How dangerous are drinking drivers now? Replicating and updating Levitt and Porter

Richard A. Dunn
University of Connecticut

Nathan W. Tefft
Bates College

Abstract: Drinking-and-driving remains a leading cause of preventable mortality and morbidity in the United States. In this article, we apply methods introduced by Levitt and Porter (this journal, 2001) to examine how the prevalence, relative risk, and externality of drinking-and-driving has evolved over the past three decades. We begin with a thorough replication of the original Levitt and Porter analysis. Although we identify several potentially important flaws in their implementation, we find that these are ultimately innocuous and do not appreciable alter their conclusions. Finally, we introduce multiple imputation to their methodology, thereby accommodating recent advances in the available data.

I. Introduction

Although deaths from motor vehicle crashes have fallen significantly over the past three decades, drinking and driving remains a leading cause of preventable mortality and morbidity in the United States.[1] In this article, we apply the methods originally introduced by Levitt and Porter (2001, hereafter referenced as *LP*) to examine how the prevalence and relative risk of drinking and driving has evolved over the past three decades to better understand the causes of this decline and provide up-to-date estimates of the external cost associated with drinking and driving.

In contrast to the overwhelming majority of studies that investigate drinking and driving behavior using data collected from surveys, *LP* demonstrated how a census of fatal motor vehicle crashes—data that is readily available in the United States through the Fatality Analysis Reporting System—could be used to recover both the prevalence and relative risk of drinking and driving. By employing a census of crashes, their methods overcome widely recognized concerns of sample selection and misreporting bias in survey instruments.

In a preview of our main results, we find that the prevalence of drinking and driving has fallen significantly. Though this pattern is consistent with statistics generated from various survey-based approaches, we continue to find that rates of drinking and driving are substantially higher than those recovered from surveys, suggesting that selection and misreporting biases remain significant methodological challenges that researchers who adopt such data sources must address.

Further, at the same time that the prevalence of drinking and driving has decreased, we find that the relative risk of drinking drivers has increased. Given the overall decline in motor vehicle fatalities, this finding indicates that while driving has become much less dangerous over the past three decades, improvements in automobile technology, road design, and traffic enforcement have had proportionally larger effects on reducing the harm caused by non-drinking drivers. Despite the increase in relative risk, however, the externality imposed by drinking drivers on other road users has actually declined.

Because we seek to document changes in the prevalence and risk associated with drinking and driving over time, an additional contribution of this empirical exercise is providing the first published effort at replicating the original results of *LP*. We find that their results are not exactly replicable based on the data and methods presented in their article. We identify two principal sources of replication error: ambiguity in a sample refinement step that omits an indeterminate

[1] The Centers for Disease Control and Prevention estimate that for every 100,000 individuals under age 75 in 2013, 387 years of potential life were lost because of injuries sustained in motor vehicle crashes, compared to a rate of 913 years in 1980 and 717 years in 1990 (Centers for Disease Control and Prevention, 2015). For reference, heart disease, malignant neoplasms, and cerebrovascular disease were associated with 952, 1329, and 158 years of potential life lost before age 75, respectively. Among males, mortality from motor vehicle crashes accounted for more than eleven times as many years of potential life lost before age 75 than prostate cancer: 552 versus 47 per 100000 males under age 75. Among females, mortality from motor vehicle crashes accounted for nearly as many years of potential life lost before age 75 as breast cancer: 219 versus 250 per 100000 females under age 75.

set of state-year observations from their analysis and misinterpretation of a FARS variable definition that *LP* use to determine the alcohol-impairment status of drivers. In addition, we find that the method *LP* employed to calculate the prevalence of drinking and driving leads to standard errors an order of magnitude too small.

Nonetheless, when we estimate the prevalence and relative-risk of drinking drivers between 1983 and 1993 using the measure of alcohol status that *LP* initially intended, we find that these replication issues cause only relatively minor differences with the original results. Based on our analysis, we conclude that the methods proposed by *LP* are fairly straightforward to implement and provide both researchers and policymakers an important, but underutilized tool for understanding the causes and consequence of drinking and driving behavior.

II. Drinking and driving in the United States

Figure 1 plots fatality counts from crash reports collected in FARS from 1994 to 2013 based on the alcohol-impairment status of the drivers involved. Over that period, the number of deaths associated with motor vehicle crashes involving at least one driver with a blood alcohol content (BAC) greater than 0.01 g/dL (alcohol-involved) fell 23.9 percent between 1994 and 2013. Concomitantly, the number of fatalities associated with motor vehicle crashes involving at least one driver above the *per se* legal BAC of 0.08 g/dL (alcohol-impaired) fell 24.8 percent.

Despite these impressive improvements, the mortality cost of drinking and driving remains large: in 2013 there were 11,896 fatalities associated with alcohol-involved driving (36% of all motor vehicle crash fatalities) and 10,076 deaths (31% of motor vehicle crash fatalities) associated with alcohol-impaired driving (NHTSA, 2014).

A reduction in the number of alcohol-related motor vehicle crash fatalities could reflect a decline in the risk associated with drinking and driving or a decrease in the number of drinking drivers. Indeed, both factors have likely been salient over the past thirty years.

Significant advances in automobile technology, including widespread adoption of airbags, introduction of electronic stability control systems (ESC), improvements in tire design, stronger steel, and GPS navigation systems, have made driving less dangerous regardless of the alcohol impairment status of drivers. As is evident in Figure 1, at the same time that alcohol-related motor vehicle crash fatalities were decreasing, the number of fatalities associated with motor vehicle crashes in which all drivers exhibit a BAC=0.0 g/dL similarly declined from 24,928 in 1994 to 20,713 in 2013—a 17.0 percent decrease.

Nevertheless, survey evidence suggests that the incidence of drinking and driven has fallen, reflecting a combination of factors, including changes in social norms toward driving after consuming alcohol, increased vigilance on the part of law enforcement agencies, and imposition of stiffer penalties on those convicted of driving under the influence of alcohol. For example, the National Roadside Survey (NRS), a periodic survey of active road-users who are administered oral breathalyzer examinations indicates that the prevalence of drinking and driving has fallen from 22.3 percent in 1973 to 7.9 percent in 2007 (Lacey, et al., 2009b).

While suggestive, the validity of point estimates constructed from survey data is questionable. As we elucidate in the subsequent section, there are serious concerns about the tendency of individuals to misreport their drinking and driving behavior, as well as the potential for non-random selection of drivers into survey samples. And, because there is little reason to expect that the nature of these issues is time-invariant, survey-based approaches are unlikely to provide consistent estimates of either prevalence at a given point in time or consistent estimates of changes in prevalence over time.

III. The Reliability of Survey-based Estimates of Prevalence and Risk

Although survey instruments are the most common source of data in empirical studies of drinking and driving behavior, there are widely acknowledged concerns about the validity of estimates from such data.

Surveys that rely on self-reported driving and driving behavior, e.g., the Behavioral Risk Factor Surveillance System (BRFSS), are considered unsuitable because respondents tend to underreport drinking and driving (Lund and Wolfe, 1991; Voas, et al., 1998; Mullahy and Sindelar, 2007), biasing contemporaneous prevalence estimates downwards. Furthermore, survey respondents are typically unable to accurately assess their level of inebriation (Vogel-Sprott, 1975; Greenfield and Rogers, 1999; Johnson, et al., 2008). As a result, estimating the prevalence of alcohol-impaired driving above specified thresholds is all but impossible.

Although BRFSS has been used to study trends in drinking and driving behavior, the implicit assumption of time-invariant measurement error is unlikely to hold. If the stigma associated with drinking and driving is causing prevalence to decrease, it is almost surely influencing the decision of whether to report truthfully. Thus, declines in the prevalence of self-reported drinking and driving most likely overstate the true decrease in prevalence. Using self-reported drinking and driving behavior to study to make comparisons across different geographic areas and population subgroups are subject to similar criticism (Greenfield & Rogers, 1999; Caetano & McGrath, 2005).

An alternative information source for estimates of prevalence is the NRS, which periodically seeks a nationally representative sample of drivers who are asked to provide breathalyzer and blood samples (Lund and Wolfe, 1991; Voas, et al., 1998; Zador, Krawchuk and Voas, 2000; Compton and Berning, 2009; Lacey, et al., 2009b; Voas, et al., 2012; Berning, Compton and Wochinger, 2015).[2] Although the NRS overcomes the tendency of individuals to systematically under-report their alcohol-impaired driving behaviors by directly measuring BAC, drivers cannot be compelled to participate in the NRS, leading to potential sample selection bias.

For example, in the 2007 NRS, 1016 of 9553 drivers that were signaled to enter the testing site at night did not do so (10.6%). And out of the 9553 drivers originally signaled, only 7159 voluntarily submitted to a valid breathalyzer exam (74.9%). If alcohol-impaired drivers are less likely to enter the NRS testing site and less likely to submit a breath sample, then prevalence

[2] Because each NRS costs several million dollars, they are conducted infrequently (roughly once every decade: 1973, 1986, 1996, 2007, and 2013)

rates will tend to be under-estimated and, as a result, the relative-risk of alcohol-impaired driving will be over-estimated.

The NRS research team has developed various strategies to impute the BAC level of drivers who refuse to participate, but these come with their own set of assumptions, most notably that drivers who choose not to enter the NRS testing site do not differ in BAC from those that do enter.[3] Further, as strategies to improve participation and address driver selection are adopted and discarded, comparability of NRS estimates across time can become more difficult.

The potential for non-random selection of drivers into NRS has long been recognized, but the NRS potentially suffers from a less well understood source of sample selection. Recent versions of the NRS utilize a stratified-sampling approach based on the 60 primary sampling units (city/county regions) from the National Automotive Sampling System/General Estimates System (NASS/GES). Conducting the NRS survey protocol at a particular primary sampling unit (PSU) requires the cooperation of all law-enforcement agencies with jurisdiction over the area (city, county, state, etc), yet cooperation is not compulsory. For example, during the 2007 NRS, lack of cooperation required 25-30% of PSUs to be substituted with other jurisdictions willing to participate (Lacey, et al., 2009a). To the extent that law enforcement agencies in cities and regions have different procedures that influence the prevalence of alcohol-impaired driving and that these procedures are associated with willingness to participate in the NRS, it is difficult to determine whether the final sample of PSUs are truly representative of the US driving environment.

It is also worth noting that one of the reasons that jurisdictions have refused to participate in the NRS stems from efforts to increase driver participation rates. At survey sites, drivers are directed into the testing area by uniformed police-officers. Although there are always signs indicating that entering the testing site and participating in the survey is voluntary, this aspect of the data collection design has led to protests from the American Civil Liberties Union, as well as several lawsuits claiming violation of 4th Amendment protections against unreasonable search and seizure.[4] Based on the response, a number of jurisdictions have decided they will no longer participate in the NRS protocol, including Fort Worth, TX and Reading, PA.

---

[3] The assumption that no systematic differences exist between those who provided a BAC sample and those who refused was examined in the 2007 NRS by offering financial incentives to a sample of initial NRS refusers in an attempt to reverse their decisions. Although they find that drivers who initially refuse participation, but accept the incentive and provide a breath sample, are no more likely to be alcohol-impaired than those who had agreed to participate initially, this does not address the possibility that initial refusers who accept the financial incentive differ systematically from those who continue to refuse or those who decided not to enter the testing site.

The 2013 NRS introduced passive alcohol monitors that provided BAC measures using ambient air prior to drivers providing consent for a more precise breathalyzer method, i.e., before they could refuse participation. Coupled with the possible confusion of testing sites with sobriety check-points, the US House of Representative responded by passing an amendment prohibiting the National Highway Traffic and Safety Administration from using funds to support the NRS.

[4] Regarding one such case, an attorney representing the the independent contractor hired to implement the NRS on behalf of the National Highway Traffic and Safety Administration, replied that the plaintiff, "was in no way

IV. The LP Method

Given the potential bias associated with survey-based approaches to data collection, the statistical approach developed by *LP* provides researchers a potentially valuable alternative empirical framework. The application of statistical theory to indirectly identify the prevalence and relative risk of drinking and driving using data on motor vehicle crashes predates *LP*, e.g. induced and quasi-induced exposure methods (Cuthbert, 1994; Stamatiadis and Deacon, 1997; Kirk & Stamadiatis, 2001; Jiang and Lyles, 2010). Within this larger set of literature, their primary contribution is demonstrating that the distribution of driver types involved in two-vehicle fatal crashes is sufficient to identify both the prevalence of each type on the road and their risk of causing a fatal crash relative to a reference type.

Because the composition of drivers involved in two-vehicle fatal accidents is likely very different from the composition of drivers on the road, the conclusion that the distribution of driver types involved in such accidents is sufficient to identify both prevalence and relative risk may initially seem puzzling. Therefore, before recapitulating the theory, it is worth appreciating the elegant intuition behind their results.

Suppose two types of drivers exist, *A* and *B*. We can classify crashes involving two vehicles according to the types of drivers involved: *AA*, *BB*, and *AB*. If there is only one *A* driver in the population, then there can be no *AA* crashes. But, if half the crashes are *AB*, then that one type *A* driver must be very dangerous. Hence, the observed distribution of driver types involved in two-vehicle crashes reveals both the prevalence of type *A* drivers (very small) and their relative risk of causing a crash (very large).

More generally, if we believe type *A* drivers are more dangerous than type *B* drivers, given a fixed number of *AA* and *BB* fatal crashes, the larger the number of AB crashes, the more dangerous type *A* drivers must be. Similarly, given a fixed number of *AB* crashes, the larger the number of *AA* crashes, the greater the prevalence of type *A* drivers. Thus, even though drivers involved in two vehicle fatal crashes are not representative of the population of drivers, the distribution of driver types in such accidents provides all the necessary information to identify both prevalence and relative risk.

The methodology developed by *LP* is exceptionally flexible, allowing for any number of mutually exclusive driver categories. In this study, we restrict attention to the risk of drinking and legally impaired drivers relative to non-drinking drivers. Hence, we employ the simplest case possible: two driver types.[5] Following *LP*, classify drivers as *sober* (denoted *S*) if they have a BAC=0 and *drinking* (denoted *D*) if they have a BAC≥.01 g/dL with N_D and N_S denoting the number of each operating a vehicle within a given geographic area and time period (*sober* and

compelled to stop, and, indeed, hundreds of other vehicles completely ignored the civilian data collector and continued on their merry way." This response recalls the concerns about driver selection raised previously.

[5] *LP* extend their analysis to consider risk interacting driver characteristics and drinking status leading to four driver types, e.g., {male, drinking; male, non-drinking; female, drinking; female, non-drinking}. This extension is beyond the scope of the current study.

drinking were the terms employed by *LP* and we maintain their terminology and notation to make reference to their work easier).

For now, assume that both the number of interactions a driver has with other vehicles (passing in the same or opposite direction, following or leading, meeting at an intersection, etc.) and the composition of the drivers encountered is independent of driver type. Imposing both simultaneously is termed the equal-and-independent-mixing (*EIM*) assumption.[6] Then, the probability of interacting with a driver of type i conditional on an interaction occurring is $Pr(i|I=1)=N_i/(N_D+N_S)$ and the probability of a driver of type i interacting with a driver of type j is simply the product of these probabilities: $Pr(i,j|I=1)=N_iN_j/(N_D+N_S)^2$.

Further assume that a fatal crash occurs when a driver makes a fatal error, θ_i, the likelihood of which depends upon driver type (allowing for heterogeneity within driver type, θ_i is the mean fatal error probability for drivers of type i). Thus, the probability that a fatal crash occurs when a driver of type i interacts with a driver of type j is $Pr(A=1|I=1, i, j)=\theta_i+\theta_j+\theta_i\,\theta_j$. As the chance of a fatal crash occurring with any given interaction between vehicles is extremely small, the final term can be ignored, so that $Pr(A=1|I=1, i, j)\approx\theta_i+\theta_j$. The probability of a fatal crash between drivers of type i and j is:

$$\Pr(A, i, j|I=1) = \Pr(A|I=1, i, j)\Pr(i, j|I=1) = \frac{N_iN_j(\theta_i+\theta_j)}{(N_D+N_S)} \qquad [1]$$

Notice, this model does not assume that a two-car motor vehicle crash that involves one drinking driver is caused by the drinking driver. In this framework, it is possible that a drinking driver makes no errors (with probability $1-\theta_D$), but is struck and killed by a sober driver (with probability θ_S).

These probabilities are conditional on an interaction occurring, but driver interactions are not recorded in FARS. Applying Bayes' Rule, however, yields:

$$\mathrm{P_{ij}} = \Pr(i, j|A=1) = \frac{N_iN_j(\theta_i+\theta_j)}{2[\theta_D(N_D)^2+(\theta_D+\theta_S)N_DN_S+\theta_S(N_S)^2]} \qquad [2]$$

This provides the probability of observing two driver types conditional on a crash occurring, precisely the data structure of FARS. These probability expressions yield two linearly independent equations (the equations for P_{SD} and P_{DS} capture observationally identical outcomes and $P_{SS} + P_{DD} + 2P_{DS} = 1$) in four unknowns: N_D, N_S, θ_D and θ_S. Let $N=N_D/N_S$ denote the ratio of drinking drivers to sober drivers and $\theta=\theta_D/\theta_S$ denote the relative risk of drinking drivers. Then, the probabilities of observing the three different combinations of driver types in a fatal two-vehicle crash are:

$$\mathrm{P_{DD}} = \frac{\theta N^2}{\theta N^2+(\theta+1)N+1}, \qquad \mathrm{P_{DS}} = \frac{(\theta+1)N}{\theta N^2+(\theta+1)N+1}, \qquad \mathrm{P_{SS}} = \frac{1}{\theta N^2+(\theta+1)N+1} \qquad [3]$$

Assuming that the composition of drivers is independent across crashes, the joint distribution of two-vehicle crashes characterized by driver type is multinomial:

[6] *LP* demonstrate how the EIM assumption can be relaxed and find a minimal effect on their estimates.

$$\Pr(A^{DD}, A^{DS}, A^{SS}) = \frac{(A^{DD}+A^{DS}+A^{SS})!}{A^{DD}!A^{DS}!A^{SS}!}(P_{DD})^{A^{DD}}(P_{DS})^{A^{DS}}(P_{SS})^{A^{SS}} \quad [4]$$

where A^{ij} denotes the number of two-vehicle crashes involving one type *i* and one type *j* driver. Substituting the probabilities from [3] into [4] yields the likelihood function to be maximized over θ and *N*.[7]

The *EIM* assumption requires that the distribution of driver types is constant. Clearly, the prevalence of drinking and driving can differ across both geography and time (hour-to-hour, day-to-day, and year-to-year). Thus, *N* must be defined flexibly within the likelihood function to accommodate the level at which equal mixing is assumed. *LP* operationalize this by defining *N* as a fully-interacted function of dummy variables that characterize the spatiotemporal level of aggregation over with *EIM* is assumed, e.g., hour × year.

To summarize, under the *EIM* assumption the distribution of driver-type involved in fatal two-vehicle crashes is sufficient to identify both the prevalence of drivers with BAC≥.01 and their risk of causing a fatal crash relative to a driver with BAC=0. A straightforward application of Bayes' Rule allows us to identify the proportion of drivers with BAC≥.01 using the observed BAC levels of drivers involved in fatal two-vehicle crashes.

Single-vehicle crashes The distribution of driver interactions provides the source of identification in the *LP* approach. With single-vehicle crashes, however, no interaction is observed, and thus they do not provide additional identification. To see this, define λ_D and λ_S as the probabilities that a drinking driver or sober driver commits an error that causes a fatal one-vehicle crash (comparable to the definitions of θ_D and θ_S above). If C is an indicator for a one-car crash occurring, then the probabilities of each driver type's involvement in a crash (defined as Q_i) are:

$$Q_D = \Pr(i = D|C = 1) = \frac{\lambda_D N_D}{\lambda_D N_D + \lambda_S N_S}, \; Q_S = \Pr(i = S|C = 1) = \frac{\lambda_S N_S}{\lambda_D N_D + \lambda_S N_S}$$

If the relative risk for drinking and sober drivers is $\lambda = \lambda_D/\lambda_S$, then $Q_D/Q_S = \lambda N$ and the distribution of driver types involved in one-vehicle accidents cannot be used to separately identify the proportion of drinking drivers on the road and their risk of causing a fatal accident relative to sober drivers. Intuitively, a high proportion of single-vehicle accidents involving

---

[7] With *T* driver types, the probability that drivers of types *i* and *j* will interact is $Pr(i,j|I=1)=N_iN_j/N^2$, where *N* is the total number of drivers. Defining type 1 as the reference, the probability of observing types *i* and *j* in a fatal crash is:

$$P(i,j|A = 1) = \frac{N_{i,1}N_{j,1}(\theta_{i,1}+\theta_{j,1})}{\sum_{k=1}^{T}\sum_{l=1}^{T}\left(N_{k,1}N_{l,1}(\theta_{k,1}+\theta_{l,1})\right)} \quad \text{[Error! Main Document Only.]}$$

where $N_{i,1} = N_i/N_1$ and $\theta_{i,1} = \theta_i/\theta_1$. Based on these probabilities, the maximum likelihood function for *T* types is a straight-forward generalization of equation [4]:

$$P(\{A_{ij}\}|A_{Total}) = \frac{(A_{Total})!}{\prod_{i=1}^{T}\prod_{j=i}^{T}(A_{ij}!)}\prod_{i=1}^{T}\prod_{j=i}^{T}\left(P_{ij}{}^{A_{ij}}\right) \quad \text{[Error! Main Document Only.]}$$

drinking drivers could arise because there are many drinking drivers on the road or because their relative risk of causing a fatal accident is large.

Although single-vehicle accidents do not provide additional identification, they are included in the likelihood function to improve the precision of prevalence estimates, increase the rate of model convergence, and separately identify the relative risks of causing a one-vehicle fatal crash. Their inclusion also introduces a way to simplify the maximum likelihood routine.

Legally impaired drivers The method proposed by *LP* relies on defining mutually exclusive driver types. Thus, to estimate the relative risk and prevalence of drivers with a BAC above a specified threshold, three driver types must be defined: {BAC=0, 0>BAC>*L*, BAC≥*L*}, where L is the per se legal BAC limit for alcohol impairment. This leads to six possible driver combinations for an observed two-vehicle fatal crash.

One could estimate the likelihood function derived in footnote 7 over three driver types. As the relative risk and prevalence of drivers who are drinking, but not legally impaired are not a specific concern in their article, however, they operationalize estimation by discarding any crashes involving such drivers. This reduces the problem to the simple case of two driver types described above. We follow their approach in our subsequent analysis of legally impaired drivers.

V. Bringing LP to data

Three issues deserve specific discussion with respect to operationalizing the *LP* methodology using the data and computing technology currently available to researchers: different possible alcohol-impairment measures within FARS; estimating the prevalence alcohol-involved and alcohol-impaired driving at the national level; and the introduction of multiple imputation to accommodate drivers with missing BAC test results. While the lattermost is noteworthy because the multiply imputed BAC values currently supplied by in FARS were not available to *LP*, the first stands out as a potentially important source of error in the original analysis.

Alcohol-impairment measures

The approach developed by *LP* provides a valuable alternative to survey-based methods, as it relies not on a survey of drivers, but a census of fatal of motor vehicle crashes. If driver impairment level in FARS were measured without error, i.e., if all drivers in a fatal motor vehicle accident were tested for BAC, identification of their method is based upon assumptions regarding driving behavior rather than reporting behavior. Of course, BAC testing is not universal, leading to potentially a different variety of sample selection bias based on the decisions of law enforcement officers, rather than drivers.

LP address this concern in two ways. First, they restrict their analysis to state-year pairs in which at least 95% of drivers in fatal motor vehicle crashes are tested for BAC. Although this reduces concern about non-random testing within states, there is the real possibility that states with high testing rates different systematically in either the prevalence of drinking and driving or the risk associated with drinking and driving. This generates some concerns about the potential generalizability of results based on this sample.

Second, *LP* define driver alcohol-impairment status using the police officer's evaluation of whether or not a driver had been drinking, arguing that selection is unlikely to apply because it is available for virtually every driver involved in a fatal crash. Besides the obvious concern that law enforcement officers systematically misjudge whether a driver had consumed alcohol, our replication analysis reveals that *LP* misinterpreted the alcohol-impairment variables published in FARS. Far from being universal, almost one-quarter of drivers in their sample lack a subjective measure of alcohol involvement.

Determining the effect of using different alcohol impairment measures contained in FARS on estimation results is therefore a key contribution of the replication analysis that follows.

Estimation of nationwide prevalence

As noted earlier, *N* is defined in the likelihood function as a fully interacted set of dummy variables over the spatiotemporal level of aggregation over which EIM is assumed. Hence, the maximum likelihood routine does not generate a nation-level estimate of prevalence. *LP* construct a nation-wide estimate by running the maximum likelihood routine twice. The estimates of θ and λ from the first estimation (when *N* is fully interacted) are inserted directly into the likelihood function, which is then maximized treating *N* as a constant, i.e., assuming equal mixing nationally.

As an alternative, we propose to solve for a nation-wide estimate of *N* using the relationship derived earlier for single-vehicle fatal crashes: $N = Q_D^{Total}/\hat{\lambda}Q_S^{Total}$, where Q_D^{Total}/Q_S^{Total} is the ratio of drinking to non-drinking drivers involved in one-vehicle fatal crashes nationally and $\hat{\lambda}$ is the estimated relative risk parameter. The standard error is recovered by applying the delta method.[8]

The calculation of *N* above suggests another approach to specifying the maximum likelihood function that greatly reduces the computational burden of estimation. Rather than direct estimation of a fully-interacted function of dummy variables, we substitute for N_i in the likelihood function. To do so, we aggregate crash counts at the level *EIM* is assumed and then substitute $Q_D^i/\lambda Q_D^i$ for N_i, where i denotes the unit of observation defined by the *EIM* assumption. By greatly reducing the number of parameters, substitution decreases the amount of time required for the likelihood function to converge to a maximum, as well as the time required for the delta method to calculate the standard error on *N*.[9]

Multiple imputation

Since *LP* conducted their analysis of FARS data, reporting of driver alcohol impairment status has undergone two important changes. First, the proportion of drivers killed in a fatal motor vehicle crash who are tested to determine BAC has increased from 54% in 1982 to 68% in 1997

---

[8] When EIM is assumed to hold at the most disaggregated level (state-year-hour-weekend), the number of estimated coefficients in the function that defines *N* is exceptionally large and thus the delta method is computationally intensive. The delta method required roughly one-week to converge to a solution for *N*.

[9] With *EIM* assumed at the state-year-hour-weekend, substitution reduces the estimation step from twenty-two hours to less than one-minute and the delta method calculation from seven days to less than one second.

to 76% in 2008. For surviving drivers, the proportion has increased from 16% in 1982 to 26% in 1997 to 29% in 2008 (Hedlund, Ulmer, Northrup, 2004; Cassanova, Hedlund and Tilson, 2012).

Second, researchers at the Department of Transportation have developed a multiple imputation strategy to accommodate missing BAC values in FARS that can be applied to all available incident reports from 1982 onward (Subramanian, 2002). Thus, we can compare results using the full sample of fatal motor vehicle crashes based on the multiply imputed alcohol impairment variables with the results when we restrict the sample to only those state-year pairs with at least 95% testing rates. Similarity in parameter estimates will support use of the former to improve precision.

VI. Replicating Levitt and Porter

As a first step in our analysis, we attempt to replicate the results reported by *LP*, who provide numerous waypoints that direct the replication exercise. We dedicate substantial space to this exercise for the following reasons. First, although *LP* are commonly cited in the literature, their methods are rarely adopted. Difficulty in replicating their results may be one cause and we hope to clarify for others how and why it is difficult to match the portfolio of results *LP* report. Second, the methods developed by *LP* are very powerful and we hope this replication exercise establishes a baseline approach to its implementation in addressing other research questions.

It is worth noting here that *LP* offer a number of extensions to their approach, including allowing risk to vary by (discrete) observable driver characteristic, as well as numerous specification checks, e.g., allowing departures from the *EIM* assumption and differential crash avoidance probabilities. These are all valuable contributions, but are beyond the scope of the current article, which focuses on replicating and updating the main results of *LP*. In light of the robustness of the results reported by *LP*, we believe this is a reasonable approach given the complexity of the exercise and space available in a single article.

Table 1 summarizes the sample selection procedures and statistics we will seek to match enumerated in the order of presentation within the original manuscript. Two issues arose during this exercise and we treat each in turn.

Issue 1: Ambiguity in sample selection criterion

For the purposes of replication, *Item 6* is clearly the most problematic, as *LP* provide neither additional information as to which state-year pairs were discarded nor a discussion of the specific data problems that would have led to omission. Without this information, we can only hope to approximate the sample utilized by *LP* by imposing our own, reasonable standards vis-a-vis data quality.

This problem is compounded by the ambiguity in defining the sample being used to construct certain statistics [*Items 3*, *5*, and *7f*]. For example, after noting that they restrict the sample to fatal crashes that occur from 1983 to 1993 [*Item 1*] between 8:00 P.M. and 5:00 A.M [*Item 2*], *LP* report that their sample includes over 100 000 one-car crashes and over 40 000 two-car crashes [*Item 3*]. Within FARS, of the 180 632 fatal motor vehicle crashes that occur from 1983 to 1993 between 8:00 P.M. and 5:00 A.M, 127 239 involve one driver and 47 987 involve two drivers. Although these figures do not contradict those offered by *LP*, the magnitudes seem rather different. We suspect these figures correspond to the estimation sample after additional restrictions [*Items 6* and *7*] have been imposed: in Table 1, *LP* report 103 077 fatal one-car crashes and 39 470 fatal two-car crashes.

While *Item 3* seems to anticipate yet-to-be-presented refinements, *Item 5* appears to ignore refinements described previously. Of the fatal motor vehicle crashes that occur from 1983 to 1993 between 8:00 P.M. and 5:00 A.M, only 2.99% involve three or more drivers (7.27% of all drivers involved in a fatal crash during those years and hours). Although this is less than six percent, and thus not contradictory, it is difficult to believe that *LP* chose random inequality constraints. In comparison, out of all the fatal motor vehicle crashes that occur from 1983 to 1993, 5.35% involve three or more drivers (12.02% of all drivers involved in a fatal crash during those years). Thus, the statistic cited by *LP* is true in either case, but based on magnitudes, was likely calculated relative to all fatal crashes from 1983 to 1993 [*Item 1*], not only those fatal crashes 8:00 P.M. and 5:00 A.M. [*Items 1* and *2*].

It is also worth mentioning that *LP* seem to use the terms *car* and *motor-vehicle* interchangeably. There is no mention in their manuscript of restricting the analysis according to the type of vehicle driven and we continue with that practice here.

Issue 2: Incorrectly defining a drinking driver

LP exclude observations with missing demographic and crash information [*Item 7*] noting subsequently that the police officer's evaluation of whether or not a driver had been drinking is that it is available for virtually every driver involved in a fatal crash [*Item 8*]. The latter statistic is incorrect and suggests that *LP* misinterpreted the alcohol-impairment status variables contained in FARS.

The FARS PERSON Data File includes the variable DRINKING with the label "Police Reported Alcohol Involvement." The FARS codebook provides the following definition for this variable: "This data element records whether alcohol was involved for this person and reflects the judgment of law enforcement." From the 1975 version onwards, the variable is coded 0 if the law enforcement officer judged that alcohol was not involved; 1 if the law enforcement officer judged that alcohol was involved; 8 if the crash report does not include the judgement of the law enforcement officer; and 9 if the law enforcement officer judges that it is unknown if alcohol was involved.

Of the 223,213 drivers in fatal crashes involving either one or two vehicles from 1983 to 1993 between 8:00 P.M. and 5:00 A.M., 67,770 (30.4%) were either categorized by a police officer as unknown alcohol involvement (DRINKING=9) or this information was not provided (DRINKING=8). Although values of 8 and 9 are not "missing" in the sense that a valid entry exists, in a practical sense, we would tend to think of such values as failing to provide a clear assessment of alcohol involvement. As a result, of the 175,226 fatal motor vehicle crashes involving either one or two vehicles from 1983 to 1993 between 8:00 P.M. and 5:00 A.M., 34.8% included at least one driver that lacked either an affirmative or negative assessment of alcohol involvement from a law enforcement officer. This suggests that a different FARS variable (or combination of variables) is being utilized by *LP*.

In addition to the judgement of the police officer, FARS includes information about whether a blood alcohol content (BAC) test was conducted, the type of test procedure used, and the test result. This variable is missing for 101,610 (45.5%) of drivers in fatal crashes involving either one or two vehicles from 1983 to 1993 between 8:00 P.M. and 5:00 A.M. Table 2 summarizes the distribution of drivers in fatal crashes involving either one or two vehicles from 1983 to 1993 between 8:00 P.M. and 5:00 A.M. according to the reported judgement of the responding police officer and BAC test result.

It is worth recognizing that the decision on the part of the responding police-officer to provide their judgement of alcohol involvement may depend upon whether they anticipate a BAC test occurring. Officers who know that a test will be conducted may be more likely to withhold their subjective judgement, deferring instead to the objective test result. If true, the potential selection issues connected with the BAC testing of drivers would then spill over to the judgements provided by police officers. Therefore, a measure of alcohol-involvement that utilized both pieces of information—officer judgement and BAC test results—would be preferred.

Such a variable exists in FARS, and we suspect this is the measure actually employed by *LP*. Specifically, the FARS VEHICLE Data File includes the variable DR_DRINK with the following description: “This data element records whether the driver was drinking and is derived from data elements in the Vehicle and Person data files. Data from the Vehicle and Person data files are analyzed and if there is "sufficient information" to conclude that a driver was drinking, i.e., positive BAC data or police-reported alcohol involvement, then a driver is classified as drinking.”

Unlike police reported alcohol involvement, the constructed variable DR_DRINK is available for nearly all drivers, because lack of sufficient information to assign alcohol involvement includes both cases where alcohol involvement can be rejected (BAC=0.0 and/or judgement by the responding officer that alcohol was not involved), as well as for drivers where alcohol status is unknown (neither a BAC test result nor a law enforcement judgement). This definition seems unsatisfactory in that drivers lacking any information about alcohol involvement status are assigned to the non-drinking category by default and such drivers account for 15 percent of drivers involved fatal crashes involving either one or two vehicles from 1983 to 1993 between 8:00 P.M. and 5:00 A.M. Nevertheless, our subsequent results will strongly suggest that this is the alcohol-involvement variable employed by *LP*.

There are reasonable alternatives, however, that use both police-officer judgement and BAC test results to define alcohol involvement status while allowing drivers lacking this information to remain categorized as missing. Table 3 presents four possible definitions of a drinking driver based on police officer judgement and BAC test results. The first definition is the one described by *LP*, the second definition is the one we believe *LP* actually implement, the third definition uses police reported alcohol involvement as the primary measure of alcohol involvement (BAC test result is used only when reported alcohol involvement is unknown or not reported); and the fourth definition uses BAC test as the primary measure of alcohol involvement (police reported alcohol involvement is used only when a BAC test result is missing). Also reported are the unconditional and conditional drinking-driver rates (conditional on the defined alcohol measure taking a non-missing value). For the remainder of the replication exercise, we will report results using each of the four definitions described.

It is important to recognize that the choice of definition for alcohol-involved driving affects the state-year selection criteria when using BAC level to determine whether drivers were legally impaired by alcohol. *LP* seek to omit state-year observations for which less than 95% of drivers that are judged to be drinking by the responding police officer receive a BAC test (*Item 9*). But, if they mistakenly employ *Definition 2* in lieu of *Definition 1*, they will under-omit. From Table 2, there are 83 561 drivers who are judged to be drinking by a police officer, but 109 507 drivers classified as drinking according to *Definition 2*. With 19 504 drivers who are not tested for BAC among those that are judged to be drinking by a police officer, the choice of latter definition implies a non-testing rate of 17.8 percent, rather than the actual non-testing rate of 23.3 percent.

Given the possible interaction between BAC testing procedures and the decision on the part of the responding officer to report their subjective opinion of the alcohol involvement status of drivers, it is not altogether clear that the 95 percent threshold proposed by *LP* is meaningful. Nevertheless, we will repeat their analysis using each of the four definitions of alcohol-involved driving described above.

Replicating summary statistics

Table 4 compares the summary statistics for the sample of fatal crashes involving either one or two vehicles from 1983 to 1993 between 8:00 P.M. and 5:00 A.M. based upon the four alcohol involvement measured defined above, along with those reported by *LP* for their sample. Crashes that include drivers with missing information about age, gender, driving history and drinking involvement status (using the respective definitions) are omitted. In columns 2-5 No attempt is made to identify the state-year observations that are discarded by *LP*.

Based on these results, it seems most likely that *Definition 2* based on the FARS variable DR_DRINK was used to conduct the baseline analysis of the original article, rather than *Definition 1* based upon the FARS variable DRINKING. Of the four possible definitions, only *Definition 2* yields driver counts greater than those reported by *LP* (recall, in their final sample some drivers will have been dropped because of omitting an undetermined number state-year observations). Each of the other definitions also produce drinking driving rates several percentage points above those reported by *LP*, while the rate using *Definition 1* is only 0.6 percentage points below that originally reported. Similarly. The distribution of crashes across driver type are much more closely matched using *Definition 2* compared to definition actually described by *LP* (*Definition 1*) or the other alternatives.

The last column present summary statistics when we attempt to objectively determine the data quality omission criteria imposed, but not defined, by *LP* (*Item 6*). Specifically, we omit state-year observations in which driver or crash characteristics are missing for more than t percent of fatal one- or two-vehicle crashes from 1983 to 1993 between 8:00 P.M. and 5:00 A.M. To determine a value for t, we employ a grid search in one percentage point intervals to find the threshold above which omitting state-year observations with missing driver and crash characteristics would most closely match the crash counts reported by *LP*. This occurs when state-year observations with missing information for more than 13 percent of fatal one- or two-vehicle crashes. The summary statistics for this sample are reported in the final column and match those reported by *LP* relatively well.

Replicating estimation results

By assumed level of equal and independent mixing Table 5 reports estimates of θ (relative risk of a drinking driver causing a fatal two-vehicle crash) and λ (relative risk of a drinking driver causing a fatal one-vehicle crash) allowing the level at which *EIM* is assumed to vary. The unit of observation is a state-year-hour-weekend and N is treated as a function of a fully-interacted set of dummy variables that define the spatiotemporal unit of *EIM*.

Based on the comparison of summary statistics described above, we also restrict the sample to only those state-years in which less than 13 percent of crashes exhibit missing driver or crash characteristics. For each set of regression results, we report the number degrees of freedom used in estimation, which equals two plus the number of units over which equal and independent mixing is assumed.

Three patterns clearly stand out. First, defining alcohol involvement based solely on the reported judgement of police officers (*Definition 1*) leads to estimates of θ and λ significantly lower than those reported by *LP* and utilize far too few degrees of freedom.

Second, employing BAC test results as the primary measure of alcohol involvement supplemented with police reported alcohol involvement when a BAC test result is missing (*Definition 4*) yields estimates of θ and λ that are consistently greater than those reported in *LP*.

Third, results using either the FARS variable DR_DRINK as the measure of alcohol involvement (*Definition 2*) or police reported alcohol involvement as the primary measure of alcohol involvement supplemented with BAC test result when reported alcohol involvement is unknown or not reported (*Definition 3*) yields estimates of θ and λ that are relatively close to the values reported by *LP*, though they do not match exactly.

By specification of likelihood function Table 6 reports estimates of θ, λ, and the fraction of drinking drivers[10] under different specifications of the likelihood function and alcohol-involvement definitions when *EIM* is assumed at the state-hour-year-weekend level. The choice of estimating N as a function of fully-interacted dummy variables or substituting for N_i in the likelihood function with $Q_D^i/\lambda Q_D^i$ appears to have a relatively minor effect on parameter estimates. The former specification leads to risk estimates that are somewhat larger than the latter, but estimates of prevalence are only marginally different. Given the computational benefits, we adopt the substitution specification in the subsequent set of analyses.

By year and hour Figures 2 and 3 summarize estimation results when relative risk is allowed to vary by year and hour, respectively. The results when employing *Definitions 2* and *3* appear to most closely match those reported by *LP*. This impression is confirmed when comparing mean

[10] Recall that $N=N_D/N_S$, so that the fraction of drinking drivers, $N_D/(N_D+N_S)$, equals $N/(1+N)$.

squared prediction error (MSPE) under different possible alcohol impairment definitions (not reported).

By level of alcohol impairment The first two columns of Table 7 report estimates of θ and λ for drivers with BAC $\geq$ 0.1g/dL, the *per se* legal limit for alcohol-impaired driving during the period under consideration[11], using the sample of state-year observations in which at least 95 percent of drivers who are categorized as drinking receive a BAC test. Again, the estimates using *Definition 2* to define drinking status—and thus establish the base over which the 95 percent testing rate is calculated—are close to the figures reported by *LP*.

Summary

Based on the results of the replication analysis reported above, we conclude that while there are several notable issues with the analysis undertaken by *LP*, these are ultimately innocuous flaws.

It seems likely that the measure of alcohol-involvement described by *LP* is different from the measure they employ. As a result, their argument for employing the subjective judgement of police officer is factually incorrect: this information is missing for a significant number of drivers. The alcohol involvement definition they actually use assumes that drivers lacking a BAC test and a clear (affirmative or negative) subjective judgement are not drinking drivers (*Definition 2*). On its face, this seems like an untenable assumption.

Further, it is possible that the absence of a subjective police officer judgement may be related to BAC testing procedures, namely officers may be more likely to withhold judgement if they know a BAC test will be conducted. For these reasons, we would argue that using the subjective judgement of the responding police officer, supplemented by the BAC test result in cases where the judgement is either inconclusive or unreported (*Definition 3*), is the preferred approach for categorizing the alcohol involvement status of drivers.

For practical purposes, however, our results indicate that the definition operationalized by *LP* yields relative risk estimates that are nearly identical to those that arise when using the definition that seems to have the strongest *a priori* rationale. Hence, while *LP* do not do what they say they did, the approach they adopt nevertheless arrives at approximately the "most correct" result.

Of course, because we are unable to impose the same data quality restrictions as *LP*, we are estimating the maximum likelihood function over a different estimation sample. It is therefore not terribly surprising that we cannot exactly match their parameter estimates. Nonetheless, our very simple attempt to find an object data quality threshold yields estimates that are still relatively close those reported by *LP*. Thus, while we cannot replicate their results using the information presented in their article, the conclusions they reach are broadly confirmed when applying the most appropriate definition of alcohol involvement status.

[11] The *per se* limit in every US state is 0.08 g/dL as of May 2016.

VII. Multiple imputation for missing BAC information

Restricting the estimation sample for legally impaired drivers to only those state-year observations in which at least 95 percent of drivers who are categorized as drinking receive a BAC test greatly reduces the sample size. This not only reduces precision of estimates, but also raises concerns about the generalizability of findings to the omitted states.

Researchers at the Department of Transportation have developed a multiple imputation strategy to accommodate missing BAC values in FARS that can be applied to all available incident reports from 1982 onward (Subramanian, 2002). The regression equations used to construct imputed BAC in FARS include characteristics of the crash, including the time of day, the number of vehicles, and whether any deaths occurred (Subramanian, 2002). In our analysis, this generates a circularity that is generally avoided when using imputed values. Nevertheless, NHTSA has high confidence that these imputations are valid, particularly when researchers use them to categorize drivers by BAC thresholds.

For every driver in FARS, there are ten imputed BAC values, effectively generating ten datasets. These equal the measured BAC level when a test was conducted. Therefore, we generate estimated values of θ_i and λ_i for each of the i={1,…,10} datasets substituting for N_i in the likelihood function with $Q_D^i/\lambda Q_D^i$. As these estimates are asymptotically normal, we apply the following general formulae for a multiply imputed regression coefficient:

$$\hat{p} = \frac{1}{M}\sum_{i=1}^{M} p_i \qquad [7]$$

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\sigma_i^2 + \left(1+\frac{1}{M}\right)\left(\frac{1}{M-1}\right)\sum_{i=1}^{M}(p_i-\hat{p})^2} \qquad [8]$$

where M is the number of imputations, p_i is the coefficient estimate in dataset i, and σ_i is the standard error on the estimate of p_i (Rubin 1987).

For comparison with other possible drinking definitions based on a combination of subjective judgements from police officer reports and BAC test results, the bottom panel of Table 5 presents estimates of the relative risk of drinking drivers using the MI BAC values provided in FARS at various levels of the *EIM* assumption. For each of the ten MI replications, a driver is classified as drinking if their imputed BAC level is greater than zero. The estimated parameter values are generally similar to those recovered under *Definition 2*, though slightly smaller as the *EIM* is relaxed at the state-level, which would be consistent with police-officers failing to assign a positive drinking status to drivers with relatively low BAC levels. Nonetheless, the magnitude of this bias appears relatively small.

Figures 4 and 5 compare the year- and hour-specific estimates of the relative risk of drinking drivers from MI estimation with the results reported by *LP* and our estimates using *Definition 3*. These track the overall pattern of the estimates from *LP* and are very similar in magnitude.

Estimates of θ and λ for drivers with BAC>0.1 based on the FARS multiple imputations of BAC are reported in the final column of Table 7. The relative risk estimates using MI are smaller than those reported by LP, possibly reflecting that BAC testing rates tend to be higher in states with more dangerous impaired drivers.[12] Again, however, these magnitude differences are fairly small, particularly when compared to the estimates of relative risk generated from survey-based methods.

Summary

Based on the results of the analysis using the multiply imputed BAC records constructed by NHTSA, we conclude that researchers should adopt MI estimation. The difference in estimated parameters are relatively minor compared to samples that only use actual test results from state-year observations with high testing rates. This could simply reflect a small degree of sample variation. Alternatively, the MI approach may induce a great deal of specification bias that is coincidentally offset by a large degree of sample selection bias that arises when relying on only states with high testing rates. We cannot disentangle these explanations, but our suspicion is that the former is more likely than the latter.

VIII. Updated estimates of the relative risk and prevalence of drinking and driving

In this section, we will document how the relative risk and prevalence of drinking and driving has evolved over the past three decades. To do so, we estimate values of θ and λ over five-year intervals assuming equal and independent mixing at the state-year-hour-weekend level. We estimate separate values for all drinking drivers and those with BAC>0.08, the current *per se* legal limit for alcohol-impaired driving.

For the former, we use two definitions to define drinking and non-drinking drivers. First, we use the subjective judgement of the responding police officer, supplemented by the BAC test result in cases where the judgement is either inconclusive or unreported (*Definition 3*). Second, we follow a multiple imputation approach in which drivers with imputed BAC values greater than 0 are classified as drinking.

For the latter, we also follow two procedures. First, as in *LP*, we restrict the analysis to the sample of state-year observations in which at least 95 percent of drivers who are categorized as drinking under *Definition 3* receive a BAC test. Second, we implement the multiple imputation assigning drivers with imputed BAC values greater than or equal to 0.8 as legally impaired.

Figure 6 presents the time series of the estimates of θ and λ that do not utilize multiple imputation. The relative risk of both drinking and legally impaired drivers appears to have increased over time. In contrast, the prevalence of drinking and driving has generally fallen, though these decreases are concentrated in the first half of the study period (1982-1997). In the second half (1998-2012), however, prevalence has remained fairly constant. Figure 7 presents analogous estimates when employing multiple imputation and the general patterns is similar, though notably smoother.

---

[12] Conditional on BAC, drivers in the state-years omitted by Levitt and Porter may be just as dangerous as drivers in their estimation sample. Rather, the distribution of BAC in the latter may be more rightly skewed than the former.

IX. The externality to drinking and driving

In order to quantify the drinking and driving externality, we calculate, relative to the counterfactual scenario in which the risk of drinking driving was reduced to that of sober driving, a) the number of avoidable fatalities and b) a value of statistical life (VSL) estimate of the per mile external cost. For comparison with *LP* we first estimate counterfactual incremental fatalities and the externality per mile using the 1994 FARS data, *LP*'s estimates of θ and λ in 1993, and a VSL of $3 million.

Our summary statistics of the 1994 sample restrictions track *LP*'s very closely. In that year, 40,716 people died in motor vehicle crashes in the United States. We calculate the number of fatalities due to drinking driving by adding up fatalities in which the deceased was not in the drinking driver's vehicle (assuming that all passengers internalize their driver's behavior yields lower bound estimates). Avoidable fatalities are then calculated by multiplying fatalities by the proportion of crashes caused by drinking drivers minus the proportion caused by sober drivers according to a crash's composition of drivers, as implied by the model, e.g. multiplying drinking-sober (two-vehicle) crash fatalities by $(\theta - 1)/(\theta + 1)$. There were 292 avoidable fatalities in drinking-drinking (two-vehicle) crashes (the total was multiplied by 0.5, in this case); 1,053 avoidable fatalities in vehicles driven by sober drivers in drinking-sober (two-vehicle) crashes; 213 avoidable fatalities in vehicles driven by sober drivers who died in three-plus-vehicle crashes with at least one drinking driver; and 583 avoidable pedestrian fatalities due to drinking drivers in single-vehicle crashes. There were therefore 2,140 total external deaths attributable to drinking drivers (the sub-categories do not add to this total due to rounding). The only category where our estimate meaningfully differs from *LP* is avoidable fatalities in vehicles driven by sober drivers in drinking-sober (two-vehicle) crashes, but the result of this discrepancy is to only modestly lower the externality per mile estimates, as is evident below.

Using our estimate for drinking driving prevalence in 1994, and restricting attention only to night driving miles according to *LP*'s procedure, we estimate that there were approximately 49 billion drinking driving miles in 1994. Combined with the total VSL value of avoidable fatalities ($6.4 billion), we estimate that the external cost of drinking drivers was approximately 13 cents per mile in 1994. This is quite close to *LP*'s estimate of 15 cents per mile.

We next update the externality per mile estimate and show its change over time in Figure 8. For each year reported, we maintain the same assumptions and sample restrictions described above except we use our estimates for θ and λ in 5-year trailing windows, a tripled VSL of $9.1 million to conform with recent Department of Transportation policy (DOT, 2013), and an updated annual estimate of vehicle miles traveled for each reported year (DOT, 2012). In addition to displaying drinking driving externality estimates, Figure 8 also reports the drunk driving externality (BAC threshold 0.08) for every five-year window. There is a clear downward trend in both external cost estimates, with the drinking driving externality decreasing from 55 to 35 cents per mile between 1987 and 2012, and the drunk driving externality decreasing from 80 to 50 cents per mile, accordingly.

X. Conclusion

This study achieved three objectives: replicate the results reported by Levitt and Porter (2001), demonstrate how data more recently available to researchers can be incorporated into their methodology, and provide updated estimates of the relative risk and prevalence of drinking and driving in the United States.

Although we were unable to exactly replicate the results of Levitt and Porter, we consider the replication exercise an unqualified success. We believe our analysis has identified the key issues that prevent a successful replication. To the extent that difficulties in matching the statistics reported by Levitt and Porter have discouraged researchers from adopting their approach to answer other research questions, a potentially powerful analytical tool has been underutilized. By demonstrating that estimation results are largely robust to the specification choices that are either ambiguously defined or incorrectly described in the original article, we have offered a number of reasonable approaches that can be applied by other researchers who wish to implement their methodology. Indeed, we hope that this contribution leads to much wider use of the methods first derived by Levitt and Porter.

We also believe that integrating the maximum likelihood estimation approach developed by Levitt and Porter into the multiple imputation framework is an important contribution given the current structure of the FARS data. Using the multiply imputed values for BAC when test results are unavailable, as is now available in FARS, may not be appropriate for all research questions. Nevertheless, researchers cannot ignore the obvious misreporting and sample selection bias that arise in existing survey instruments—consistent with Levitt and Porter, our estimates of prevalence are substantially larger than those constructed from the National Roadside Survey. At least for the purposes of recovering prevalence and relative-risk, our results clearly demonstrate that use of the imputed data yields estimation results that are very similar to those from a subset of states with exceptionally high BAC testing rates, but with the benefit of much greater precision.

Our updated estimates of prevalence and relative-risk are consistent with patterns observed in other data, though, like Levitt and Porter, our estimate of prevalence is nearly three times higher than that reported in the most recent NRS. Our results indicate that the prevalence of drinking and driving has been relatively flat since at least the early 2000's, while the relative risk of drinking and driving has increased significantly. Hence, the recent plateau in the alcohol-related driving fatality rate after nearly 40 years of steady declines can be explained by two distinct mechanisms: a change in the trajectory of prevalence and a change in the trajectory of risk.

This pattern suggests that the current group of individuals who operate a motor vehicle while impaired by alcohol are more dangerous than the historical average (the most intoxicated when they drive and/or travel the most miles while intoxicated) and may be the least likely to respond to existing policies. Hence, while these policies have been effective at greatly reducing the incidence of drinking and driving in the United States, further reductions among the remaining group of drinking drivers may require innovative policy interventions.

Finally, our externality calculations suggest that the external cost of drinking and driving has steadily decreased since the 1980s. Although we cannot identify the absolute risk of causing a fatal crash based on the methods applied here, this result indicates that, on average, drinking drivers are less dangerous today than they once were. This is not unexpected given the steady advances in automobile and roadway safety technology over the past three decades. Given the increase in relative risk of drinking driving, however, these advances appear to be "sober-biased."

These finding surely present as many questions as they answer. Identifying the relative importance of different technological improvements and policy interventions in reducing the prevalence and social cost of drinking and driving will be a central research question moving forward. The tools developed by Levitt and Porter (2001) and extended here can play a valuable role in this process.

References

1. Berning A, Compton R, & Wochinger K (2015). Results of the 2013-2014 National Roadside Survey of Alcohol and Drug Use by Drivers. National Highway Traffic Safety Administration, DOT HS 812 118. Washington, DC.

2. Caetano R & McGrath C (2005). Driving under the influence (DUI) among US ethnic groups. *Accident Analysis and Prevention*, 37(2): 217-224.

3. Centers for Disease Control and Prevention, National Centers for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS) [online]. (2015). Available from: www.cdc.gov/injury/wisqars

4. Compton R & Berning A (2009). Traffic Safety Facts: Results of the 2007 National Roadside Survey of Alcohol and Drug Use by Drivers. National Highway Traffic Safety Administration, DOT HS 811 175. Washington, DC.

5. Cuthbert JR (1994). An extension of the induced exposure method of estimating driver risk. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 177-190.

6. (DOT) United States Department of Transportation, Federal Highway Administration, Office of Highway Policy Administration, (2012). Traffic Volume Trends, December 2012.

7. (DOT) United States Department of Transportation, Office of the Secretary of Transportation (2013). Guidance on Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analyses. February 28, 2013.

8. Greenfield TK & Rogers JD (1999). Alcoholic beverage choice, risk perception and self-reported drunk driving: effects of measurement on risk analysis. *Addiction*, 94(11): 1735-1743.

9. Hedlund JH, Ulmer RG, & Northrup VS (2004). State laws and practices for BAC testing and reporting drivers involved in fatal crashes National Highway Traffic Safety Administration, DOT HS 809 756. Washington, DC.

10. Jiang X & Lyles RW (2010). A review of the validity of the underlying assumptions of quasi-induced exposure. *Accident Analysis & Prevention*, 42(4): 1352-1358.

11. Johnson MB, Voas RB, Kelley-Baker T, & Furr-Holden CD (2008). The consequences of providing drinkers with blood alcohol concentration information on assessments of alcohol impairment and drunk-driving risk. *Journal of Studies on Alcohol and Drugs*, 69(4).

12. Kelley-Baker T, Lacey JH, Voas RB, Romano E, Yao J, & Berning A (2013). Drinking and driving in the United States: Comparing results from the 2007 and 1996 National Roadside Surveys, *Traffic Injury Prevention*, 14(2): 117-126.

13. Kirk A & Stamatiadis N (2001). Evaluation of the Quasi-induced Exposure. US DOT Transportation Center: Southeastern Transportation Center.

14. Lacey JH, Kelley-Baker T, Furr-Holden D, Voas RB, Moore C, Brainard K, Tippetts AS, Romano E, Torres P & Berning A (2009a). 2007 National Roadside Survey of Alcohol and Drug Use by Drivers: Methodology. National Highway Traffic Safety Administration, DOT HS 811 237. Washington, DC.

15. Lacey JH, Kelley-Baker T, Furr-Holden D, Voas RB, Moore C, Brainard K, Tippetts AS, Romano E, Torres P & Berning A (2009b). 2007 National Roadside Survey of Alcohol and Drug Use by Drivers: Alcohol Results. National Highway Traffic Safety Administration, DOT HS 811 248. Washington, DC.

16. Levitt SD & Porter J (2001). How dangerous are drinking drivers? *Journal of Political Economy*, 109(6):1198-1237.

17. Lund AK & Wolfe AC (1991). Changes in the incidence of alcohol-impaired driving in the United States, 1973-1986. *Journal of Studies on Alcohol*, 52(4): 293–301.

18. Mullahy J & Sindelar JL (2007). Do drivers know when to say when? an empirical analysis of drunk driving. *Economic Inquiry*, 32(3): 383-394.

19. NHTSA. 2014. Traffic Safety Facts: 2013 Data. Alcohol-Impaired Driving. National Highway Traffic Safety Administration, DOT HS 812 102. Washington, DC.

20. Rubin D. 1987. Multiple imputation for non-response in surveys. New York: John Wiley & Sons.

21. Santos A, McGuckin N, Nakamoto HY, Gray D, and Liss S (2011). Summary of Travel Trends: 2009 National Household Travel Survey. FHWA-PL-ll-022.

22. Stamatiadis N & Deacon JA (1997). Quasi-induced exposure: methodology and insight. *Accident Analysis & Prevention*, 29(1): 37-52.

23. Subramanian R (2002). Transitioning to Multiple Imputation: A New Method to Estimate Missing Blood Alcohol Concentration (BAC) values in FARS. National Highway Traffic Safety Administration, DOT-HS-809-403. Washington DC.

24. Voas RB, Wells J, Lestina D, Williams A, & Greene M (1998). Drinking and driving in the United States: The 1996 National Roadside Survey. *Accident Analysis and Prevention*, 30(2): 267–275.

25. Voas RB, Torres P, Romano E, & Lacey JH (2012). Alcohol-related risk of driver fatalities: an update using 2007 data. *Journal of Studies on Alcohol and Drugs*, 73(3): 341.

26. Vogel-Sprott M (1975). Self-evaluation of performance and the abiity to discriminate blood alcohol concentrations. *Journal of Studies on Alcohol*, 36(1): 1-10.

27. Zador PL, Krawchuk SA, & Voas RB (2000). Alcohol-related relative risk of driver fatalities and driver involvement in fatal crashes in relation to driver age and gender: an update using 1996 data. *Journal of Studies on Alcohol and Drugs*, 61(3):387.

Figure 1: Annual motor vehicle crash fatalities in the United States: 1994-2013

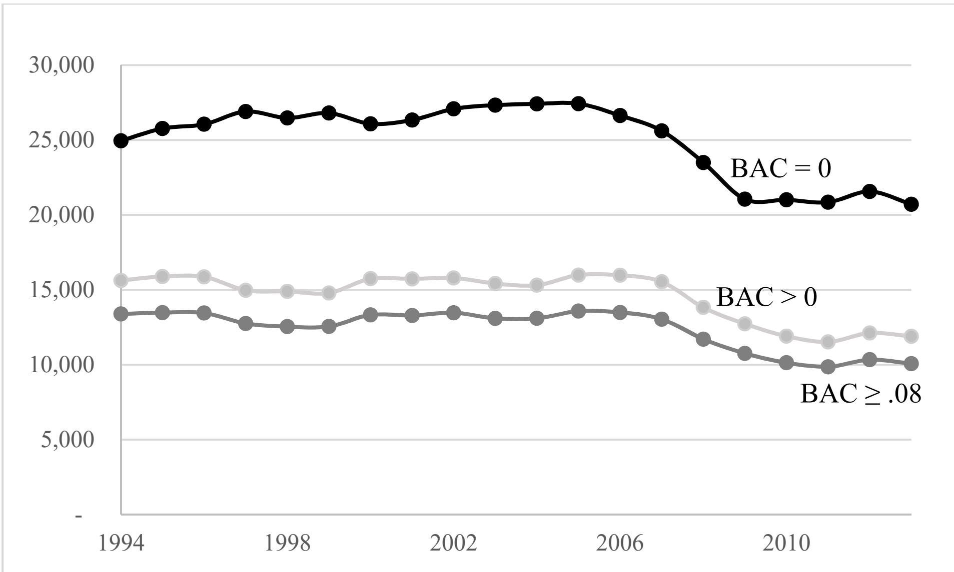
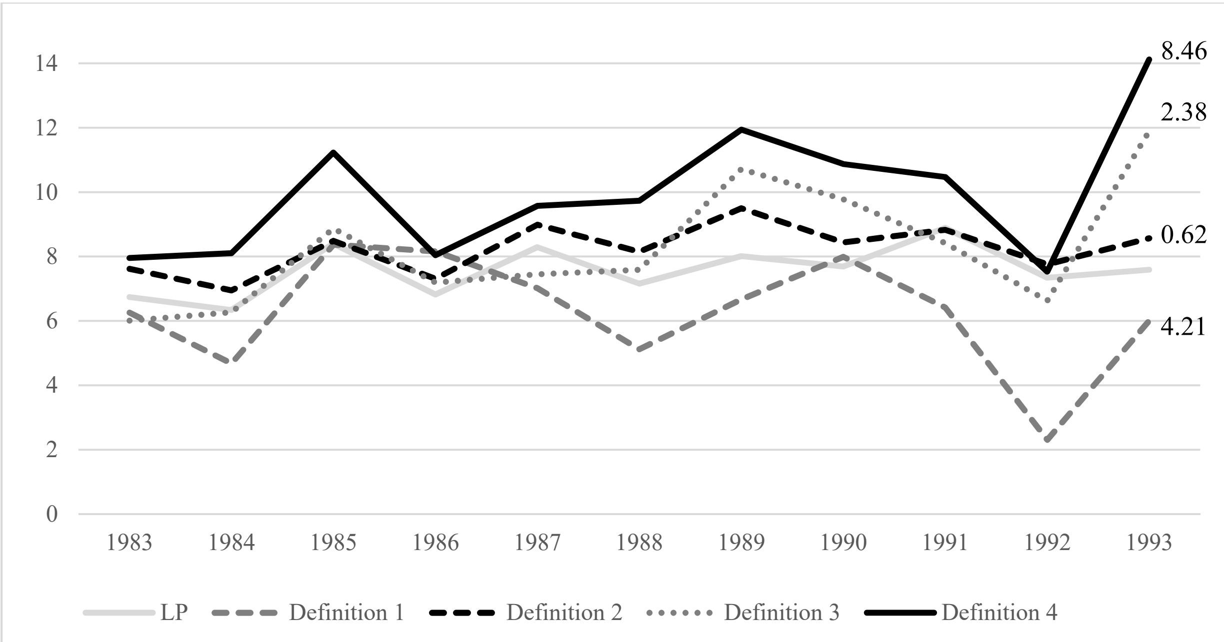
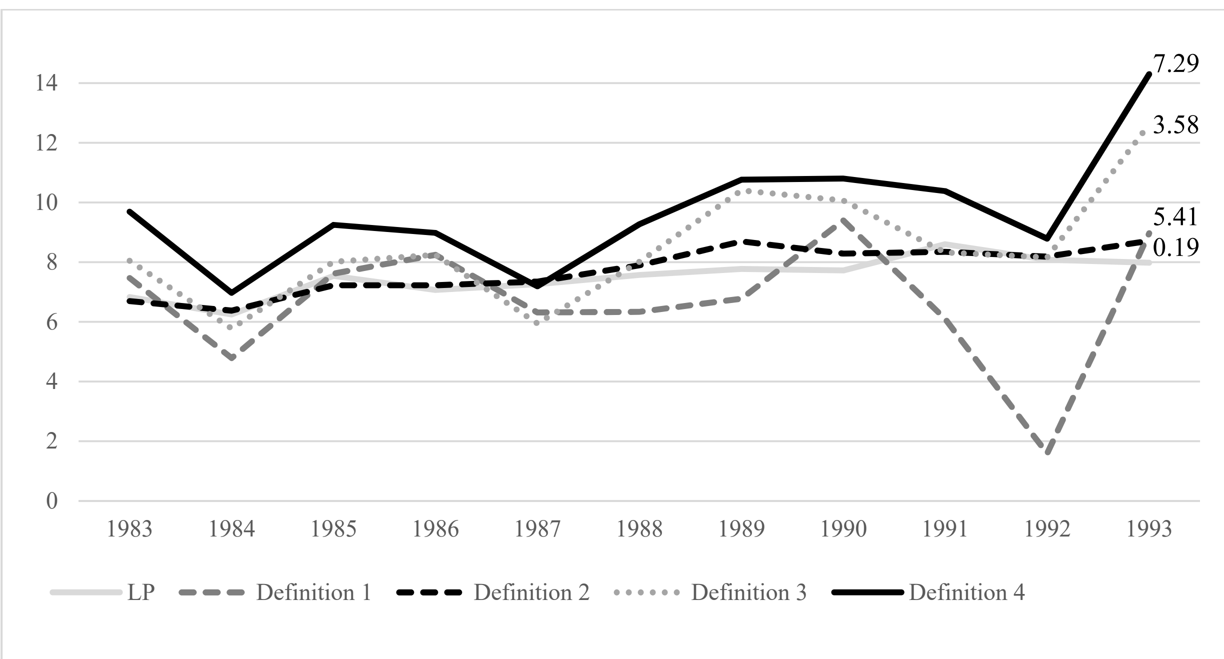


Figure 2: Year-specific estimates of θ and λ under different alcohol involvement definitions

Panel A: θ



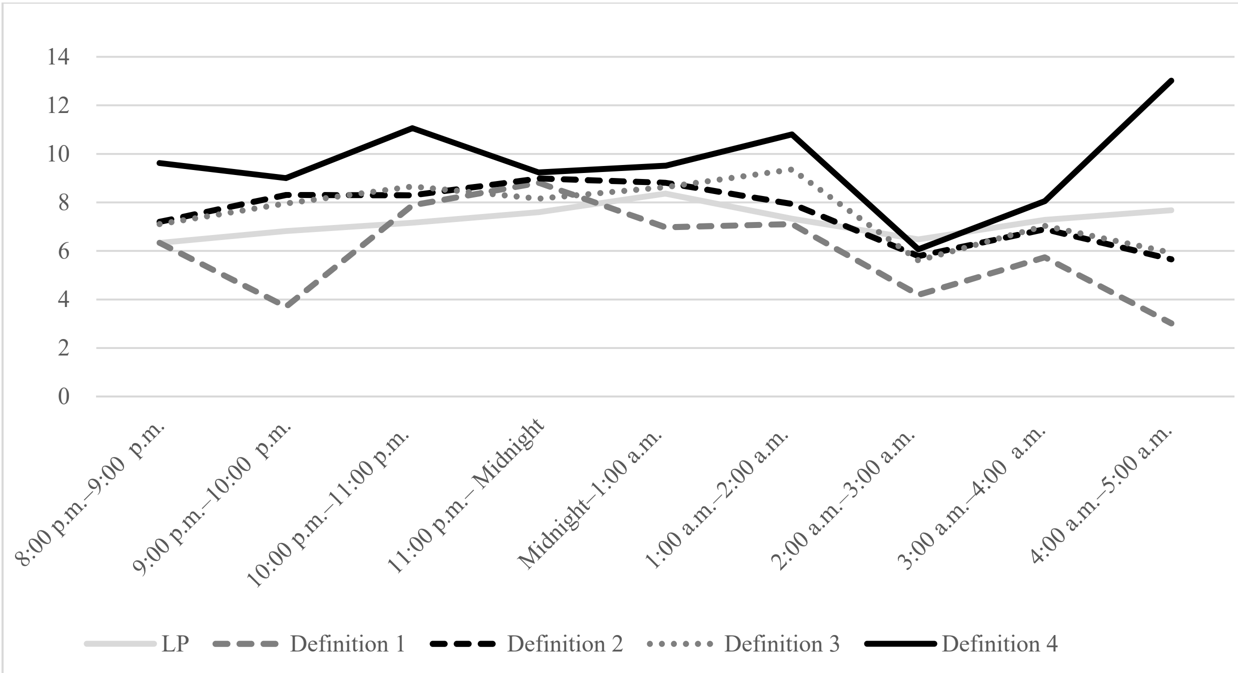
Panel B: λ



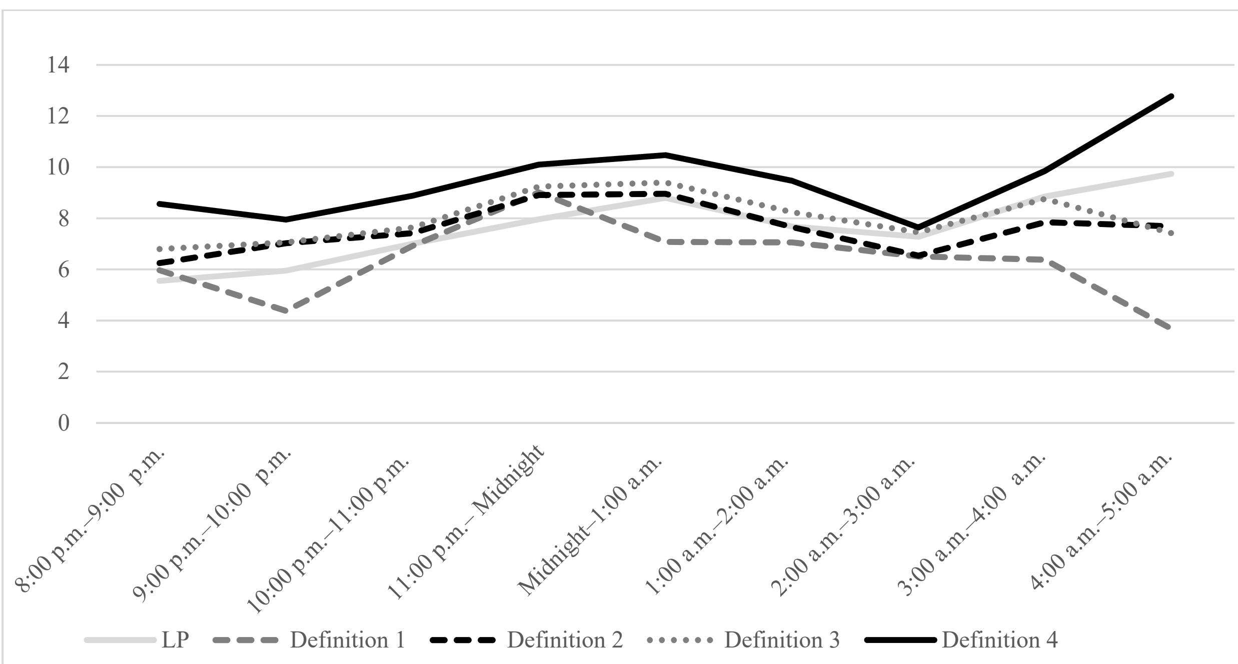
Note: Maximum likelihood estimates substituting for N with $Q_D^i/\lambda Q_D^i$ in the likelihood function assuming EIM at the state-hour-year-weekend level. Mean squared prediction error relative to estimates reported by Levitt and Porter (2001) mark the right-hand terminal point of each series.

Figure 3: Hour-specific estimates of θ and λ under different alcohol involvement definitions

Panel A: θ



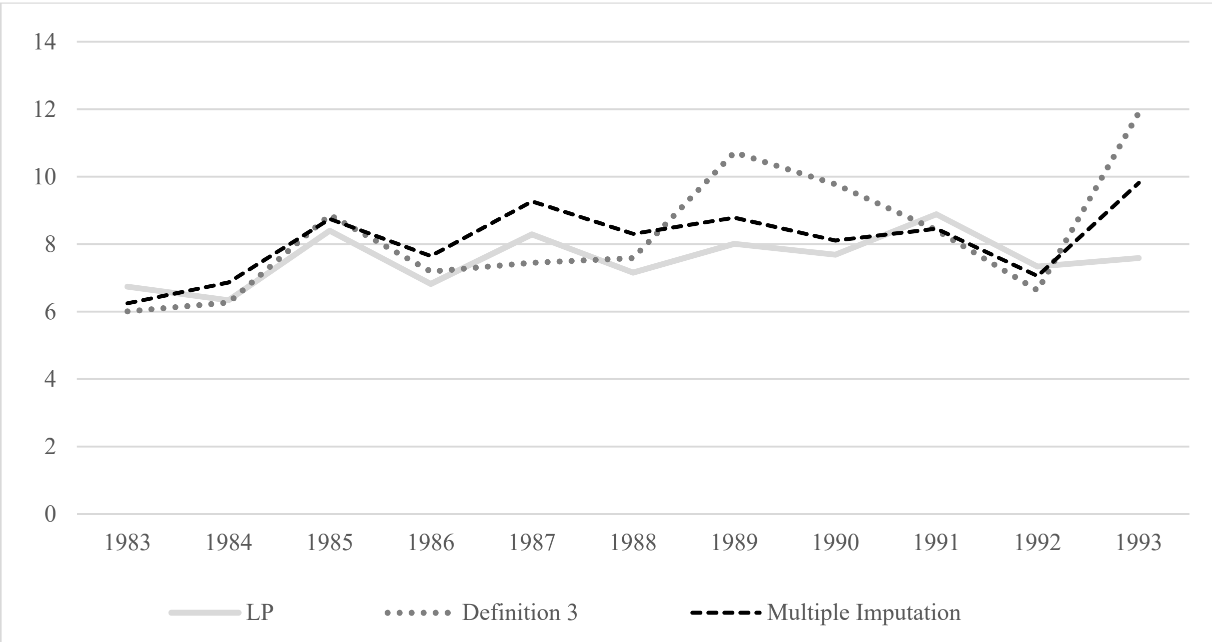
Panel B: λ



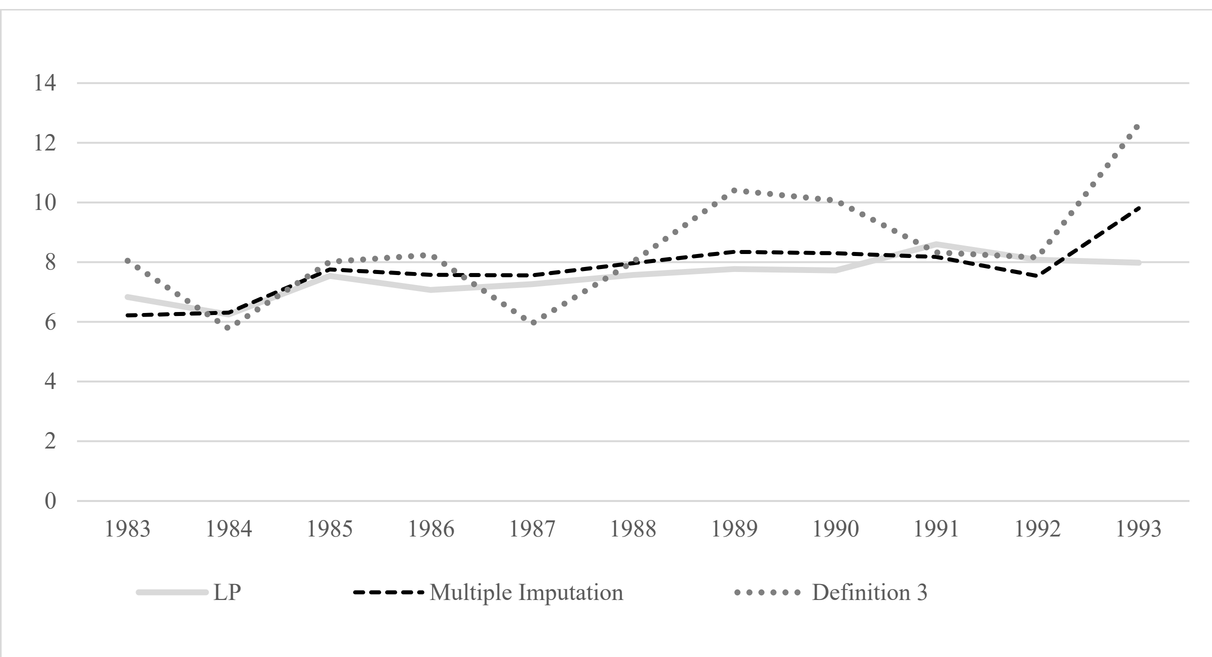
Note: Maximum likelihood estimates substituting for *N* with $Q_D^i/\lambda Q_D^i$ in the likelihood function assuming *EIM* at the state-hour-year-weekend level.

Figure 4: Year-specific estimates of θ and λ applying FARS MI BAC values

Panel A: θ



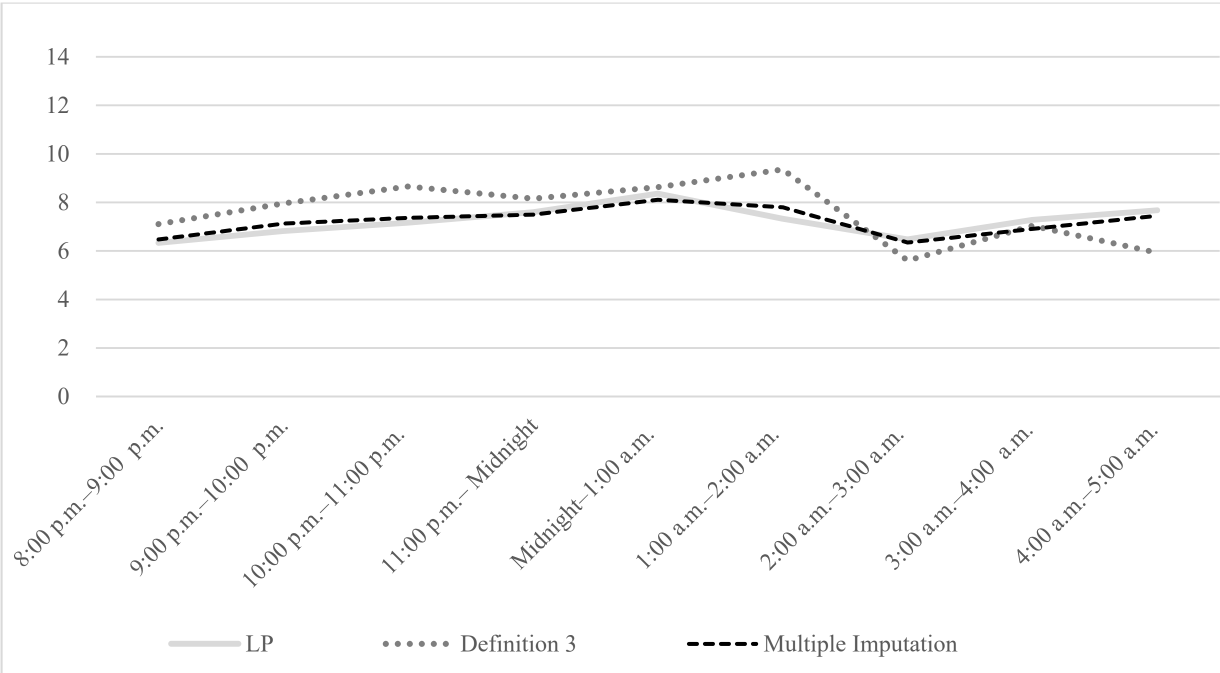
Panel B: λ



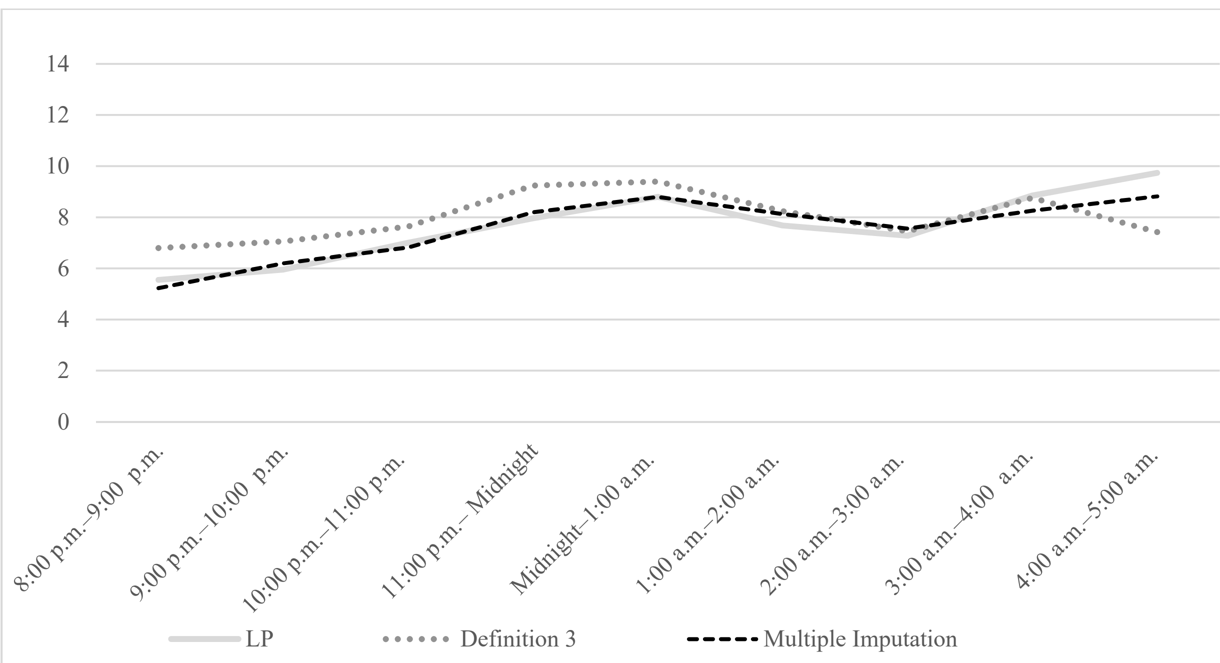
Note: Maximum likelihood estimates substituting for N with $Q_D^i/\lambda Q_D^i$ in the likelihood function assuming *EIM* at the state-hour-year-weekend level. Multiple imputation estimates are based on ten iterations.

Figure 5: Hour-specific estimates of θ and λ apply FARS MI BAC values

Panel A: θ



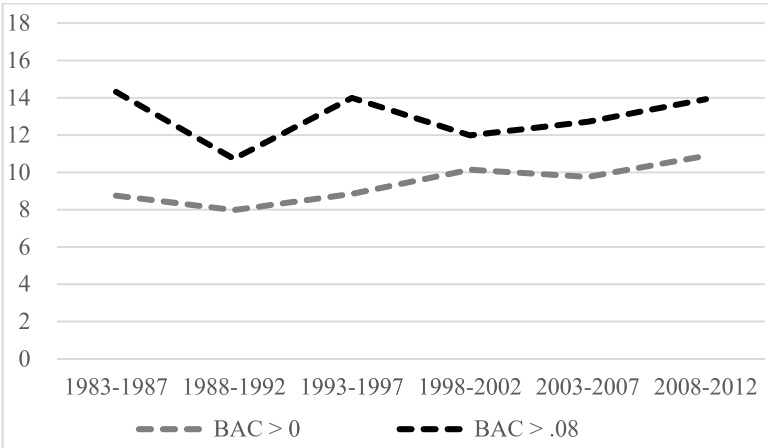
Panel B: λ



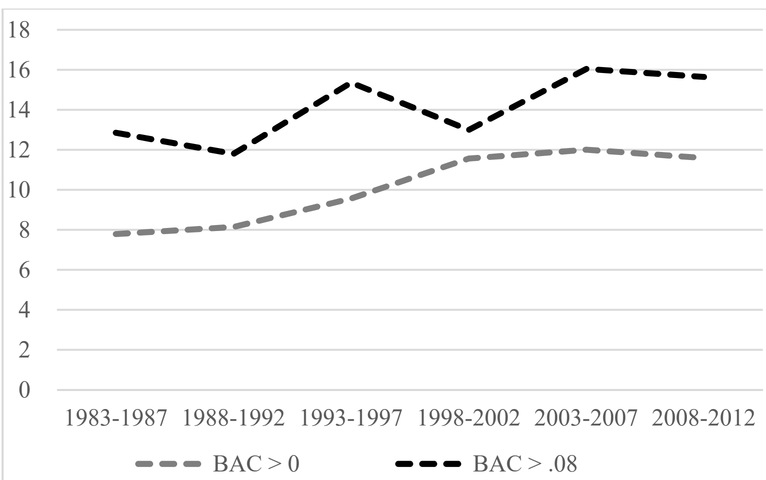
Note: Maximum likelihood estimates substituting for N with $Q_D^i/\lambda Q_D^i$ in the likelihood function assuming *EIM* at the state-hour-year-weekend level. Multiple imputation estimates are based on ten iterations.

Figure 6: Estimates of prevalence and relative risk of driving over specified BAC thresholds: 1983-2012

Panel A: θ



Panel B: λ



Panel C: Prevalence

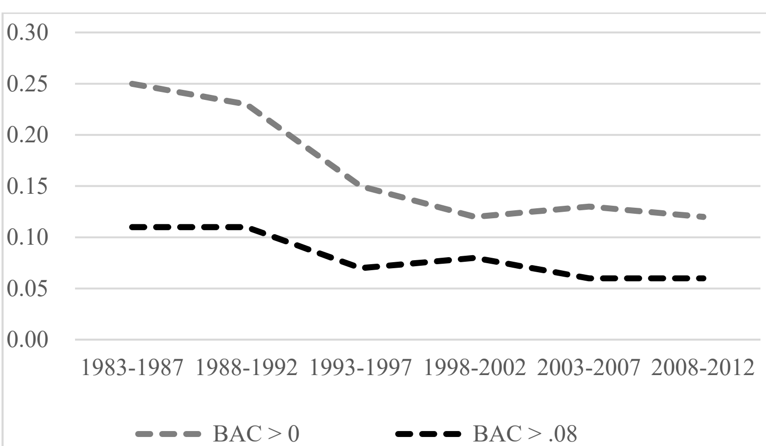
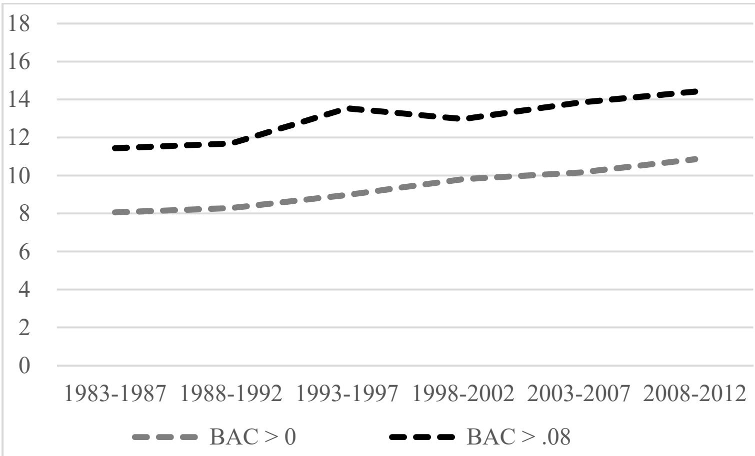
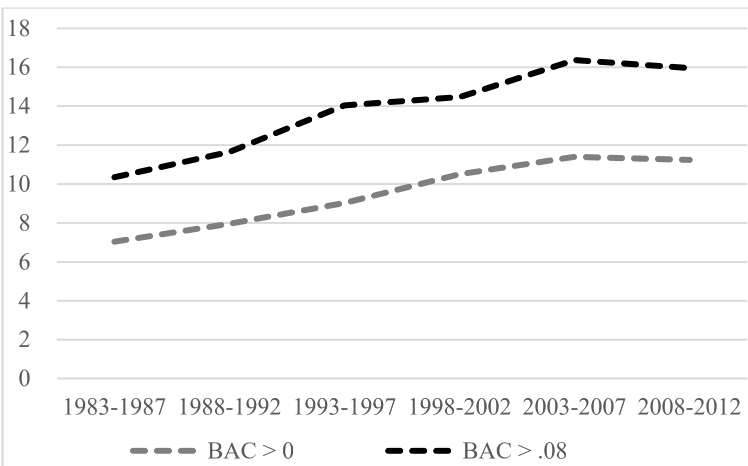


Figure 7: Estimates of prevalence and relative risk of driving over specified BAC thresholds applying FARS MI BAC values: 1983-2012

Panel A: θ



Panel B: λ



Panel C: Prevalence

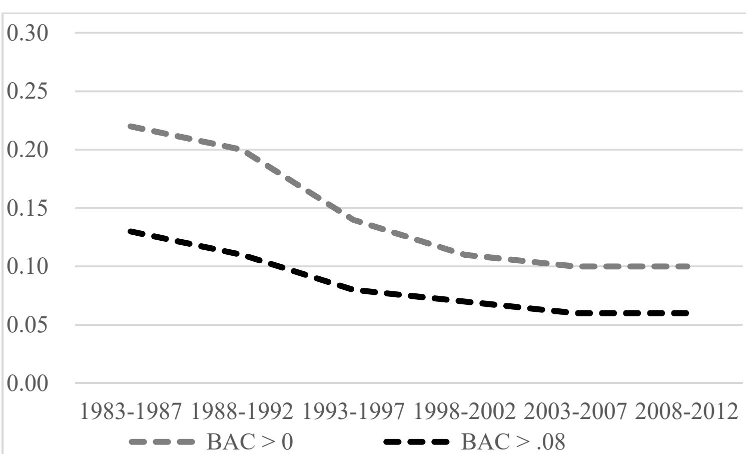
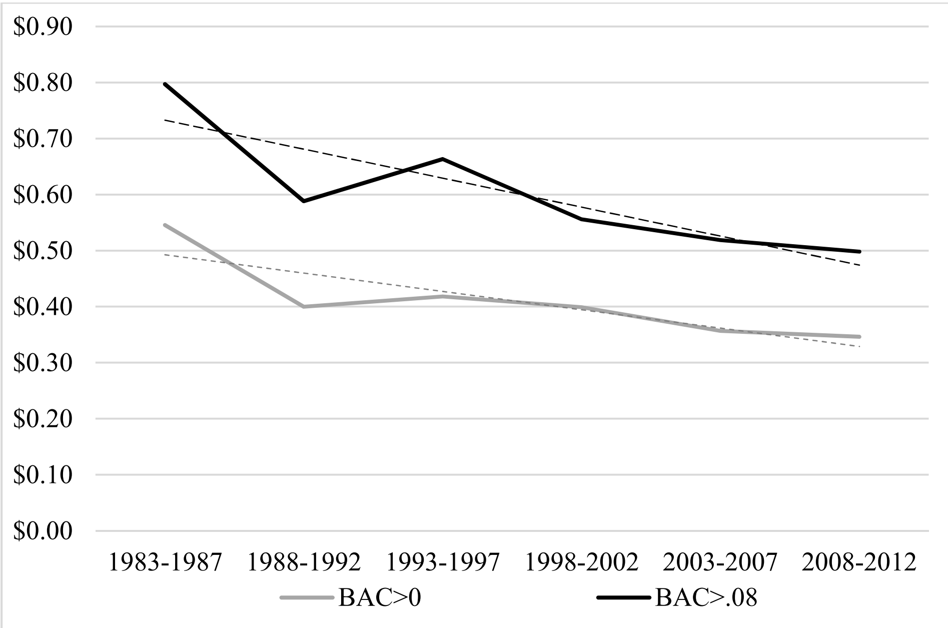


Figure 8: External cost of increased mortality risk per mile driven by alcohol impairment level.



Note: Calculations based upon prevalence estimates from MI estimation (see Figure 7) assuming $9.1 million VSL for all periods, and using the final year for each period as the population of accidents and drivers.

Table 1: Outline of LP Replication Exercise

	Result or sample restriction from LP (2003)	Location (page, ¶/table)	Replication result
Sample restrictions			
1.	“restrict the analysis...to the years 1983 to 1993"	1213, 2	428,338 fatal crashes 633,178 vehicles with driver
2.	“limit our sample to those hours (8:00 P.M.-5:00 A.M.)"	1213, 2	180,632 crashes (240,752 vehicles) Missing hour: 3,334 crashes (3,419 vehicles)
3.	“sample includes over 100,000 one-car crashes and over 40,000 two-car crashes"	1213, 2	127,239 one-vehicle crashes 47,987 two-vehicle crashes
4.	“during these hours, almost 60 percent of drivers involved in fatal crashes have been drinking"	1213,2	Depends on definition of drinking (see Table 3)
5.	“Crashes involving three or more drivers...less than 6 percent of fatal crashes...are dropped ”	1213, 2	8PM-5AM: 2.99% of crashes (7.27% of drivers) All hours: 5.35% of crashes (12.02% of drivers)
6.	“All crashes from a handful of state-year pairs with obvious data problems are also eliminated.”	1213, 2	N/A
7.	“exclude from the sample any crash in which one or more of the drivers are missing information about:		
7a.	time or location of the accident		See sample restrictions 2 and 6 above
7b.	police-reported drinking status		34.8% of crashes (see Table)
7c.	age	1213, 3	3.9% of crashes
7d.	sex		3.3% of crashes
7e.	past driving record		6.9% of crashes
7f.	combined...roughly 8 percent of all crashes.”		Including police-reported status: 36.8% of (3) Excluding police-reported status: 7.1% of (3)
8.	"the police officer’s evaluation…is available for virtually every driver"	1214, 1	28% of drivers lack police evaluation (see Table 2)
Supplementary analysis based on tested BAC			
9.	“exclude all crashes occurring in states that do not test at least 95% of those judged to be drinking in that year…more than 80% of fatal crashes”	1214, 2	162,879 (84.3%) of crashes are dropped
Results			
10.	Summary statistics	1215, Table 1	see Table 4
	Relative risk and prevalence based on police-report		
11.	by assumed level of equal and independent mixing	1217, Table 2	see Tables 5 and 6
12.	by year	1219, Table 3	see Figure 2
13.	by time of day	1221, Table 4	see Figure 3
14.	Relative risk and prevalence based on BAC test	1222, Table 5	see Table 7

Table 2: Distribution of police officer judgement of alcohol involvement and BAC test results

	Status according to Police Officer (DRINKING)				
Status according to BAC test	Not drinking	Drinking	Not reported [Missing]	Unknown	Total
BAC=0	16,444 (7.4%)	3,060 (1.4%)	4,671 (2.1%)	7,172 (3.2%)	31,347 (14.0%)
BAC>0	4,855 (2.2%)	63,269 (28.3%)	6,421 (2.9%)	15,711 (7.0%)	90,256 (40.4%)
No test available	47,445 (21.3%)	20,370 (9.1%)	21,024 (9.4%)	12,771 (5.7%)	101,610 (45.5%)
Total	68,744 (30.8%)	86,699 (38.8%)	32,116 (14.4%)	35,654 (16.0%)	**223,213**

Table 3: Potential definitions of drinking driver

				% Drinking	
Definition	Drinking = 1	Drinking =0	Missing	Uncon.	Con.
1 (LP described)	Officer judges alcohol involved	Officer judges alcohol not involved	Officer determination missing	38.8%	55.8%
2 (LP actual)	BAC>0, or Officer judges alcohol involved	BAC=0 or missing, and Officer judges alcohol not involved	none	50.9%	50.9%
3 (Preferred)	Officer judges alcohol involved, or BAC>0 if officer determination missing	Officer judges alcohol not involved, or BAC=0 if officer determination missing	Officer determination missing, and BAC test missing	48.8%	57.5%
4	BAC=1, or Officer judges alcohol involved if no BAC test	BAC=0, or Officer judges alcohol not involved if no BAC test	BAC test missing, and Officer determination missing	49.6%	58.4%

Table 4: Summary statistics for fatal crashes by definition of drinking driver

		Definition of drinking driver employed:				Definition 2, omitting state-years with missing data for more than:
	LP	1	2	3	4	13% of crashes
Sample of crashes between 8:00 PM and 5 AM						
Number of fatal one-car crashes	103,077	85,103	118,500	105,180	105,180	104,445
Number of fatal two-car crashes	39,470	25,677	44,379	34,500	34,500	39,082
Percentage of all drivers in fatal crashes:						
Reported to be drinking by police	53.4%	56.4%	52.8%	57.5%	58.5%	53.4%
Male	82.2%	82.8%	82.8%	83.0%	83.0%	82.8%
Under age 25	44.0%	40.8%	40.1%	40.2%	40.2%	40.3%
Bad previous driving record	37.2%	37.3%	36.9%	37.4%	37.4%	36.9%
Reported drinking and male	45.2%	48.9%	45.7%	49.8%	50.7%	46.2%
Reported drinking and under age 25	24.6%	24.3%	22.1%	24.2%	24.4%	22.5%
Reported drinking and bad driving record	23.5%	24.5%	22.7%	24.8%	25.2%	22.9%
Percentage of fatal one-car crashes with:						
One drinking driver	63.0%	65.6%	62.3%	67.2%	68.6%	63.1%
One sober driver	37.0%	34.4%	37.8%	32.8%	31.4%	37.0%
Percentage of fatal two-car crashes with:						
Two drinking drivers	14.2%	16.2%	14.0%	16.6%	16.0%	14.3%
One drinking, one sober driver	53.2%	49.8%	52.7%	52.2%	54.2%	52.7%
Two sober drivers	32.6%	34.0%	33.4%	31.3%	29.8%	33.1%
Restricted sample with 95% BAC reporting rate						
Percentage of fatal one-car crashes with:						
One legally drunk driver	55.9%		57.4%			58.8%
One sober driver	44.1%		42.6%			41.2%
Percentage of fatal one-car crashes with:						
Two legally drunk drivers	6.2%		6.4%			6.5%
One legally drunk, one sober driver	47.8%		49.3%			50.3%
Two sober drivers	46.0%		44.3%			43.2%

Notes: See Table 3 for drinking driver definitions.

Table 5: Relative likelihood of causing a fatal crash under different levels of equal mixing and alcohol involvement definitions

	LP					
Relative two-car fatal crash risk for drinking drivers (θ)	3.79 (0.14)	4.87 (0.16)	4.92 (0.16)	5.14 (0.17)	6.48 (0.20)	7.51 (0.22)
Relative one-car fatal crash risk for drinking drivers (λ)	5.04 (0.11)	5.46 (0.12)	5.50 (0.12)	5.67 (0.12)	6.72 (0.14)	7.45 (0.15)
Degrees of freedom	3	11	101	200	3427	6668
	Definition 1					
Relative two-car fatal crash risk for drinking drivers (θ)	1.94 (0.34)	2.70 (0.33)	2.87 (0.34)	3.11 (0.34)	4.79 (0.43)	5.70 (0.49)
Relative one-car fatal crash risk for drinking drivers (λ)	3.28 (0.31)	3.66 (0.26)	3.82 (0.26)	4.00 (0.27)	5.62 (0.33)	6.26 (0.36)
Degrees of freedom	3	11	101	200	845	1365
	Definition 2					
Relative two-car fatal crash risk for drinking drivers (θ)	3.57 (0.14)	4.49 (0.16)	4.55 (0.16)	4.74 (0.16)	6.18 (0.20)	6.98 (0.22)
Relative one-car fatal crash risk for drinking drivers (λ)	4.70 (0.11)	5.04 (0.11)	5.10 (0.12)	5.25 (0.12)	6.33 (0.14)	6.90 (0.15)
Degrees of freedom	3	11	101	200	4039	6910
	Definition 3					
Relative two-car fatal crash risk for drinking drivers (θ)	2.91 (0.23)	3.84 (0.26)	3.94 (0.26)	4.16 (0.27)	6.05 (0.34)	7.10 (0.39)
Relative one-car fatal crash risk for drinking drivers (λ)	4.41 (0.20)	4.80 (0.20)	4.89 (0.20)	5.07 (0.20)	6.75 (0.26)	7.52 (0.29)
Degrees of freedom	3	11	101	200	1946	3127
	Definition 4					
Relative two-car fatal crash risk for drinking drivers (θ)	4.02 (0.26)	5.29 (0.31)	5.43 (0.31)	5.70 (0.32)	7.34 (0.39)	8.49 (0.45)
Relative one-car fatal crash risk for drinking drivers (λ)	5.53 (0.22)	6.02 (0.23)	6.15 (0.23)	6.35 (0.24)	7.74 (0.29)	8.51 (0.32)
Degrees of freedom	3	11	101	200	1934	3097
	Multiple Imputation					
Relative two-car fatal crash risk for drinking drivers (θ)	3.71 (0.56)	5.01 (0.16)	5.15 (0.19)	5.38 (0.20)	7.06 (0.23)	7.84 (0.25)
Relative one-car fatal crash risk for drinking drivers (λ)	5.35 (0.75)	5.83 (0.29)	5.95 (0.16)	6.12 (0.17)	7.29 (0.17)	7.53 (0.17)
Degrees of freedom	3	11	101	200	3980	6737
Unit of observation over which "equal mixing" of drivers is imposed	all data	hour	hour x year	hour x year x weekend	hour x state x year	hour x state x year x weekend

Note: *N* is estimated as a function of a fully interacted set of dummy variables defined by the spatiotemporal unit over which *EIM* is assumed, except under multiple imputation, when it is substituted (see text).

Table 6: Relative likelihood of causing a crash and prevalence of drinking and driving under different estimation specifications and alcohol involvement definitions

	LP	Definition 2		Definition 3	
Relative two-car fatal crash risk for drinking drivers (θ)	7.51 (0.22)	8.10 (0.26)	6.98 (0.22)	7.93 (0.46)	7.10 (0.39)
Relative one-car fatal crash risk for drinking drivers (λ)	7.45 (0.15)	7.57 (0.19)	6.90 (0.15)	7.99 (0.34)	7.52 (0.29)
Proportion of drinking drivers	0.186 (0.004)	0.189 (0.003)		0.197 (0.006)	
Specification of *N*:	estimated	estimated	substituted	estimated	substituted
Degrees of freedom:	6668	6910		3127	
Equal-and-independent mixing assumed:	hour x state x year x weekend				

Table 7: Estimates of the relative risk of alcohol-impaired driving

	LP	Definition 2	Multiple Imputation
Relative two-car fatal crash risk for drivers with BAC>0.1 (θ)	13.24 (1.17)	14.37 (1.36)	12.9 (0.48)
Relative one-car fatal crash risk for drivers with BAC>0.1 (λ)	14.39 (1.05)	14.7 (1.14)	12.69 (0.39)
Degrees of freedom	Unreported	1496	6832