Accident Analysis and Prevention 45 (2012) 180–186

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999–2008

Brian Christopher Tefft*

AAA Foundation for Traffic Safety, 607 14th Street NW, Suite 201, Washington, DC 20005, United States

ARTICLE INFO

Article history:
Received 21 December 2010
Received in revised form 5 March 2011
Accepted 24 May 2011

Keywords:
Accidents
Traffic
Drowsiness
Imputation

ABSTRACT

The proportion of motor vehicle crashes that involve a drowsy driver likely is greater than existing crash databases reflect, due to the possibility that some drivers whose pre-crash state of attention was unknown may have been drowsy. This study estimated the proportion of crashes that involved a drowsy driver in a representative sample of 47,597 crashes in the United States from 1999 through 2008 that involved a passenger vehicle that was towed from the scene. Multiple imputation was used to address missing data on driver drowsiness. In the original (non-imputed) data, 3.9% of all crashes, 7.7% of non-fatal crashes that resulted in hospital admission, and 3.6% of fatal crashes involved a driver coded as drowsy; however, the drowsiness status of 45% of drivers was unknown. In the imputed data, an estimated 7.0% of all crashes (95% confidence interval: 4.6%, 9.3%), 13.1% of non-fatal crashes that resulted in hospital admission (95% confidence interval: 8.8%, 17.3%), and 16.5% of fatal crashes (95% confidence interval: 12.5%, 20.6%) involved a drowsy driver. Results suggest that the prevalence of fatal crashes that involve a drowsy driver is over 350% greater than has been reported previously.

1. Introduction

Operator drowsiness, sleepiness, or fatigue (hereafter referred to as drowsiness) has been documented as a causal or contributing factor in aviation, maritime, and trucking accidents (National Transportation Safety Board, 1990, 1991, 1994). However, estimates of the proportion of motor vehicle crashes that involve drowsy drivers vary widely depending upon sources of data and analytical methods used.

Knipling and Wang (1994) analyzed data from years 1989 through 1993 and reported that 0.9% of all police-reported crashes and 3.6% of fatal crashes in the United States involved a driver coded as drowsy. The authors cited several other studies of large crash databases that reported results in this range, but also noted that studies of crash databases were likely to underestimate the prevalence of crashes that involved drowsy drivers due to data limitations. Wang et al. (1996) analyzed a national sample of crashes that occurred in 1995 in which a passenger vehicle was towed and estimated that 2.6% of these crashes involved a drowsy driver. The authors noted that the role of drowsiness was unknown in 46% of crashes and that their estimate of the proportion of crashes involving drowsy drivers was likely conservative.

An Australian study classified crashes as drowsiness-related if drowsiness was cited by the police or if the crash involved departure from the roadway in the absence of other causes or contributing factors suggestive of attentive driving, and estimated that 6% of all reported crashes and 15% of fatal crashes in 1992 in New South Wales involved a drowsy driver (Fell, 1994). Knipling and Wang (1995) used a similar method to refine their earlier estimates (1994), and estimated that 1.2%–1.6% of all reported crashes involved a drowsy driver. Masten et al. (2006) used data from a sample of crashes in North Carolina to develop a statistical model to classify crash-involved drivers as drowsy or not drowsy, applied this model to national data on fatal crashes, and estimated that 15–33% of drivers involved in fatal crashes nationwide from 2001 through 2003 were drowsy.

In a study in which 109 vehicles were equipped with cameras and other data collection equipment for a period of 12–13 months, Klauer et al. (2006) reviewed pre-event video and estimated that 22% of crashes and near crashes in the study population involved moderate to severe drowsiness. However, the majority of outcomes in the study population were near crashes and unreported minor crashes (Dingus et al., 2006); the extent to which these results may be generalized to more severe crashes is unknown.

The aim of the current study was to improve upon past estimates of the proportion of crashes that involve a drowsy driver, overall and relation to crash severity, using multiple imputation to address missing data on driver drowsiness.

Abbreviations: CDS, Crashworthiness Data System; CI, Confidence interval; GES, General Estimates System; FARS, Fatality Analysis Reporting System; NASS, National Automotive Sampling System; NHTSA, National Highway Traffic Safety Administration.

* Tel.: +1 2026385944; fax: +1 2026385943.
E-mail address: bteff t@aaafoundation.org

0001-4575/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2011.05.028

2. Methods

2.1. Data

The data used here was obtained from the National Highway Traffic Safety Administration (NHTSA)'s National Automotive Sampling System (NASS) Crashworthiness Data System (CDS) (2009), a sample of police-reported crashes in which a passenger vehicle (car, light truck, utility vehicle, or van) was towed. Data reported in NASS CDS is compiled by teams of investigators who obtain information from police crash reports, medical records, crash reconstructions, and interviews with drivers and passengers involved in crashes. Records are weighted to produce nationally representative estimates (NHTSA, 2009).

NASS CDS data files from years 1999 through 2008 were analyzed using statistical software Stata (StataCorp, 2007). Vehicles coded as having no driver, vehicles not eligible for investigation in NASS CDS (large trucks, motorcycles, construction equipment, etc.), and vehicles of unknown type were excluded. The final data set contained records of 80,821 vehicles involved in 47,597 crashes.

2.2. Main outcome measure

The main outcome measure was driver drowsiness. Information regarding drowsiness was obtained from NASS CDS variable DRIVDIST, which NASS CDS uses to classify a driver's attention to driving. This variable is coded on the basis of information from police reports and from interviews conducted by NASS CDS investigators with crash-involved drivers and passengers (NHTSA, 2009b). Values of DRIVDIST include: attentive or not distracted, looked but did not see, sleepy or fell asleep, distracted (13 separate codes are used for specific distractions), and unknown. To avoid having categories with small cell sizes in the analysis in the current study, the variable DRIVDIST was collapsed into the following categories: attentive/not distracted; looked but did not see; distracted by other occupant; distracted by outside person, object, or event; distracted by secondary task (e.g., cell phone, radio, climate control, eating); other/unknown distraction; and sleepy or fell asleep (hereafter *drowsy*). This collapsed variable is referred to hereafter as *attention*.

2.3. Missing data

The variable DRIVDIST was coded as unknown for 45% of all drivers, 42% of drivers involved in non-fatal crashes that resulted in hospitalization, and 73% of drivers involved in fatal crashes. Multiple imputation (Rubin, 1987) was used to estimate the proportion of these drivers who were drowsy. Imputation was performed using the method of chained equations (Van Buuren et al., 1999) implemented in Stata (Royston, 2004, 2005, 2009). Imputation was performed at the driver level. Variables that were strongly associated with drowsiness or with missing values of attention, variables that were to be used in subsequent analysis of the imputed data, and variables that were needed to reflect the NASS CDS sample design for accurate variance estimation were included in the imputation model (Table 1). The data were weighted using the NASS CDS record weights.

Other variables in the imputation model were imputed when they were missing in the original data. Imputation was not performed on records in which variables considered critical to the determination of drowsiness (e.g., time of day, crash type) or variables with several categories (e.g., trafficway flow, surface condition) had missing values; missing values of attention were not imputed in 693 records (1.0% of weighted drivers) for these reasons. Attention also was not imputed for drivers whose injury severity was coded as *fatal-ruled disease* because of ambiguity regarding whether these drivers died prior to the crash or subsequently ($n = 116$, 0.02% of weighted drivers).

Imputation was performed in two stages. All variables except attention were imputed in the first stage. After the first stage, maximum injury severity in crash – the only imputed variable that was coded at the crash level – had different imputed values for different drivers in the same crash in 269 crashes (0.9% of weighted crashes). When this occurred, the value imputed for the driver in the same crash who had the fewest total missing values was copied to the records of all other drivers in that crash. Attention was imputed in the second stage. Ten independent imputations were performed. Imputation of attention was performed separately for crashes that occurred during the hours 11:00 PM–6:59 AM vs. 7:00 AM–10:59 PM, and for crashes that resulted in hospitalization or death vs. those that did not. Only 238 of 2768 records of fatally injured drivers contained non-missing values for attention. Due to concern that any potential misclassification of attention among these drivers could severely bias the imputed results with respect to fatally injured drivers, the injury severity variable used in imputation was collapsed into three categories: not injured/not treated, treated for injury/not hospitalized, and hospitalized/fatal. This is equivalent to assuming that crashes resulting in death are a random sample of crashes resulting in hospitalization conditional upon the other variables in the imputation model. Deaths were

Table 1
Variables used in imputation of driver attention (derived from National Automotive Sampling System Crashworthiness Data System variable[a] in parentheses).

Variable name	Values
Attention (DRIVDIST)	Attentive/not distracted; looked but did not see; distracted by other occupant; distracted by outside person, object, or event; distracted by secondary task; other/unknown distraction; drowsy
Maximum injury severity in crash (ATREAT); driver injury severity (TREATMNT)	Not injured/not treated; treated for injury/not hospitalized; hospitalized/fatal
Number of vehicles in crash (VEHFORMS)	1; 2 or more
Pre-event maneuver (PREMOVE)	Going straight; other active driving maneuver
Crash type (ACCTYPE)	Road departure; other
Day of week (DAYWEEK)	Sunday; Monday; Tuesday; Wednesday; Thursday; Friday; Saturday
Hour of day (TIME)	1–2:59 AM; 3–4:59 AM; ...; 9–10:59 PM; 11 PM–12:59 AM
Trafficway flow (TRAFFLOW)	Undivided; divided w/o barrier; divided with barrier; one-way
Number of passengers (OCUPANTS)	0; 1 or more
Driver age (years) (AGE)	16–24; 25–39; 40–59; 60+
Driver sex (SEX)	Female; male
Light condition (LGTCOND)	Daylight; dawn/dusk; dark-lighted; dark
Relation to intersection/junction (RELINTER)	Intersection related; not intersection related
Roadway alignment (ALIGNMNT)	Curve; tangent
Speed limit (miles per hour) (SPLIMIT)	<30; 30–35; 40–45; 50–55; 60+
Number of travel lanes (LANES)	1; 2; 3; 4+
Surface condition (SURCOND)	Dry; not dry (wet/ice/slush)
Pre-crash critical event (PREEVENT)	Departed travel lane; other
Vehicle disposition (TOWPAR)	Not towed; towed
Year (YEAR)	Binary indicator for each year
Stratum (PSUSTRAT)	Binary indicator for each stratum
Primary sampling unit (PSU)	Binary indicator for each PSU (nested within strata)

[a] For definitions and coding of variables, see National Automotive Sampling System Crashworthiness Data System 2008 Coding and Editing Manual. Washington, DC: National Highway Traffic Safety Administration; 2009.