



Published in final edited form as:

*Accid Anal Prev.* 2013 January ; 50: 1244–1251. doi:10.1016/j.aap.2012.09.025.

## Inverse propensity weighting to adjust for bias in fatal crash samples

David E. Clark<sup>a,\*</sup> and Edward L. Hannan<sup>b</sup>

<sup>a</sup>Department of Surgery, Maine Medical Center, 887 Congress Street, Portland ME 04102 USA

<sup>b</sup>School of Public Health, State University of New York at Albany, Rensselaer NY 12144 USA

### Abstract

**Background**—The Fatality Analysis Reporting System (FARS) has data from all areas of the United States, but is limited to fatal crashes. The National Automotive Sampling System – General Estimates System (NASS-GES) includes all types of serious traffic crashes, but is limited to a few sampling areas. Combining the strengths of these two samples might offset their limitations.

**Methods**—Logistic regression (allowing for sample design, and conditional upon selected person-, event-, and geographic-level factors) was used to determine the propensity ( $P_{FC}$ ) for each injured person in 2002–2008 NASS-GES data to be in a fatal crash sample. NASS-GES subjects injured in fatal crashes were then reweighted by a factor of  $W_{FC} = (1/P_{FC})$  to create a “pseudopopulation”. The weights ( $W_{FC}$ ) derived from NASS-GES were also applied to injured subjects in 2007 FARS data to create another pseudopopulation. Characteristics and mortality predictions from these artificial pseudopopulations were compared to those obtained using the original NASS-GES sample. The sum of  $W_{FC}$  for FARS cases was also used to estimate the number of crash injuries for rural and urban locations, and compared to independently reported data.

**Results**—Compared to regression results using the original NASS-GES sample, unadjusted models based on fatal crash samples gave inaccurate estimates of covariate effects on mortality for injured subjects. After reweighting using  $W_{FC}$ , estimates based upon the pseudopopulations were similar to results obtained using the original NASS-GES sample. The sum of  $W_{FC}$  for FARS cases gave reasonable estimates for the number of crash injuries in rural and urban locations, and provided an estimate of the rural effect on mortality after controlling for other factors.

**Conclusions**—Weights derived from analysis of NASS-GES data (the inverse propensity for selection into a fatal crash sample) allow appropriate adjustment for selection bias in fatal crash samples, including FARS.

### Keywords

Traffic crash; mortality; fatality; FARS; NASS-GES; inverse propensity; weight; rural; urban

---

© 2012 Elsevier Ltd. All rights reserved.

\*Corresponding author: Telephone 207-774-2381, Fax 207-774-0459.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## 1. Introduction

The Fatality Analysis Reporting System (FARS) is a detailed census of all traffic crashes in the United States resulting in the death of at least one person within thirty days. It has been maintained since 1975 by the National Highway Traffic Safety Administration (NHTSA, 2005), and has enabled much valuable research within NHTSA and by external investigators (NHTSA, 2008). FARS includes geographic information about fatal crashes, and information about persons who survived a crash in which another person died. However, FARS by definition does not include any information about survivors of comparable nonfatal crashes, which limits the usefulness of FARS data.

In order to provide more general information about traffic crashes, NHTSA has also developed the National Automotive Sampling System – General Estimates System (NASS-GES), an annual population-based stratified random sample of police-reported crashes that includes non-fatal crashes as well as fatal crashes. However, NASS-GES is based upon a relatively small number of sampled locations and does not have as much detail as FARS. This limits the usefulness of NASS-GES data.

Using NASS-GES, it is possible to estimate the probability that an injured person with certain identifiable conditions would be involved in a fatal crash. This information theoretically might enable inferences about the characteristics of all injured persons, even in situations when only a fatal crash database is available. Appropriate adjustment of the sampling bias of a fatal crash database might then result in more valid estimates of covariate effects on outcomes in the entire population of injured subjects. We investigated one approach to adjusting the sampling bias in FARS using information derived from NASS-GES, in order to study geographic and other factors that influence mortality among persons injured in traffic crashes.

## 2. Methods

This study was part of a larger project presented to an Institutional Review Board at Maine Medical Center, which judged it exempt from further review because it only uses anonymous public data. Data management and statistical analyses were carried out using Stata (Version 11, StataCorp, College Station TX) and statistical methods that allowed for sampling design and weights.

### 2.1. Management of NASS-GES data

NASS-GES is a weighted stratified sample of police-reported traffic crashes resulting in personal injury or property damage (NHTSA, 2005). Data are obtained from 60 Primary Sampling Units (PSU) within the United States, identified as a city, county, or group of counties (together including a total of 77 out of the 3142 US counties), selected to represent 14 predetermined types of geographic areas (PSU Strata). Next, a sample of police jurisdictions (PJ) is selected from each PSU. Finally, within the selected PJs, cases are selected from each of six types (Case Strata). The overall probability that a case will be selected is the product of the probabilities of selection at each level. From the inverse of this overall probability, NASS-GES derives a weight, which will be designated here as  $W_{GES}$ , to allow estimation of the actual totals and proportions in the US population. Further details about the sample design are given in a NASS-GES Technical Note (NHTSA, 1991).

Data for 2002–2008 were downloaded from the NASS-GES internet site. Analysis for this study included only injured subjects, namely those who had been described in a police crash report as having “possible injury”, “injury of unknown severity”, “non-incapacitating injury”, “incapacitating injury” or “fatal injury”. Characteristics of subjects involved in a

crash from which at least one person died (“fatal crash”) were compared to subjects involved in nonfatal crashes. Characteristics of the persons, vehicles, and events included age, sex, pedestrian/occupant, safety belt use, ejection, vehicle speed (high/low), vehicle damage (severe/other), head-on collision, number of vehicles involved, number of persons injured, time of crash, and individual outcome (lived/died). Missing values were replaced with imputed values provided by NASS-GES. Vehicle speeds were considered high if the police estimate of the speed was at least 50 miles per hour; if there was no police estimate, vehicle speed was considered high if the posted speed limit was at least 50 miles per hour (Clark, 2003).

The NASS-GES Technical Note also lists the cities or counties in each of the 60 PSUs (NHTSA, 1991). For each severely injured person, a crash county was assigned as: 1) The PSU county if the PSU contained a single county; 2) The driver’s residence county if the PSU contained more than one county and the driver’s ZIP code matched one of the included counties; or 3) The most frequent residence county for the PSU if it contained more than one county and the driver’s ZIP code did not match any included county. This approach resulted in 77 crash counties, which could be linked using the Federal Information Processing Standard (FIPS) county codes (US Census, 2011) to county-level data in the Area Resource File (HRSA, 2011). In particular, variables were created indicating the county Rural-Urban Continuum Code (RUCC), which classifies US counties into nine groups from most urban (1) to most rural (9).

The complex sampling design of NASS-GES was accommodated using the “svy” commands available in Stata, with the first stage consisting of PSU grouped within PSU Strata, a second stage consisting of PJ, and a third stage consisting of individual cases grouped within Case Strata. The “subpopulation” option was used to restrict analysis to injured persons. The probability of death for injured persons was modeled using the command “svy:logit” and the independent variables described above. The analysis was repeated for the smaller subsample of injured persons involved in crashes where at least one person died (a fatal crash sample).

## 2.2. Creation of a pseudopopulation from NASS-GES fatal crash subsample

For each crash in which one or more subjects had been injured, one injured subject was randomly selected as an “index person”. Logistic regression with the independent variables described above, including variables relating to the index person, was then used to estimate the probability that one or more subjects in the crash had been fatally injured. For each crash, this produced a “propensity score” ( $P_{FC}$ ) estimating the probability that the crash would be included in a subsample limited to fatal crashes.

An “inverse propensity” weighting factor ( $W_{FC}$ ) was then calculated as the inverse of the propensity score ( $1/P_{FC}$ ), rounded to the nearest integer. This weight was applied to all injured persons in the fatal crash sample, whether or not they were the randomly-selected index person. Using the Stata “expand” command, each subject in the fatal crash subsample was replicated once, and a weight of  $W_{GES} * (W_{FC} - 1)$  was assigned to each of the replicated cases. Identical results could be obtained by creating  $(W_{FC} - 1)$  replications of each subject and retaining the original weight of  $W_{GES}$  for each case. Replicated cases (which will also be referred to as “pseudosubjects”) retained the characteristics of the original cases from which they had been replicated, except that pseudosubjects were all recorded as survivors, whether or not the original subject had survived. The pseudosubjects were appended to the original cases in the fatal crash sample, creating a “pseudopopulation”, theoretically having the same characteristics as the population of injured persons who would have been included in the fatal crash sample if their crash resulted in one or more deaths.

Figure 1 depicts graphically the rationale for obtaining an “inverse propensity weight”. For each possible covariate pattern, subjects have a propensity ( $P_{FC}$ ) to be involved in a fatal crash (and therefore included in a sample limited to fatal crashes). Subjects involved in a fatal crash include a number  $D$  who died and a number  $L$  who lived. Subjects not involved in a fatal crash consist of a number  $X$ , all of whom by definition survived. Since  $P_{FC}$  can be estimated as  $(D+L)/(D+L+X)$ , simple algebra gives

$$X! \left| \frac{1}{P_{FC}} - 1 \right| !D! L!$$

Figure 2 depicts the steps in creating a “pseudopopulation”. “Pseudosubjects” are created by replicating each original subject (those categorized  $D$  or  $L$  in Figure 1) in the original fatal crash sample; characteristics of each original subject (indexed by the letter  $i$ ) are retained by the pseudosubjects it generates, except that all of the pseudosubjects are considered survivors. The inverse propensity weight is defined as  $W_{FCi} = 1/P_{FCi}$ . The pseudopopulation consists of the pseudosubjects (with weight  $W_{FCi} - 1$ ) appended to the original subjects (with weight 1).

Characteristics of the pseudopopulation resulting from appending pseudosubjects to the NASS-GES fatal crash subsample were compared to the characteristics of the original NASS-GES stratified random sample. This pseudopopulation was analyzed as a one-stage unstratified sample with the weights adjusted as described above. The probability of death for injured subjects in the pseudopopulation was modeled using the command “svy:logit” and the independent variables described above, and compared to the results obtained from the original NASS-GES sample.

### 2.3. Creation of a pseudopopulation from FARS fatal crash sample

Data from 2007 were then downloaded from the FARS internet site (NHTSA, 2008) and variables required for the weighting equations were modified to be as similar as possible to those in NASS-GES. FARS is essentially a census (attempted 100% sample) of all fatal crashes in the United States, so weights and sampling design do not need to be considered. County crash locations identified in FARS were classified by RUCC (HRSA, 2011). Frequencies and regression results (using ordinary logistic regression) were compared to those obtained from the fatal crash subsample that had been created from NASS-GES. FARS cases were then replicated using the inverse-propensity weights that had been derived from the NASS-GES sample: Weights of 1 were assigned to original FARS cases and  $(W_{FCi} - 1)$  to replicated cases. Subjects in this FARS-based pseudopopulation were analyzed using the Stata “svy” commands, as if they had been selected in a single stage with probability  $P_{FCi}$  from a population of injured subjects. The probability of death for injured persons in the pseudopopulation was modeled using the command “svy:logit” and the independent variables described above. Results were compared to those obtained from the original NASS-GES sample.

Finally, the total of the weights applied to the FARS data (pseudosubjects in addition to original subjects) was used to estimate the number of persons who had been injured in each county during 2007. These estimates were compared to the 2007 totals reported for US counties on available state websites, which are provided as an addendum to the report of Goldstein and colleagues (2011). For comparison, weighted estimates were also derived from NASS-GES data and a propensity score that did not include RUCC codes. For each RUCC category, the observed case fatality rate was compared to the case fatality rate that

would be predicted from a model that did not include RUCC variables in calculating the propensity score.

### 3. Results

#### 3.1. Description of NASS-GES samples and propensity scores

NASS-GES for years 2002–2008 contained records on over a million persons, of whom 65% (weighted proportion 82%) did not suffer any injury. There were 354,557 injured subjects, corresponding (after weighting) to an injured population of 18,578,787. These subjects were injured in 227,014 crashes, corresponding to 12,608,943 injury crashes after weighting. In the NASS-GES sample, 17,737 subjects were injured in one of 8,595 crashes in which at least one person died (“fatal crashes”), corresponding (after weighting) to a population of 374,026 injured in one of 193,322 fatal crashes. Thus, the probability that an injured person was in a fatal crash was approximately  $374,026/18,578,787 = 2.01\%$ , while the probability that an injury crash was a fatal crash was approximately  $193,322/12,608,943 = 1.53\%$ .

Table 1 demonstrates that injured persons in fatal crashes were more likely to be older, male, and either pedestrians, motorcyclists, or unbelted (and sometimes ejected) occupants of vehicles that were travelling at high speeds, rolled over, or suffered severe vehicle damage. Fatal crashes were also more frequent with head-on collisions, after midnight, and in rural areas. The more persons who were injured, the more likely one of them died. The relative importance of these factors, controlling for each of the others, is reflected in the logistic regression equations predicting the probability that a subject would be in a fatal crash (Table 2, “Model 1”).

Table 2 can be used to calculate the probability that a given subject with an injury in the NASS-GES sample would be included in a fatal crash subsample (i.e., the propensity score). For example, a 50-year-old male driver wearing a belt during the daytime, who crashed into a tree in the most rural county at high speed with severe vehicle damage, but injuring nobody else, would have an odds of being in a fatal crash (in this case, with no other person injured, the same as dying himself) equal to

$$0.000700 * 1.74 * 1.70 * 2.68 * 2.92 * 1.91 = 0.0309.$$

Like any probability, the corresponding propensity score can be obtained as odds/(odds+1), giving 0.03. The weighting factor determining the number of cases in the pseudopopulation would be the nearest whole number to  $1/(0.03)$ , namely 33. Therefore the pseudopopulation would contain the original subject from the fatal crash subsample plus 32 replicated cases; whether or not the original subject survived, all the replicated cases (pseudosubjects) would be analyzed as survivors.

An alternative propensity score was also obtained without including county RUCC information (Table 2, “Model 2”). The other coefficients were virtually the same as in Model 1. A potential use for the weights derived from Model 2 will be described below.

#### 3.2. Comparison of pseudopopulations to full NASS-GES sample

Creation of a pseudopopulation from the NASS-GES fatal crash subsample using Model 1 allowed characteristics of the full NASS-GES sample of injured persons to be closely approximated. Logistic regression models predicting mortality demonstrate the biased results obtained from a non-random fatal crash sample, the results after adjustment of the bias using the replication method, and the results obtained from the original population-

based sample (Table 3). Odds ratios in Table 3 approximate the risk ratio with respect to the referent category of females aged 15–39, who were belted non-ejected occupants of vehicles that were not traveling at high speed, did not roll over, and did not suffer severe damage, in crashes that did not involve head-on collision, occurred between 6AM and midnight, and were in urban or suburban areas (RUCC 1–3). Without reweighting, the effects of high speed, vehicle rollover, and rural location are not seen, and there is even an apparent “protective effect” for head-on collisions. After reweighting, covariate effects were similar to those seen in the original population-based sample.

Alternative specifications of the sample design did not affect point estimates of regression coefficients (and therefore did not affect weights). However, specification of the number of sampling levels and strata made an important difference in the calculation of standard errors. Standard errors for the completely specified GES sample were most closely approximated by considering the replicated sample as a simple unstratified random sample.

The characteristics of FARS subjects were similar to those in the NASS-GES fatal crash subsample (Table 1). A model predicting mortality using unweighted FARS data gave biased results similar to those obtained from the NASS-GES fatal crash subsample, but a model after replication and weighting gave results similar to those obtained from the original NASS-GES population-based sample (Table 3).

### 3.3. Other estimates based on FARS-based pseudopopulation

Table 4 demonstrates some other potential uses of weighted FARS data, using data from counties that independently reported their numbers of traffic injuries.(Goldstein et al., 2011) Predictions using weights derived from Model 1 or Model 2 both underestimated the number of injuries in urban counties. However, predictions from Model 1 were much closer than those from Model 2 for rural counties. Comparison of the observed case fatality rate to the case fatality rate predicted from Model 2 (not including RUCC) demonstrates the increased mortality in rural counties.

## 4. Discussion

Previous researchers have noted that FARS is a biased sample of crashes, and have developed methods to correct for this bias by comparing persons within the same vehicle, or vehicles within the same crash event (Cummings, McKnight, & Weiss, 2003; Evans, 1986; Islam & Goetzke, 2009; Lui, McGee, Rhodes, & Pollack, 1988; O'Neill & Barry, 1995). However, these methods cannot be used to compare outcomes among crashes occurring in different locations. Elliot and colleagues (2006) have suggested combining FARS with the NASS Crashworthiness Data System (NASS-CDS), and their suggestion could be extended to NASS-GES, but this approach would still be limited to the relatively small number of areas sampled by either component of NASS.

### 4.1. Possible methods for bias correction using inverse propensity weighting

Persons injured in fatal crashes are clearly different from those in nonfatal crashes. However, if the most important sampling biases can be identified and quantified, the inverse of the propensity for selection may be used to estimate the characteristics of the original population even when only a non-random sample is available (Haneuse et al., 2009; Kang & Schafer, 2007; Lu, Jin, Chen, & Gluer, 2006). Similar methods have been used to adjust for non-random treatment classification or survey nonresponse (Hernan & Robins, 2006; Rao, Sigurdson, Doody, & Graubard, 2005), and resulting weights have been applied with some success for regression analyses (Howe, Cole, Chmiel, & Munoz; Pan & Schaubel, 2009).

Other applications of propensity scores in traffic safety research have been recently described by Sasidharan and Donnell (2012).

The inverse probability weighting method depends upon identifying enough characteristics of the subjects in the population and in the biased subsample so that, after controlling for these characteristics, the selection of the subject into the subsample is approximately random. In our example, when a person with given age, sex, belt use, and crash characteristics is injured, there is a certain probability that someone will die as a result of the crash. Additional randomness is introduced because the person who dies may or may not be the person whose characteristics are part of the model.

A simpler approach to adjusting for sampling bias might stratify by one or more characteristics and give a weight to each of these. However, in order to allow for interactions between different factors, this would require a number of categories rising exponentially with the number of factors and would quickly become cumbersome. In essence, the propensity score method extends the categorical method, because it creates a number of categories corresponding to the number of covariate patterns in the subsample that are possible in the given regression equation.

The methods described here assume that the characteristics of persons with injuries are similar (after controlling for identifiable factors) whether or not they die. For some applications, it might be useful to limit the subpopulation to persons with “incapacitating” or “fatal” injuries. Another alternative might be to estimate the probability that a person with nonfatal injury would be included as a survivor in the fatal crash sample. However, since a large proportion of the injured persons in fatal crashes die, this alternative would ignore much of the information in the fatal crash sample, including all of the information from crashes where the only injured person also dies.

It would be possible to calculate a propensity score for each individual person without regard to their being in the same vehicle or crash with others, although this would ignore the correlations within these clusters. Some method other than the random identification of an “index person” (e.g., multilevel modeling) could also be used to account for person-level effects on the probability of inclusion in a fatal crash sample. However, it would become rather complex to include information on more than one individual person and then to devise an appropriate system of adjusting the weights based upon these results.

When using survey data, such as NASS-GES, it is relatively easy to obtain point estimates using sample weights. However, it is not easy to account properly for the sample design and other adjustments when calculating standard errors, and this is an area of ongoing theoretical research (Gelman, 2007). The general principles of error estimation for complex survey data have been described elsewhere (Li & Levy, 2009; Roberts, Rao, & Kumar, 1987). In practice, considering the pseudopopulation as a one-stage unstratified random sample in this study provided reasonable approximations to the regression standard errors obtained using the original NASS-GES sample and design specifications.

#### **4.2. Potential applications of pseudopopulations derived from fatal crash samples**

If a database for an entire population is available, then it is obviously unnecessary to analyze a pseudopopulation, and we created a NASS-GES pseudopopulation from its fatal crash subsample only for the purpose of validation. However, if a database for a population involved in non-fatal as well as fatal traffic crashes is not available, then a pseudopopulation based on a fatal crash sample may be the best way to estimate the characteristics of the larger population. For example, if policy makers in some region of the United States wanted to estimate the increased mortality among their citizens attributable to not wearing safety

belts, a reasonable estimate could be made using a pseudopopulation based upon FARS data from their own region and weights obtained from Table 2.

Weighting FARS data may also have value for estimating the number of persons injured due to traffic crashes, as demonstrated in Table 4. The Federal Highway Administration formerly collected nonfatal injury data from the various states, but stopped doing this after 1996 because “these data have been erroneous and can be misleading” (OHIM, 1999). It is understandable that thousands of police jurisdictions may have considerable variability in the identification of crashes and injury severities. However, if it is assumed that fatal crashes are identified similarly throughout the United States, and that NASS-GES is a reasonably accurate representation of the country, then an estimate using the methods described here may actually be more accurate than relying on state reports of varying quality and completeness.

Conversely, if the reported data from a state, county, or other geographic unit are considered reasonably accurate, Table 4 also shows that they can be used to formulate an observed-to-expected ratio for that geographic area, in this case using weights derived from a propensity score that does not include any geographic data. A similar approach could be undertaken to investigate the possibility of residual confounding due to causes other than geographic diversity.

Previous studies using NASS-GES have demonstrated the predictive effect of person-level, event-level, and geographic characteristics on the probability of death given injury (Clark, 2003; NHTSA, 2008). In particular, the increased risk of mortality in rural areas has been demonstrated, even controlling for other determinants of injury severity that might be more common in rural areas. The methods described here provide further evidence of this relationship.

### 4.3. Limitations

The methods we describe depend upon the assumption that, conditional upon the covariates used in the estimation of propensity, inclusion of injured persons in a fatal crash sample is essentially random. This assumption is more likely to be satisfied if the population for which the propensity is calculated has been in relatively severe crashes, which is why we restricted it to crashes involving personal injury; if the population were restricted only to crashes with incapacitating or fatal injury, the propensity should be even more accurate. However, this might restrict the applicability of the pseudopopulation that would be generated.

The degree to which other outcomes or characteristics of the pseudopopulation resemble those of the original population also depends upon the randomness of the propensity for being in the fatal crash sample. This argues in favor of using as many variables as possible to determine the propensity score. However, there are limitations in the similarity and completeness of variables used both in NASS-GES and in FARS, and probably in any other combination of databases to which these methods could be applied.

Weighting methods can also be unstable if other assumptions are violated, for example when a propensity score is very low (perhaps due to a small number of anomalous cases) and therefore results in an excessive weight or number of replications. One proposed *ad hoc* approach to this problem would truncate excessive weights or replications. Certainly, any inferential method based on sampling assumptions should be verified by population data whenever possible.



#### 4.4. Conclusion

As a general concept, the inverse propensity weighting method simply adds another sampling level to the inverse probability of selection weighting method of NASS-GES, and adjusts the weights accordingly. If the procedures proposed here can be validated using other sources that combine fatal and nonfatal crash data, they may be a useful basis for making inferences from samples that include only fatal crashes, in particular the Fatality Analysis Reporting System.

#### Acknowledgments

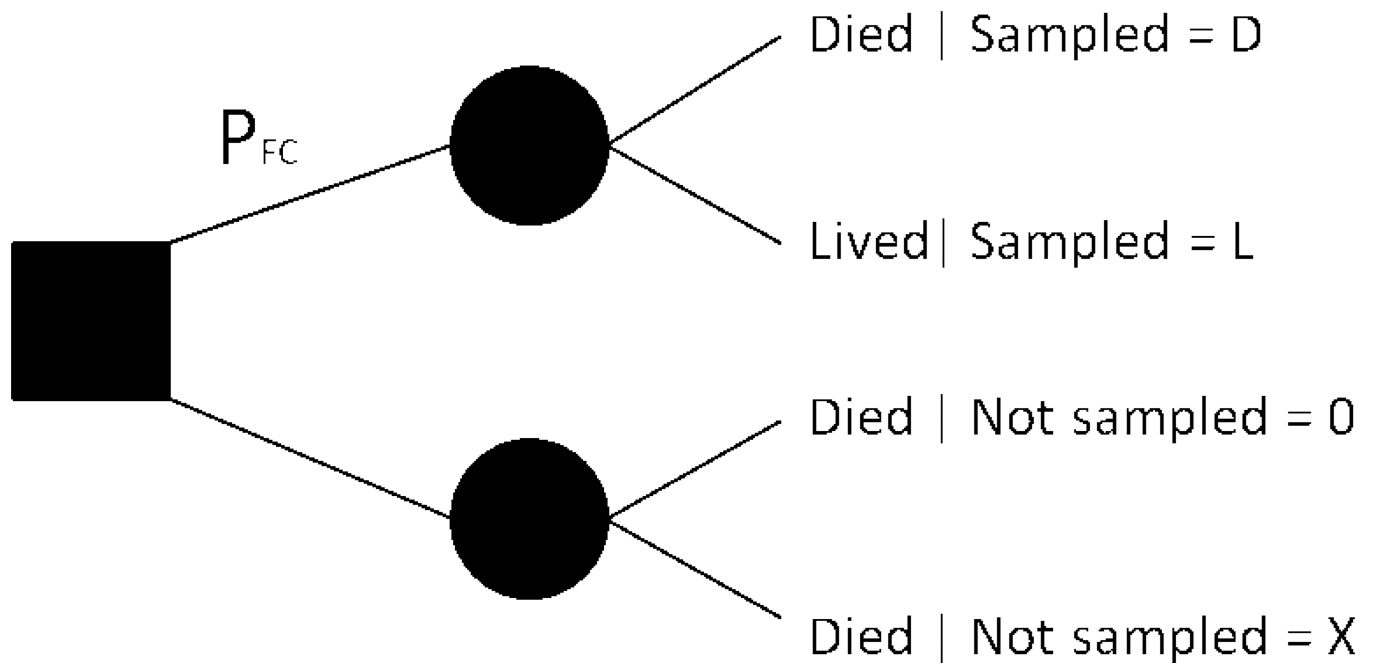
Supported by NIH grant R21HD061318 and CDC grant R01CE001594.

#### References

- Clark DE. Effect of population density on mortality after motor vehicle collisions. *Accident Analysis and Prevention*. 2003; 35(6):965–971. [PubMed: 12971931]
- Cummings P, McKnight B, Weiss NS. Matched-pair cohort methods in traffic crash research. *Accident Analysis and Prevention*. 2003; 35(1):131–141. [PubMed: 12479904]
- Elliott MR, Kallan MJ, Durbin DR, Winston FK. Effectiveness of child safety seats vs seat belts in reducing risk for death in children in passenger vehicle crashes. *Archives of Pediatric and Adolescent Medicine*. 2006; 160(6):617–621.
- Evans L. Double pair comparison--a new method to determine how occupant characteristics affect fatality risk in traffic crashes. *Accident Analysis and Prevention*. 1986; 18(3):217–227. [PubMed: 3730096]
- Gelman A. Struggles with survey weighting and regression modeling. *Statistical Science*. 2007; 22(2): 153–164.
- Goldstein GP, Clark DE, Travis LL, Haskins AE. Explaining regional disparities in traffic mortality by decomposing conditional probabilities. *Injury Prevention*. 2011; 17(2):84–90. [PubMed: 21212443]
- Haneuse S, Schildcrout J, Crane P, Sonnen J, Breitner J, Larson E. Adjustment for selection bias in observational studies with application to the analysis of autopsy data. *Neuroepidemiology*. 2009; 32(3):229–239. [PubMed: 19176974]
- Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health*. 2006; 60(7):578–586. [PubMed: 16790829]
- Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *American Journal of Epidemiology*. 2011; 173(5):569–577. [PubMed: 21289029]
- HRSA (Health Resources and Services Administration). [Accessed 29 Nov 2011] Area Resource File. 2011. <http://arf.hrsa.gov>.
- Islam S, Goetzke F. Correcting sample selection in FARS data to estimate seatbelt use. *Journal of Safety Research*. 2009; 40(5):389–393. [PubMed: 19932321]
- Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*. 2007; 22(4):523–539.
- Li L, Levy PS. Regression modeling of combined data from multiple sample surveys. *Statistics in Medicine*. 2009; 28(16):2160–2169. [PubMed: 19462413]
- Lu Y, Jin H, Chen MH, Gluer CC. Reduction of sampling bias of odds ratios for vertebral fractures using propensity scores. *Osteoporosis International*. 2006; 17(4):507–520. [PubMed: 16369725]
- Lui KJ, McGee D, Rhodes P, Pollack D. An application of a conditional logistic regression to study the effects of safety belts, principal impact points, and car weights on drivers' fatalities. *Journal of Safety Research*. 1988; 19:197–203.
- NHTSA (National Highway Traffic Safety Administration). [Accessed 29 Nov 2011] National Automotive Sampling System General Estimates System. Technical Note, 1988 to 1990. 1991. <http://www-nrd.nhtsa.dot.gov/Pubs/807796.pdf>.

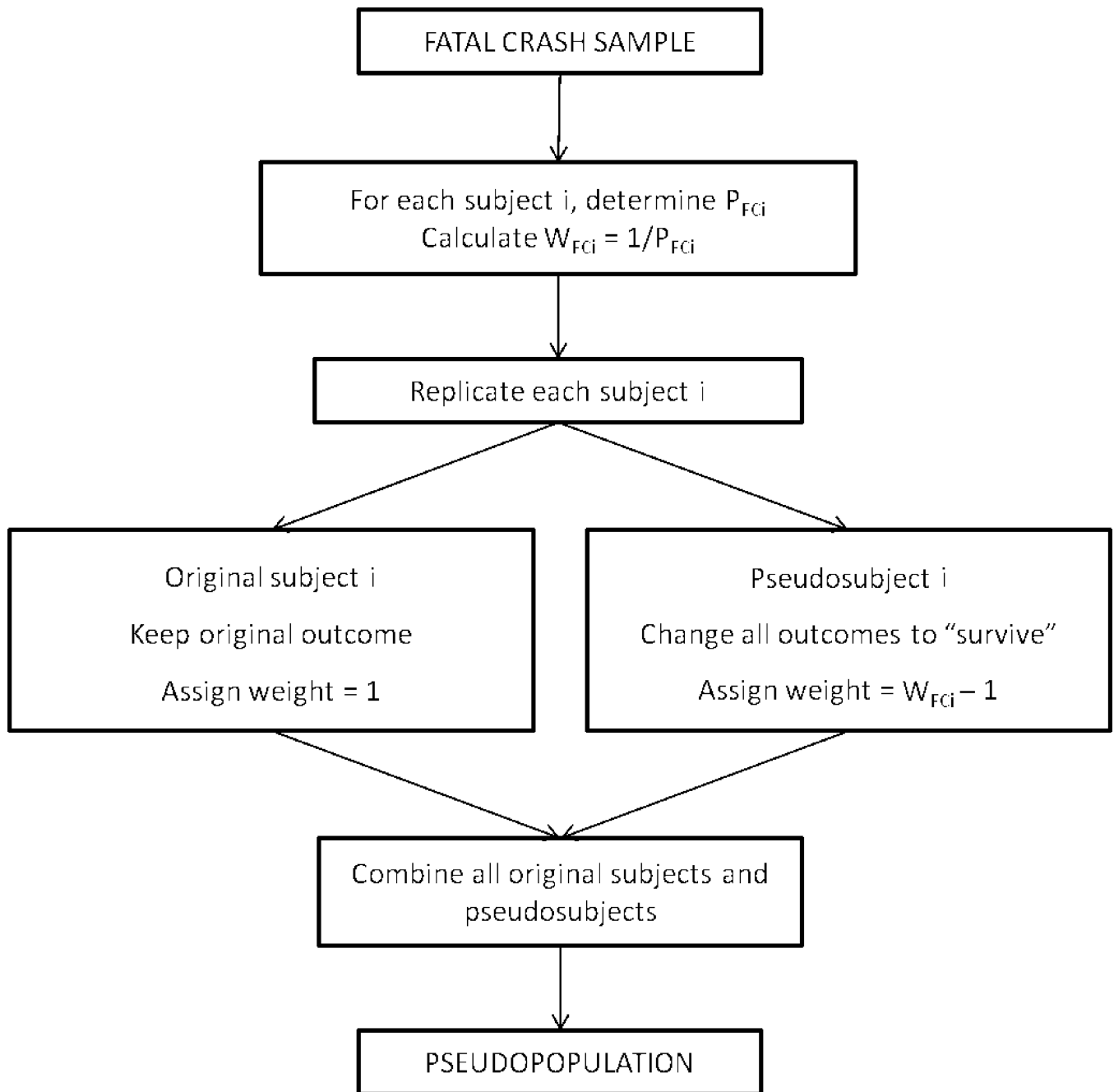
- NHTSA. [Accessed 29 Nov 2011] National Automotive Sampling System (NASS) General Estimates System (GES) Analytical User's Manual 1988–2005. 2005. <http://www-nrd.nhtsa.dot.gov/Pubs/AUM05.pdf>.
- NHTSA. [Accessed 29 Nov 2011] Traffic Safety Facts: 2006 Data. 2008. <http://www-nrd.nhtsa.dot.gov/Pubs/810809.pdf>.
- O'Neill TJ, Barry SC. Truncated logistic regression. *Biometrics*. 1995; 51(2):533–541. [PubMed: 7662842]
- OHIM (Office of Highway Information Management). Highway Performance Monitoring System Reassessment: Final Report, Revised April 1999: USDOT FHWA-PL-99-001. 1999
- Pan Q, Schaubel DE. Evaluating bias correction in weighted proportional hazards regression. *Lifetime Data Analysis*. 2009; 15(1):120–146. [PubMed: 18958616]
- Rao RS, Sigurdson AJ, Doody MM, Graubard BI. An application of a weighting method to adjust for nonresponse in standardized incidence ratio analysis of cohort studies. *Annals of Epidemiology*. 2005; 15(2):129–136. [PubMed: 15652718]
- Roberts G, Rao JNK, Kumar S. Logistic regression analysis of sample survey data. *Biometrika*. 1987; 74(1):1–12.
- Sasidharan L, Donnell ET. Application of propensity scores and potential outcomes to estimate effectiveness of traffic safety countermeasures: Exploratory analysis using intersection lighting data. *Accident Analysis and Prevention*. 2012 in press.
- US Census. [Accessed 29 Nov 2011] 2011. <http://www.census.gov/geo/www/tiger>

- A sample limited to fatal crashes will be severely biased with respect to factors that might affect mortality.
- NASS-GES can be used to estimate the propensity for a subject with given characteristics to be included in a fatal crash sample.
- Weighting a fatal crash sample (e.g., FARS) by the inverse of the derived propensity score allows appropriate adjustment for selection bias.



**Figure 1.**

Derivation of an inverse propensity weight. For a given covariate pattern, each subject has a certain probability  $P_{FC}$  to be involved in a fatal crash, which can be considered a sampling “propensity”. Subjects involved in a fatal crash include a total  $D$  who died and a total  $L$  who lived. The method described in this study assumes that for each person classified as  $D$  or  $L$ , there are theoretically  $((1/P_{FC}) - 1)$  similar persons classified  $X$ , all of whom survived.



**Figure 2.** Creation of a pseudopopulation. Using the inverse propensity score weights, the fatal crash sample can be expanded to a “pseudopopulation” that allows inferences about the theoretical population from which the fatal crash sample is taken.

**Table 1**

Characteristics of injured persons in non-fatal crashes or fatal crashes in NASS-GES for 2002–2008, and FARS for 2007.

	NASS-GES 2002–08 Non-fatal injury crashes (weighted proportions)	NASS-GES 2002–08 Fatal crashes (weighted proportions)	FARS 2007
Sampled persons	336,820	17,737	72,769
Weighted population	18,203,707	374,026	72,769
Sampled crashes	218,419	8,595	37,435
Weighted crash total	12,415,622	193,321	37,435
Person / Vehicle			
Died	0%	56.5%	56.7%
Age 0–14	8.4%	7.9%	7.6%
Age 15–39	54.6%	52.1%	51.9%
Age 40–64	29.4%	29.1%	28.8%
Age 65–79	5.8%	7.6%	7.5%
Age 80+	1.7%	3.1%	4.2%
Male	47.6%	66.5%	66.0%
Pedestrian	4.2%	6.9%	8.1%
Motorcyclist	3.1%	7.1%	11.1%
Unbelted	17.6%	49.6%	44.0%
Ejected	0.7%	15.1%	12.3%
Speed >50MPH	25.3%	60.0%	64.3%
Vehicle rolled	8.5%	28.6%	26.2%
Severe damage	43.6%	77.7%	82.0%
Crash / Location			
>=2 vehicles	69.1%	42.9%	40.8%
Head-on crash	4.8%	12.2%	10.3%
12PM – 6AM	8.4%	20.0%	21.7%
>=2 injured persons	30.8%	48.2%	46.9%
>=3 injured persons	9.9%	22.6%	22.8%
>=4 injured persons	3.6%	10.9%	11.4%
>=5 injured persons	1.4%	5.6%	5.7%
County RUCC 1–3	87.4%	73.9%	69.9%
County RUCC 4–6	9.1%	15.5%	20.4%
County RUCC 7–9	3.6%	10.7%	9.6%

RUCC=Rural-Urban Continuum Code (1=most urban, 9=most rural).

**Table 2**

Odds ratios from weighted logistic regression models predicting whether a crash associated with personal injury in NASS-GES (2002–2008) would also be associated with fatal injury.

Characteristic	Model 1: Including county RUCC	Model 2: Excluding county RUCC
Sampled crashes	227,014	227,014
Weighted crash total	12,608,943	12,608,943
Baseline odds	0.000700	0.000737
	OR (95% CI)	OR (95% CI)
Age 0–14 vs. 15–39	0.71 (0.56–0.91)	0.72 (0.56–0.92)
Age 40–65 vs. 15–39	1.74 (1.53–1.99)	1.75 (1.53–2.00)
Age 65–79 vs. 15–39	3.04 (2.60–3.55)	3.05 (2.62–3.56)
Age 80+ vs. 15–39	3.78 (2.99–4.79)	3.79 (2.94–4.89)
Male	1.70 (1.56–1.84)	1.70 (1.56–1.85)
Pedestrian	20.60 (16.3–26.1)	19.93 (15.3–26.0)
Motorcyclist	2.65 (2.14–3.28)	2.59 (2.05–3.26)
Unbelted	3.14 (2.76–3.57)	3.23 (2.80–3.71)
Ejected	8.31 (7.17–9.63)	8.24 (7.10–9.57)
High speed	2.68 (2.18–3.29)	2.80 (2.45–3.20)
Vehicle rolled	1.52 (1.36–1.70)	1.52 (1.38–1.67)
Severe damage	2.92 (2.59–3.30)	2.96 (2.64–3.32)
>=2 vehicles vs. 1 vehicle	0.88 (0.78–0.99)	0.85 (0.76–0.95)
Head-on crash	2.58 (2.15–3.09)	2.71 (2.36–3.11)
12PM – 6AM	1.73 (1.59–1.88)	1.66 (1.50–1.85)
>=2 injured persons	2.24 (2.03–2.49)	2.25 (2.03–2.50)
>=3 injured persons	1.66 (1.49–1.85)	1.67 (1.50–1.86)
>=4 injured persons	1.29 (1.05–1.60)	1.28 (1.03–1.59)
>=5 injured persons	1.82 (1.41–2.35)	1.80 (1.41–2.29)
RUCC 4–5 vs. RUCC 1–3	1.26 (1.01–1.58)	
RUCC 6–7 vs. RUCC 1–3	1.42 (0.88–2.30)	
RUCC 8–9 vs. RUCC 1–3	1.91 (1.65–2.21)	

OR=Odds Ratio, CI=Confidence Interval, RUCC=Rural-Urban Continuum Code.

**Table 3**

Results of logistic regression models predicting mortality for injured persons.

Characteristic	NASS-GES Fatal crash sample (Original weights)	NASS-GES Fatal crash sample (Adjusted weights)	NASS-GES Population-based (Original weights)	FARS Fatal crash sample (Unweighted)	FARS Fatal crash sample (Adjusted weights)
Sample size	17,737	17,737	354,557	72,768	72,768
Population size	374,026	16,051,138	18,577,733	72,768	2,445,019
Baseline odds	0.352	0.000645	0.000574	0.264	0.000630
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Age 0-14	0.41 (0.34-0.51)	0.65 (0.52-0.82)	0.66 (0.52-0.83)	0.41 (0.38-0.44)	0.71 (0.67-0.76)
Age 40-65	1.93 (1.58-2.36)	1.82 (1.64-2.02)	1.86 (1.63-2.12)	1.67 (1.61-1.74)	1.78 (1.73-1.84)
Age 65-	4.01 (3.05-5.27)	3.80 (3.19-4.52)	4.01 (3.41-4.71)	2.96 (2.77-3.16)	3.57 (3.40-3.75)
Age 80+	4.34 (3.38-5.56)	5.21 (4.13-6.58)	5.74 (4.68-7.04)	3.30 (3.02-3.60)	4.79 (4.49-5.11)
Male	1.29 (1.11-1.49)	1.56 (1.42-1.73)	1.49 (1.38-1.61)	1.36 (1.32-1.41)	1.57 (1.53-1.62)
Pedestrian	18.27 (12.1-27.5)	23.82 (19.4-29.1)	27.81 (23.1-33.5)	18.46 (16.3-20.9)	21.96 (20.6-23.4)
Motorcyclist	5.09 (3.60-7.20)	2.35 (1.96-2.81)	2.42 (1.97-2.98)	3.88 (3.66-4.12)	2.63 (2.52-2.73)
Unbelted	2.52 (1.94-3.28)	4.14 (3.71-4.62)	4.34 (3.57-5.29)	1.92 (1.85-2.00)	3.58 (3.47-3.69)
Ejected	3.50 (2.67-4.59)	10.06 (8.71-11.6)	9.12 (7.62-10.9)	2.92 (2.75-3.10)	9.84 (9.41-10.9)
High speed	1.06 (0.93-1.22)	2.60 (2.34-2.90)	2.66 (2.20-3.22)	1.00 (0.97-1.04)	2.75 (2.68-2.84)
Vehicle rolled	0.84 (0.66-1.06)	1.34 (1.18-1.51)	1.37 (1.22-1.54)	0.97 (0.93-1.01)	1.37 (1.32-1.42)
Severe damage	1.00 (0.81-1.22)	3.21 (2.78-3.71)	3.38 (2.93-3.90)	1.83 (1.73-1.93)	3.53 (3.37-3.71)
Head-on crash	0.74 (0.64-0.86)	2.66 (2.29-3.09)	2.67 (2.21-3.22)	0.77 (0.73-0.80)	2.81 (2.70-2.93)
12PM - 6AM	1.40 (1.20-1.63)	1.70 (1.52-1.89)	1.74 (1.61-1.88)	1.22 (1.17-1.27)	1.69 (1.64-1.75)
RUCC 4-6	1.06 (0.87-1.29)	1.28 (1.11-1.48)	1.25 (0.98-1.60)	1.02 (0.98-1.07)	1.31 (1.27-1.36)
RUCC 7-9	1.03 (0.85-1.24)	1.53 (1.21-1.92)	1.61 (1.29-2.01)	1.07 (1.01-1.13)	1.55 (1.48-1.62)

OR=Odds Ratio, CI=Confidence Interval, RUCC=Rural-Urban Continuum Code.



**Table 4**

Traffic injury and death statistics for counties with available data (Goldstein et al., 2011).

	Counties in US	Counties with data	A: Deaths recorded	B: Injuries recorded	C: Injuries Predicted (Model 1)	D: Injuries Predicted (Model 2)	Observed to predicted mortality (A/B) / (A/D)
Total	3,142	1,752	26,024	1,822,417	1,564,884	1,588,167	0.87
RUCC							
1	414	236	10,560	984,437	792,973	763,678	0.78
2	325	189	5,011	373,409	315,420	300,885	0.81
3	351	200	3,048	189,722	195,378	185,445	0.98
4	218	134	1,983	96,011	83,052	97,775	1.02
5	105	57	482	24,563	22,315	26,209	1.07
6	609	341	2,515	83,747	89,568	117,110	1.40
7	450	254	1,266	42,780	39,809	51,922	1.21
8	235	131	519	12,864	12,415	21,342	1.66
9	435	210	640	14,884	13,954	23,801	1.60

RUCC = Rural-Urban Continuum Code (1=most urban, 9=most rural).