Published in final edited form as:

Stat Med. 2017 October 30; 36(24): 3763–3771. doi:10.1002/sim.7404.

Driving In Search of Analyses

Bruce Simons-Morton, EdD, MPH

Senior Investigator, Health Behavior Branch, DIPHR, NICHD, NIH, 6710B Rockledge Room 3166, Bethesda, MD 20892-7000, 301-496-5674

Abstract

While transportation safety has greatly improved over the past two decades, motor vehicle crash injuries remain a leading cause of morbidity and mortality, particularly among young drivers. Driver errors and behaviors such as speeding and distraction contribute disproportionately to crashes among inexperienced novices, who develop safe driving judgment only with substantial driving experience, commonly described as the “young driver problem”. Research on young drivers has applied a range of research methods, including analyses of national archival data (mainly from police reports), crash analyses, observation of driver behavior, surveys of driver behavior and dispositions, and experimental research on driver behavior and vehicle crash worthiness. Prominent research questions regarding young driver safety include what and how do novices learn to drive safely, what are the predictors of young driver crashes, what is the variability and over-time trajectories of young driver performance and outcomes, and to what extent is the young driver problem due mainly to average population risk or high-risk groups? Current research on young drivers is complicated by small sample sizes, relatively rare events, high within and between group variability, missing data, the need to estimate exposure, and the lack of longitudinal and experimental designs, problems that require complex analytic methods. In this paper we provide an overview of driving research methods, examples of research addressing the young driver problem, and examples of statistical collaboration on young driver research, focusing particularly on estimating prediction of crash risk and estimating variability in young driver performance and outcomes.

Keywords

injury; adolescence; distraction; risk; variability; count data; naturalistic driving; Poisson distribution; ordinal latent variable model; trajectory classes; recurrent event modeling; marginal analyses

INTRODUCTION

As one of the leading causes of death and disability in the US and the world, motor vehicle crashes are an important public health concern [1, 2]. Globally, the high prevalence of injuries due to motor vehicle crashes (MVC) is likely to increase in the future, with the rapid growth in vehicle sales and driving exposure in many rapidly-developing countries [2, 3]. Even in Western countries where motor vehicle crash rates and resulting morbidity and mortality have been declining over the past decade or so, owing to safer vehicles and roads, MVC-related morbidity and mortality remains a significant problem [2]. Indeed, the rapid

evolution toward self-driving (automated) vehicles, which promises to reduce the contribution of driver behavior to crashes, poses unique safety challenges as both the technology and policies to regulate the technology attempt to resolve issues related to when the driver or autonomous vehicle is responsible, when and how the transition from driver-to-automation and from automation-to-driver can be made safely, and just how quickly and how smart vehicles can become [4]. Despite the advent of automated vehicles, drivers, particularly young drivers who tend to drive older, less technologically sophisticated vehicles, are likely to retain substantial autonomy for the foreseeable future. Meanwhile, unsafe driver behaviors, including impairment due to alcohol, marijuana, and other substances, lack of sleep, and distracting secondary tasks, as well as judgment errors and aggressive driving, will contribute to crashes as they always have [5]. Hence, there remains a compelling need for and interest in applied research that would guide transportation safety policy and programs within the context of evolving technology and a highly dynamic environment involving roads, vehicles, drivers, passengers, and other vulnerable road users such as pedestrians and bicyclists [6,7].

Among other topics transportation researchers are interested in the assessment of crash prevalence, trends, variability, and contributing factors. What is the prevalence of crashes overall and among groups of interest, and are crash rates increasing or decreasing over time? How can crash rates be calculated accurately? What groups are at greatest risk and what driver characteristics, driver behaviors, and road conditions contribute to crash risk? How can driver behavior be improved? How can vehicles and roadways be designed to reduce crashes and injuries among vehicle occupants and other road users? Basically, policy makers need research data and findings that will enable them to determine the correct balance between safety and mobility and to determine which policies and practices would be best in this regard.

TRANSPORTATION RESEARCH METHODS

There are many sources of information about crashes that can contribute to an understanding of the problem. Archival data from police reports on crash prevalence and causes are widely available in the U.S. Direct observation can provide data on the point prevalence of seat belt use, driver second task, and other driver behaviors. Vicarious observation in the form of video recordings from cameras installed (along with other instrumentation) in vehicles is a popular source of “naturalistic” information on driver behavior and related crash critical incidents. Survey research is popular for assessing self-reported events, driving behavior, and respondent attitudes and related characteristics. Experimental research is conducted on test tracks and in simulators. Hence, there are a range of useful sources of crash data, each with advantages for particular research questions, but also limitations. Here briefly we describe the primary sources of data.

Archival Data Bases

In the U.S. States are required to report fatal and other crashes attended by police and these data make up the Fatality Analysis Reporting System (FARS) [8] and the National Auto Sampling System (NASS) [9] of non-fatal crashes, both widely-used national data sources.

These data sets include information on crash severity and characteristics, for example, single or multiple vehicle, day or night, weather, driver and passenger characteristics (age and sex). NHTSA also maintains a data base of traffic records from 36 participating states that includes licensing, crash incidence, and injury surveillance [10], which researchers frequently use to estimate the number of licensed drivers (as a measure of exposure). (Presumably, insurance companies have good data on crash prevalence, consequences, and causes of policy holders, but this information is not generally available for external research purposes.) The quality of accident reporting systems varies considerably from country to country [2].

Professional crash investigators conduct special crash investigations for NHTSA that include vehicle and driver evaluations [11]. Case selection is based on priorities that vary over time, for example, to assess vehicle controls, safety devices, driver distraction, and vehicle-pedestrian interactions. Crash analyses provide detailed information about a sample of crashes based on the selection criteria from police, insurance, and detailed accident investigations, including interviews with the occupants in vehicles involved in crashes. This rich data source is limited by the small samples, lack of objective exposure information, and participant recall.

Among the many sources of crash data, FARS is one of the most important and frequently-used data sets because fatal crashes are of greatest concern, the data set is complete, and the cases are well documented, particularly relative to non-fatal crash data sets which tend to be less complete [9]. However, both NASS and FARS data are commonly used to examine prevalence, trends over time, crash conditions, and contributing factors. Analyses can identify groups at risk for crashing (e.g., age and sex), particular types of crashes (e.g., single vehicle crashes consistent with loss of control, rear-end crashes consistent with close following and inattention), crash conditions (e.g., geography, road type, weather, time of day, road conditions, and passenger presence), and some driver behaviors (e.g., speed, blood alcohol level) [5]. A major limitation of NASS and FARS data is the lack of exposure information. Therefore, rates are often calculated on the bases of licensed drivers, for example from the National Driver Register [9] or mileage estimates, for example, from the National Household Travel Survey [12].

Hence, useful national data sets are available for the analyses of crashes and their contributing factors.

Direct and Vicarious Observation

Driver behavior can be observed directly or vicariously using video technology. Direct observation of safety belt use, secondary (to driving) tasks (such as phone use), and other behaviors can be conducted usefully by trained observers to provide point prevalence estimates [13]. Video data are sometimes collected at intersections and other locations to examine how vehicles interact with road and traffic conditions [14]. In one interesting application of direct and vicarious observation, my research team sought to determine if the presence of teenage passengers was associated with risky driving among teenage drivers [15]. We stationed observers at the exits of high school parking lots, who noted the age (teen or adult) and sex of the front seat occupants. Across the way another observer video

recorded the vehicle so that it could be identified later. At nearby locations where the vehicles could get up to speed, another researcher video recorded traffic and assessed vehicle speed using a radar device. Later in the laboratory, we viewed the video and calculated the distance between vehicles (gap) and speed of the vehicles we had identified exiting the high school parking lots, and all other traffic, termed usual traffic. We found that teenage drivers on average drove slightly faster and closer to the vehicle ahead than usual traffic, but in the presence of a male teenage passenger, both male and female teenage drivers drove significantly faster and allowed significant shorter gap than usual traffic. Teenage male drivers in the presence of female teenage passengers drove more safely, similar to usual traffic.

Video is one of several companion vehicle sensors in naturalistic driving research. Naturalistic driving methods use multiple sensors installed as a package in the vehicles of participating driver volunteers. Instrumentation includes GPS, accelerometers, cameras, and lane trackers, which collectively can provide extensive data on drivers over time [16]. To limit data storage some systems only save data triggered by an elevated gravitational-force event (g-force), which capture most events of interests because maneuvers that trigger such events exert high g-forces and are common in crashes and near misses. This reduces the amount of data collected and store, or continuously recorded and stored, and later reduced and coded. Other systems continuously save driving data from the sensors and download it to a remote server. Later, the data can be linked by time stamp or other means, providing comprehensive information about each trip, vehicle, and driver, including mileage, speed, elevated g-force events, lane deviation, secondary task engagement, passenger presence, and driving conditions. In a laboratory designed for this purpose the data can be reduced to identify events of interest, for example, speeding, crashes, near crashes, elevated g-force events, and coded according to driving conditions and driver behavior. Naturalistic driving methods provide extensive observational data on drivers over time, but are limited by the small number of drivers who can be assessed (due to the expense), the small number of meaningful events that occur, the large accumulation of data, and the high expense of data reduction, coding, and storage. However, each event can be well examined and patterns of behavior evaluated, and rates per mileage calculated accurately, providing substantial information about crash causality.

Several small naturalistic driving studies have been conducted or are underway in Canada, Australia, Europe, and elsewhere. The 100-Car study was the first, moderately large study, involving 100 high mileage drivers age 18 for 12-months each [16]. Findings from the study included the risk of secondary task engagement and contributing factors about rear-end collisions and run-off-road crashes. The 40-car study examined driving behavior of newly-licensed teenagers over the first 18 months of independent driving [17]. Analyses focused on the variability in crash and near crash rates, risky driving, and speeding; the association of elevated gravitational force events and crash/near crash rates; and the trajectories of risky driving and crash/near crash rates. The largest naturalistic study conducted, the Strategic Highway Research Program (SHRP2) study, completed data collection in 2015. SHRP2 employs a unique public use data system allowing researchers access to de-identified, summarized data and with some qualifications to assure privacy allows researchers access to original data [18].

Survey Research

Surveys are frequently conducted to collect information about transportation and road safety. Some important surveys are government sponsored and used by many researchers, for example, the National Household Travel Survey is conducted to assess the driving habits of U.S. households [12]. Similarly, to obtain population estimates of driving behavior, NHTSA conducts surveys on safety belt use [19] and speeding [20]. Surveys are generally included as one of several methods because some data is easier to collect or can only be obtained through surveys, for example, risky and aggressive driving behavior, motivation (attitudes and beliefs) for risky driving, and the like. Surveys can provide a wealth of information on potential contributing factors and are a rare source of information about driver attitudes and other dispositions. In on-going research my team annually examines in a national cohort of adolescents and young adults self-reported risky driving behavior, including unsafe driving practices, potentially distracting secondary task engagement, and impaired driving [21, 22].

Test Tracks and Simulators

Test tracks are actual roads absent public traffic, generally located at transportation research centers. Some test tracks include signalized intersections, turn outs, crosswalks, roundabouts, and the like; some have lighting for night driving and can manufacture precipitation in various forms. Test tracks are often used to evaluate vehicle equipment, such as the utility of the layout of the dashboard and technology innovations, and driving performance under various driving conditions, such as at night and with precipitation. They are sometimes also used to conduct controlled driver behavior experiments. For example, in recent research my team evaluated the intersection management performance of novice and adult drivers when also managing a telephone task [23]. As participants drove the mile-long test track they encountered a signalized intersection. After several passes, as they approached the intersection the participants were given a cell phone by the research assistant in the front seat passenger seat and told to dial a number and gather specific information, at which time the light was turned amber. At Time 1, when the novices had been driving on their own for only a few weeks, about 30% failed to look up and ran the red light, while 100% of the experienced adults looked up and stopped. At Time 2, six months later, the results were similar. The novice teenagers were very good at the cell phone task, attending to it faithfully until completed, while the adult participants were less facile with the phone, but looked up after a few initial manipulations of the phone in time to see the amber light and stop the vehicle for the red light.

Simulators are another popular transportation research tool. Simulations come in many forms and complexities that provide various level of realism. The simplest sort of simulator is a computer with a joystick or steering wheel extensions. Most serious research employs laboratory simulators with an actual vehicle with braking, turning, and other vehicle operations linked to sophisticated software so that the driver can move the vehicle within the simulated driving scenarios projected on screens around the vehicle [24]. Simulation studies can address almost any driving topic, from the utility of vehicle technology to aspects of driver behavior. For example, in a recent set of experiments, my research team evaluated the effects of teenage passengers on the simulated risky driving behavior of male teenage drivers. In one experiment we randomized participants to drive with a young passenger, a

study confederate, who in one condition expressed risk-accepting and in the other condition risk-averse attitudes during a pre-drive exercise where the participant and passenger watched videos of dangerous or ultra-conservative driving. Not surprisingly, participants who drove with the risk-accepting confederate passenger drove in a more-risky manner than those who drove with the risk-averse confederate passenger, suggesting that passenger norms can affect young driver behavior [25].

STATISTICAL COLLABORATION ON DRIVING RESEARCH

Statistical collaboration in driving research, as in other areas of research, is needed to provide the best analyses of complicated data sets. As noted, there is a range of data sources, each with unique strengths and limitations. Notably, exposure must often be estimated, data are sometimes missing, the calculation and comparisons of rates can be complicated, distributions are not always normal, and statistical modeling is complex. While there are conventional methods for dealing with missing data, when it can be imputed and how best to do so warrant careful methodological attention. Missing data is common in archival, observational, and survey methods; simulation studies have a relatively high proportion of drop outs or incomplete drives due to simulator sickness (10%–20% depending on driver age); and in naturalistic driving GPS data, important for many calculations, is sometimes missing while the vehicle starts up or in hybrid vehicles when their batteries are not online while the vehicle idles. One example of a complicated imputation of missing data occurred in an analysis of FARS data in which we sought to determine the possible increase in fatal crash risk when teenage passengers were present [26]. We wished to calculate risk based on exposure using FARS, which includes fatalities among drivers and passengers, but does not provide exposure data on miles driven and ridden. Fortunately, the NHTS provides this information for family members, but not for other passengers, hence, the mileage of teenage drivers carrying a non-family, teenage passenger is not available. However, using NHTSA data we could estimate teen passenger exposure from the amount that teens in the participating households reported riding as passengers with non-family teenage drivers. Accordingly, we found that fatal crash risk of teenage drivers was greater in the presence of teenage passengers, particularly male teenage passengers.

Driving research sometimes challenges conventional statistical methods. For example, the small samples in naturalistic driving studies with many hours and miles of driving data collected daily over a year or more provide an interesting data structure that may defy conventional statistical methods, as described in another paper in this series by Albert [27]. The following examples of statistical contributions to driving research are based on research conducted by my research team and are not necessarily representative of the range of such collaborations and may not even be the most common such collaborations, but should provide useful, if selective, examples. These examples include analyses of archival data, observing of driving behavior, naturalistic driving, simulation, and survey research.

THE YOUNG DRIVER PROBLEM

Novice drivers have high crash rates that decline rapidly from licensure for about 6 months and 1000 miles and then more slowly for a period of years before reaching the level of

experienced adults [28]. While both age and inexperience are implicated, there is substantial variability in risk, limited understanding about what novices learn that improves their safety, and uncertainty about how best to facilitate safety during this critical early stage. The young driver problem is that the more novices drive, the greater their risk of crashing, but the only way they develop safe driving judgment is by driving. Notably, the development of competence for any complicated task requires substantial practice and feedback from errors, and driving is no exception [29]. While driver education and supervised practice driving are undoubtedly helpful, there is no evidence that training is associated with independent driving safety [30]. Like learning chess or skiing, much of the learning comes from trial and error and making driving errors can be dangerous. My research team has been studying teenage driving using data from archival, simulation, test track, and naturalistic methods. Here we describe some of this research to illustrate the importance of statistical collaboration.

Statistical Methods and the Naturalistic Teenage Driving Study

The Naturalistic Teenage Driving Study (NTDS), conducted from 2002–2007, was the first naturalistic study to focus on teenage drivers [31]. Teenage drivers ($n=42$) and at least one of their parents were recruited within 3 weeks after obtaining a license for independent driving and followed for 18 months. The vehicle the teen drove most of the time was instrumented with a Data Acquisition System that included a variety of sensors. Cameras were mounted strategically to monitor the driver's face, hand and arm movements, dashboard, and forward and rear roadway; an overhead camera took a blurred snapshot (not all passengers would have consented to participate in the study) of the occupants. Accelerometers assessed g-force events, GPS assessed mileage, and a laser assessed lane tracking and vehicle following. The data were recorded continuously, stored on a computer, and downloaded periodically for reduction and coding in a lab designed for these purposes. Hence, the data were not viewed in real time, but the recorded data were viewed some weeks or months later. Measures of primary interest included crashes and near crashes (CNC), elevated g-force events (kinematic risky driving – KRD), speeding, and distraction. Questionnaires were administered periodically to assess participant characteristics.

Findings from the NTDS indicated that CNC rate was 3.9 times higher among the teens than their parents driving the same vehicles on similar roadways and driving conditions [29]. CNC rates among parents were low and invariable and among teenagers were substantially higher and highly variable. While about a third of teenage drivers had few or no CNC, another proportion had numerous events. Among some teenagers high CNC rates during the first months of driving diminished over time, accounting for most of the average decline in the sample [32, 33]. Kinematic risky driving, measured as the rate per miles driven of elevated g-force events due to hard braking, fast starting, sharp turns, and over-correction maneuvers, were 5 times higher on average among teenagers than adults, and did not decline over time. However, there was substantial variability among teenage drivers, with a low and a high trajectory class, but with substantial within driver variability [33]. The best predictor of both CNC rates and KRD rates was the teenage drivers reported social norms, their perceptions about the risky driving propensity of their friends [34].

Distraction was also of analytic interest. Crash risk due to distraction from driver engagement in tasks secondary to driving, such as using a phone, were evaluated as follows: the video of each CNC was coded according to the behavior of the driver the seconds before the event. Then, non-CNC road segments (termed baselines) were randomly sampled proportional to each driver's mileage and coded in similar fashion. The likelihood of a CNC when engaging in a secondary task was calculated by comparing the rate of secondary task engagement proximal to each CNC event with the rate of secondary task engagement in the non-crash baseline road segments. Analyses of the NTDS data were compared to similar analyses of the 100 Car Study of high mileage adult drivers [35]. While NTDS teenagers engaged in secondary tasks about as much as adult drivers, their crash risk was significant when texting (OR=3.87), dialing (OR= 8.3), reaching for a phone (OR=7.05), reaching for objects other than a phone (OR = 8.0), and eating (OR=2.99). Adult drivers had higher likelihood of a crash only while dialing (OR=2.99). Notably, each of these are tasks take the driver's eyes off the forward roadway. Additional analyses of eye glance, found that the longer the teenage driver's eyes were off the forward roadway while engaging in a secondary task, the greater the CNC likelihood. It remains unclear why secondary task engagement resulted in greater risk for teenagers than adults, but a few likely possibilities include teenagers being less adept than adults at selecting relatively safe driving conditions, failure to modify speed, or greater residual cognitive distraction when turning back to the roadway after secondary task engagement. These straight-forward analyses were enabled by the unique video data that provide objective assessment of the CNC event and the presence of the secondary tasks, and the objective measure of exposure (miles driven) from GPS.

NTDS Methods Research

While naturalistic driving provides unique and valuable information about crash causes, with relatively few study participants, frequent observations over time on each participant, and relatively rare outcomes of interest, the data structure was somewhat unusual. Standard statistical methods for longitudinal data analysis might have provided poor statistical properties and biased estimates of association. Hence, the research required non-standard methods. For CNC and KRD counts, Poisson distributions were assumed (with mileage offsets) and data were aggregated over 1, 3, or 6 month periods to obtain stable prevalence estimates.

It has been conventional in naturalistic driving studies, given the rarity of crashes, to lump crashes and near crashes in a single measure of near crashes. Near crashes are defined as events that are similar to crashes, except actual contact was avoided by a last-second evasive maneuver (close call); near crashes tend to occur about 4 times more frequently than actual crashes. Previous research has evaluated the extent to which near crashes can be considered surrogates for crashes. Principally, Guo *et al.* [36] examined this question using data from the 100 Car study and found that analyses of CNC increased precision and underestimated the risk of contributing factors. (A criticism of this approach, and of naturalistic driving in general, is that most crashes in naturalistic studies are not severe and are qualitatively different from serious crashes, which tend to involve high speeds and result in serious injuries and property damage, unlike most crashes in naturalistic driving research that involve minor injuries and property damage.)

While CNC rates, like crash rates obtained from archival data sources, are higher among novice and young drivers than older drivers and decline over time, the distribution of CNC rates and change over time can best be examined in naturalistic studies. *Guo et al.* [32] used k-sort cluster analyses to identify three CNC groups, one with persistently high CNC rates, one with persistently low CNC rates, and one with initially high rates that declined after 6 months. It seems that some novices may have exercised good driving judgement even as beginners, while others seemed to learn from their driving experience over a period of months, and others did not seem to learn from experience how not to crash. Building on this work Guo and colleagues [37] proposed recurrent-event models (identical rate and varying rate) based on non-homogeneous Poisson process with constant piecewise intensity function for detecting the CNC change-points of novice teenage driving risk and applied to the NTDS data (the varying-rate model with one change-point provided the best fit). Currently, Guo and colleagues [37] are evaluating the CNC change-point in the NTDS using a Bayesian finite mixture model. A latent variable identified potential clusters and a Markov chain Monte Carlo algorithm sampled from the posterior distributions, with simulations to evaluate the performance of the model. Application to the NTDS data indicate three change-points (52, 108, 150 hours of driving after licensure), with differing overall intensity rates and patterns of change.

Using CNC as the outcome of interest in the NTDS, we examined its association with KRD. In particular, we were interested in determining if the pattern of KRD predicted the likelihood of a CNC in the following month. The KRD rates for multiple months predicted the likelihood of a CNC in the following month relatively accurately with an area under the Receiver Operating Curve = 0.76 [38], findings that were similar to analyses of the 100 Car Study [39]. We noted that KRD rates declined immediately briefly after a CNC, suggesting that drivers moderated their driving behavior after a crash. To evaluate this more formally, we used the SHRP2 naturalistic driving data set that includes some 245 young drivers [18]. We calculated the rates of KRD before and after each crash and found that the rates declined for at least 2 months, while the rates among those who did not crash remained stable, consistent with an effect of learning from crash experience [40].

As with CNC, the variability in KRD among novice teenage drivers could only be examined in longitudinal research using naturalistic driving methods, hence, it was not known prior to the NTDS. Using latent class models of KRD over the 18-month period, Simons-Morton *et al.* [33] found two trajectory classes best characterized the overtime distribution, with substantial variability in within groups. Zhang, Albert, & Simons-Morton [41], examined the long sequence of count data using marginal analyses, accounting for gender and other participant characteristics and time-dependent covariates such as time of licensure and passenger presence. Subsequently, Kim *et al.* [42] showed that the within-subject variation is approximately the same as the between-subject variation in KRD over time, using a Poisson regression model with stochastic random effects for these trip-level count data.

Jackson *et al.* [43] examined the utility of ordinal latent variable models for jointly modeling the association of CNC and KRD. These analyses made it possible to make inferences about the overtime dependence of the two measures of risk and to predict CNC from KRD and other co-variables.

DISCUSSION

Driving research provides many opportunities for analyses of important public health questions and interesting methodological challenges, particularly but not exclusively among young drivers. Recent research has contributed to better understanding of crash risk, but many important questions remain, including how best to characterize risk and the importance of average population risk relative to the small sample of high-risk drivers that can distort the population risk profile. While each source of driving data and method of assessing crash risk provides useful information, naturalistic driving methods are unique in providing objective longitudinal information about events and contributing factors. However, given the small sample sizes and rare events, the data pose analytic challenges to resolving critical questions of interest. Unresolved research questions on driving performance and outcomes that might be addressed through analyses of driving data from various sources, including naturalistic, archival, observational, and experimental research, include the following: (1) What do novices learn from experience that contributes to the rapid reduction in crash rate the first year of independent driving? (2) Do near crashes serve as useful surrogates for crashes? (3) Can crash risk be predicted from KRD, speed, distraction, and other measures, solely or in combination? (4) Is there a group of high risk drivers and if so how much of the problem do they account for and what characterizes them? (5) What is the variability in crash performance and outcomes due to individual and environmental sources? (6) Are there trajectory classes of driving behavior and crash risk? (7) Do drivers learn from their mistakes? (8) What is the variability in road conditions in which drivers elect to engage in secondary tasks? (9) Do some teenage passengers increase and others decrease crash risk and does this vary under certain driving conditions? Creative application of statistical methods will be required to resolve these questions.

Acknowledgments

This research was supported in part by the intramural program of the *Eunice Kennedy Shriver* National Institute of Health and Human Development

References

1. National Highway Traffic Safety Administration. Young drivers. National Highway Traffic Safety Administration; Washington, DC: 2015. Traffic safety facts: 2013 data. Report No. DOT HS 812 200 Available from: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812200> (accessed 15 March 2017).
2. World Health Organization. Global Status Report on Road Safety 2015. World Health Organization; Geneva, Switzerland: 2015. Available from: http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/ (accessed 15 March 2017).
3. World Health Organization. Global Status Report on Road Safety 2013: Supporting a Decade of Action. 2013. http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/ (accessed 15 March 2017).
4. Insurance Information Institute. Self-driving cars and insurance. 2015. Available from: <http://www.iii.org/issue-update/self-driving-cars-and-insurance>. (accessed 15 March 2017)
5. National Highway Traffic Safety Administration. Traffic Safety Facts 2012: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. National Highway Traffic Safety Administration; Washington, DC: 2014. Report No. US DOT HS 812 032 Available from: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812032> (accessed 15 March 2017).

6. National Highway Traffic Safety Administration. Pedestrians. National Highway Traffic Safety Administration; Washington, DC: 2015. Traffic Safety Facts: 2013 Data. Report No. DOT HS 812 124 Available from <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812124> (accessed 15 March 2017).
7. National Highway Traffic Safety Administration. Bicyclists and Other Cyclists. National Highway Traffic Safety Administration; Washington, DC: 2015. Traffic Safety Facts: 2013. Report No. U.S. DOT HS 812 151 Available from <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812151> (accessed 15 March 2017).
8. National Highway Traffic Safety Administration. Fatality Analysis Reporting System (FARS). 2017. Available from: [https://one.nhtsa.gov/Data/Fatality-Analysis-Reporting-System-\(FARS\)](https://one.nhtsa.gov/Data/Fatality-Analysis-Reporting-System-(FARS)) (accessed 15 March 2017).
9. National Highway Traffic Safety Administration. National Automotive Sampling System (NASS). 2017. [https://one.nhtsa.gov/Data/National-Automotive-Sampling-System-\(NASS\)](https://one.nhtsa.gov/Data/National-Automotive-Sampling-System-(NASS)) (accessed 15 March 2017).
10. National Highway Traffic Safety Administration. National Driver Register. 2017. Available from: [https://one.nhtsa.gov/Data/National-Driver-Register-\(NDR\)](https://one.nhtsa.gov/Data/National-Driver-Register-(NDR)) (accessed 15 March 2017)
11. National Highway Traffic Safety Administration. Special Crash Investigations. 2017. Available from: <https://www.nhtsa.gov/research-data> (accessed 15 March 2017).
12. NHTSA. National Household Travel Survey (NHTS). 2014. <https://www.fhwa.dot.gov/policyinformation/nhts.cfm>
13. National Highway Traffic Safety Administration. Traffic Safety Facts: Guidelines to observe and estimate statewide seat belt use at night. National Highway Traffic Safety Administration; Washington, DC: 2010. Traffic Technology Transfer Series # 383 Available from: <https://pdfs.semanticscholar.org/3fb2/616bf8577d5e704e760d80656d264d327678.pdf> (accessed 15 March 2017).
14. Lofgren, M. Measuring seatbelt usage through existing traffic surveillance systems: Pilot project. Rural Transportation Safety and Security Center, North Dakota State University; 2009. Available from: <http://www.ugpti.org/pubs/pdf/DP213.pdf> (accessed 15 March 2017)
15. Simons-Morton BG, Lerner N, Singer J. The observed effects of teenage passengers on the risky driving behavior of teenage drivers. Accident Analysis & Prevention. 2005; 37(6):973–982. DOI: 10.1016/j.aap.2005.04.014 [PubMed: 15921652]
16. Neale, VL., Dingus, TA., Klauer, SG., Knipling, RR., Holbrook, GT., Petersen, A. The 100-car naturalistic driving study: Experimental design. National Highway Traffic Safety Administration; Washington, DC: 2002. Report No. DOT HS 809 536 Available from: <https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/100carphase1report.pdf> (accessed 15 March 2017).
17. Simons-Morton BG, Klauer SG, Ouimet MC, Guo F, Albert PS, Lee SE, Ehsani JP, Pradhan AK, Dingus TA. Naturalistic teenage driving study: Findings and lessons learned. Journal of Safety Research. 2015; 54:41–44. DOI: 10.1016/j.jsr.2015.06.010 [PubMed: 26403899]
18. Transportation Research Board The second Strategic Highway Research Program. 2006–2015. Available from: <http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Blank2.aspx> (accessed 15 March 2017).
19. National Highway Traffic Safety Administration. Guidelines to Observe and Estimate Seat Belt Use at Night. National Highway Traffic Safety Administration; Washington, DC: 2010. Report No. DOT HS 811 288 Available from: <https://pdfs.semanticscholar.org/3fb2/616bf8577d5e704e760d80656d264d327678.pdf> (accessed 12 December 2016).
20. Schroeder, P., Kostyniuk, L., Mack, M. National Survey of Speeding Attitudes and Behaviors. National Highway Traffic Safety Administration; Washington, DC: 2011. Report No. DOT HS 811 865 Available from: https://www.nhtsa.gov/staticfiles/nti/pdf/2011_N_Survey_of_Speeding_Attitudes_and_Behaviors_811865.pdf (accessed 15 March 2017).
21. Simons-Morton B, Li K, Ehsani J, Vaca FE. Covariability in three dimensions of teenage driving risky behavior: Impaired driving, risky and unsafe driving behavior, and secondary task engagement. Traffic Injury Prevention. 2016; 17(5):441–446. DOI: 10.1080/15389588.2015.1107183 [PubMed: 26514232]

22. Vaca FE, Li K, Hingson R, Simons-Morton BG. Transitions in riding with an alcohol/drug-impaired driver from adolescence to emerging adulthood in the US. *Journal of Studies on Alcohol and Drugs*. 2016; 77(1):77–85. DOI: 10.15288/jsad.2016.77.77 [PubMed: 26751357]
23. Lee SE, Olsen ECB, Simons-Morton BG. Eyegance behavior of novice teen and experienced adult drivers. *Transportation Research Board Journal*. 2006; 1980:57–64. DOI: 10.3141/1980-10
24. Caird, JK., Horrey, WJ. *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*. CRC Press, Taylor and Francis Group; Boca Raton, FL: 2011. p. 5-1-5-17.
25. Simons-Morton BG, Bingham CR, Ouimet MC, Pradhan A, Falk E, Li K, Green P, Almani F, Shope J. Experimental effects of injunctive norms on simulated risky driving among teenage males. *Health Psychology*. 2014; 33(7):616–627. DOI: 10.1037/a0034837 [PubMed: 24467258]
26. Ouimet MC, Simons-Morton BG, Zador PL, Lerner ND, Freedman M, Duncan GD, Wang J. Using the U.S. National Household Travel Survey to estimate the impact of passenger characteristics on young drivers' relative risk of fatal crash involvement. *Accident Analysis & Prevention*. 2010; 42(2):689–694. DOI: 10.1016/j.aap.2009.10.017 [PubMed: 20159095]
27. Albert P. Innovative statistical modeling for naturalistic driving. (submitted to this issue, TBA).
28. Williams AF. Teenage drivers: Patterns of risk. *Journal of Safety Research*. 2003; 34(1S):5–15. DOI: 10.1016/S0022-4375(02)00075-0 [PubMed: 12535901]
29. Simons-Morton BG, Ehsani JP. Learning to drive safely: Reasonable expectations and future directions for the learner period. *Safety*. 2016; 2(4):20.doi: 10.3390/safety2040020
30. Williams, AF., Preusser, DF., Ledingham, KA. *Feasibility Study on Evaluating Driver Education Curriculum*. National Highway Traffic Safety Administration; Washington, DC: 2009. Report No. DOT HS 811 108 Available from: <https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/811108.pdf> (accessed 15 March 2017).
31. Lee SE, Simons-Morton BG, Klauer SG, Ouimet MC, Dingus TA. Naturalistic assessment of novice teenage crash experience. *Accident Analysis & Prevention*. 2011; 43(4):1472–1479. DOI: 10.1016/j.aap.2011.02.026 [PubMed: 21545880]
32. Guo F, Simons-Morton BG, Klauer SG, Ouimet MC, Dingus T, Lee SE. Variability in crash and near crash risk among novice teenage drivers: A naturalistic study. *Journal of Pediatrics*. 2013; 163(6):1670–1676. DOI: 10.1016/j.jpeds.2013.07.025 [PubMed: 23992677]
33. Simons-Morton BG, Cheon K, Guo F, Albert P. Trajectories of risky driving among novice teenagers. *Accident Analysis & Prevention*. 2012; 51C:27–32. DOI: 10.1016/j.aap.2012.10.011
34. Simons-Morton BG, Ouimet MC, Zhang Z, Lee SL, Klauer SE, Wang J, Chen R, Albert PE, Dingus TE. The effect of passengers and risk-taking friends on risky driving and crashes/near crashes among novice teenagers. *Journal of Adolescent Health*. 2011; 49(6):587–593. DOI: 10.1016/j.jadohealth.2011.01.009 [PubMed: 22098768]
35. Klauer S, Guo F, Simons-Morton BG, Dingus T. Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*. 2014; 370(1):54–59. DOI: 10.1056/NEJMs1204142 [PubMed: 24382065]
36. Guo F, Klauer SG, Hankey JM, Dingus TA. Near crashes as crash surrogate for naturalistic driving studies. *Transportation Research Record: Journal of the Transportation Research Board*. 2010; 2147:66–74. DOI: 10.3141/2147-09
37. Chen C, Guo C. Evaluating the influence of crashes on driving risk using recurrent event models and naturalistic driving study data. *Journal of Applied Statistics*. 2016; 43(12):2225–2238. DOI: 10.1080/02664763.2015.1134449
38. Simons-Morton BG, Zhang Z, Jackson JC, Albert PS. Do elevated gravitational-force events while driving predict crashes and near crashes? *American Journal of Epidemiology*. 2012; 175(10): 1075–1079. DOI: 10.1093/aje/kwr440 [PubMed: 22271924]
39. Fang Y, Guo F. Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*. 2013; 61:3–9. DOI: 10.1016/j.aap.2012.06.014 [PubMed: 22785091]
40. O'Brien F, Simons-Morton BG, Bible J, Liu D. Do young drivers become safer after being involved in a collision? *Psychological Science*. 2017; doi: 10.1177/0956797616688118
41. Zhang Z, Albert PS, Simons-Morton BG. Marginal analysis of longitudinal count data in long sequences: Methods and application to a driving study. *Annals of Applied Statistics*. 2012; 6(1): 27–54. DOI: 10.1214/11-AOAS507 [PubMed: 27087885]

42. Kim SD, Chen Z, Zhang Z, Simons-Morton BG, Albert P. Bayesian hierarchical Poisson regression models: An application to a driving study with kinematic events. *Journal of the American Statistical Association*. 2013; 108(502):494–503. DOI: 10.1080/01621459.2013.770702 [PubMed: 24076760]
43. Jackson J, Albert PS, Zhang Z, Simons-Morton BG. Ordinal latent variable models and their application in the study of newly licensed teenage drivers. *Journal of the Royal Statistical Society: Series C(Applied Statistics)*. 2012; 62(3):435–450. DOI: 10.1111/j.1467-9876.2012.01065.x