Accident Analysis and Prevention 45 (2012) 646–653

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

# Analysis of crash severities using nested logit model—Accounting for the underreporting of crashes

Sunil Patil[a,*], Srinivas Reddy Geedipally[b], Dominique Lord[c]

[a] RAND Europe, Westbrook Center, Milton Road, Cambridge – CB4 1YG, UK
[b] Center for Transportation Safety, Texas Transportation Institute – The Texas A&M University System, College Station, TX – 77843-3135, USA
[c] Zachry Department of Civil Engineering, Texas A&M University, College Station, TX – 77843, USA

ARTICLE INFO

*Article history:*
Received 16 July 2011
Received in revised form 22 August 2011
Accepted 18 September 2011

*Keywords:*
Crash injury severity
Nested logit
Underreporting

ABSTRACT

Recent studies in the area of highway safety have demonstrated the usefulness of logit models for modeling crash injury severities. Use of these models enables one to identify and quantify the effects of factors that contribute to certain levels of severity. Most often, these models are estimated assuming equal probability of the occurrence for each injury severity level in the data. However, traffic crash data are generally characterized by underreporting, especially when crashes result in lower injury severity. Thus, the sample used for an analysis is often outcome-based, which can result in a biased estimation of model parameters. This is more of a problem when a nested logit model specification is used instead of a multinomial logit model and when true shares of the outcomes-injury severity levels in the population are not known (which is almost always the case). This study demonstrates an application of a recently proposed weighted conditional maximum likelihood estimator in tackling the problem of underreporting of crashes when using a nested logit model for crash severity analyses.



## 1. Introduction

Vehicle crash is one of the most common causes of death and injury around the world. This has resulted in the proclamation of the current decade (2011–2020) as the *Decade of Action for Road Safety* by the United Nations (United Nations, 2010). Highway safety professionals often aim at reducing the number and associated severity of traffic crashes through the analysis of existing reported data. Studies focusing on analyzing traffic crash severities aim at identifying and quantifying the effects of the factors, which affect different crash injury severities. Recently, researchers have used logit models for this kind of analyses.

Crash injury severities are often classified as follows: fatal, incapacitating, non-incapacitating, slight or possible injury and no injury/property damage only. The data used in these studies are often collected by the police from reported crashes. However, less severe crashes are often underreported due to various reasons which include avoidance of reporting by the driver(s) involved (Hauer and Hakkert, 1988; Elvik and Mysen, 1999; Blincoe et al., 2002). Thus, studies, which assume a random sampling strategy, are likely to result in producing biased results of parameter estimation (Savolainen et al., 2011). While sampling bias can be corrected when the population shares of these severity levels are known, it is very rare for crash data to obtain these true shares.

Recently, Bierlaire et al. (2008) have proposed a new estimator, which can account for outcome (choice)-based sampling when the population shares of the outcomes are unknown. This estimator is proposed for the model structures, such as the nested logit (NL) model, which belong to the wider generalized extreme value (GEV) family. The main objective of our study is to investigate if this estimator can be useful in addressing the crash underreporting problem when a NL model is used for crash injury severity analysis.

In the next section, we present a review of relevant literature. This is followed by the sections on methodology, details of data, results, and conclusions.

## 2. Literature review

Various model structures have been used to model crash injury severities. A detailed review and assessment of these models is presented in a recent paper (Savolainen et al., 2011). Many of the studies including those carried out by Shankar et al. (1996), Chang and Mannering (1998), Chang and Mannering (1999), Lee and Mannering (2002), Abdel-Aty and Abdelwahab (2004), Holdridge et al. (2005), Savolainen and Mannering (2007), Haleem and Abdel-Aty (2010), and Hu and Donnell (2010) have used the NL model specification to analyze the crash injury severity data.

* Corresponding author. Tel.: +44 1223 353 329; fax: +44 1223 358 845.
*E-mail addresses:* spatil@rand.org, patilnsunil@gmail.com (S. Patil), srinivas-g@ttimail.tamu.edu (S.R. Geedipally), d-lord@tamu.edu (D. Lord).

0001-4575/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2011.09.034