Published in final edited form as:

J Safety Res. 2019 December ; 71: 243–249. doi:10.1016/j.jsr.2019.09.024.

Validation of Not-at-fault Driver Representativeness Assumption for Quasi-Induced Exposure Using U.S. National Traffic Databases

Shen Sijun, N. Pope Caitlin

The Center for Injury Research and Policy, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH

Department of Pediatrics, College of Medicine, The Ohio State University, Columbus

Stamatiadis Nikiforos,

Department of Civil Engineering, College of Engineering, University of Kentucky, Lexington, KY

Zhu Motao

The Center for Injury Research and Policy, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH

Department of Pediatrics, College of Medicine, The Ohio State University, Columbus

Division of Epidemiology, College of Public Health, The Ohio State University, Columbus 700 Children's Drive, Columbus, Ohio, 43205

Abstract

Introduction: The quasi-induced exposure (QIE) method has been widely implemented into traffic safety research. One of the key assumptions of QIE method is that not-at-fault drivers represent the driving population at the time of a crash. To validate the representativeness assumption of the QIE method, recent studies have utilized the remaining not-at-fault drivers in three-or-more-vehicle crashes (D3_other) after excluding the first not-at-fault drivers as the reference group in a single state crash database. However, it is unclear if the QIE representativeness assumption is valid on a national scale and is a representative sample of driving population in the United States (US). The aim of this study was to assess the QIE representativeness assumption on a national scale and to evaluate if D3_other could serve as a representative sample of the US driving population.

Method: Using the Fatality Analysis Reporting System (FARS) and the National Occupant Protection Use Survey (NOPUS), distributions of driver gender, age, vehicle types, times, and roadway types among the not-at-fault drivers in clean two-vehicle crashes, the first not-at-fault drivers in three-or-more-vehicle crashes, and the remaining not-at-fault drivers in three-or-more vehicle crashes were compared to the driver population observed in NOPUS.

Results: The results showed that with respect to driver gender, vehicle types, time, and roadway types, drivers among D3_other did not show statistical significant difference from NOPUS

observations. The age distribution of D3_other driver was not practically different to NOPUS observations.

Conclusions: Overall, we conclude that D3_other drivers in FARS represents the driving population at the time of the crash.

Practical Applications: our study provides a solid foundation for future studies to utilize D3_other as the reference group to validate the QIE representativeness assumptions and has potential to increase the generalizability of future FARS studies.

Keywords

clean multiple vehicle crashes; quasi-induced exposure; representativeness assumption; Fatality Analysis Reporting System

1. Introduction

Traffic crashes are one of the leading causes of unintentional death in United States (US) (Fenelon, Chen, & Baker, 2016). Considerable investigation has gone into evaluating traffic safety and measuring crash risks for various age groups and environmental conditions. One of the most common measures of crash risk is crash rate, or dividing the absolute crash frequency by the driver's risk exposure. Traditionally, researchers have measured driver's risk exposure using metrics such as number of residents, number of licensed drivers, vehicle miles driven, and registered vehicles.

However, these commonly used metrics of exposure introduce different amounts of bias (Regev, Rolison, & Moutari, 2018). For example, when using number of residents, these numbers are representative of the general population and may not fully represent people "at-risk". Additionally, using the number of licensed drivers also presents bias as numbers may be dependent on certain roadway and environmental characteristics. For example, compared to younger drivers, older drivers are less likely to travel on high-speed roads and are more likely to limit their driving to daylight hours compared to drivers of younger ages (Langford & Koppel, 2006). To add to this concern, the accuracy of US licensed driver data obtained from the Federal Highway Administration has also been challenged and well documented (e.g., the number of licensed young drivers are under reported) (Foss & Martell, 2013; Curry, Kim, & Pfeiffer, 2014a), further restricting the use of licensed drivers as an estimate of driving exposure. Researchers were also restricted by the accessibility or feasibility of collecting required disaggregated estimates of vehicle miles driven for a particular driver or vehicle subgroup and driving conditions. The unavailability of disaggregated estimates also applies to registered vehicles.

To overcome the aforementioned limitations associated with traditional driving exposure metrics, an alternative approach to measuring driving exposure is the quasi-induced exposure (QIE) method (Stamatiadis & Deacon, 1997). This method uses the number of not-at-fault drivers for a specific driver and vehicle subgroup and driving condition from the crash database itself, instead of retrieving traditional estimates (e.g., vehicle miles driven and number of licensed drivers) from additional, outside databases. By estimating the number of not-at-fault drivers, it allows researchers to calculate the crash risk for a driver/vehicle

subgroup under a particular driving condition which is difficult or even impossible to evaluate using the traditional driving exposure (e.g., number of licensed drivers and vehicle miles driven). The crash risk for a driver/vehicle subgroup was defined as the ratio of the number of at-fault drivers to not-at-fault drivers in a specific condition (Stamatiadis & Deacon, 1997). Drivers with higher ratios of being at-fault compared to not being at-fault in a specific crash type were at greater crash risk than those with lower odds ratios. The underlying assumption of QIE is that the not-at-fault drivers for a particular group in clean two-vehicle crashes (i.e., one and only one driver was responsible for the crash) were drawn randomly from the general driving population, suggesting not-at-fault drivers serve as a reasonable representative sample of drivers under a particular driving condition.

Although many studies have applied the QIE method to examine crash risks for various driver/vehicle groups and crash location/time conditions (Aldridge et al., 1999; Stamatiadis & Puccini, 1999; McIntosh & Katcher, 2000; Rice, Peek-Asa, & Kraus, 2003; Curry, 2017), the validity of the QIE and its assumption of representativeness to the driving population at the time of crash (i.e., representativeness assumption) remains a topic of discussion. Previously, using Michigan crash data from 1982–1988, Lyles, Stamatiadis, and Lighthizer (1991) compared the distributions of non-at-fault drivers with the distributions obtained from field observations conducted within southwestern Michigan. Lyles et al. (1991) further tested the assumption by examining the independency of at-fault and non-at-fault driver distributions. Findings revealed that not-at-fault driver distributions were independent of at-fault driver distributions, suggesting not-at-fault drivers were truly randomly selected from the general driving population (Lyles et al., 1991). In slight contradiction, Stamatiadis and Deacon (1997) found similarities between at-fault and non-at-fault drivers using a Kentucky crash database from 1990–1992. The authors found that younger at-fault drivers were more likely to hit younger not-at-fault drivers, concluding the correlation of age groups between at-fault and not-at-fault drivers may be explained by similarities in driving patterns among drivers in a same age group (e.g., driving time and roadway types frequently driven) (Stamatiadis & Deacon, 1997).

Building on these findings, recent studies have used alternative approaches to evaluate the QIE method by comparing non-at-fault driver distributions in clean two-vehicle crashes with not-at-fault driver distributions involved in clean three-or-more-vehicle crashes, with respect to driver age and gender and vehicle types (Chandraratna & Stamatiadis, 2009; Jiang & Lyles, 2010; Curry, Pfeiffer, & Elliott, 2016). Those studies argued that if the randomness assumption among not-at-fault drivers was indeed true in clean two-vehicle crashes, it was expected that the not-at-fault drivers in clean three-or-more-vehicle crashes should be at greater randomness and not-at-fault drivers would be similar to later not-at-fault drivers in three-or-more vehicle crashes. Overall, these studies report that distributions of driver/vehicle group and crash time/locations between two-vehicle and three-or-more-vehicle crashes were not statistically different, in support for the validity of the QIE representativeness assumption.

Nonetheless, all of the above-mentioned studies (Chandraratna & Stamatiadis, 2009; Jiang & Lyles, 2010; Curry et al., 2016) only used a single state crash database (e.g., Michigan, Kentucky, and New Jersey, respectively). Yet to be studied is the generality of QIE

representativeness assumption on a national scale. More importantly, as the previous approaches mentioned was conducting validations via comparing drive/vehicle and crash time/location distributions within one crash database, the validation results may still be subject to system bias. That is, if any subgroup of not-at-fault drivers in that crash database was a biased sample from the driving population, comparisons between subgroups of non-at-fault drivers may not be able to demonstrate the validity of the QIE representativeness assumption. Furthermore, those studies were guided by the notion that if the QIE representative assumption was true in clean two-vehicle crashes, such assumption would be valid among not-at-fault drivers in three-or-more vehicle crashes (which should have greater randomness than those in earlier two-vehicle crashes). Thus, if the not-at-fault drivers in two-vehicle crashes were similar to those in later three-or-more vehicle crashes, the representative assumption of the QIE in clean two-vehicle crashes was valid. However, those studies might be subjected to circular reasoning issues. That is, they began with the representativeness assumption of the QIE method in clean two-vehicle crashes (building on which not-at-fault drivers in three-or-more vehicles could be representative to driving population and utilized as the reference group) and ended with the demonstration of the representativeness assumption in clean two-vehicle crashes.

To address the aforementioned limitations, this study aimed to determine if the QIE representativeness assumption was valid on a national scale using US traffic databases, the Fatality Analysis Reporting System (FARS) and the National Occupant Protection Use Survey (NOPUS). In order to avoid system bias introduced by comparing driver/vehicle distributions within a single crash database, we compared fatal crashes from FARS with nationally representative field observations from NOPUS. Not-at-fault drivers in clean two-vehicle crashes, the first not-at-fault drivers in clean three-or-more-vehicle crashes, and all other not-at-fault drivers excluding the first not-at-fault drivers in clean three-or-more-vehicle crashes were directly compared to NOPUS. Furthermore, in efforts to improve reliability and avoid circular reasoning, such comparisons did not make any assumptions about the randomness among drivers in FARS.

2. Materials and methods

2.1. Data sources

Drivers from NOPUS between the years of 2015–2017 (National Highway Safety Administration, 2016b) were the primary data source for estimating the general driving population. NOPUS is the only national representative and probability-based survey to examine driver seatbelt use and electronic device use in the US. (National Highway Safety Administration, 2016b). The survey was conducted at probabilistically sampled interactions (e.g., traffic stops signs and express exit ramps) and expressways in June each year between 7 am to 6 pm (National Highway Safety Administration, 2016b). Trained observers recorded vehicle occupants' gender, age, and vehicle types based on their subjective, roadside or in-vehicle assessments (National Highway Safety Administration, 2016b). The NOPUS data assigned a final weight for each occupant so that observation data can be normalized to the US national level. It also provided each occupant with 28 replicate weights which can be used to estimate the standard errors for national estimates using jackknife variance

estimation method. As the number of replicate weights assigned to each occupant varied between 2010 and 2017 (62 for 2010, 56 for 2011–2014, and 28 for 2015–2017), data before 2014 were not included in this study. The 2015–2017 NIPUS data classified drivers into three age groups: younger drivers (16–25 years old), middle-aged drivers (25–69 years old), and older drivers (70 and older years) and provided three subtypes of passenger vehicle: passenger cars, pickup trucks, and other vehicles (e.g., SUVs and vans). NIPUS also classified their observation time into three categories (weekend, workday rush hours, and workday non-rush hours). Rush hour was comprised of 7–9:30 a.m. and 3:30–6 p.m. during workdays. The roadway type in NIPUS was coded in two categories: 1) Expressways (roadways with limited access) and 2) surface street (all other roadways).

The not-at-fault driver data were obtained from the 2015–2017 FARS database. The FARS is a census of all crashes on US public roadways that result in at least one fatality within 30 days of the crash (National Highway Safety Administration, 2018). Following the QIE method, we selected all crashes involving two or more vehicles with a clear assignment of the responsibility to only one driver. The crash responsibility was determined using driver-level crash-related factors recorded in FARS. From 2015–2017, up to 4 driver-level crash related factors are recorded for crashes within FARS. Following previous studies which have used driver-level crash-related factors between 20 and 60 from FARS to indicate crash responsibility (Blower, 1998; Dubois et al., 2015), factors below 20 and greater than 60 in the FARS codebook were not considered as indicators of crash responsibility as they were not considered as direct contributors to crashes (Blower, 1998). Those factors (< 20 or > 60) included physical/mental conditions (e.g., emotional factors), some miscellaneous factors (e.g., traveling on prohibited traffic ways), and possible in-vehicle distractions (e.g., cell phone use) (Blower, 1998). We did not use police issued violations or citations to determine crash responsibility, as using citation data to determine crash responsibility might be subjected to statistical bias in the characterizations of not-at-fault drivers (Jiang et al., 2012; Curry et al., 2014b).

We identified all clean multivehicle crashes in the 2015–2017 FARS data and the not-at-fault drivers of those crashes. As driver details of NIPUS are limited to observations, crash characteristics from FARS were selected to best match characteristics from NIPUS. These changes included selecting not-at-fault drivers from crashes in FARS that occurred between 7:00 am and 6:00 pm, those who were in passenger vehicles, and drivers who were 16 years or older. As there was no classification of other vehicle type in FARS, passenger vehicles in FARS were coded into two types: pickup trucks and passenger vehicles excluding pickup trucks. Thus, for NIPUS, the vehicle types of passenger cars and other vehicles (including vans and SUVs) were combined to match FARS classification, named as passenger cars (including vans and SUVs). As a result, the vehicle types in this study were passenger cars (including vans and SUVs) and pickup trucks.

2.2. Statistical tests

To test the validity of the QIE representative assumption, the distributions of driver gender, age, vehicle type, time periods, and roadway types from NIPUS were compared to the distributions of not-at-fault drivers in FARS. Following Jiang and Lyles (2010) and Curry et

al. (2016), the not-at-fault drivers in the FARS were classified into three groups: 1) not-at-fault drivers in clean two-vehicle crashes (D2); 2) the first not-at-fault drivers in clean three-or-more-vehicle crashes (D3_1); and 3) all other not-at-fault drivers excluding the first not-at-fault drivers in clean three-or-more-vehicle crashes (D3_other).

We estimated driver/vehicle and time/location (roadway types) distributions for FARS and NOPUS by calculating their percentage points. With the replicate weights in NOPUS, the 95% confidence limits (CLs) for national percentage point estimates were computed using jackknife variance estimation method. For FARS, the 95% CLs for percentage point estimates were regular CL for a proportion. If the CLs for a distribution (e.g., gender and roadway types) between FARS and NOPUS did not overlap, the distribution between them were considered statistically significantly different. With respect to the practical difference, Jiang and Lyles (2010) have previously defined a 4 percentage point difference as a meaningful threshold. Given lack of empirical justification, we did not use the 4 percentage point difference as the major difference threshold. Instead, we presented the actual percentage point difference and referred to the practical difference threshold for interpretation. As Stamatiadis and Deacon (1997) suggested, we also separated crash data into homogenous conditions by time and location to further test the validity of the representativeness assumption of QIE for individual time and location by driver age groups.

All of the analyses were conducted in SAS Enterprise Guide 7.1 (SAS Institute, 2015) and R 3.5.0. In SAS, *Proc Survey* was used to calculate the weighted frequency of driver gender and age groups, vehicle types, observation time and locations in NOPUS and the *binconf* function in R was used to calculate the CLs for estimates in FARS.

NOPUS is a probability-based observational survey and the weight of every observation to the distributions of driver population varies. Therefore, it limited our choice conducting the parametric (e.g., Chi-square test) or nonparametric statistical tests (e.g., the Wilcoxon Mann-Whitney and Wilcoxon signed-rank tests) adopted by Curry et al. (2016), Chandraratna and Stamatiadis (2009), and Jiang and Lyles (2010). These conventional statistical tests did not differentiate observations with different weights.

3. Results

From 2015–2017, there were 15,134 not-at-fault drivers involved in clean multivehicle fatal crashes, including 9,997 drivers in clean two-vehicle crashes (D2) and 5,137 drivers in three-or-more vehicle crashes (D3_1 and D3_other). Among those involved in three-or-more-vehicle fatal crashes, 2,481 drivers were the first not-at-fault drivers (D3_1) and the remaining 2,656 were all other not-at-fault drivers excluding the first not-at-fault drivers (D3_other).

Table 1 shows the distribution of age, gender, vehicle type, time, and roadway types among all subgroups of not-at-fault drivers. Except for age, all the distributions among D3_other were not significantly different from NOPUS observations. However, in addition to age, vehicle type distributions among D2 and D3_1 were statistically significantly different from NOPUS. The distribution of roadway type among D2 was also statistically significantly

different from NOPUS. In D2, approximately 28% drivers were in pickup trucks and 26% drove on the expressways at the time of crash, while NOPUS observed less than 20% pickup drivers on the road and almost 40% drove on the expressways. Among D3_1, the proportion of drivers in pickup trucks were also slightly larger than their counterparts in NOPUS (22.5 [20.9, 24.2] versus 18.9 [17.5, 20.3]). With respect to age, younger drivers (16–24 years old) had higher proportions in D2, D3_1, and D3_other than those in NOPUS. However, it is worthwhile to note that percentage point differences in age between D3_other and NOPUS were smaller than 4, the practical difference threshold previously.

Additionally, Figure 1 and 2 show the comparisons of driver age groups by time and roadway types. During weekend, the distribution of age between D3_other and NOPUS were not significantly different. All other distributions of age by time were statistically significantly different between not-at-fault drivers and NOPUS. The largest percentage point difference of age between D3_other and NOPUS was 5.7 for drivers aged 25–69 years old on weekdays during non-rush hours. However, compared to D2 and D3_1, the driver age distributions among D3_other were less deviated from those among NOPUS. Similarly, although age distributions by roadway types in D2, D3_1, and D3_other were different from those in NOPUS, D3_other were less deviated from NOPUS observations than D2 and D3_1. The largest percentage point difference was 4.8% for drivers 25–69 years old on expressways.

4. Discussion

The objective of this study was to validate the representativeness assumption of quasi-induced exposure (QIE) in the Fatality Analysis and Reporting System (FARS) database without any randomness assumptions of the not-at-fault drivers in D2, D3_1, and D3_other before demonstration on a national scale by comparing it a national representative field observation survey, the National Occupant Protection Use Survey (NOPUS). With respect to driver gender, vehicle types, crash time, and roadway type, D3_other in FARS did not show statistically significant differences from NOPUS. However, the distributions of roadway type among D2 was statistically significantly different from that of NOPUS observations, and the proportion of D2 drivers on expressways were smaller than that of NOPUS. One possible explanation for this finding may be because severe multiple-vehicle crashes that occurred on expressways are more likely to involve more than two vehicles. The insignificant difference in roadway types between not-at-fault drivers in three-or-more-vehicle crashes (D3_1 and D3_other) and the NOPUS data further supports this explanation. In addition, drivers among D2 and D3_1 were more likely to sit in pickup trucks than drivers observed by NOPUS. Based on NOPUS observations, drivers in pickup trucks were less likely to wear a seatbelt compared to those in passenger cars and SUVs (National Highway Safety Administration, 2016a). Compared to belted drivers, unbelted drivers may be more likely to exhibit aggressive driving behaviors, resulting in their increased likelihood of being involved into severe crashes (Wilson, 1990). It may also partially explain the overrepresentation of the pickup trucks drivers in D2 and D3_1. Although they were not assigned a specific driver-related contributing behavior, they might show more aggressive driving behaviors at the time of crash.

Related to driver age, all not-at-fault drivers among D2, D3_1, and D3_other showed statistically significant difference to NOPUS. Younger and older not-at-fault drivers (16–24 and 70 years old) in FARS, particularly those among D2 and D3_1, had higher proportions than their corresponding proportions observed by NOPUS. This finding could be due in part to younger drivers' driving inexperience and older drivers' reduced motor, perceptual, and cognitive capabilities resulting in poorer defensive driving skills in a crash. It also should be noted that the classification of age in NOPUS was subjectively assessed by observers, while driver age in FARS was recorded by police at the time of the crash based on legal identity documents (e.g., driver license) (National Highway Safety Administration, 2018). Therefore, in addition to the potential deviations of not-at-fault drivers in FARS to the driving population, the inconsistencies in the distribution of driver age group between FARS and NOPUS might also result from the imperfect assessment of age from NOPUS data collectors.

However, if the practical difference defined by Jiang and Lyles (2010) was applied, the distribution of driver age group between D3_other and NOPUS was not practically different (the percentage point difference for each age group was smaller than 4) (Table 1). By slicing the data by time and roadway types (Figure 1 and 2), compared to D2 and D3_1, the driver age distribution in D3_other was more similar to that of NOPUS across all the time periods and roadway types. During weekends, there was no statistically significant difference in driver age distributions between FARS and NOPUS. Additionally, for D3_other drivers, their age distributions were less deviated to that of NOPUS on surface streets than on expressways. This could potentially be explained by lower traffic density on weekends compared to weekdays, possibly resulting in observers' better assessments of driver characteristics. In the case of expressways, observers collected data while riding in a vehicle (National Highway Safety Administration, 2016a), which potentially increased their difficulties to correctly estimate the driver age group, relatively to roadside observations. Thus, in improved conditions which could facilitate NOPUS data collectors' observations, the inconsistencies between FARS and NOPUS in driver age distribution resulting from NOPUS age classification imperfections could be reduced. The percentage point differences in age group between D3_other and NOPUS by time and roadway types were within (weekend and surface streets) or nearly within (weekday rush/non-rush hour and expressways) 4-percentage-point-difference threshold defined by Jiang and Lyles (2010). More importantly, there was no practical difference in the driver age group between D3_other and NOPUS overall.

Our study suggests that in a biased traffic crash database, such as FARS, D3_other could serve as a reasonable representative sample of the driving population with respect to driver age, gender, vehicle type, crash time, and location in the US. Many previous studies that applied the QIE principles in FARS were subject to the generalizability issues as no analysis had been conducted to test the validation of the representativeness of the selected not-at-fault drivers in their studies (Dubois et al., 2015; Li, Chihuri, & Brady, 2017; Chihuri & Li, 2019). Furthermore, Curry et al. (2016) argues that validation of the representativeness assumption of the not-at-fault drivers is a critical step in applying QIE methods. Building on the results of our study, instead of the not-at-fault drivers in D2 and D3_1, those among D3_other in FARS could be used as an alternate sample representing the driving population

at the time of crash. Our study increases the credibility of QIE method with applications in FARS and has potential to improve the generalizability of future FARS studies. Furthermore, our results provide support to previous studies which assumed D3_other drivers were randomly selected from the general driving population beforehand and used them as the reference group to validate the QIE representativeness assumption among D2 and D3_1 (Jiang & Lyles, 2010; Curry et al., 2016). As D3_other drivers in FARS could serve as a reasonable sample of the driving population in the US, D3_other drivers in other crash databases which included more general crash outcomes than FARS could provide even better fit. This provides a solid foundation for future studies to validate the QIE representativeness assumption by providing evidence towards argument about the randomness of D2, D3_1, and D3_other.

Along with many strengths, this study also has some notable limitations. First, the data collecting procedures between FARS and NOPS differ in methodology. Although we did our best to match observations between FARS and NOPS, some inconsistencies between FARS and NOPS resulted from differences in data collecting procedures could not be avoided completely. For example, the difference of younger and older drivers may partially be attributed to the difference of age classification between NOPS and FARS. The imperfect classification of age is associated with most NOPS studies (Zhu et al., 2016). Additionally, NOPS data collection is limited to favorable driving conditions and may not be completely generalizable to extreme weather conditions across regions of the US. For example, NOPS observations were conducted during June but, due to sample size restrictions, we were not able to restrict FARS observations to crashes that happened only within June. Also, as NOPS observations were only conducted in June, seasonality may influence individual driving patterns which must be considered as it may introduce some level of unknown bias into the analysis. Laapotti et al. (2006) found that younger drivers are more likely to spend their time engaging in leisure-time driving in comparison to middle-aged counterparts who spend more time driving to or from work. As younger drivers are typically out of school during the month of June and summer holidays may increase travel congestion (Memmott & Young, 2008), these seasonality differences directly impact traffic patterns during the observation window for NOPS.

In addition, as NOPS is comprised of field observation survey, the data collectors were not able to interact with drivers on the road. This restricts the detail of driver and vehicles description available when compared to the data available within the FARS database. Thus, it limits the possibilities to validate the QIE representativeness assumption into more detailed driver, vehicle, time, and crash location groups. Lastly, the crash database we used for this study only included severe fatal crashes. Future studies should incorporate national representative crash database with all levels of severities (e.g., the Crashworthiness Data System and General Estimates System) to validate the QIE representativeness assumptions further.

5. Conclusions

Overall, we demonstrated that, on the US national scale, the remaining not-at-fault drivers after excluding the first not-at-fault driver in three-or-more-vehicle fatal crashes is

reasonable sample of the driving population at the time of crash. **Practical Application:** Our results provide opportunities for future studies to identify an unbiased driving population sample in FARS which only includes severe fatal crashes and improve result generalizability. More importantly, our study provides a solid foundation for future traffic-related studies to validate the quasi-induced exposure technique and to ensure appropriate QIE method application to the crash database.

References:

- Aldridge B, Himmler M, Aultman-Hall L, & Stamatiadis N (1999). Impact of passengers on young driver safety. *TRANSPORT RES REC*(1693), 25–30.
- Blower D (1998). The relative contribution of truck drivers and passenger car drivers to two-vehicle, truck-car traffic crashes (UMTRI 98–25). Retrieved from <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/1246/91183.0001.001.pdf?sequence=2>
- Chandraratna S, & Stamatiadis N (2009). Quasi-induced exposure method: evaluation of not-at-fault assumption. *Accid Anal Prev*, 41(2), 308–313. doi:10.1016/j.aap.2008.12.005 [PubMed: 19245890]
- Chihuri S, & Li G (2019). Use of Prescription Opioids and Initiation of Fatal 2-Vehicle Crashes. *JAMA Netw Open*, 2(2), e188081. doi:10.1001/jamanetworkopen.2018.8081 [PubMed: 30768194]
- Curry AE (2017). Estimating young novice drivers' compliance with graduated driver licensing restrictions: A novel approach. *Traffic Inj Prev*, 18(1), 35–40. doi:10.1080/15389588.2016.1171857 [PubMed: 27064815]
- Curry AE, Kim KH, & Pfeiffer MR (2014a). Inaccuracy of Federal Highway Administration's licensed driver data: implications on young driver trends. *J Adolesc Health*, 55(3), 452–454. doi:10.1016/j.jadohealth.2014.05.011 [PubMed: 25151055]
- Curry AE, Pfeiffer MR, & Elliott MR (2016). Validation of quasi-induced exposure representativeness assumption among young drivers. *Traffic Inj Prev*, 17(4), 346–351. doi:10.1080/15389588.2015.1091072 [PubMed: 26376230]
- Curry AE, Pfeiffer MR, Myers RK, Durbin DR, & Elliott MR (2014b). Statistical implications of using moving violations to determine crash responsibility in young driver crashes. *Accid Anal Prev*, 65, 28–35. doi:10.1016/j.aap.2013.12.006 [PubMed: 24398139]
- Dubois S, Mullen N, Weaver B, & Bedard M (2015). The combined effects of alcohol and cannabis on driving: Impact on crash risk. *Forensic Sci Int*, 248, 94–100. doi:10.1016/j.forsciint.2014.12.018 [PubMed: 25612879]
- Fenelon A, Chen LH, & Baker SP (2016). Major Causes of Injury Death and the Life Expectancy Gap Between the United States and Other High-Income Countries. *Jama*, 315(6), 609–611. doi:10.1001/jama.2015.15564 [PubMed: 26864416]
- Foss R, & Martell C (2013). Did graduated driver licensing increase the number of newly licensed 18-year-old drivers in North Carolina. Paper presented at the Transportation Research Board Annual Meeting, Washington, DC
- Jiang X, & Lyles RW (2010). A review of the validity of the underlying assumptions of quasi-induced exposure. *Accid Anal Prev*, 42(4), 1352–1358. doi:10.1016/j.aap.2010.02.016 [PubMed: 20441852]
- Jiang X, Qiu Y, Lyles RW, & Zhang H (2012). Issues with using police citations to assign responsibility in quasi-induced exposure. *Safety Sci*, 50(4), 1133–1140.
- Laapotti S, Keskinen E, Hatakka M, Hernetkoski K, Katila A, Peraaho M, & Salo I (2006). Driving circumstances and accidents among novice drivers. *Traffic Inj Prev*, 7(3), 232–237. doi:10.1080/15389580600668537 [PubMed: 16990237]
- Langford J, & Koppel S (2006). Epidemiology of older driver crashes—identifying older driver risk factors and exposure patterns. *Transp Res Part F Traffic Psychol Behav*, 9(5), 309–321.
- Li G, Chihuri S, & Brady JE (2017). Role of alcohol and marijuana use in the initiation of fatal two-vehicle crashes. *Ann Epidemiol*, 27(5), 342–347.e341. doi:10.1016/j.annepidem.2017.05.003 [PubMed: 28595738]

- Lyles RW, Stamatiadis P, & Lighthizer DR (1991). Quasi-induced exposure revisited. *Accid Anal Prev*, 23(4), 275–285. [PubMed: 1883467]
- McIntosh G, & Katcher M (2000). Graduated driver licensing in Wisconsin: a new law for teens and parents. *Wmj*, 99(9), 31–33, 30.
- Memmott J, & Young P (2008). Seasonal variation in traffic congestion: A study of three US cities (TR-005). Retrieved from <https://rosap.ntl.bts.gov/view/dot/34970>
- National Highway Safety Administration. (2016a). Seat Belt Use in 2016 - Overall Results (DOT HS 812 351). Retrieved from <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812351>
- National Highway Safety Administration. (2016b). Traffic safety facts: Driver electronic device use in 2015 (DOT HS 812 326). Retrieved from <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812326>
- National Highway Safety Administration. (2018). Fatality Analysis Reporting System (FARS). Retrieved from <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>. Accessed 02/15, 2018
- Regev S, Rolison JJ, & Moutari S (2018). Crash risk by driver age, gender, and time of day using a new exposure methodology. *J Safety Res*, 66, 131–140. doi:10.1016/j.jsr.2018.07.002 [PubMed: 30121099]
- Rice TM, Peek-Asa C, & Kraus JF (2003). Nighttime driving, passenger transport, and injury crash rates of young drivers. *Inj Prev*, 9(3), 245–250. [PubMed: 12966014]
- SAS Institute, C., NC,. (2015). SAS Enterprise Guide 7.11 Cary, NC, USA: SAS Institute Inc.
- Stamatiadis N, & Deacon JA (1997). Quasi-induced exposure: methodology and insight. *Accid Anal Prev*, 29(1), 37–52. [PubMed: 9110039]
- Stamatiadis N, & Puccini G (1999). Fatal crash rates in the Southeastern United States: why are they higher? *TRANSPORT RES REC*(1665), 118–124.
- Wilson RJ (1990). The relationship of seat belt non-use to personality, lifestyle and driving record. *Health Educ Res*, 5(2), 175–185.
- Zhu M, Rudisill TM, Heeringa S, Swedler D, & Redelmeier DA (2016). The association between handheld phone bans and the prevalence of handheld phone conversations among young drivers in the United States. *Ann Epidemiol*, 26(12), 833–837.e831. doi:10.1016/j.annepidem.2016.10.002 [PubMed: 27894566]

Figure 1.
Comparison of the distributions of age by time periods

Figure 2.
Comparison of the distribution of age by roadway types

Table 1.
Descriptive statistics of D2, D3_1, and D3_other distributions versus NOPUS distribution, 2015–2017

Variables	n	D2 ^a % (95% CL)	n	D3_1 ^b % (95% CL)	n	D3_other ^c % (95% CL)	n ^d	NOPUS % (95% CL) ^e	D2	D3_1	D3_other
Gender											
Male	6208	62.1 (61.1, 63.0)	1488	60.0 (58.0, 61.9)	1542	58.1 (56.2, 59.9)	25,363,928	60.0 (58.7, 61.3)	2.1	0.0	2.0
Female	3789	37.9 (37.0, 38.9)	993	40.0 (38.1, 42.0)	1114	41.9 (40.1, 43.8)	16,895,948	40.0 (38.7, 41.3)	2.1	0.0	2.0
Age group											
16–24 years	1343	13.4 (12.8, 14.1)	338	13.6 (12.3, 15.0)	318	12.0 (10.8, 13.3)	3,629,053	8.6 (7.6, 9.5)	4.8	5.0	3.4
25–69 years	7335	73.4 (72.5, 74.2)	1880	75.8 (74.1, 77.4)	2151	81.0 (79.4, 82.4)	35,874,004	84.9 (83.6, 86.2)	11.5	9.1	3.9
> 70 years	1319	13.2 (12.5, 13.9)	263	10.6 (9.4, 11.9)	187	7.0 (6.1, 8.1)	2,756,818	6.5 (5.8, 7.3)	6.7	4.1	0.5
Vehicle types											
Passenger cars	7188	71.9 (71, 72.8)	1923	77.5 (75.8, 79.1)	2122	79.9 (78.3, 81.4)	34,267,419	81.1 (79.7, 82.5)	9.2	3.6	1.2
Pickup trucks	2809	28.1 (27.2, 29)	558	22.5 (20.9, 24.2)	534	20.1 (18.6, 21.7)	7,992,456	18.9 (17.5, 20.3)	9.2	3.6	1.2
Time periods											
Weekend	2728	27.3 (26.4, 28.2)	547	22.0 (20.5, 23.7)	594	22.4 (20.8, 24.0)	11,431,890	27.1 (23.3, 30.8)	0.2	5.0	4.7
Weekday rush	3458	34.6 (33.7, 35.5)	943	38.0 (36.1, 39.9)	1025	38.6 (36.8, 40.5)	14,854,264	35.1 (31.7, 38.6)	0.6	2.9	3.4
Weekday non-rush	3811	38.1 (37.2, 39.1)	991	39.9 (38.0, 41.9)	1037	39.0 (37.2, 40.9)	15,973,721	37.8 (34.3, 41.3)	0.3	2.1	1.2
Roadway types											
Surface streets	7354	73.6 (72.7, 74.4)	1553	62.6 (60.7, 64.5)	1618	60.9 (59.0, 62.8)	25,678,814	60.8 (56.1, 65.4)	12.8	1.8	0.2
Expressways	2643	26.4 (25.6, 27.3)	928	37.4 (35.5, 39.3)	1038	39.1 (37.2, 41.0)	16,581,061	39.2 (34.6, 43.9)	12.8	1.8	0.2

Note:

^aD2 refers to a not-at-fault driver in two-vehicle crash;

^bD3 refers to the first not-at-fault driver in a three-or-more-vehicle crash;

^cThe remaining not-at-fault drivers excluding the first not-at-fault drivers in a three-or-more-vehicle crash;

^dThe frequency was normalized to the US national scale using the assigned weight for each observation;

^eThe confidence limits (CLs) of the estimates on national scale was obtained using the jackknife variance estimation method with replicate weights; Values in bold indicates that the percentage point difference in proportions between statistics from FARS and NOPUS was not significantly different