001.

Chaudhary, Neil K., Tara Casanova-Powell, Linda Cosgrove, Ian Reagan, Allan Williams, and Inc. Preusser Research Group. 2012. "Evaluation of NHTSA Distracted Driving Demonstration Projects in Connecticut and New York." DOT HS 811 635. https://rosap.ntl.bts.gov/view/dot/1959.

Chen, Tiantian, Oscar Oviedo-Trespalacios, N. N. Sze, and Sikai Chen. 2022. “Distractions by Work-Related Activities: The Impact of Ride-Hailing App and Radio System on Male Taxi Drivers.” *Accident Analysis & Prevention* 178 (December): 106849. https://doi.org/10.1016/j.aap.2022.106849.

Choi, Yoon-Young, Dong-Kyu Kim, Seung-Young Kho, and Peter Y. Park. 2020. “Safety Evaluation of Four Safety Countermeasures on Freeways in South Korea.” *Journal of Transportation Safety & Security* 12 (7): 945–58. https://doi.org/10.1080/19439962.2018.1564945.

Dibaj, Samira, Aryan Hosseinzadeh, Miloš N. Mladenović, and Robert Kluger. 2021. “Where Have Shared E-Scooters Taken Us So Far? A Review of Mobility Patterns, Usage Frequency, and Personas.” *Sustainability* 13 (21): 11792. https://doi.org/10.3390/su132111792.

Donmez, Birsen, Linda Ng Boyle, and John D. Lee. 2007. “Safety Implications of Providing Real-Time Feedback to Distracted Drivers.” *Accident Analysis & Prevention* 39 (3): 581–90. https://doi.org/10.1016/j.aap.2006.10.003.

Donmez, Birsen, Maryam Merrikhpour, and Mehdi Hoseinzadeh Nooshabadi. 2021. “Mitigating Teen Driver Distraction: In-Vehicle Feedback Based on Peer Social Norms.” *Human Factors* 63 (3): 503–18. https://doi.org/10.1177/0018720819891285.

Dumitru, Adrian Iulian, Teodora Girbacia, Răzvan Gabriel Boboc, Cristian-Cezar Postelnicu, and Gheorghe-Leonte Mogan. 2018. “Effects of Smartphone Based Advanced Driver Assistance System on Distracted Driving Behavior: A Simulator Study.” *Computers in Human Behavior* 83 (June): 1–7. https://doi.org/10.1016/j.chb.2018.01.011.

Durbin, Dennis R, Daniel V McGehee, Donald Fisher, and Anne McCartt. 2014. “Special Considerations in Distracted Driving with Teens.” *Annals of Advances in Automotive Medicine* 58 (March): 69–83.

Ehrlich, Peter F., Beth Costello, and Amy Randall. 2020. “Preventing Distracted Driving: A Program from Initiation through to Evaluation.” *The American Journal of Surgery* 219 (6): 1045–49. https://doi.org/10.1016/j.amjsurg.2019.07.043.

Fisher, D L, A P Pollatsek, and A Pradhan. 2006. “Can Novice Drivers Be Trained to Scan for Information That Will Reduce Their Likelihood of a Crash?” *Injury Prevention* 12 (Suppl 1): i25–29. https://doi.org/10.1136/ip.2006.012021.

Flaherty, Michael R., Alexander M. Kim, Michael D. Salt, and Lois K. Lee. 2020. “Distracted Driving Laws and Motor Vehicle Crash Fatalities.” *Pediatrics* 145 (6): e20193621. https://doi.org/10.1542/peds.2019-3621.

Fournier, Angela K., Thomas D. Berry, and Sarah Frisch. 2016. “It Can W8: A Community Intervention to Decrease Distracted Driving.” *Journal of Prevention & Intervention in the Community* 44 (3): 186–98. https://doi.org/10.1080/10852352.2016.1166814.

García-Herrero, Susana, Juan Diego Febres, Wafa Boulagouas, José Manuel Gutiérrez, and Miguel Ángel Mariscal Saldaña. 2021. “Assessment of the Influence of Technology-Based Distracted Driving on Drivers’ Infractions and Their Subsequent Impact on Traffic Accidents Severity.” *International Journal of Environmental Research and Public Health* 18 (13): 7155. https://doi.org/10.3390/ijerph18137155.

Gershon, Pnina, Kellienne R. Sita, Chunming Zhu, Johnathon P. Ehsani, Sheila G. Klauer, Tom A. Dingus, and Bruce G. Simons-Morton. 2019. “Distracted Driving, Visual Inattention, and Crash Risk Among Teenage Drivers.” *American Journal of Preventive Medicine* 56 (4): 494–500. https://doi.org/10.1016/j.amepre.2018.11.024.

Groeger, J. A., and A. P. Banks. 2007. “Anticipating the Content and Circumstances of Skill Transfer: Unrealistic Expectations of Driver Training and Graduated Licensing?” *Ergonomics* 50 (8): 1250–63. https://doi.org/10.1080/00140130701318723.

Hassani, Sahar, Erin H. Kelly, Jennifer Smith, Sara Thorpe, Fatima H. Sozzer, Paul Atchley, Elroy Sullivan, Dean Larson, and Lawrence C. Vogel. 2017. “Preventing Distracted Driving among College Students: Addressing Smartphone Use.” *Accident Analysis and Prevention* 99 (Part A): 297–305. https://doi.org/10.1016/j.aap.2016.12.004.

Hedlund, James. 2011. "Distracted Driving: What Research Shows and What States Can Do." https://trid.trb.org/view/1106493.

Hill, Linda, Jill Rybar, Jana Jahns, Tanya Lozano, and Sara Baird. 2020. "'Just Drive': An Employee-Based Intervention to Reduce Distracted Driving." *Journal of Community Health* 45 (2): 370–76. https://doi.org/10.1007/s10900-019-00752-4.

Hill, Linda, Jill Rybar, Tara Styer, Ethan Fram, Gina Merchant, and Amelia Eastman. 2015. "Prevalence of and Attitudes About Distracted Driving in College Students." *Traffic Injury Prevention* 16 (4): 362–67. https://doi.org/10.1080/15389588.2014.949340.

Hossain, Md Mahmud, Huaguo Zhou, Subasish Das, Xiaoduan Sun, and Ahmed Hossain. 2023. "Young Drivers and Cellphone Distraction: Pattern Recognition from Fatal Crashes." *Journal of Transportation Safety & Security* 15 (3): 239–64. https://doi.org/10.1080/19439962.2022.2048763.

Huisingh, Carrie, Cynthia Owsley, Emily B Levitan, Marguerite R Irvin, Paul MacLennan, and Gerald McGwin. 2019. "Distracted Driving and Risk of Crash or Near-Crash Involvement Among Older Drivers Using Naturalistic Driving Data With a Case-Crossover Study Design." *The Journals of Gerontology: Series A* 74 (4): 550–55. https://doi.org/10.1093/gerona/gly119.

Hurwitz, David S., Erika Miller, Mafruhatul Jannat, Linda Ng Boyle, Shane Brown, Ahmed Abdel-Rahim, and Haizhong Wang. 2016. "Improving Teenage Driver Perceptions Regarding the Impact of Distracted Driving in the Pacific Northwest." *Journal of Transportation Safety & Security* 8 (2): 148–63. https://doi.org/10.1080/19439962.2014.997329.

Joseph, Bellal, Ansab Haider, Ahmed Hassan, Narong Kulvatunyou, Sandeep Bains, Andrew Tang, Bardiya Zangbar, et al. 2016. "Injury Prevention Programs against Distracted Driving: Are They Effective?" *Journal of Trauma and Acute Care Surgery* 81 (1): 144–48. https://doi.org/10.1097/TA.0000000000001022.

Karl, J. Bradley, and Charles Nyce. 2019. "How Cellphone Bans Affect Automobile Insurance Markets." *Journal of Risk and Insurance* 86 (3): 567–93. https://doi.org/10.1111/jori.12224.

———. 2020. "The Effect of Distracted Driving Laws on Automobile Liability Insurance Claims." *North American Actuarial Journal* 24 (4): 593–610. https://doi.org/10.1080/10920277.2019.1683041.

Karlsson, R. 2005. "Evaluating Driver Distraction Countermeasures." *VTI NOTAT*, no. 38A-2005. https://trid.trb.org/view/814188.

Khan, Abdul Basit, Rajat Agrawal, S. S. Jain, and Ankit Choudhary. 2022. "Review of Distracted Driving in Young Drivers: Strategies for Management of Behavioural Patterns." *International Journal of Crashworthiness* 27 (5): 1532–44. https://doi.org/10.1080/13588265.2021.1926847.

Kidd, David G., and Neil K. Chaudhary. 2019. "Changes in the Sources of Distracted Driving among Northern Virginia Drivers in 2014 and 2018: A Comparison of Results from Two Roadside Observation Surveys." *Journal of Safety Research* 68 (February): 131–38. https://doi.org/10.1016/j.jsr.2018.12.004.

Krishnan, Akhilesh, Siby Samuel, Matthew Romoser, and Donald Fisher. 2015. "A Study to Evaluate the Effectiveness of a Tablet-Based Training Program for Younger Drivers: Addressing Distraction Head On." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59 (1): 1651–55. https://doi.org/10.1177/1541931215591357.

Li, Wanjun, Konstantina Gkritza, and Chris Albrecht. 2014. "The Culture of Distracted Driving: Evidence from a Public Opinion Survey in Iowa." *Transportation Research Part F: Traffic Psychology and Behaviour*, SI: Traffic safety culture, 26 (September): 337–47. https://doi.org/10.1016/j.trf.2014.01.002.

Ma, Yanli, Gaofeng Gu, Biqing Yin, Shouming Qi, Ke Chen, and Chingyao Chan. 2022. "Support Vector Machines for the Identification of Real-Time Driving Distraction Using in-Vehicle Information Systems." *Journal of Transportation Safety & Security* 14 (2): 232–55. https://doi.org/10.1080/19439962.2020.1774019.

Meuleners, Lynn, and Michelle Fraser. 2015. "A Validation Study of Driving Errors Using a Driving Simulator." *Transportation Research Part F: Traffic Psychology and Behaviour* 29 (February): 14–21. https://doi.org/10.1016/j.trf.2014.11.009.

Mitran, Elisabeta, Tess Ellender, Dortha Cummins, and Louisiana Transportation Research Center. 2019. "Distracted Driving: Strategies and State of the Practices." FHWA/LA.18/17-02TA-SA. https://rosap.ntl.bts.gov/view/dot/44237.

Monzer, Dana, Assem Abou Ali, Maya Abou-Zeid, and Nadine Marie Moacdieh. 2022. "Voice Messaging While Driving: Effects on Driving Performance and Attention." *Applied Ergonomics* 101 (May): 103692. https://doi.org/10.1016/j.apergo.2022.103692.

Neale, V. L., S. G. Klauer, R. R. Knipling, T. A. Dingus, G. T. Holbrook, and A. Petersen. 2002. "THE 100 CAR NATURALISTIC DRIVING STUDY, PHASE I - EXPERIMENTAL DESIGN," December. https://trid.trb.org/view/643552.

NHTSA. 2020. "Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices." https://www.ghsa.org/sites/default/files/2021-09/Countermeasures%20That%20Work%2C%2010th%20Edition.pdf.

———. 2021. "Distracted Driving | NHTSA." Text. 2021. https://www.nhtsa.gov/risky-driving/distracted-driving.

Nowosielski, Robert J., Lana M. Trick, and Ryan Toxopeus. 2018. "Good Distractions: Testing the Effects of Listening to an Audiobook on Driving Performance in Simple and Complex Road Environments." *Accident Analysis & Prevention* 111 (February): 202–9. https://doi.org/10.1016/j.aap.2017.11.033.

Osman, Osama A., Mengqiu Ye, and Sherif Ishak. 2017. "Crash Risk Analysis of Distracted Driving Behavior: Influence of Secondary Task Engagement and Driver Characteristics." *Driving Assesment Conference* 9 (2017). https://doi.org/10.17077/drivingassessment.1621.

Oviedo-Trespalacios, Oscar, Mark King, Atiyeh Vaezipour, and Verity Truelove. 2019. "Can Our Phones Keep Us Safe? A Content Analysis of Smartphone Applications to Prevent Mobile Phone Distracted Driving." *Transportation Research Part F: Traffic Psychology and Behaviour* 60 (January): 657–68. https://doi.org/10.1016/j.trf.2018.11.017.

Oviedo-Trespalacios, Oscar, Atiyeh Vaezipour, Verity Truelove, Sherrie-Anne Kaye, and Mark King. 2020. "'They Would Call Me, and I Would Need to Know Because It Is like Life and Death': A Qualitative Examination of the Acceptability of Smartphone Applications Designed to Reduce Mobile Phone Use While Driving." *Transportation Research Part F: Traffic Psychology and Behaviour* 73 (August): 499–513. https://doi.org/10.1016/j.trf.2020.06.007.

Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *International Journal of Surgery* 88 (April): 105906. https://doi.org/10.1016/j.ijsu.2021.105906.

Peng, Yongxin, Guohua Song, Manze Guo, Lingtao Wu, and Lei Yu. 2023. "Investigating the Impact of Environmental and Temporal Features on Mobile Phone Distracted Driving Behavior Using Phone Use Data." *Accident Analysis & Prevention* 180 (February): 106925. https://doi.org/10.1016/j.aap.2022.106925.

Prevedouros, Panos, M. Mintu Miah, and Eftihia Nathanail. 2020. "Effects of Reading Text While Driving: A Driving Simulator Study," February. https://scholarworks.alaska.edu/handle/11122/11070.

(Raymond) Huang, Zirui, Ali Arian, Yuqiu (Rachael) Yuan, and Yi-Chang Chiu. 2020. "Using Conditional Generative Adversarial Nets and Heat Maps with Simulation-Accelerated Training to Predict the Spatiotemporal Impacts of Highway Incidents." *Transportation Research Record* 2674 (8): 836–49. https://doi.org/10.1177/0361198120925069.

Regan, Michael, and Oscar Oviedo-Trespalacios. 2022. "Driver Distraction: Mechanisms, Evidence, Prevention, and Mitigation." In *The Vision Zero Handbook : Theory, Technology and Management for a Zero Casualty Policy*, edited by Karin Edvardsson Björnberg, Matts-Åke

Belin, Sven Ove Hansson, and Claes Tingvall, 1–62. Cham, Switzerland: Springer. https://eprints.qut.edu.au/235465/.

Retting, Richard, Karen Sprattler, Heather Rothenberg, Thomas Sexton, and M. Davis and Company. 2017. "Evaluating the Enforceability of Texting Laws: Strategies Tested in Connecticut and Massachusetts." DOT HS 812 367. https://doi.org/10.21949/1525904.

Riaz, Malik Sarmad, Ariane Cuenen, Stijn Dhondt, Helen Craps, Davy Janssens, Geert Wets, Tom Brijs, and Kris Brijs. 2019. "Evaluation of a Road Safety Education Program Based on Driving Under Influence and Traffic Risks for Higher Secondary School Students in Belgium." *Safety* 5 (2): 34. https://doi.org/10.3390/safety5020034.

Richard, Christian M., Kelly Magee, Paige Bacon-Abdelmoteleb, and James L. Brown. 2018. *Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices, Ninth Edition, 2017*. Edited by Battelle Memorial Institute. NHTSA BSR Countermeasures That Work. https://doi.org/10.21949/1525943.

Rudisill, Toni M., and Motao Zhu. 2015. "The Association between States' Texting Regulations and the Prevalence of Texting While Driving among U.S. High School Students." *Annals of Epidemiology* 25 (12): 888–93. https://doi.org/10.1016/j.annepidem.2015.09.004.

Rupp, Michael A., Marc D. Gentzler, and Janan A. Smither. 2016. "Driving under the Influence of Distraction: Examining Dissociations between Risk Perception and Engagement in Distracted Driving." *Accident Analysis & Prevention* 97 (December): 220–30. https://doi.org/10.1016/j.aap.2016.09.003.

Sajid Hasan, Ahmed, Mohammad Jalayer, Eric Heitmann, and Joseph Weiss. 2022. "Distracted Driving Crashes: A Review on Data Collection, Analysis, and Crash Prevention Methods." *Transportation Research Record*, April, 03611981221083917. https://doi.org/10.1177/03611981221083917.

Simmons, Sarah M., Anne Hicks, and Jeff K. Caird. 2016. "Safety-Critical Event Risk Associated with Cell Phone Tasks as Measured in Naturalistic Driving Studies: A Systematic Review and Meta-Analysis." *Accident Analysis & Prevention* 87 (February): 161–69. https://doi.org/10.1016/j.aap.2015.11.015.

Singh, Santokh. 2010. "Distracted Driving and Driver, Roadway, and Environmental Factors," September. https://trid.trb.org/view/1116580.

Tontodonato, Pamela, and Allyson Drinkard. 2020. "Social Learning and Distracted Driving among Young Adults." *American Journal of Criminal Justice* 45 (5): 821–43. https://doi.org/10.1007/s12103-020-09516-6.

Usami, Davide Shingo, Brayan González-Hernández, Kristaq Pepkolaj, and Luca Persia. 2023. "A Methodology to Explore the Road Safety Impact of Fitness to Drive Solutions for Commercial Drivers: The PANACEA Project." *Transportation Research Procedia*, AIIT 3rd International Conference on Transport Infrastructure and Systems (TIS ROMA 2022), 15th-16th September 2022, Rome, Italy, 69 (January): 823–30. https://doi.org/10.1016/j.trpro.2023.02.241.

Vegega, Maria, Brian Jones, Chris Monk, and United States. Office of Impaired Driving and Occupant Protection. 2013. "Understanding the Effects of Distracted Driving and Developing Strategies to Reduce Resulting Deaths and Injuries: A Report to Congress." DOT HS 812 053. https://doi.org/10.21949/1525817.

Xue, Qingwan, Xingyue Wang, Yinghong Li, and Weiwei Guo. 2023. "Young Novice Drivers' Cognitive Distraction Detection: Comparing Support Vector Machines and Random Forest Model of Vehicle Control Behavior." *Sensors* 23 (3): 1345. https://doi.org/10.3390/s23031345.

Yager, Christine and Texas A&M Transportation Institute. 2013. "An Evaluation of the Effectiveness of Voice-to-Text Programs at Reducing Incidences of Distracted Driving." SWUTC/13/600451-00011-1. https://rosap.ntl.bts.gov/view/dot/25912.

Zangi, Noa, Rawan Srour-Zreik, Dana Ridel, Hadas Chassidim, and Avinoam Borowsky. 2022. "Driver Distraction and Its Effects on Partially Automated Driving Performance: A Driving Simulator

Study among Young-Experienced Drivers." *Accident Analysis & Prevention* 166 (March): 106565. https://doi.org/10.1016/j.aap.2022.106565.

Zhu, Motao, Sijun Shen, Donald A. Redelmeier, Li Li, Lai Wei, and Robert Foss. 2021. "Bans on Cellphone Use While Driving and Traffic Fatalities in the United States." *Epidemiology (Cambridge, Mass.)* 32 (5): 731–39. https://doi.org/10.1097/EDE.0000000000001391.

Appendix. 1

Table A-1. Summary of reviewed studies

	Author and year	Keywords	Evaluation Metrics	Countermeasure	Data Collection Tool
1	Chaudhary et al. 2012	-*	Number of distracted driving cases	Enforcement	On-site observation
2	Bommer 2018	-	Number of distracted driving cases	Enforcement	On-site observation
3	Kidd and Chaudhary 2019	Cellphone; Distracted driving; Distraction; Observation survey; Texting	Number of distracted driving cases	Enforcement	On-site observation
4	Retting et al. 2017	-	Number of distracted driving cases	Enforcement	On-site observation
5	Fournier, Berry, and Frisch 2016	Community intervention; distracted driving; prompting; texting	Number of distracted driving cases	Campus intervention	On-site observation
6	Joseph et al. 2016	Distracted driving; educational campaign; texting and driving; talking and driving	Number of distracted driving cases	Campus intervention	On-site observation
7	Chan et al. 2010	Driving Simulation; Novice Drivers; Eye Movements; Attention Maintenance; Distraction; Hazard Anticipation; Speed Management	Individual driving behavior	Driver training	Driving simulator
8	Atchley and Chan 2011	Countermeasures, vigilance, monotony, concurrent task, lane-keeping performance, driver safety	Individual driving behavior	Real-time feedback	Driving simulator
9	Prevedouros, Miah, and Nathanail 2020	CSET Minorities, Race, Fatality Analysis, FARS, Hawaiians, Pacific Islanders, Minorities, Alcohol, Speeding, Restraint Use, Age, Gender, Urban, Rural	Individual driving behavior	Real-time feedback	Driving simulator
10	Bamney 2022	Driver distraction, Case-crossover, Naturalistic driving	Individual driving behavior	Real-time feedback	Driving simulator
11	Alessandro Calvi, Benedetto, and D'Amico 2018	Driving performance; driving simulator; hands-free mobile; hand-held mobile; mobile phone use; road safety	Individual driving behavior	Real-time feedback	Driving simulator
12	Karlsson 2005	Driving Behavior, Accident Prevention, Attention, Distraction, Simulators, Eye Movements	Individual driving behavior	Real-time feedback	Driving simulator
13	A. Calvi and D'Amico 2013	Driving simulation; road tunnel; driving performance; tunnel effects	Individual driving behavior	Real-time feedback	Driving simulator
14	Meuleners and Fraser 2015	Validation study; Driving simulator; On-road testing	Individual driving behavior	Real-time feedback	Driving simulator
15	Yager and Texas A&M	Texting, Voice-to-Text, Speech-to-Text, Distracted Driving, Distraction, Mobile Device Use, Impairment	Individual driving behavior	Real-time feedback	Driving simulator

	Transportation Institute 2013				
16	Chen et al. 2022	Driver distraction; Taxi driver; Ride-hailing system; Driving simulator; Human factor; Professional driver	Individual driving behavior	Real-time feedback	Driving simulator
17	Donmez, Boyle, and Lee 2007	Driver distraction; Eye movements; Real-time feedback; In-vehicle information systems; Driving simulator	Individual driving behavior	Real-time feedback	Driving simulator
18	Zangi et al. 2022	Partially automated driving; Driving simulator; Driver distraction; Non-driving related task; Eye movements; Heart rate variability	Individual driving behavior	Real-time feedback	Driving simulator
19	Xue et al. 2023	Distraction driving; random forest model; support vector machines; vehicle control behavior; young novice driver.	Individual driving behavior	Real-time feedback	Driving simulator
20	Monzer et al. 2022	Voice messaging; TextingDriving distraction; Driving simulator; Eye tracking	Individual driving behavior	Real-time feedback	Driving simulator
21	Nowosielski, Trick, and Toxopeus 2018	Attention; Boredom; Distraction; Driving simulator; Executive function; Fatigue.	Individual driving behavior	Real-time feedback	Driving simulator
22	Aljasim and Kashef 2022	Deep learning; stacking; ensemble learning; distracted driving	Individual driving behavior	Real-time feedback	Mobile phone
23	Ma et al. 2022	Distraction detection; driver distraction; driving performance; in-vehicle information system; kernel function; SVM model	Individual driving behavior	Real-time feedback	Mobile phone
24	Krishnan et al. 2015	-	Individual driving behavior	Real-time feedback	Mobile phone
25	Dumitru et al. 2018	Advanced driver assistance system; Driving simulator; Smartphone driver distraction; Visual attention	Individual driving behavior	Real-time feedback	Mobile phone
26	Regan and Oviedo-Trespalacios 2022	Driver distraction · Distracted driving · Road safety · Theory · Impact · Countermeasures · Mitigation · Vision zero	Individual driving behavior	Vehicle technology	Mobile phone
27	Peng et al. 2023	Distracted driving; Mobile phone use; Structural equation modeling; Driving behavior; Safety	Glance	-**	Eye tracker
28	Durbin et al. 2014	-	Glance	Real-time feedback	Eye tracker
29	Usami et al. 2023	Impact assessment; commercial health toolkits; fitness to drive; road safety	Glance	Real-time feedback	Eye tracker
30	Flaherty et al. 2020	Distracted driving, statutes and laws, text messaging, mobile devices, traffic accidents	Fatality	Enforcement	Fatality Analysis Reporting System

31	Zhu et al. 2021	Automobile driving; Cause of death; Cell phone use; Traffic crashes; Text messaging	Fatality	Enforcement	Fatality Analysis Reporting System
32	Hossain et al. 2023	Young driver; correspondence analysis; fatal crash; crash data analysis	Fatality	Enforcement	Fatality Analysis Reporting System
33	Mitran et al. 2019	-	Fatality	Enforcement	Fatality Analysis Reporting System
34	Choi et al. 2020	Safety evaluation; automated speed enforcement cameras; chevron signs; delineator posts; shoulder rumble strips; empirical Bayes before and after method	Fatality	Enforcement	Crash database
35	García-Herrero et al. 2021	Road traffic accidents; technology-based distractions; aberrant infractions; speed infractions; bayesian network; traffic accidents severity	Fatality	-	Crash database
36	Gershon et al. 2019	-	Individual driving behavior	-	Real-world driving data
37	Osman, Ye, and Ishak 2017	-	Crash risk	-	Naturalistic Driving Study
38	Huisingh et al. 2019	Case-crossover; Driver distraction; Naturalistic driving.	Near crash	-	Naturalistic Driving Study
39	Simmons, Hicks, and Caird 2016	Meta-analysis; Research synthesis; Cell or mobile phones; Talking; Dialing; Texting; Naturalistic driving; Distracted driving	Near crash risk	-	Naturalistic Driving Study database
40	Bakhit, Guo, and Ishak 2018	-	Near crash risk	-	Naturalistic Driving Study database
41	Singh 2010	Associated factor, Internal sources, Cognitive activities, Inattention, Descriptive analysis, Bivariate analysis	Near crash risk	-	National Motor Vehicle Crash Causation Survey
42	Karl and Nyce 2020	-	Financial losses	Enforcement	Insurance database
43	Karl and Nyce 2019	-	Financial losses	Enforcement	Insurance database
44	Adeola, Omorogbe, and Johnson 2016	Distracted driving, Motor vehicle crash (MVC), Teenagers, Teens	Self-report attitudes	Educational programs	Survey
45	Hurwitz et al. 2016	Distracted driving; teenage drivers; transportation safety; education; outreach	Self-report attitudes	Educational programs	Survey

46	Carter et al. 2014	Motor vehicle crash; Adolescents; Distracted driving	Self-report attitudes	Educational programs	Survey
47	Rudisill and Zhu 2015	Adolescent; Automobile driving; Epidemiology; Text messaging	Self-report attitudes	Educational programs	Survey
48	Buczek et al. 2022	Distracted driving; driver safety; educational program; impaired driving; injury prevention; motor vehicle crash; teen drivers.	Self-report attitudes	Educational programs	Survey
49	Riaz et al. 2019	Traffic safety; driving under influence; traffic risks; education; pre-drivers; evaluation	Self-report attitudes	Educational programs	Survey
50	Fisher, Pollatsek, and Pradhan 2006	Novice drivers, risk recognition, driver training, eye movements, driving simulators	Self-report attitudes	Educational programs	Survey
51	Groeger and Banks 2007	Driving; Driver training; Transfer of training; Skill acquisition	Self-report attitudes	Educational programs	Survey
52	Rupp, Gentzler, and Smither 2016	Risk perception; Driver distraction; Structural equation modeling	Self-report attitudes	Behavior intervention	Survey
53	Oviedo-Trespalacios et al. 2020	User-centered design; Risky behaviors; Inattention; Ergonomics; Road safety	Self-report attitudes	Real-time feedback	Survey
54	Donmez, Merrikhpour, and Nooshabadi 2021	Driving simulator; gender; in-vehicle information systems; normative feedback; teenage driver distraction.	Self-report attitudes	Behavior intervention	Survey
55	Hassani et al. 2017	Distracted driving; College students; Prevention; Workshop; Smartphones	Self-report behavior	Educational programs	Survey
56	Berlin et al. 2021	Distracted driving; Health education; Intervention; Prevention; Undergraduate students.	Self-report behavior	Behavior intervention	Survey
57	Tontodonato and Drinkard 2020	Distracted driving; Texting while driving; Cell phone use while driving; Social learning theory	Self-report behavior	Behavior intervention	Survey

Notes:

*Some studies did not have keywords

** Some studies focused on developing performance metrics without specifically applying them for evaluating a certain countermeasure

Tables

Table 1. Keywords and operations used for systematic search

Constraints	Keywords & Operations
Document Title	"distracted driving"
	AND
General Content	"effectiveness" OR "evaluation" OR “metric”
	AND
General Content	“behavior” OR “crash” OR “loss” OR “attitude”

Table 2. Summary of the performance metrics

No.	Metric	Category	Specific metrics	Countermeasure	Data collection tool	Sample size	Secondary evaluation dimension	# Articles
1	Number of distracted driving cases	Direct behavior	Number of distracted driving cases within a certain time period	Enforcement	On-site observation	3000~31,000 (observations)	Demographics (gender); geographical area; time of day	5
2	Individual driving behavior	Direct behavior	Deviation of lane position; steering wheel angle; accelerator release time; speed	Driver training / real-time feedback	Driving simulator; mobile phone	20~40 (participants)	Demographics (age)	20
3	Glance	Direct behavior	Glance duration in a certain area; number of glances; number of blinks	Driver training / real-time feedback	Eye tracker	20~40 (participants)	Demographics (age)	3
4	Crash	Behavior impact	Crash odds ratio; severity; fatality	Enforcement	Crash database; Fatality Analysis Reporting System	40,000 (events)	Demographics (age)	6
5	Near crash	Behavior impact	Crash odds ratio	Enforcement	Naturalistic Driving Study database	200~25,000 (events)	Demographics (age, gender); vehicle characteristics; roadway; secondary task	6
6	Financial losses	Behavior impact	Claim frequency; loss cost	Enforcement	Insurance database	950~70,000 (observations)	Demographics (income; age; gender; unemployment)	2
7	Self-report attitudes	Indirect behavior	Knowledge/perceived usefulness of a countermeasure; perceived risk/capacity for distracted driving	Educational programs/campaigns / as supplements of other metrics	Survey	35~6000 (participants)	Demographics (age)	11
8	Self-report behavior	Indirect behavior	Number of distracted driving cases; number of citations	Educational programs/campaigns / as supplements of other metrics	Survey	35~6000 (participants)	Demographics (age, gender, ethnicity)	3

Table 3. Representative Examples of Survey Evaluation Method

Sources	Aspects	Questions (examples)
(Retting et al. 2017)	Knowledge about enforcement countermeasures	Do you think the texting law in [your state] is enforced? (yes/no)
(Hill et al. 2020)	Knowledge about the consequences	While driving, it is safer to talk with a hands-free device than with a handheld device. (true/false)
(Berlin et al. 2021)	Attitudes about the distracted driving behavior	I believe that it is alright to program a navigation system while driving. (true/false)
(Hill et al. 2015)	Perceived capability of distracted driving (Dunning-Kruger effect)	How capable are you of talking on a cell phone and driving? How capable are others talking on a cell phone and driving? (Likert scale)
(Riaz et al. 2019; Donmez, Merrikhpour, and Nooshabadi 2021)	Perceived distraction approval (social norm theory)	“My parents would disapprove if I ride with a driver who has drunk alcohol.” (true/false)
(Rudisill and Zhu 2015)	Self-report distracted driving behavior	During the past 30 days, on how many days did you text or e-mail while driving a car or other vehicle?”
(Retting et al. 2017)	Self-report number of citations	Have you received a ticket for talking on a handheld cellphone while driving IF EVER / IN THE PAST MONTH?

Table 4. Summary of the advantages and disadvantages of the metrics

No.	Metric	Advantages	Disadvantages
1	Number of distracted driving cases	Well-established; easy to compare the data among years	Costly
2	Individual driving behavior	Accurate; detailed information can be collected; more demographic information can be collected in the process	Costly
3	Glance	Essential indicators to crash; accurate; detailed information could be collected; more demographic information can be collected in the process	Costly; lack of realness
4	Crash	Direct outcome of distracted driving	Inaccurate reporting of crash data
5	Near crash	Rich data source from Naturalistic Driving Study	Less direct outcome of distracted driving
6	Financial losses	Direct outcome of distracted driving	Inaccurate reporting of crash data
7	Self-report attitudes	Cheaper and faster to collect; can facilitate the development of behavioral interventions; more demographic information can be collected using the same tool	Might not be related to actual behavior
8	Self-report behavior	Cheaper and faster to collect; more demographic information can be collected using the same tool	Might not reflect actual behavior

Table 5. Summary of the example studies

Label	Study	Author/year	Metrics
A	Enforcement evaluation	(Chaudhary et al. 2012)	(1) Enforcement hours (2) Number of texting drivers (3) Number of hand-held phone uses (4) Number of citations (5) If heard of the cell phone enforcement/slogan (6) Perceived importance for the police to enforce (7) Perceived chance of being ticketed if using a hand-held phone (8) Perceived strictness of the police to enforce this law (9) Ever got a ticket for using a hand-held phone (10) Ever got a ticket for using a hand-held phone in the past month
B	Smartphone app evaluation	(Yager and Texas A&M Transportation Institute 2013)	(1) Self-reported texting behavior (2) Feelings about texting while driving (3) Perceived voice-to-text capabilities and behavior (4) Opinions on texting while driving (5) Driver response time (6) Missed response events (7) Speed (8) Standard deviation of lane deviation (9) Eye gazes (10) Task completion (11) Subject performance rating
C	Training evaluation	(Buczek et al. 2022)	(1) Okay to talk on the phone while driving with both hands on the wheel? (2) Okay to read text messages ONLY at a stop light? (3) How likely you would stop a friend from practicing the following behaviors? (example choice: Driving 30 min after consuming two alcoholic drinks) (4) How likely you would stop a family member from practicing the following behaviors? (example choice: Driving 30 min after consuming two alcoholic drinks) (5) Self-reported distracted driving frequency

Figures

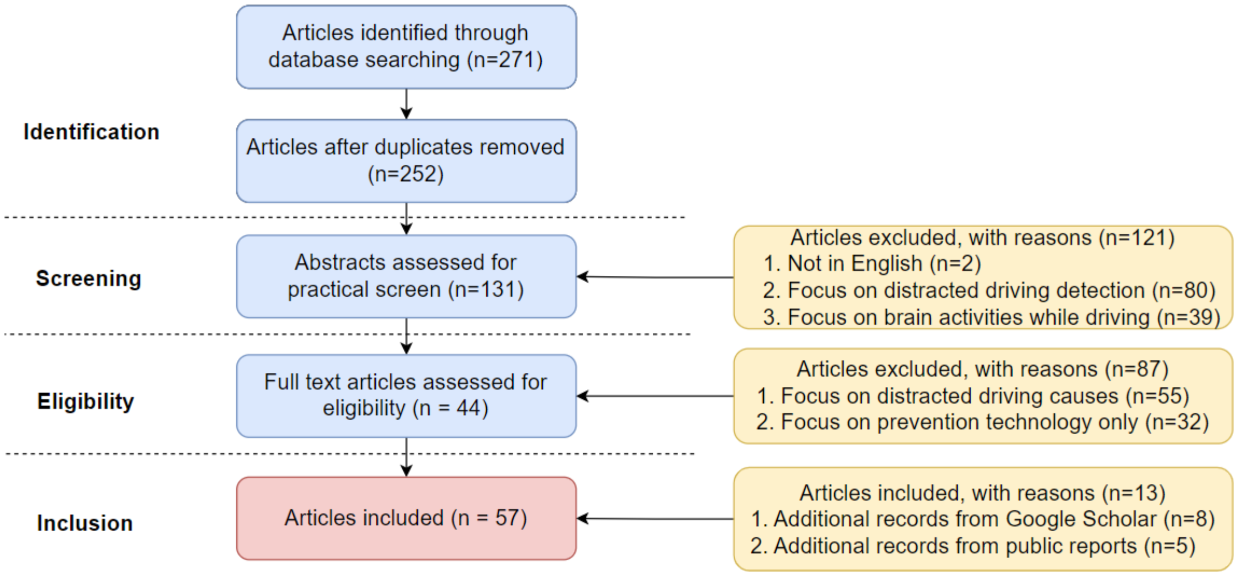

Figure 1. Literature review process