Analytic Methods in Accident Research 1 (2014) 1–22

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Analytic methods in accident research: Methodological frontier and future directions

Fred L. Mannering [a,*], Chandra R. Bhat [b]

[a] School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, USA
[b] Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 301 E. Dean Keeton St. Stop C1761, Austin, TX 78712, USA

A R T I C L E I N F O

Article history:
Received 14 August 2013
Received in revised form
4 September 2013
Accepted 5 September 2013
Available online 18 October 2013

Keywords:
Highway safety
Crash frequency
Crash severity
Econometric methods
Statistical methods
Accident analysis

A B S T R A C T

The analysis of highway-crash data has long been used as a basis for influencing highway and vehicle designs, as well as directing and implementing a wide variety of regulatory policies aimed at improving safety. And, over time there has been a steady improvement in statistical methodologies that have enabled safety researchers to extract more information from crash databases to guide a wide array of safety design and policy improvements. In spite of the progress made over the years, important methodological barriers remain in the statistical analysis of crash data and this, along with the availability of many new data sources, present safety researchers with formidable future challenges, but also exciting future opportunities. This paper provides guidance in defining these challenges and opportunities by first reviewing the evolution of methodological applications and available data in highway-accident research. Based on this review, fruitful directions for future methodological developments are identified and the role that new data sources will play in defining these directions is discussed. It is shown that new methodologies that address complex issues relating to unobserved heterogeneity, endogeneity, risk compensation, spatial and temporal correlations, and more, have the potential to significantly expand our understanding of the many factors that affect the likelihood and severity (in terms of personal injury) of highway crashes. This in turn can lead to more effective safety countermeasures that can substantially reduce highway-related injuries and fatalities.

1. Introduction

Worldwide, more than 1.2 million people die annually in highway-related crashes and as many as 50 million more are injured and, by 2030, highway-related crashes are projected to be the 5th leading cause of death in the world (World Health Organization, 2009, 2013). In addition to the statistics on death and injuries, highway-related crashes result in immeasurable pain and suffering and many billions of dollars in medical expenses and lost productivity. The enormity of the impact of highway safety on human societies has resulted in massive expenditures on safety-related countermeasures, laws governing highway use, and numerous regulations concerning the manufacturing of highway vehicles. While the success of many of these efforts in reducing the likelihood of highway crashes and mitigating their impact cannot be denied, the toll that highway crashes continue to extract on humanity is clearly unacceptable.

* Corresponding author. Tel.: +1 765 496 7913.
E-mail addresses: flm@ecn.purdue.edu, flm@purdue.edu (F.L. Mannering), bhat@mail.utexas.edu (C.R. Bhat).

2213-6657/$ - see front matter
http://dx.doi.org/10.1016/j.amar.2013.09.001

Critical to the guidance of ongoing efforts to improve highway safety is research dealing with the statistical analysis of the countless terabytes of highway-crash data that are collected worldwide every year. The statistical analysis of these crash data has historically been used as a basis for developing road-safety policies that have saved lives and reduced the severity of injuries. And, while the quality of data has not always progressed as quickly as many safety researchers would have liked, the continual advance in statistical methodologies has enabled researchers to extract more and more information from existing data sources.

With this said, as in most scientific fields, a dichotomy has evolved between what is used in practice and what is used by front-line safety researchers, with the methodological sophistication of some of the more advanced statistical research on roadway accidents having moved well beyond what can be practically implemented to guide safety policy. However, it is important that the large and growing methodological gap between what is being used in practice and what is being used in front-line research not be used as an excuse to slow the methodological advances being made, because the continued development and use of sophisticated statistical methodologies provides important new inferences and ways of looking at the underlying causes of highway-crashes and their resulting injury severities. Continuing methodological advances, in time, will undoubtedly help guide and improve the practical application of statistical methods that will influence highway-safety policy. Thus, while the intent of this paper is to focus on the current frontier of methodological research (after reviewing current methodological issues), it is important that readers recognize the different objectives between applied and more fundamental research, and the role that sophisticated methodological applications have in ultimately improving safety practice and developing effective safety policies.

The current paper begins by quickly reviewing traditional sources of highway-accident data (Section 2) and the evolution of statistical methods used to analyze these data (Section 3). It then moves on to present some critical methodological issues relating to the analysis of highway-accident data (Section 4). This is followed by a discussion of some emerging sources of crash data that have the potential to significantly change methodological needs in the safety-research field (Section 5). The paper concludes with a discussion of some of the more promising methodological directions in accident research (Section 6), and a summary and insights for the future methodological innovation in accident research (Section 7).

2. Traditional highway crash data

Most existing highway-accident studies have extracted their data from police crash reports. These reports are used to establish the frequency of crashes at specific locations and the associated injury-severities of vehicle occupants and others involved in these crashes. In the U.S., common injury severities are assessed by police officers at the scene of the crash such as no injury, possible injury, evident injury, disabling injury, fatality (within 30 days of the crash).[1] Police-reported data also include a great deal of information that can serve as explanatory variables in modeling injury-severity outcomes, including information on time of day, age and gender of vehicle occupants, road-surface conditions, weather conditions, possible contributing factors to the crash, roadway type, roadway lighting, speed limits, basic roadway geometrics (curve, grade, etc.), type of crash (rollover, rear end, etc.) type of object(s) struck, driver sobriety, safety belt usage, airbag deployment, and so on. This information can be quickly expanded further by linking the data with government-provided roadway information (including traffic volumes, pavement friction, detailed roadway geometric characteristics, traffic-signal details) and detailed weather-related data (including temperature ranges, specific precipitation types and accumulations).

While the occurrence of a crash and the severity levels reported by police data have been used in many previous studies to provide insights relating to the factors affect highway safety, the inaccuracies of police-reported data are well documented. For example, it has been well established in the literature that less severe crashes are less likely to be reported to police and thus less likely to appear in police databases (Yamamoto et al., 2008; Ye and Lord, 2011). With regard to the severity of crashes, considerable inaccuracies have been found when comparing police severity reports with the severity assessment made by medical staff at the time of admission to the hospital (Compton, 2005; McDonald et al., 2009; Tsui et al., 2009). Also, with regard to traditional police data, a study by Shin et al. (2009), showed that the medical costs associated with the “no injury” compared to the “evident injury” severity categories were higher due to subsequent hospital admissions (injuries sustained were not reported or observed at the scene). Despite the limitations of traditional crash data (such as police-reported data), these data have supported countless research efforts that have attempted to improve our understanding of the factors that influence the occurrence of crashes and the personal injuries that result. A wide variety of methodological approaches have been used to explore traditional crash data, and these methodologies have become increasingly sophisticated over time as researchers seek to address the many less obvious characteristics of the data in the hope of uncovering important new inferences relating to highway safety.

[1] Other types of injury-severity measurement data that have been used include the Abbreviated Injury Scale (AIS) which was originally developed by the American Association for Automotive Medicine, the Organ Injury Scales (OIS) proposed by the American Association for the Surgery of Trauma and the Injury Severity Score (ISS) used by hospitals.

3. Evolution of methodological approaches in accident research

Two relatively recently published papers provide a comprehensive review of current methodological approaches for studying crash frequencies, the number of crashes on a roadway segment or intersection over some specified time period (Lord and Mannering, 2010), and crash severities, usually measured by the most severely injured person involved in the crash (Savolainen et al., 2011). The intent of this paper is not to replicate the detailed discussions of the methodological alternatives provided in those papers, but instead to focus on discussing the methodological evolution, the current methodological frontier and remaining methodological issues (the interested reader is referred to those papers for a review of previously used methodological approaches). However, because several important methodological developments and applications have been undertaken since those previous review papers were published, Tables 1 and 2 are provided to give an update of the literature (by methodological-approach category) previously presented in Lord and Mannering (2010) and Savolainen et al. (2011) (please see those papers, if necessary, for detailed descriptions of the methodological approaches listed in these tables). Tables 1 and 2 list the methodological approaches in the approximate chronological order that they have first appeared in the accident-research literature.

With regard to the evolution of methodological alternatives in accident research, the frequency of crashes has been studied with a wide variety of methods over the years. Because crash frequencies (the number of crashes occurring on a roadway entity over some time period) are count data (non-negative integers), the Poisson regression approach to count data has served as a basis for some initial research efforts that have sought to determine factors that influence crash frequencies so that effective crash-mitigation designs and policies could be determined. As research progressed, the limitations of the simple Poisson regression model quickly became obvious and Poisson variants became the dominant methodological approach. For example, the negative binomial model (or Poisson–Gamma) became widely used because it can handle overdispersed data (data where the mean of the frequencies is much greater than the variance, see Lord and Mannering, 2010). And, because crash-frequency data bases were often found to have many observations with no observed crashes, researchers considered zero-inflated Poisson and negative binomial regressions, which attempt to account for the preponderance of zeros by splitting roadways into two separate states, a zero state and a normal count state. Similarly, a variety of other count-data models and variations have also been considered over the years including the Gamma model, Conway–Maxwell–Poisson model, the negative binomial-Lindley model, and so on. Still other work has looked at crashes not as count data per se, but instead as the duration of time between crashes (duration models), which in turn can be used to generate crash frequencies over specified time periods. Recently, a series of studies (see Castro et al., 2012; Narayanamoorthy et al., 2013; Bhat et al., 2014) have recast count models as a restrictive case of a generalized ordered-response model, with a latent long-term risk propensity for crashes coupled with thresholds that determine the translation of that risk to the instantaneous probability of a crash outcome. Such a generalized ordered-response approach to count data has several potential advantages, including making it much easier to extend univariate count models to multivariate count models and accommodating spatial and temporal dynamics.

Other methodological advances models have sought to address what might be considered as more subtle issues with crash-frequency data. Issues such as the effect of unobserved factors on crash frequencies, spatial and temporal correlations among crash-count data, the possibility of roadway segments shifting among multiple crash states (discrete crash situations (states) that fundamentally shift roadway safety, and others) have all been addressed in the steady progression of methodological advances in the field.

A similar path has been followed by studies that have addressed the severity of crashes (see Table 2). Starting with simple binary discrete outcome models such as binary logit and probit models, models evolved to consider multiple discrete outcomes (to consider a variety of injury-severity categories such as no injury, possible injury, evident injury, disabling injury and fatality). For the multiple discrete outcome models, multinomial models that do not account for the ordering of injury outcome (that is, from no-injury to fatality) have been widely applied from the simple multinomial logit model, to the nested logit model, and to the random parameters logit model (which can account for the effect of unobserved factors across crash observations). Modeling approaches that do consider the ordering of injury severities, such as the ordered probit and logit model, have also been applied with increasingly sophisticated forms to overcome possible restrictions imposed by traditional ordered-modeling approaches. Also, as with count-data models, crash-severity models have been extended to consider the existence of multiple crash-severity states (discrete crash situations that fundamentally shift injury severity) and unobserved differences in injury severity outcomes across the population using finite-mixture/latent-class approaches (see Table 2).[2]

4. Some important ongoing methodological considerations

In spite of the steady progression of methodological innovation in the crash analysis field, as reflected in the papers presented in Tables 1 and 2, there remain many fundamental issues that have not been completely addressed or are often

[2] Most crash-severity models are based on data that are conditional on a crash having occurred. This permits the use of detailed crash data including the age and physical characteristics of people involved in the crash, the possible deployment of airbags, and so on. However, there have also been efforts to model crash frequencies and severities simultaneously (these efforts have been led by the bivariate/multivariate research efforts listed in Table 1), although these approaches cannot use the detailed post-crash data that is available in an injury-severity model that is conditioned on the crash having occurred.

Table 1
Summary of previous research analyzing crash-frequency data[a].

Methodological approach	Previous research
Poisson regression model	Gustavsson and Svensson (1976), Joshua and Garber (1990), Jones et al. (1991), Miaou and Lum (1993), Miaou (1994), Kumara and Chin (2005), Ma (2009), Ye et al. (2013), Li et al. (2013)
Negative binomial/Poisson–gamma models	Maycock and Hall (1984), Brüde and Larsson (1993), Bonneson and McCoy (1993), Miaou (1994), Kumala (1995), Shankar et al. (1995), Poch and Mannering (1996), Maher and Summersgill (1996), Mountain et al. (1996, 1998), Milton and Mannering (1998), Brüde et al. (1998), Karlaftis and Tarko (1998), Persaud and Nguyen (1998), Turner and Nicholson (1998), Heydecker and Wu (2001), Carson and Mannering (2001), Miaou and Lord (2003), Amoros et al. (2003), Hirst et al. (2004), Abbas (2004), Lord et al. (2005a), El-Basyouny and Sayed (2006), Lord (2006), Kim and Washington (2006), Lord and Mahlawat (2009), Malyshkina and Mannering (2010b), Daniels et al. (2010), Cafiso et al. (2010), Geedipally and Lord (2010), Lao et al. (2011b), Geedipally and Lord (2011), Lord and Kuo (2012), Meng and Qu (2012), Park et al. (2012), Viera Gomes et al. (2012), Pirdavani et al. (2013), Ye et al. (2013)
Duration models	Jovanis and Chang (1989), Chang and Jovanis (1990), Mannering (1993), Chung (2010), Jovanovic et al. (2011)
Bivariate/multivariate models	Maher (1990), Miaou and Lord (2003), Miaou and Song (2005), Bijleveld (2005), Song et al. (2006), Ma and Kockelman (2006), Park and Lord (2007), Bonneson and Pratt (2008), Geedipally and Lord (2010), Ma et al. (2008), Depaire et al. (2008), Ye et al. (2009), Aguero-Valverde and Jovanis (2009), El-Basyouny and Sayed (2009a), Park et al. (2010a); Wang et al. (2011), Lao et al. (2011a), Pei et al. (2011), Anastasopoulos et al. (2012a), Chiou and Fu (2013), Caliendo et al., (2013), Yu and Abdel-Aty (2013b), Castro et al. (2012), Narayanamoorthy et al. (2013)
Zero-inflated Poisson and negative binomial models	Miaou (1994), Shankar et al. (1997, 2003), Carson and Mannering (2001), Lee and Mannering (2002), Kumara and Chin (2003), Qin et al. (2004), Lord et al. (2005b, 2007), Malyshkina and Mannering (2010a)
Random effects models, spatial and temporal correlation models	Johansson (1996), Shankar et al. (1998), Miaou and Lord (2003), Flahaut et al. (2003), MacNab (2004), Miaou et al. (2003, 2005), Wang and Abdel-Aty (2006), Aguero-Valverde and Jovanis (2006, 2008), Li et al. (2008a), Quddus (2008), Guo et al. (2010), Aguero-Valverde and Jovanis (2010), Mitra and Washington (2012), Castro et al. (2012), Narayanamoorthy et al. (2013), Aguero-Valverde (2013), Mohammadi and Samaranayake (in preparation), Xie et al. (in preparation)
Generalized estimating equation models	Lord and Persaud (2000), Lord et al. (2005a), Wang and Abdel-Aty (2008), Lord and Mahlawat (2009)
Neural network, Bayesian Neural network, and vector machine models	Abdelwahab and Abdel-Aty (2001), Chang (2005), Riviere et al. (2006), Xie et al. (2007), Li et al. (2008b), Yu and Abdel-Aty (2013c)
Hierarchical/multilevel models	Jones and Jørgensen (2003), Kim et al. (2007a), Aguero-Valverde and Jovanis (2010), Ahmed et al. (2011), Usman et al. (2012), Yu et al. (2013), Deublein et al. (2013), Yu and Abdel-Aty (2013a, 2013b)
Negative multinomial model	Ulfarsson and Shankar (2003), Hauer (2004), Caliendo et al. (2007)
Poisson-lognormal and Poisson–Weibull models	Miaou et al. (2005), Lord and Miranda-Moreno (2008), Aguero-Valverde and Jovanis (2008), Cheng et al. (2013)
Gamma model	Oh et al. (2006), Daniels et al. (2010)
Conway–Maxwell–Poisson model	Lord et al. (2008), Sellers and Shmueli (2010), Lord et al. (2010), Geedipally and Lord (2011), Giuffre et al. (2011), Francis et al. (2012), Lord and Guikema (2012)
Censored regression models	Anastasopoulos et al. (2008, 2012a, 2012b)
Generalized additive models	Xie and Zhang (2008), Li et al. (2009)
Random parameters count models	Anastasopoulos and Mannering (2009), El-Basyouny and Sayed (2009b), Granowski and Manner (2011),Venkataraman et al. (2011, 2013, in preparation), Ukkusuri et al. (2011), Mitra and Washington (2012), Wu et al. (2013), Bullough et al. (2013), Castro et al., 2012, Narayanamoorthy et al. (2013), Bhat et al. (2014), Chen and Tarko (this issue)
Finite-mixture/latent-class and Markov switching models	Malyshkina et al. (2009), Park and Lord (2009), Malyshkina and Mannering (2010a), Park et al. (2010b), Peng and Lord (2011), Zou et al. (2013), Zou et al. (2014)
Negative binomial-Lindley model	Lord and Geedipally (2011), Geedipally et al. (2012)
Count model recast as a generalized ordered-response system	Castro et al. (2012), Narayanamoorthy et al. (2013), Bhat et al. (2014)

[a] *Source*: Updated from Lord and Mannering (2010).

Table 2
Summary of previous research analyzing crash-injury severities[a].

Methodological approaches	Previous research
Binary logit/probit models	Shibata and Fukuda (1994), Farmer et al. (1997), Khattak et al. (1998), Krull et al. (2000), Al-Ghamdi (2002), Bedard et al. (2002), Toy and Hammitt (2003), Ballasteros et al. (2004), Chang and Yeh (2006), Sze and Wong (2007), Lee and Abdel-Aty (2008), Pai (2008), Rifaat and Tay (2009), Haleem and Abdel-Aty (2010), Peek-Asa et al. (2010), Kononen et al. (2011), Moudon et al. (2011), Santolino et al. (2012)
Multinomial logit models	Shankar and Mannering (1996), Carson and Mannering (2001), Abdel-Aty and Abdelwahab (2004), Ulfarsson and Mannering (2004), Khorashadi et al. (2005), Islam and Mannering (2006), Kim et al. (2007b); Malyshkina and Mannering (2008), Savolainen and Ghosh (2008), Schneider et al. (2009), Malyshkina and Mannering (2010a, 2010b), Rifaat et al. (2011), Ye and Lord (2011), Schneider and Savolainen (2011), Eluru (2013), Yasmin and Eluru (2013), Ye and Lord (2014)
Nested logit models	Shankar et al. (1996), Chang and Mannering (1998, 1999), Lee and Mannering (2002), Abdel-Aty and Abdelwahab (2004), Holdridge et al. (2005), Savolainen and Mannering (2007), Haleem and Abdel-Aty (2010), Hu and Donnell (2010), Patil et al. (2012), Wu et al. (2013); Yasmin and Eluru (2013)
Sequential logit/probit models	Saccomanno et al. (1996), Dissanayake and Lu (2002a, 2002b), Helai et al. (2008), Yamamoto et al. (2008), Jung et al. (2010), Xu et al. (2013)
Heteroskedastic ordered logit/probit models	O'Donnell and Connor (1996), Wang and Kockelman (2005), Lemp et al. (2011)
Ordered logit/probit models	Khattak et al. (1998, 2002), Klop and Khattak (1999), Renski et al. (1999), Khattak (2001), Kockelman and Kweon (2002), Quddus et al. (2002), Abdel-Aty (2003), Austin and Faigin (2003), Kweon and Kockelman (2003), Zajac and Ivan (2003), Khattak and Rocha (2003), Yamamoto and Shankar (2004), Donnell and Mason (2004), Khattak and Targa (2004), Abdel-Aty and Keller (2005), Lee and Abdel-Aty (2005), Shimamura et al. (2005), Garder (2006), Lu et al. (2006), Oh (2006), Siddiqui et al. (2006), Pai and Saleh (2007), Das et al. (2008), Gray et al. (2008), Wang and Abdel-Aty (2008), Chimba and Sando (2009), Wang et al. (2009); Pai (2009), Xie et al. (2009), Haleem and Abdel-Aty (2010), Jung et al. (2010), Quddus et al. (2010), Ye and Lord (2011), Zhu and Srinivasan (2011), Ferreira and Couto (2012), Abay (2013a), Jiang et al. (2013a, 2013b), Eluru (2013), Mergia et al. (2013), Yasmin and Eluru (2013), Ye and Lord (2014)
Log-linear models	Chen and Jovanis (2000)
Generalized ordered outcome models	Srinivasan (2002), Eluru et al. (2008), Quddus et al. (2010), Castro et al. (2013), Eluru (2013), Abay et al. (2013), Yasmin and Eluru (2013), Yasmin et al. (2014)
Simultaneous binary logit model	Ouyang et al. (2002)
Bivariate/multivariate binary probit models	Winston et al. (2006), Lee and Abdel-Aty (2008)
Bivariate/multivariate ordered probit models	Yamamoto and Shankar (2004), de Lapparent (2008), Eluru et al. (2010), Rana et al. (2010), Abay et al. (2013), Chiou et al. (2013a), Yasmin et al. (2013), Russo et al. (in preparation)
Artificial neural networks	Abdelwahab and Abdel-Aty (2001), Delen et al. (2006), Chimba and Sando (2009)
Mixed joint binary ordered logit model	Eluru and Bhat (2007)
Mixed logit model (random parameters logit model)	Milton et al. (2008), Kim et al. (2008, 2010, 2013), Malyshkina and Mannering (2010b), Kim et al. (2010), Altwaijri et al. (2011), Anastasopoulos and Mannering (2011), Moore et al. (2011), Ye and Lord (2011), Morgan and Mannering (2011), Chiou et al. (2013b), Aziz et al. (2013), Abay (2013a); Manner and Wunsch-Ziegler (2013), Yasmin and Eluru (2013), Ye and Lord (2014)
Partial proportional odds model	Wang and Abdel-Aty (2008), Wang et al. (2009), Quddus et al. (2010)
Finite-mixture/latent-class and Markov switching models	Malyshkina and Mannering (2009), Xie et al. (2012), Eluru et al. (2012), Xiong and Mannering (2013), Xiong et al. (2013), Yasmin et al. (2014)
Heterogeneous outcome model	Quddus et al. (2010)
Mixed ordered probit (random parameters probit) model	Zoi et al. (2010), Paleti et al. (2010), Xiong et al. (2013)
Spatial and temporal correlations	Castro et al. (2013)

[a] *Source*: Updated from Savolainen et al. (2011).

overlooked.[3] These include issues relating to: parsimonious vs. fully specified models; unobserved heterogeneity; selectivity-bias/endogeneity; risk compensation; choice of methodological approach; under-reporting of crashes with less severe injuries; and spatial and temporal correlations. Each of these can substantially influence findings and the inferences drawn from the analysis of data. Table 3 provides a listing of some research efforts that have addressed these issues in the past, and a discussion of these issues is provided below.

[3] See the review articles by Lord and Mannering (2010) and Savolainen et al. (2011) for some additional discussions on fundamental issues in existing crash-frequency and crash-severity research.

Table 3
Research that has addressed identified ongoing methodological considerations in highway-accident research.

Methodological consideration	Previous research
Parsimonious vs. fully specified models[a]	Jovanis et al. (2011), Mitra and Washington (2012)
Unobserved heterogeneity	Eluru and Bhat (2007), Milton et al. (2008), Eluru et al. (2008, 2010), Kim et al. (2008, 2010, 2013), Malyshkina et al. (2009), Park and Lord (2009), Anastasopoulos and Mannering (2009), El-Basyouny and Sayed (2009b), Malyshkina and Mannering (2009, 2010a, 2010b), Park et al. (2010a), Zoi et al. (2010); Paleti et al. (2010), Peng and Lord (2011); Granowski and Manner (2011), Venkataraman et al. (2011, in preparation), Ukkusuri et al. (2011), Altwaijri et al. (2011), Anastasopoulos and Mannering (2011), Moore et al. (2011), Ye and Lord (2011), Peng and Lord (2011); Morgan and Mannering (2011), Xie et al. (2012), Mitra and Washington (2012), Wu et al. (2013), Chiou et al. (2013b), Aziz et al. (2013), Zou et al. (2013), Castro et al. (2013), Abay et al. (2013), Yasmin and Eluru (2013), Xiong and Mannering (2013), Xiong et al. (2013), Bhat et al. (2014), Shaheed et al. (in preparation), Yasmin et al. (2014)
Selectivity bias/endogeneity	Winston et al. (2006), Eluru and Bhat (2007), Paleti et al. (2010), Rana et al. (2010), Abay et al. (2013), Bhat et al. (2013)
Risk compensation	Winston et al. (2006)
Choice of methodological approach	Abdel-Aty (2003), Lord et al. (2005b), Anastasopoulos and Mannering (2011), Geedipally et al. (2010), Geedipally and Lord (2011), Ye and Lord (2011, 2014), Anastasopoulos et al. (2012a), Abay (2013a), Ye et al. (2013), Eluru (2013), Yasmin and Eluru (2013)
Under-reporting of crashes with less severe injuries	Kumara and Chin (2005), Yamamoto et al. (2008), Ma (2009), Ye and Lord (2011), Patil et al. (2012), Yasmin and Eluru (2013)
Spatial and temporal correlation	Flahaut et al. (2003), MacNab (2004), Miaou and Song (2005), Song et al. (2006), Wang and Abdel-Aty (2006), Aguero-Valverde and Jovanis (2006, 2008, 2010), Guo et al. (2010), Peng and Lord (2011), Castro et al. (2012, 2013), Abay (2013a), Narayanamoorthy et al. (2013), Chiou et al. (in preparation), Mohammadi and Samaranayake (in preparation), Xie et al. (in preparation)

[a] The bias introduced by omitting a significant variable is discussed and demonstrated in any standard econometrics text (see for example, Greene, 2012).

4.1. Parsimonious vs. fully specified models

The data available to researchers is often limited, and many variables known to significantly affect the frequency and severity of crashes may not be available. There may also be a need to develop relatively simplistic models using only explanatory variables that can be gathered and projected for use in practice, where practitioners may have access to little data or technical expertise. Given these data limitations or the need to specify models with a few simplistic explanatory variables, parsimonious models are often estimated.[4] An example would be estimating a model of crash frequency using only the volume of traffic as an explanatory variable. Clearly many other factors affect the frequency of crashes such as environmental conditions, roadway geometrics, the vehicle mix of traffic, lane widths, and so on. The problem with just using traffic volume as the explanatory variable is that the model will be excluding significant explanatory variables and the model-estimated parameter for traffic volume will be estimated with bias (this is referred to as an omitted variables bias) and application of the model will be fundamentally flawed because changes in the omitted variables (environmental conditions, roadway geometrics, etc.) cannot be captured and the predicted crash frequencies will be incorrect. In addition, a model with only traffic volume is limited in its value for designing countermeasures, precisely because the impacts of design features that can be controlled by traffic engineers (such as roadway curvature or pavement surface type) are not considered. In summary, the real problem with parsimonious models is that practitioners, and even researchers, do not fully grasp, or often conveniently overlook, the limitations of these simplistic models in terms of biased parameter estimates and policy value. For practitioners, the application of such models can easily produce erroneous estimates and provide lesser information for countermeasure design relative to a more fully specified model that includes variables that are amenable to changes in design. Researchers often extend simplistic parsimonious models with more sophisticated statistical methods often not realizing that the omitted variable bias present in their model compromises all of the conclusions that they are likely to draw. Thus, it is extremely important to recognize the limitations of parsimonious models, avoid them if at all possible, and consider more sophisticated statistical approaches to mitigate their adverse consequences. This is particularly important because parsimonious specifications can lead to more susceptibility to the econometric considerations listed and discussed below.

[4] Examples of this include the models in the Highway Safety Manual (2010), where many practical compromises have to be made to arrive at usable models of highway safety.

4.2. Unobserved heterogeneity

Some of the many factors affecting the frequency and severity of crashes are not observable, or the necessary data may be nearly impossible to collect. If these unobserved factors (often referred to as unobserved heterogeneity) are correlated with observed factors, biased parameters will be estimated and incorrect inferences could be drawn. For example, consider a statistical model of crash-injury severity that has age as one of the explanatory variables. Age is correlated with many underlying factors that are likely to affect crash-injury severity such as physical health, susceptibility of bones to breakage, body positioning at the time of crash, reaction times that may mitigate the severity of the crash, and so on. By including only age, age is acting as a proxy variable for many underlying factors that are likely to vary considerably across crash-injury observations because people of the same age are likely to have differences in these unobserved factors. By assuming that age has the same effect on injury severity across the population, the analyst is placing a potentially significant restriction on the model that may affect not only the inferences drawn from the age-variable parameter estimate, but also from other parameter estimates in the model. There are statistical corrections for dealing with this problem (see Table 3), but many researchers have overlooked this issue in the past.

4.3. Selectivity-bias/endogeneity

One of the most often overlooked elements in model estimation can be generally termed as selectivity-bias/endogeneity. This can take many forms, some of which are obvious and some of which are more subtle. As an example, consider a model that seeks to determine the effectiveness of ice-warning signs in reducing the frequency of crashes during icy conditions. The most common approach to studying this problem would be to collect crash-frequency data (crashes occurring during icy conditions) for roadway segments with ice-warning signs and roadway segments without. Then, using a naïve approach, estimate a model that has the presence of an ice-warning sign as an indicator variable – which takes a value of one if an ice-warning sign is present and zero otherwise (there are other statistical approaches to evaluating this phenomenon including the estimation of completely separate models for ice-warning sign and non-ice-warning sign roadway segments). If one were to estimate such a model, it is quite likely that the parameter estimate for the ice-warning sign indicator variable would have a substantial downward bias – seriously understating the effectiveness of ice warning signs. This is because ice-warning signs are likely to be placed on roadway segments with a history of a large number of ice crashes. Thus, the presence of an ice-warning sign (and its indicator variable in the model) will be correlated with unobserved factors that affect the frequency of ice-related crashes. These unobserved factors could include things such local micro-climate conditions that make some roadway segments more likely to accumulate moisture and freeze relative to others, making them more susceptible to high ice-crash frequencies. There have been countless studies that have likely arrived at erroneous inferences by ignoring such effects and not undertaking the proper statistical techniques for correcting such a selectivity effect.

Often times, the selectivity-bias/endogeneity can be more subtle. An example would be a study to determine the effectiveness of a new vehicle safety feature (such as side-impact airbags) in reducing the injury severity in crashes. The naïve approach would be to look at vehicles with the safety feature and those without, and assess the safety feature's effectiveness in reducing injury severity by, for example, using an indicator variable (one if the vehicle has the safety feature present and zero otherwise). The problem with this approach is that the drivers owning the vehicles with the safety feature are not likely to be a random sample of the driver population. In fact, studies have shown that the safest drivers are most likely to own cars with advanced safety features (Winston et al., 2006). Thus, the parameter estimate for the indicator variable for the presence of the safety feature will capture all the unobserved heterogeneity relating to its driver (which is more likely to be a safe driver) that will tend to result in less severe crashes (unobserved factors such as those relating to risk aversion and so on). This in turn will tend to impart a serious upward bias in the parameter estimate that would substantially overstate the effectiveness of the safety feature in reducing injury severity. Again, there are statistical corrections for this (see Winston et al., 2006), but they are often overlooked in model estimation.

Yet another example would be an attempt to capture the true effect of a posted speed limit on the frequency and severity of crashes. However, again there is a self-selectivity present in that speed limits may be set as a function of road classification or may be influenced by past crash histories. For example, a 70 mi/h maximum speed will likely only be observed on full-access-controlled rural interstates, so all of the unobserved characteristics (unobserved heterogeneity) of such roads may end up being captured by the model's parameter estimate of the speed-limit variable, which may then tend to over or under estimate the true effect of the speed limit. Similarly, highways with many crashes (for whatever reason) may be given lower speed limits to improve safety, but a poorly specified model (with potentially important missing variables that truly explain why the highway is dangerous) may conclude that lower speed limits are less safe because the roads with low speed limits will be correlated with a higher than expected number of crashes.

Resolving the self-selectivity/endogeneity issue can be achieved through various statistical corrections, but this is not done nearly enough in accident-related research and there is an urgent need for future studies to give full consideration to this issue.[5]

[5] It is also worthy to note that a skeptical view of this issue would be that almost every variable can be hypothesized to be endogenous in some way, which would make model estimation cumbersome if not impossible. The key to addressing endogneity, then, is to carefully consider the context and potential impact of the endogeneity of specific variables in the model.

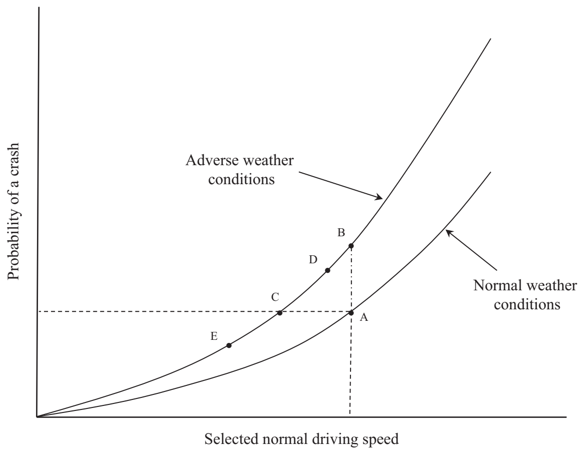

Fig. 1. Driver adaptation to changing weather conditions – the trade-off between speed and safety.

4.4. Risk compensation

The likelihood that drivers respond to changing road conditions by altering their behavior makes understanding the effect of these changing road conditions extremely difficult. An example would be a model that may find that the frequency of crashes declines during inclement weather. There are a number of explanations for this, including the possibility that the drivers self-select so that the safest drivers are more likely to drive in inclement weather and less-safe drivers may avoid inclement weather. But there is the very real possibility that each driver will compensate for the adverse conditions by altering their driving behavior to keep an acceptable level of risk. A simple illustration of this process is given in Fig. 1 with approximate speed/crash probability curves.[6] In looking at Fig. 1, under normal weather conditions each individual driver makes a trade-off between their selected speed and what they consider to be an acceptable level of safety (represented by the probability of a crash in this figure), resulting in Point A. Under adverse weather conditions, the relationship between speed and the probability of a crash shifts the curve upward. If the driver continues at the same speed as driven in normal weather conditions, Point B is reached and the probability of a crash increases accordingly. If the driver were to maintain the same crash probability, slowing down to Point C would be required. It is reasonable to speculate that all drivers will adapt to the adverse weather condition to some degree, likely resulting in a speed/crash probability equilibrium somewhere between Points B and C on the adverse weather-conditions curve (for example, Point D). There is also the possibility that some drivers may over compensate for the adverse weather conditions driving much slower resulting in equilibrium at Point E where the probability of a crash is even lower than it was before the adverse weather conditions.

From a statistical perspective, risk compensation presents a very difficult problem because the equilibrium point of each driver is not known (some may be at Point B, some at Point C, some at Point D, and so on) and the equilibrium point may not be stable over time. With regard to time stability, consider driver reactions to snowy weather conditions. In areas that experience snowy conditions frequently, driver experience will enable them to reach a snowy-condition equilibrium point that is more likely to be stable over time. However, in regions with infrequent snow fall, the spread of driver equilibrium points is likely to be over a much broader range of the speed/crash-probability curve because drivers do not have the experience to accurately assess crash probabilities under these conditions. And, as the frequency of snowfall changes over time, the resulting impacts on the frequency and severity of crashes will also change. So the effects of the same adverse weather conditions are likely to be both temporally and spatially (across geographic regions) unstable.

More recently applied statistical and econometric methods such as random parameters models and finite-mixture/latent-class/Markov-switching models can potentially provide some insight into the effects of risk compensation on the true impact of phenomena such as adverse weather conditions, but much additional methodological work is needed to move beyond simple statistical applications in order to seek fundamentally new insights.

4.5. Choice of methodological approach

Researchers have expended considerable energy in trying to determine which general methodological approaches are best suited to crash-related data. For example, with regard to crash-injury severities, there have been countless studies and discussions as to which general discrete-outcome approach is most appropriate: models that do not consider the natural ordering of injury severity data (ranging from no injury to fatality) such as the multinomial logit, nested logit and random parameters (mixed) logit; or models that do consider the natural ordering of data such as traditional ordered probit and

[6] In fact, many other elements could easily be considered in this graph (for example, risky behaviors beyond speed such as the decision to engage in distracted or impaired driving, following other cars too closely, and so on) but only speed and crash probability are used here for illustrative purposes.

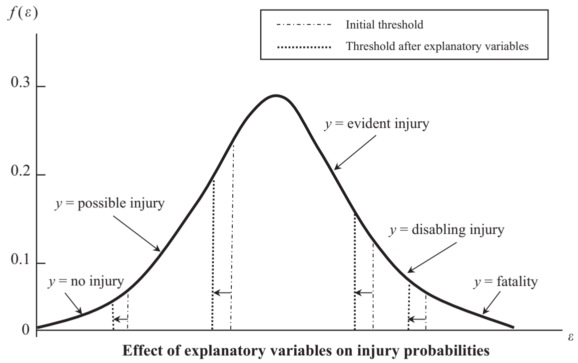

Fig. 2. Illustration of the limitations of the standard ordered probability model as applied to crash-injury severity.
Source: Adapted from Washington et al., (2011).

logit models (see Table 2). Because the data are ordered, many researchers have assumed without much empirical exploration that ordered models are the preferred methodological approach (see Washington et al. (2011) for a discussion of this point). However, all methodological approaches have inherent limitations and the superiority of one model over another can often not be proven mathematically and, in fact, even empirical generalizations cannot be made because the overall model fit may vary from one database to the next.

To provide an illustration of the trade-offs that must sometimes be made in applying competing methodological approaches, consider the inherent limitations of the traditional ordered probit model when applied to crash-injury data (see for example Eluru et al. (2008), who discuss this in detail when proposing a generalized ordered probit model for injury severity). Traditional ordered probability models are derived by defining an unobserved variable, z, which is typically specified as a linear function of a vector of explanatory variables ($\mathbf{X}$) and the associated vector of parameters ($\boldsymbol{\beta}$) is estimated by assuming a distribution of is an independently randomly distributed disturbance terms (ε). The probability of specific crash-injury severity outcomes is then determined by integration of the area under the density function as shown in Fig. 2, with the vertical lines in this figure (the vertical dash-dot lines) being thresholds separating discrete injury-severity categories and these are also determined as part of the estimation process. In standard ordered probability models, the effect of explanatory variables is to shift the thresholds as shown in Fig. 2 (from the dash-dot vertical lines to the dot vertical lines). A visual inspection of this figure reveals a severe limitation of ordered probability models in that is impossible for an explanatory variable to simultaneously increase or decrease both of the extreme severity categories (no injury and fatality).

To see how this is a problem, consider the following example provided in Washington et al. (2011). Suppose that one of the explanatory variables in determining injury severity is whether or not an airbag was deployed in the crash. The airbag-deployment indicator variable in a standard ordered model would move the thresholds shown in Fig. 2 to either increase the probability of a fatality (and subsequently decrease the probability of no injury) or decrease the probability of fatality (and subsequently increase the probability of property damage only). But the reality may be that the deployment of an airbag not only reduces the probability of a fatality but also reduces the probability of no-injury since airbag deployment itself could cause minor injuries. If this situation exists, a traditional ordered probability model is not appropriate because it does not have the flexibility to allow the extreme categories to simultaneously increase or decrease.[7] Estimation with a standard ordered model in this case will produce biased parameter estimates that could easily lead to incorrect inferences.

In an unordered discrete-modeling framework (such as a multinomial logit, nested logit or random parameters logit), accounting for the fact that an explanatory variable can simultaneously increase or decrease extreme severity categories is a total non-issue since this can be readily handled by including the airbag-deployment indicator in specific equations that determine individual severity-category probabilities. Thus, in choosing between ordered and unordered models, researchers often must make a tradeoff between considering the ordered nature of the data and restricting how explanatory variables affect outcome probabilities.[8]

Developing a general rule that establishes the superiority of one methodological approach over another has understandably eluded both crash-frequency and injury-severity researchers. Empirical evidence from many studies suggest that the superiority of one methodological approach over another can be very data-dependent[9] and, even with the same data,

[7] In recognition of this important limitation, there has been a body of recent work using generalized ordered outcome models which relax this restriction (see, for example, Eluru et al., 2008; Castro et al., 2013).

[8] Similar issues arise when considering how best to model crash-frequency analysis. For models that can be statistically compared, such as the simple Poisson and negative binomial models, a specific model can be justified using simple statistical tests such as the likelihood ratio test. However, models that do not lend themselves to direct statistical comparison, such as modeling frequencies as a count process vs. modeling them as duration data using the time between successive crashes, often lead to ambiguous statistical justifications.

[9] For example, in injury-severity models that are nested and can be directly compared statistically (such as the standard fixed-parameters multinomial logit and nested logit models), depending on the source of the injury-severity data, some studies have found the simple multinomial logit model to be justified whereas others have found the more involved nested logit model to be justified (see, for example, Savolainen and Mannering, 2007).

comparison of models which are often non-nested (such as is the case for ordered and unordered probability models) can leave much to be desired in terms of defensible statistical evidence. With this said, there have been a number of recent efforts that have undertaken empirical comparisons of alternate injury-severity model structures (Abay, 2013a; Yasmin and Eluru, 2013; Ye and Lord, 2014) and, although there will always be questions relating the generalizability of the results across multiple databases, these studies provide at least some evidence for model comparisons.

4.6. Under-reporting of crashes with less severe injuries

It is well documented that crashes resulting in no injuries, or less severe injuries, are more likely to be under-reported and thus do not appear in crash databases (Yamamoto et al., 2008; Ye and Lord, 2011; Yasmin and Eluru, 2013). In the presence of such under-reporting, the observed distribution of crashes (from reported crashes) among the injury-severity categories will differ from the actual distribution of crashes among the severity categories. For modeling crash-injury severities with traditional model-estimation techniques, the consequence will be a potentially severe bias in model-estimated parameters that could lead to incorrect inferences.[10] The matter of under-reporting has been extensively studied in discrete-outcome model literature, and is just a variation of outcome-based sampling. There are numerous corrective estimation techniques such as the weighted conditional maximum likelihood estimator and others (Ye and Lord, 2011; Patil et al., 2012). While several researchers have addressed the under-reporting problem in crash-severity analyses (see Abay, 2013b), there is a need to continue work in this area, particularly with more advanced methodologies such as random parameters and multiple-state models.

Under-reporting of less severe crashes obviously also affects crash-frequency models, but the effect of under-reporting on crash frequencies has been studied less often than it has been studied on crash severities. The consequence of omitting minor crashes from frequency models can be problematic in that locations with a large number of minor crashes may not show up as the safety hazard that they are, and minor changes in conditions (weather events, traffic volumes, etc.) could quickly move a roadway location with seemingly no major safety concern, into a very serious safety-deficient location as many of the unreported minor crashes become more severe reported crashes. The complexity of issues involved with under-reporting in count-data models can be formidable, but ignoring under-reporting in these models can also lead to erroneous inferences.

4.7. Spatial and temporal correlation

Both crash frequency and severity data often have observations that are in close spatial or temporal proximity. All data are likely to have unobserved factors that may influence the frequency and/or severity of crashes and, because these unobserved factors are likely to be correlated over space and time, ignoring the spatial and temporal correlation of data will almost certainly result in inefficient and possibly inconsistent parameter estimates. Examples of such unobserved factors could be pavement irregularities that may not be observed but may extend over time or space, micro-climate effects that may result in reduced friction over time and space, local sight-distance restrictions that again may extend over time and space. There have been numerous efforts that have begun to explicitly address spatial and temporal correlation (see the section on methodological frontiers later in this paper).

5. Emerging data sources

Traditional crash frequency and severity are based on data that is collected after a crash has occurred. This is highly restrictive in many ways. First, there are many near-crashes that contain potentially important information regarding crash generation and severity that do not appear in traditional crash data bases. Second, as discussed above, many minor crashes are not recorded through traditional sources leading to a loss of potentially important information. Third, many important contributing factors to crash occurrence and resulting severity are not collected (for example, vehicle speed, driver braking and maneuvering responses, etc.) leading to considerable unobserved heterogeneity that complicates modeling and precludes important information that could be used to make significant new inferences. Fourth, police-reported measures of injury severity (no injury, possible injury, evident injury, disabling injury, fatality) are based on observations at the crash scene and can change as further medical diagnosis is undertaken.

There are several important emerging sources of data that could address some of these data concerns. One example is the recent availability of Crash Outcome Data Evaluation System (CODES) data in select U.S. states which has permitted researchers to assess crash severity with significantly greater detail. These data provide detailed information on injury levels, location of injuries, cost of injuries, and so on, but they rely on the linkage of police-reported crash records with medical records which is itself often a difficult and imprecise task.[11] However, when police crash reports are successfully

[10] An exception to this is the multinomial logit model. If the restrictive assumptions of the fixed-parameters multinomial logit model hold (the independence of irrelevant alternatives), in the presence of such under-reporting all parameters will be correctly estimated except the constants, and these can be readily corrected if the extent of under-reporting is known (see Washington et al., 2011).

[11] CODES data may also help with some of the under-reporting of crashes if those involved in a non-reported crash subsequently seek medical attention.

matched with corresponding medical data, the level of detail available in CODES data goes well beyond police-reported injury assessment and includes details on injury types (fractures, dislocations, internal organ damage, crushing, burns, etc.) and locations body (head and neck, spine and back, torso, extremities). CODES data can also allow for more detailed analysis of cost data (another potential but underutilized assessment of severity) with information on medical costs (professional, hospital, emergency department, drugs, rehabilitation, long-term care), other associated costs (police/ambulance/fire, insurance administration, loss of wages, loss of household work, legal/court costs, property damage) and possible quality-of-life costs in terms of quality-adjusted life years (Blincoe et al., 2002).

Another emerging source of data is that collected from specially equipped cars to gather so-called naturalistic driving data. In these cases, cars are equipped with video-recording technologies, onboard vehicle sensors that record a wide array of data including lateral and longitudinal acceleration, yaw rate, brake and accelerator applications, and radars to measure proximities to other vehicles and objects. Such an instrumented car generates an incredible amount of data, but many issues arise in using such data including (1) the infrequent occurrence of crash and near-crash events results and the need for very long observation periods to generate enough truly useful data; (2) drivers knowing that they are driving in an instrumented vehicle may alter their behavior; and (3) the sheer volume of data makes managing and statistically modeling a cumbersome task. Even with these issues considered, the emergence of naturalistic data offers the potential greatly expand the scope of statistical modeling and the inferences that can be drawn in years to come.

Still another promising source of data is information gathered from vehicles' Event Data Recorders (EDR's), often referred to as a “black boxes”, which record significant amounts of data prior and during the crash. Currently, EDR's are not mandatory, but many automakers include them in their cars and it has been estimated that even as early as the 2005 model year, 64% of passenger vehicles sold had the device (Insurance Institute for Highway Safety, 2013). In December of 2012, the National Highway Traffic Safety Administration (NHTSA) proposed a rule requiring the devices in all 2015 and later model vehicles. Most EDR's are built into a vehicle's airbag control module and record information about airbag deployment. However, some also record pre-crash data, like engine throttle and vehicle speed from the engine control module. For the 2013 model year, EDR's must record: change in forward crash speed; maximum change in forward crash speed; time from beginning of crash at which the maximum change in forward crash speed occurs; speed vehicle was traveling; percentage of engine throttle, percentage full (how far the accelerator pedal was pressed); whether or not brake was applied; whether or not driver was using safety belt; whether or not frontal airbag warning lamp was on; driver frontal airbag deployment; and number of impact events. Some more advanced EDR's currently record additional information such as sideways acceleration, forward or rearward acceleration, engine speed, driver steering input, right front passenger safety belt status, engagement of electronic stability control system, antilock brake activity, side airbag deployment time for driver and right front passenger and seat track positions for both the driver and right front passenger. Occupant size and position for drivers and right front passengers may also be recorded. Clearly accessibility to such information could greatly improve the specification of crash injury-severity models.

6. The methodological frontier

Given the limitations of traditional data, there have been substantial methodological developments in recent years that have led to important new inferences in the study of crash frequency and severity. Perhaps some of the most important methodological advances have dealt with ways of addressing (a) unobserved heterogeneity, and have included random parameters and multi-state models such as Markov switching and finite-mixture/latent-class models, (b) multivariate models, including spatial and/or temporal dependence effects, and (c) self-selection or endogeneity issues. Finally, there has been some effort to incorporate “soft” measures of driver personalities and attitudes in safety modeling. Each one of these issues is discussed in turn in the subsequent sections.

6.1. Unobserved heterogeneity

As shown in Tables 1 and 2 (see also the references listed in the unobserved heterogeneity category in Table 3), there has been great interest in recent years in models that incorporate unobserved heterogeneity. These modeling approaches provide important ways to address issues relating to unobserved heterogeneity. Random parameters models can potentially capture unobserved heterogeneity by allowing parameters to vary across observations (such as a roadway segment) or be fixed within group of observations but vary across groups that are specified by the analyst (such as roadway segments on the same highway route). The disadvantage of random parameters models is that the distributional assumption required to estimate the random parameters may not adequately capture unobserved group-specific features within the population (in contrast to groups of observations that the analyst may specify, there may exist homogeneous groups of data which may not be known to the analyst).

Finite-mixture/latent-class models take a somewhat different approach to addressing unobserved heterogeneity by identifying distinct subgroups of data with homogeneous attributes. In contrast to traditional random parameters models, finite-mixture/latent-class models consider unobserved heterogeneity by using a finite and specified number of mass points to identify homogeneous subgroups of data (as opposed to having the analyst identify subgroups based on some observed characteristics, such as grouping roadway segments that are along the same route). The potential advantage of this is that it does not require, as in the case of traditional random parameters models, a distributional assumption relating how parameters vary across observations (or groups of observations) or analyst determination of observation groups. The disadvantage is that it does not account for the possibility of within-group variation due its restrictive homogeneity assumption on characteristics of the within-group observations.

The combination of finite-mixture/latent-class and random-parameters models (incorporating random parameters within a finite-mixture/latent-class model) to more fully capture the unobserved heterogeneity has been considered in a number of research efforts in statistics (Verbeke and Lesaffre, 1996), econometrics (Lenk and DeSarbo, 2000), marketing research (Allenby et al., 1998), and recently in accident research (Xiong and Mannering, 2013). This hybrid modeling idea considers the possibility of observational random parameters sampled from an assumed continuous distribution within each of the groups within a finite-mixture framework. Hence it can account for group-specific heterogeneity and individual-observation heterogeneity within each group.

Models that have multiple states of safety also have the potential to address unobserved heterogeneity in exciting new ways. The idea is that fundamentally different states of safety exist and that highways may shift between these over time. This has given rise to the application of Markov switching models, in crash-count and crash-severity applications (see Tables 1 and 2), which assume highway segments switch over time, according to a Markov process, among multiple states of highway safety. The logic behind addressing unobserved heterogeneity in this way is unobserved multiple states may exist because of different environmental conditions, driver reactions and other factors that may not necessarily be available to the analyst and that these may change over time, and that these states can be identified as part of the model-estimation process.[12]

6.2. Multivariate models

Multivariate models refer to cases where there are multiple dependent variables that are inter-related with each another. In the context of crash frequencies, a simple example of a multivariate count model is the case of analyzing intersection crash-related injuries by crash type (head-on, rear-end, angular, collision with a stationary object, etc.). Analyzing crash-related injuries by type is important because of differential impacts of relevant exogenous variables on different crash types. For instance, intersections with stop signs may lead to more rear-end crashes relative to intersections controlled by signal lights, as drivers may brake suddenly when arriving at the stop sign and do not leave adequate time for the following driver to stop in time (relative to the case of a signal light), as has been observed by Kim et al. (2006). However, there may be relatively little difference between stop sign controlled intersections and signal controlled intersections in the number of head-on collisions. This is an example of a case where the control type at the intersection has a differential effect on different crash types, and ignoring this will, in general, lead to inconsistent estimates for the count of crashes of each type as well for the total count of crashes. A possible approach to consider this heterogeneity in variable effects is to estimate separate univariate count models for each crash type, but the problem is that unobserved factors are likely to impact multiple crash counts simultaneously. This necessitates the consideration of multivariate count models.

There are other motivations that also lead to multivariate models. Thus, the frequency of crashes at a particular intersection may be inter-linked with those at other intersections over space because of unobserved factors (such as land-use design features, and local variations in driver behavior) that can cause a dependence between crash occurrences at proximately located intersections. At the same time, if data are collected at each intersection over multiple years, and the unit of analysis is the annual number of crashes, intersection-specific unobserved factors (such as pedestrian walkway continuity) will cause a temporal correlation in the number of crashes at the same intersection over time. Such spatial and temporal dependencies result in multivariate models of very large dimension.

From a methodological standpoint, the field has long since matured in the area of univariate count models, but this has not been the case with multivariate count data. Current methods to deal with multivariate data are either too restrictive, relatively cumbersome and time-consuming, and/or literally infeasible in the case of high dimensionality (as often is the case when accommodating spatial and temporal dependencies). One promising approach that has been recently applied for multivariate models involves the recasting of traditional count models as a special case of a generalized ordered-response model. In this recasting, the count is the result of a latent risk propensity that gets mapped into the observed count outcomes through thresholds that are themselves functions of exogenous variables. In this formulation, the linkage across count categories is generated through the latent risk propensity, and excess probability masses (such as excess zero values) are easily handled without the need for zero-inflated and hurdle-count type devices that get very cumbersome in multivariate count settings (see the last row in Table 1).

Multivariate issues also readily arise in crash injury-severity data, such as the case of vehicle crashes in which multiple vehicles are involved, with each vehicle having one or more occupants. In such cases, the different occupants of each vehicle may experience different levels of injury severity, based on observed factors (such as seat belt use, vehicle type, and position of the occupant in the vehicle) and unobserved factors (such as vehicle condition and maintenance record, and mental and physical state of the vehicle occupant). Some of the unobserved factors may play a role in the injury severity sustained by

[12] The empirical success of zero-inflated count-data models (see Table 1) to model crash frequencies provides some empricial evidence of the presence of unobserved safety states. Multi-state models (Markov switching models) have also been successfully estimated in the safety field by Malyshkina et al. (2009), Malyshkina and Mannering (2009, 2010a) and Xiong et al. (2013).

multiple individuals. For example, the vehicle condition should affect the injury severities of all occupants of each vehicle, while the pavement condition at the location of the crash should affect the injury severities of all individuals involved in the crash. The presence of these common unobserved elements points to the need for a multivariate injury-severity model that characterizes the severity levels of all individuals involved in the crash. In contrast, most crash-related injury severity studies in the safety literature either pool all individuals across all crashes and estimate an individual-level injury severity model that completely severs the link between individuals involved in the same crash (which leads to inefficient econometric estimation at the very least, and potentially inconsistent estimation in many situations; see Abay et al., 2013), or model the injury severity of the most severely injured individual in a crash (which does not provide a comprehensive view of the nature and severity of all injuries sustained in the crash). Recently, there have been a few safety studies that have formulated and employed a multivariate injury severity model (see Table 2). These include copula-based models as introduced by Bhat and Eluru (2009) in the general transportation literature and Eluru et al. (2010) in the safety literature that allow a flexible dependency structure in the unobserved factors influencing injury risk across individuals (see, for example, Rana et al. (2010); Yasmin et al. (2013)). The concept of copulas is discussed in a little more detail in the next section.

As in the case of crash counts, a multivariate injury-severity model also arises when taking account of spatial and temporal dependencies. For example, consider the case of crashes at proximally located intersections. It is certainly possible that observed design elements at one crash location (say, for example, the presence of an island at an intersection) not only influences injury risk propensity at that location, but also have a "spatial spillover" effect on the injury propensity at proximally located crash sites. In addition, there may be common unobserved (to the analyst) location factors that may lead to a spatial-correlation effect in the error terms of the injury-risk propensity at proximally located crash locations. Ignoring such spatial dependencies will, in general, result in inconsistent and inefficient parameter estimation in non-linear models (see LeSage and Pace, 2009). There have been some recent efforts to address this concern in general, and in the safety literature in particular. For example, Castro et al. (2013) use Bhat's (2011) maximum approximate composite marginal likelihood (MACML) approach to estimate a multivariate model with spatial dependency, and the approach holds considerable potential for application in a variety of multivariate contexts.

Another related area where multivariate models should be useful is in the analysis of naturalistic driving data. Indeed, the sheer volume of the naturalistic driving data makes statistical modeling an interesting and challenging task. There are several opportunities to enhance currently used analytic methods (or even venture into alternative approaches) to deal with such massive data sets. For instance, statistical pattern recognition and machine learning may offer avenues for combination with more traditional multivariate statistical methods to deal with high dimensional data and recognize/model patterns from large data streams (National Academies, 2013).

6.3. Selectivity bias/endogeneity

The issue of selectivity bias/endogeneity has been discussed earlier in Section 4.3, and falls under the general framework of treatment-outcome models in econometrics (see Heckman and Vytlacil, 2005), with the treatment (for example, ice-warning signs and posted speed limits) and the outcome (crash frequency or injury severity) being modeled jointly. The method used in almost all of the very few earlier safety analysis studies to accommodate endogeneity is based on the use of an instrumental variable approach that involves computing the predicted probability of the treatment, and replacing the treatment variable in the outcome equation by the predicted probability. Unfortunately, the two stage estimation as just discussed is not appropriate for non-linear outcome models such as count models and injury severity models (see Greene, 2009).

There are two possible (and correct) approaches to accommodate endogeneity in non-linear models. The first, control function or two stage residual inclusion (2SRI), approach involves (a) estimating the treatment or endogenous variable (which can itself be a continuous variable or a limited-dependent variable) using appropriate techniques (with one or more instrumental variables as predictors), (b) obtaining predictions of the endogenous variable, (c) computing residuals from this first stage, and then (d) including these first stage residuals (in addition to the endogenous variable). In the case when both the first stage and second stage equations are linear relationships as opposed to one or both being non-linear relationships, this 2SRI approach is equivalent to two stage least squares or 2SLS. Terza et al. (2008) show that 2SRI is consistent for non-linear models, while other two stage approaches are not. But it can be a challenge in this 2SRI approach to find good instruments, and the approach also constitutes a limited information approach that can be fraught with econometric efficiency and collinearity problems (Puhani, 2000). In addition, the analytic correction or a bootstrapping empirical estimator for obtaining the correct standard errors can be cumbersome.

The second approach is a full information maximum likelihood (FIML) approach. When using the traditional count formulations for crash frequency, the FIML approach includes a random error term in the parameterization of the expected value of the count discrete distribution (so that the expected value is not only a function of exogenous variables and the treatment variable, but also includes a random term). A dependence structure is then specified between this random term and the random term involved in the treatment model. Then, conditional on the error term in the count model, the probability of the treatment and of the outcome can be written as the product of the individual probabilities of the treatment and of the outcome. The unconditional probability of the treatment and outcome may be obtained by integrating out the error term of the count model (see Greene, 2009, for a discussion). Similarly, in the case of an injury severity outcomes, and assuming a binary treatment variable, one needs to have a propensity equation for the treatment (this propensity translates to the observed treatment indicator, in the usual binary model fashion) and an appropriate specification for injury severity with the treatment as an indicator variable (in the form of either a single injury-severity propensity equation that is related to the observed injury severity levels through thresholds in the ordered-response or generalized ordered-response formulation, or in the form of multiple propensity equations, one for each injury severity category, in the unordered-response formulation). The error terms in the treatment and outcome propensities are then specified to have a dependency structure. After accommodating this dependency structure, the structural parameter on the treatment in the outcome model may be viewed as the “cleansed” and “true” causal effect of the treatment. In this formulation, the joint probability of the treatment and the outcome takes a bivariate truncated distribution (if an ordered-response or generalized ordered-response model is used for injury severity) or a multivariate truncated distribution (if an unordered model is used for injury severity).

A methodological frontier issue in safety analysis is then first to accommodate endogeneity considerations appropriately. For the count model outcome, the recasting as a generalized ordered-response model may be particularly effective in capturing endogeneity issues, and should open up a suite of possibilities for specifying and testing endogeneity effects. Further, there is substantial room for exploring a variety of copula structures for the error dependency between the treatment and outcome variables. A copula is a device or function that generates a stochastic dependence relationship (a multivariate distribution) among random variables with pre-specified marginal distributions (see Bhat and Eluru, 2009). The precise definition of a copula is that it is a multivariate distribution function defined over the unit cube linking uniformly distributed marginals. There are several different types of copulas, each of which provides a different probability density function for the stochastic dependence relationship. Using a copula approach, an analyst can make use of the full information content available in the data through the FIML approach, while also alleviating misspecification problems in the dependence structure.[13]

6.4. Accommodating soft psychometric measures in safety analysis

Safety analysis research, for the most part today, uses “hard” observed variables as explanatory variables in crash frequency and injury-severity modeling. However, there are many examples where “soft” attitude measures and related “values” also may be important determinants. Understanding the impact of such “soft” measures can be very helpful for the design of information campaigns and behavioral modification considerations. For example, consider the effect of driver aggressiveness on crash occurrence and injury-severity levels. The analyst can obtain indicators of aggressiveness through surveys that elicit information on self-reported frequency (per month or per week) of participating in such acts as “excessive speeding”, “making threatening maneuvers with the car”, and “failure to signal”, or through personality inventories such as the Driver Anger Expression Inventory and the Driver Angry Thoughts Questionnaire (see Benfield et al., 2007), or through naturalistic driving data. Unfortunately, these indicators typically get combined and converted into a single binary indicator of aggressiveness, and are then occasionally studied as a function of demographic/situational attributes. Rarely has there been an examination of the effect of driver aggressiveness on crash occurrence and injury severity. One area that would certainly benefit the safety literature is to consider soft latent constructs (such as driver aggressive personality in general and when driving in particular), and relate these not only to relevant demographic/situational attributes, but also to the outcome of interest in safety analysis. A useful approach for this is the integrated choice and latent variables (ICLV) framework that expands typical econometric models to allow latent constructs representing “soft” psychometric considerations (see Bolduc et al., 2005; McFadden, 2013). The ICLV approach not only can provide a deeper understanding into safety determinants, but can also potentially enhance the predictive ability of current safety models. A typical ICLV model includes a latent variable structural equations model that specifies latent constructs of safety-related personality traits and attitudes (such as aggressiveness, responsibility, nervousness under pressure, etc.) as a function of observed covariates. Further, the latent constructs (or variables) themselves are viewed as being manifested through the attitudinal and perception indicator variables in a latent measurement equation model, which recognizes the presence of measurement error in capturing the intrinsic latent constructs. Finally, the “soft” latent variables and the “hard” observed variables are used together to explain safety-relate outcomes. The ICLV approach has substantial potential for use in safety analysis, particularly with recent developments that make the estimation and application of the approach much more practical (see Bhat and Dubey, 2013).[14]

7. Summary and insights

It is clear from the above discussion that accident research has benefited greatly from the application of more appropriate, and often more sophisticated, statistical methodologies. The application of these new statistical methodologies has enabled researchers to extract important new inferences from available data. However, many important methodological

[13] Another useful research frontier is to extend consideration to treatments that are not binary (see, for example, Bhat et al., 2013).

[14] Another important issue is to accommodate several of the econometric considerations discussed in earlier sections simultaneously. For example, accommodating the multivariate nature of counts or injury-severity levels does not alleviate the problems caused by unobserved heterogeneity or endogeneity. A few recent studies (see the studies that appear in more than one row in Table 3) have started considering the multiple econometric challenges simultaneously, but such studies are far and few in between.

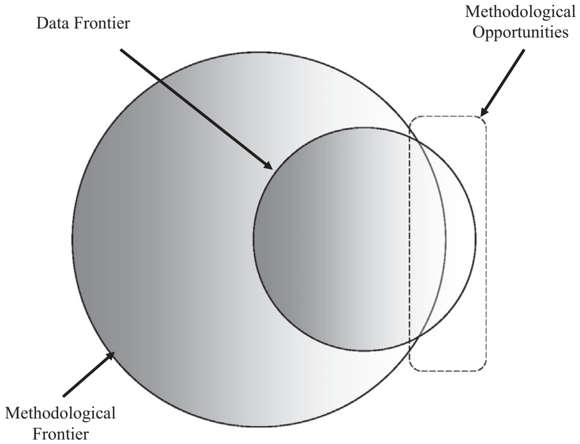

Fig. 3. State of methodological research with traditional crash data.

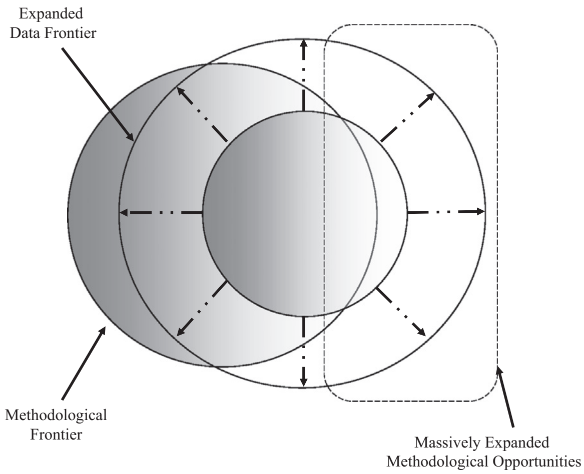

Fig. 4. State of methodological research with emerging crash-data sources.

issues remain relating to model specification, unobserved heterogeneity, selectivity-bias/endogeneity, risk compensation, missing data, addressing spatial and temporal correlations, and so on. Important new data sources, such as data from naturalistic driving, are becoming available, but many of the fundamental issues facing the statistical modeling of current data will also pervade these new data sources, and many new methodological concerns will most certainly arise from these sources. To be sure, there have been recent methodological applications such as random parameters models, finite-mixture/latent-class models, multi-state switching models, and others that hold considerable promise for improving the statistical analysis of current and future data sources.

Considering the above, the development and application of analytic methods in accident research is entering an era of unprecedented opportunities. This era that is being brought about by a combination of recent advances in methodological techniques and the availability of exciting new data sources. To show the interaction between methodology and data in the field and how it is evolving, it could be easily argued that the accident-research field has been dealing with relatively static data (quantity and quality) for decades (primarily police-reported crash data). This has kept a virtually constant "data frontier" while the "methodological frontier" has marched, in many respects, well beyond data capabilities. This is illustrated in Fig. 3, where it can be seen that the methodological opportunities have been limited by data availability from traditional sources. However, as illustrated in Fig. 4, the advent of many emerging data sources is beginning to greatly expand the data frontier, creating an urgent need for new methodological advances.

It is important to recognize that the many methodological opportunities that will present themselves in the coming years must be viewed from the perspective of what has been done in the past. Fundamental methodological issues encountered with past data (unobserved heterogeneity, selectivity-bias/endogeneity, risk compensation, missing data) will most certainly be present with new data sources and great caution must be exercised because there is often the tendency with new data (particularly data that is greatly expanded in terms volume and number of observations) to adopt methodological approaches that ignore important fundamental methodological issues.

As research relating to the statistical analysis of highway crash data (and new data that can provide information on near-crash events) progresses, it is important that researchers continue to address the fundamental methodological questions and continually strive to expand the methodological frontier. Not expanding the methodological frontier, and continuing to use methodological approaches with known deficiencies, has the potential to lead to erroneous and ineffective safety policies that may result in unnecessary injuries and loss of life.

Acknowledgments

The authors would like to thank the safety modeling specialists and reviewers who provided constructive feedback on an earlier version of this paper.

References

Abay, K., 2013a. Examining pedestrian-injury severity using alternate disaggregate models. Research in Transportation Economics 43, 123–136.
Abay, K., 2013b. Essays on Microeconometric Modeling of Road Crash Outcomes. Doctoral Dissertation, Faculty of Social science, University of Copenhagen, Denmark.
Abay, K., Paleti, R., Bhat, C., 2013. The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. Transportation Research Part B 50, 74–89.
Abbas, K., 2004. Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt. Accident Analysis and Prevention 36 (2), 149–163.
Abdel-Aty, M., Abdelwahab, H., 2004. Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. Accident Analysis and Prevention 36 (3), 447–456.
Abdel-Aty, M., Keller, J., 2005. Exploring the overall and specific crash severity levels at signalized intersections. Accident Analysis and Prevention 37 (3), 417–425.
Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research 34 (5), 597–603.
Abdelwahab, H., Abdel-Aty, M., 2001. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transportation Research Record 1746, 6–13.
Aguero-Valverde, J., 2013. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: comparing the precision of crash frequency estimates. Accident Analysis and Prevention 50, 289–297.
Aguero-Valverde, J., Jovanis, P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38 (3), 618–625.
Aguero-Valverde, J., Jovanis, P., 2008. Analysis of road crash frequency with spatial models. Transportation Research Record 2061, 55–63.
Aguero-Valverde, J., Jovanis, P., 2009. Bayesian multivariate Poisson log-normal models for crash severity modeling and site ranking. Transportation Research Record 2136, 82–91.
Aguero-Valverde, J., Jovanis, P., 2010. Spatial correlation in multilevel crash frequency models: effects of different neighboring structures. Transportation Research Record 2165, 21–32.
Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis and Prevention 43 (4), 1581–1589.
Al-Ghamdi, A., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accident Analysis and Prevention 34 (6), 729–741.
Allenby, G., Arora, N., Ginter, J., 1998. On the heterogeneity of demand. Journal of Marketing Research 35 (3), 384–389.
Altwaijri, S., Quddus, M., Bristow, A., 2011. Factors affecting the severity of traffic crashes in Riyadh City. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board. Washington, D.C.
Amoros, E., Martin, J., Laumon, B., 2003. Comparison of road crashes incidence and severity between some French counties. Accident Analysis and Prevention 35 (4), 537–547.
Anastasopoulos, P., Tarko, A., Mannering, F., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 40 (2), 768–775.
Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41 (1), 153–159.
Anastasopoulos, P., Mannering, F., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accident Analysis and Prevention 43 (3), 1140–1147.
Anastasopoulos, P., Shankar, V., Haddock, J., Mannering, F., 2012a. A multivariate tobit analysis of highway accident-injury-severity rates. Accident Analysis and Prevention 45 (1), 110–119.
Anastasopoulos, P., Mannering, F., Shankar, V., Haddock, J., 2012b. A study of factors affecting highway accident rates using the random-parameters Tobit model. Accident Analysis and Prevention 45 (1), 628–633.
Austin, R., Faigin, B., 2003. Effect of vehicle and crash factors on older occupants. Journal of Safety Research 34 (4), 441–452.
Aziz, A., Ukkusuri, S., Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York City. Accident Analysis and Prevention 50, 1298–1309.
Ballasteros, M., Dischinger, P., Langenberg, P., 2004. Pedestrian injuries and vehicle type in Maryland, 1995–1999. Accident Analysis and Prevention 36 (1), 73–81.
Bedard, M., Guyatt, G., Stones, M., Hirdes, J., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accident Analysis and Prevention 34 (6), 717–727.
Benfield, J., Szlemko, W., Bell, P., 2007. Driver personality and anthropomorphic attributions of vehicle personality relate to reported aggressive driving tendencies. Personality and Individual Differences 42 (2), 247–258.
Bhat, C., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transportation Research Part B 43 (7), 749–765.
Bhat, C., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transportation Research Part B 45 (7), 923–939.
Bhat, C., Dubey, S., 2013. A New Estimation Approach to Integrate Latent Psychological Constructs in Choice Modeling. Technical Paper. Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin. 〈http://www.ce.utexas.edu/prof/bhat/ABSTRACTS/ICLV_Modeling.pdf〉.
Bhat, C., Born, K., Sidharthan, R., Bhat, P., 2014. A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections. Analytic Methods in Accident Research (this issue).
Bijleveld, F., 2005. The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. Accident Analysis and Prevention 37 (4), 591–600.

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., Spicer, R., 2002. Economic Impact of Motor Vehicle Crashes, 2000. National Highway Traffic Safety Administration.
Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A., 2005. Hybrid choice models with logit kernel: applicability to large scale models. In: Lee-Gosselin, M., Doherty, S. (Eds.), Integrated Land-Use and Transportation Models: Behavioural Foundations, Elsevier, Oxford, pp. 275–302.
Bonneson, J., McCoy, P., 1993. Estimation of safety at two-way stop-controlled intersections on rural roads. Transportation Research Record 1401, 83–89.
Bonneson, J., Pratt, M., 2008. Procedure for developing accident modification factors from cross-sectional data. Transportation Research Record 2083, 40–48.
Brüde, U., Larsson, J., 1993. Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit? Accident Analysis and Prevention 25 (5), 499–509.
Brüde, U., Larsson, J., Hegman, K.-O., 1998. Design of Major Urban Junctions—Accident Prediction Models and Empirical Comparison. VTI, Linkööping, Sweden.
Bullough, J., Donnell, E., Rea, M., 2013. To illuminate or not to illuminate: roadway lighting as it affects traffic safety. Accident Analysis and Prevention 53, 65–77.
Cafiso, S., di Graziano, A., Di Silvestro, G., La Cava, G., Persaud, B., 2010. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. Accident Analysis and Prevention 42 (4), 1072–1079.
Caliendo, C., De Guglielmo, M., Guida, M., 2013. A crash-prediction model for road tunnels. Accident Analysis and Prevention 55, 107–115.
Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis and Prevention 39 (4), 657–670.
Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accident Analysis and Prevention 33 (1), 99–109.
Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. Transportation Research Part B 46 (1), 253–272.
Castro, M., Paleti, R., Bhat, C., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis and Prevention 52, 188–203.
Chang, H., Yeh, T., 2006. Risk factors to driver fatalities in single-vehicle crashes: comparisons between non-motorcycle drivers and motorcyclists. Journal of Transportation Engineering 132 (3), 227–236.
Chang, H.-L., Jovanis, P., 1990. Formulating accident occurrence as a survival process. Accident Analysis and Prevention 22 (5), 407–419.
Chang, L.-Y., 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Safety Science 43 (8), 541–557.
Chang, L.-Y., Mannering, F., 1998. Predicting vehicle occupancies from accident data: an accident severity approach. Transportation Research Record 1635, 93–104.
Chang, L.-Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. Accident Analysis and Prevention 31 (5), 579–592.
Chen, W., Jovanis, P., 2000. Method for identifying factors contributing to driver-injury severity in traffic crashes. Transportation Research Record 1707, 1–9.
Chen, E., Tarko, A., 2014. Modeling safety of highway work zones with random parameters and random effects models. Analytic Methods in Accident Research. (this issue).
Cheng, L, Geedipally, S., Lord, D., 2013. The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data. Safety Science 54, 38–42.
Chimba, D., Sando, T., 2009. Neuromorphic prediction of highway injury severity. Advances in Transportation Studies 19 (1), 17–26.
Chiou, Y.-C., Fu, C., 2013. Modeling crash frequency and severity using multinomial-generalized Poisson model with error components. Accident Analysis and Prevention 50, 73–82.
Chiou, Y.-C., Fu, C., Hsieh, C.-W., 2013. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic Methods in Accident Research. (in preparation).
Chiou, Y.-C., Hwang, C.-C., Chang, C.-C., Fu, C., 2013a. Modeling two-vehicle crash severity by a bivariate generalized ordered probit approach. Accident Analysis and Prevention 51, 175–184.
Chiou, Y.-C., Lan, L., Chen, W.-P., 2013b. A two-stage mining framework to explore key risk conditions on one-vehicle crash severity. Accident Analysis and Prevention 50, 405–415.
Chung, Y., 2010. Development of an accident duration prediction model on the Korean Freeway Systems. Accident Analysis and Prevention 42 (1), 282–289.
Compton, C., 2005. Injury severity codes: a comparison of police injury codes and medical outcome as determined by NASS CDS investigators. Journal of Safety Research 36 (5), 483–484.
Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accident Analysis and Prevention 42 (2), 393–402.
Das, A., Pande, A., Abdel-Aty, M., Santos, J., 2008. Urban arterial crash characteristics related with proximity to intersections and injury severity. Transportation Research Record 2083, 137–144.
de Lapparent, M., 2008. Willingness to use safety belt and levels of injury in car accidents. Accident Analysis and Prevention 40 (3), 1023–1032.
Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis and Prevention 38 (3), 434–444.
Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accident Analysis and Prevention 40 (4), 1257–1266.
Deublein, M., Schubert, M., Adey, B., Kohler, J., Faber, M., 2013. Prediction of road accidents: a Bayesian hierarchical approach. Accident Analysis and Prevention 51, 274–291.
Dissanayake, S., Lu, J., 2002a. Analysis of severity of young driver crashes, sequential binary logistic regression modeling. Transportation Research Record 1784, 108–114.
Dissanayake, S., Lu, J., 2002b. Factors influential in making an injury severity difference to older drivers involved in fixed object–passenger car crashes. Accident Analysis and Prevention 34 (5), 609–618.
Donnell, E., Mason, J., 2004. Predicting the severity of median-related crashes in Pennsylvania by using logistic regression. Transportation Research Record 1897, 55–63.
El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transportation Research Record 1950, 9–16.
El-Basyouny, K., Sayed, T., 2009a. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41 (4), 820–828.
El-Basyouny, K., Sayed, T., 2009b. Accident prediction models with random corridor parameters. Accident Analysis and Prevention 41 (5), 1118–1123.
Eluru, N., 2013. Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. Accident Analysis and Prevention 55, 1–11.
Eluru, N., Bagheri, M., Miranda-Moreno, L., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway–railway crossings. Accident Analysis and Prevention 47, 119–127.
Eluru, N., Bhat, C., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention 39 (5), 1037–1049.
Eluru, N., Bhat, C., Hensher, D., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention 40 (3), 1033–1054.
Eluru, N., Paleti, R., Pendyala, R., Bhat, C., 2010. Modeling multiple vehicle occupant injury severity: a copula-based multivariate approach. Transportation Research Record 2165, 1–11.
Farmer, C., Braver, E., Mitter, E., 1997. Two-vehicle side impact crashes: the relationship of vehicle and crash characteristics to injury severity. Accident Analysis and Prevention 29 (3), 399–406.
Ferreira, S., Couto, A., 2012. Categorical modeling to evaluate road safety at the planning level. Journal of Transportation Safety and Security 4 (4), 308–322.
Flahaut, B., Mouchart, M., San Martin, E., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accident Analysis and Prevention 35 (6), 991–1004.

Francis, R., Geedipally, S., Guikema, S., Dhavala, S., Lord, D., LaRocca, S., 2012. Characterizing the performance of the Conway–Maxwell Poisson generalized linear model. Risk Analysis 32 (1), 167–183.
Garder, P., 2006. Segment characteristics and severity of head-on crashes on two-lane rural highways in Maine. Accident Analysis and Prevention 38 (4), 652–661.
Geedipally, S., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson–gamma models. Accident Analysis and Prevention 42 (4), 1273–1282.
Geedipally, S., Lord, D., 2011. Examination of crash variances estimated by Poisson–Gamma and Conway–Maxwell–Poisson models. Transportation Research Record 2241, 59–67.
Geedipally, S., Lord, D., Dhavala, S., 2012. The negative binomial-Lindley generalized linear model: characteristics and application using crash data. Accident Analysis and Prevention 45, 258–265.
Geedipally, S., Patil, S., Lord, D., 2010. Examination of methods to estimate crash counts by collision type. Transportation Research Record 2165, 12–20.
Giuffre, O., Grana, A., Roberta, M., Corriere, F., 2011. Handling underdispersion in calibrating safety performance function at urban, four-leg, signalized intersections. Journal of Transportation Safety and Security 3 (3), 174–188.
Gray, R., Quddus, M., Evans, A., 2008. Injury severity analysis of accidents involving young male drivers in Great Britain. Journal of Safety Research 39 (5), 483–495.
Granowski, M., Manner, H., 2011. On factors related to car accidents on German Autobahn connectors. Accident Analysis and Prevention 43 (5), 1864–1871.
Guo, F., Wang, X., Abdel-Aty, M., 2010. Modeling signalized intersection safety with corridor spatial correlations. Accident Analysis and Prevention 42 (1), 84–92.
Greene, W., 2009. Models for count data with endogenous participation. Empirical Economics 36 (1), 133–173.
Greene, W., 2012. Econometric Analysis, 7th edition Prentice Hall, Upper Saddle River, NJ.
Gustavsson, J., Svensson, A., 1976. A Poisson regression model applied to classes of road accidents with small frequencies. Scandinavian Journal of Statistics 3 (2), 49–60.
Haleem, K., Abdel-Aty, M., 2010. Examining traffic crash injury severity at unsignalized intersections. Journal of Safety Research 41 (4), 347–357.
Hauer, E., 2004. Statistical road safety modeling. Transportation Research Record 1897, 81–87.
Heckman, J., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73 (3), 669–738.
Helai, H., Chor, C., Haque, M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. Accident Analysis and Prevention 40 (1), 45–54.
Heydecker, B., Wu, J., 2001. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference. Advances in Engineering Software 32 (10), 859–869.
Highway Safety Manual, 2010. The American Association of State Highway and Transportation Officials. Washington, DC.
Hirst, W., Mountain, L., Maher, M., 2004. Sources of error in road safety scheme evaluation: a method to deal with outdated accident prediction models. Accident Analysis and Prevention 36 (5), 717–727.
Holdridge, J., Shankar, V., Ulfarsson, G., 2005. The crash severity impacts of fixed roadside objects. Journal of Safety Research 36 (2), 139–147.
Hu, W., Donnell, E., 2010. Median barrier crash severity: some new insights. Accident Analysis and Prevention 42 (6), 1697–1704.
Insurance Institute for Highway Safety, 2013. Highway Safety Research and Communications. 〈http://www.iihs.org/research/qanda/edr.aspx〉. (accessed 24.06.13).
Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: some additional evidence. Journal of Safety Research 37 (3), 267–276.
Jiang, X., Huang, B., Yan, X., Zaretzki, R., Richards, S., 2013a. Two-vehicle injury severity models based on integration of pavement management and traffic engineering factors. Traffic Injury Prevention 14 (5), 544–553.
Jiang, X., Huang, B., Zaretzki, R., Richards, S., Yan, X., Zhang, H., 2013b. Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models. Accident Analysis and Prevention 57, 55–56.
Johansson, P., 1996. Speed limitation and motorway casualties: a time series count data regression approach. Accident Analysis and Prevention 28 (1), 73–87.
Jones, A., Jørgensen, S., 2003. The use of multilevel models for the prediction of road accident outcomes. Accident Analysis and Prevention 35 (1), 59–69.
Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention 23 (2), 239–255.
Joshua, S., Garber, N., 1990. Estimating truck accident rate and involvements using linear and Poisson regression models. Transportation Planning and Technology 15 (1), 41–58.
Jovanis, P., Chang, H., 1989. Disaggregate model of highway accident occurrence using survival theory. Accident Analysis and Prevention 21 (5), 445–458.
Jovanis, P., Aguero-Valverde, J., Wu, K.-F., Shankar, V., 2011. Analysis of naturalistic driving event data: omitted-variable bias and multilevel modeling approaches. Transportation Research Record 2236, 49–57.
Jovanovic, D., Backalic, T., Basic, S., 2011. The application of reliability models in traffic accident frequency analysis. Safety Science 49 (8–9), 1251–1256.
Jung, S., Qin, X., Noyce, D., 2010. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Analysis and Prevention 42 (1), 213–224.
Karlaftis, M., Tarko, A., 1998. Heterogeneity considerations in accident modeling. Accident Analysis and Prevention 30 (4), 425–433.
Khattak, A., 2001. Injury severity in multivehicle rear-end crashes. Transportation Research Record 1746, 59–68.
Khattak, A., Kantor, P., Council, F., 1998. Role of adverse weather in key crash types on limited-access: roadways implications for advanced weather systems. Transportation Research Record 1621, 10-19.
Khattak, A., Pawlovich, M., Souleyrette, R., Hallmark, S., 2002. Factors related to more severe older driver traffic crash injuries. Journal of Transportation Engineering 128 (3), 243–249.
Khattak, A., Rocha, M., 2003. Are SUVs 'Supremely Unsafe Vehicles'? Analysis of rollovers and injuries with sport utility vehicles. Transportation Research Record 1840, 167–177.
Khattak, A., Targa, F., 2004. Injury severity and total harm in truck-involved work zone crashes. Transportation Research Record 1877, 106–116.
Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. Accident Analysis and Prevention 37 (5), 910–921.
Kim, D., Washington, S., 2006. The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. Accident Analysis and Prevention 38 (6), 1094–1100.
Kim, D.-G., Washington, S., Oh, J., 2006. Modeling crash types: new insights into the effects of covariates on crashes at rural intersections. Journal of Transportation Engineering 132 (4), 282–292.
Kim, D.-G., Lee, Y., Washington, S., Choi, K., 2007a. Modeling crash outcome probabilities at rural intersections: application of hierarchical binomial logistic models. Accident Analysis and Prevention 39 (1), 125–134.
Kim, J-K., Kim, S., Ulfarsson, G., Porrello, L., 2007b. Bicyclist injury severities in bicycle–motor vehicle accidents. Accident Analysis and Prevention 39 (2), 238–251.
Kim, J.-K., Ulfarsson, G., Kim, S., Shankar, V., 2013. Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. Accident Analysis and Prevention 50, 1751–1758.
Kim, J.-K., Ulfarsson, G., Shankar, V., Kim, S., 2008. Age and pedestrian injury severity in motor-vehicle crashes: a heteroskedastic logit analysis. Accident Analysis and Prevention 40 (5), 1695–1702.

Kim, J.-K., Ulfarsson, G., Shankar, V., Mannering, F., 2010. A note on modeling pedestrian injury severity in motor vehicle crashes with the mixed logit model. Accident Analysis and Prevention 42 (6), 1073–1081.
Klop, J., Khattak, A., 1999. Factors influencing bicycle crash severity on two-lane, undivided roadways in North Carolina. Transportation Research Record 1674, 78–85.
Kockelman, K., Kweon, Y., 2002. Driver injury severity: an application of ordered probit models. Accident Analysis and Prevention 34 (3), 313–321.
Kononen, D.W., Flannagan, C., Wong, S., 2011. Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accident Analysis and Prevention 43 (1), 112–122.
Krull, K., Khattak, A., Council, F., 2000. Injury effects of rollovers and events sequence in single-vehicle crashes. Transportation Research Record 1717, 46–54.
Kweon, Y., Kockelman, K., 2003. Overall injury risk to different drivers: combining exposure, frequency, and severity models. Accident Analysis and Prevention 35 (4), 441–450.
Kumala, R., 1995. Safety at Rural Three- and Four-Arm Junctions: Development and Applications of Accident Prediction Models. VTT Publications 233, Technical Research Centre of Finland, Espoo, Finland.
Kumara, S., Chin, H., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Injury Prevention 3 (4), 53–57.
Kumara, S., Chin, H., 2005. Application of Poisson underreporting model to examine crash frequencies at signalized three-legged intersections. Transportation Research Record 1908, 46–50.
Lao, Y., Wu, Y.-J., Corey, J., Wang, Y., 2011a. Modeling animal–vehicle collisions using diagonal inflated bivariate Poisson regression. Accident Analysis and Prevention 43 (3), 220–227.
Lao, Y., Zhang, G., Wu, Y.-J., Wang, Y., 2011b. Modeling animal–vehicle collisions considering animal–vehicle interactions. Accident Analysis and Prevention 43 (6), 1991–1998.
Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. Accident Analysis and Prevention 37 (4), 775–786.
Lee, C., Abdel-Aty, M., 2008. Presence of passengers: does it increase or reduce driver's crash potential? Accident Analysis and Prevention 40 (5), 1703–1712.
Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34 (2), 149–161.
Lemp, J.D., Kockelman, K., Unnikrishnan, A., 2011. Analysis of large truck crash severity using heteroskedastic ordered probit models. Accident Analysis and Prevention 43 (1), 370–380.
Lenk, P., DeSarbo, W., 2000. Bayesian inference for finite mixture of generalized linear models with random effects. Psychometrika 65 (1), 93–119.
LeSage, J., Pace, R., 2009. Introduction to Spatial Econometrics. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton.
Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008a. The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. Accident Analysis and Prevention 40 (4), 1531–1542.
Li, X., Lord, D., Zhang, Y., 2009. Development of Accident Modification Factors for Rural Frontage Road Segments in Texas Using Results From Generalized Additive Models. Working Paper. Zachry Department of Civil Engineering, Texas A&M University, College Station, TX.
Li, X., Lord, D., Zhang, Y., Xie, Y., 2008b. Predicting motor vehicle crashes using support vector machine models. Accident Analysis and Prevention 40 (4), 1611–1618.
Li, Z., Wang, W., Liu, P., Bigham, J., Ragland, D., 2013. Using geographically weighted Poisson regression for county-level crash modeling in California. Safety Science 58, 89–97.
Lord, D., 2006. Modeling motor vehicle crashes using Poisson–gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention 38 (4), 751–766.
Lord, D., Geedipally, S., 2011. The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. Accident Analysis and Prevention 43 (5), 1738–1742.
Lord, D., Geedipally, S., Guikema, S., 2010. Extension of the application of Conway–Maxwell–Poisson models: analyzing traffic crash data exhibiting under-dispersion. Risk Analysis 30 (8), 1268–1276.
Lord, D., Guikema, S., 2012. The Conway–Maxwell–Poisson model for analyzing crash data. Applied Stochastic Models in Business and Industry 28 (2), 122–127.
Lord, D., Guikema, S., Geedipally, S., 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. Accident Analysis and Prevention 40 (3), 1123–1134.
Lord, D., Kuo, P.-F., 2012. Examining the effects of site selection criteria for evaluating the effectiveness of traffic safety countermeasures. Accident Analysis and Prevention 47, 52–63.
Lord, D., Mahlawat, M., 2009. Examining the application of aggregated and disaggregated Poisson–gamma models subjected to low sample mean bias. Transportation Research Record 2136, 1–10.
Lord, D., Manar, A., Vizioli, A., 2005a. Modeling crash-flow-density and crash-flow–v/c ratio for rural and urban freeway segments. Accident Analysis and Prevention 37 (1), 185–199.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44 (5), 291–305.
Lord, D., Miranda-Moreno, L., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson–gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46 (5), 751–770.
Lord, D., Persaud, B., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transportation Research Record 1717, 102–108.
Lord, D., Washington, S., Ivan, J., 2005b. Poisson, Poisson–gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37 (1), 35–46.
Lord, D., Washington, S., Ivan, J., 2007. Further notes on the application of zero inflated models in highway safety. Accident Analysis and Prevention 39 (1), 53–57.
Lu, G., Noyce, D., McKendry, R., 2006. Analysis of the magnitude and predictability of median crossover crashes utilizing logistic regression. In: Proceedings of the TRB 85th Annual Meeting CD-ROM, January 22–26.
Ma, J., 2009. Bayesian analysis of underreporting Poisson regression model with an application to traffic crashes on two-lane highways. Paper #09-3192. Presented at the 88th Annual Meeting of the Transportation Research Board. Washington, D.C.
Ma, J., Kockelman, K., 2006. Bayesian multivariate Poisson regression for models of injury count by severity. Transportation Research Record 1950, 24–34.
Ma, J., Kockelman, K., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40 (3), 964–975.
MacNab, Y., 2004. Bayesian spatial and ecological models for small-area crash and injury analysis. Accident Analysis and Prevention 36 (6), 1019–1028.
Maher, M., 1990. A bivariate negative binomial model to explain traffic accident migration. Accident Analysis and Prevention 22 (5), 487–498.
Maher, M., Summersgill, I., 1996. A comprehensive methodology for the fitting predictive accident models. Accident Analysis and Prevention 28 (3), 281–296.
Malyshkina, N., Mannering, F., 2008. Effect of increases in speed limits on severities of injuries in accidents. Transportation Research Board 2083, 122–127.
Malyshkina, N., Mannering, F, 2009. Markov switching multinomial logit model: an application to accident-injury severities. Accident Analysis and Prevention 41 (4), 829–838.
Malyshkina, N., Mannering, F., 2010a. Zero-state Markov switching count-data models: an empirical assessment. Accident Analysis and Prevention 42 (1), 122–130.

Malyshkina, N., Mannering, F., 2010b. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accident Analysis and Prevention 42 (1), 131–139.
Malyshkina, N., Mannering, F., Tarko, A., 2009. Markov switching negative binomial models: an application to vehicle accident frequencies. Accident Analysis and Prevention 41 (2), 217–226.
Manner, H., Wunsch-Ziegler, L., 2013. Analyzing the severity of accidents on the German Autobahn. Accident Analysis and Prevention 57, 40–48.
Mannering, F., 1993. Male/female driver characteristics and accident risk: some new evidence. Accident Analysis and Prevention 25 (1), 77–84.
Maycock, G., Hall, R., 1984. Accidents at 4-Arm Roundabouts. TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne, U.K.
McDonald, G., Davie, G., Langley, J., 2009. Validity of police-reported information on injury severity for those hospitalized from motor vehicle traffic crashes. Traffic Injury Prevention 10 (2), 184–190.
McFadden, D., 2013. The new science of pleasure: consumer choice behavior and the measurement of well-being. In: Hess, S., Daly, A.J. (Eds.), Handbook of Choice Modelling, Edward Elgar, Cheltenham, UK, 〈http://elsa.berkeley.edu/wp/mcfadden122812.pdf〉. (Forthcoming).
Meng, Q., Qu, X., 2012. Estimation of rear-end vehicle crash frequencies in urban road tunnels. Accident Analysis and Prevention 48, 254–263.
Mergia, W., Eustace, D., Chimba, D., Qumsiyeh, M., 2013. Exploring factors contributing to injury severity at freeway merging and diverging locations in Ohio. Accident Analysis and Prevention 55, 202–210.
Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis and Prevention 26 (4), 471–482.
Miaou, S.-P., Bligh, R., Lord, D., 2005. Developing median barrier installation guidelines: a benefit/cost analysis using Texas data. Transportation Research Record 1904, 3–19.
Miaou, S.-P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transportation Research Record 1840, 31–40.
Miaou, S.-P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention 25 (6), 689–709.
Miaou, S.-P., Song, J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion and spatial dependence. Accident Analysis and Prevention 37 (4), 699–720.
Miaou, S.-P., Song, J., Mallick, B., 2003. Roadway traffic crash mapping: a space–time modeling approach. Journal of Transportation and Statistics 6 (1), 33–57.
Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor vehicle accident frequencies. Transportation 25 (4), 395–413.
Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40 (1), 260–266.
Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash modeling. Accident Analysis and Prevention 49, 439–448.
Mohammadi, M., Samaranayake, V., 2013. The effect of incorporating temporal correlations into negative binomial count data models. In: Proceedings of the Fourth International Conference on Road Safety and Simulation, Rome, Italy.
Moore, D., Schneider, W., Savolainen, P., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accident Analysis and Prevention 43 (3), 621–630.
Morgan, A., Mannering, F., 2011. The effects of road-surface conditions, age, and gender on driver-injury severities. Accident Analysis and Prevention 43 (5), 1852–1863.
Moudon, A., Lin, L., Jiao, J., Hurvitz, P., Reeves, P., 2011. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. Accident Analysis and Prevention 43 (1), 11–24.
Mountain, L., Fawaz, B., Jarrett, D., 1996. Accident prediction models for roads with minor junctions. Accident Analysis and Prevention 28 (6), 695–707.
Mountain, L., Maher, M., Fawaz, B., 1998. The influence of trend on estimates of accidents at junctions. Accident Analysis and Prevention 30 (5), 641–649.
Narayanamoorthy, S., Paleti, R., Bhat, C., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation Research Part B 55, 245–264.
National Academies, 2013. Frontiers in Massive Data Analysis. The National Academies Press, Washington, DC.
O'Donnell, C., Connor, D., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention 28 (6), 739–753.
Oh, J., 2006. Development of severity models for vehicle accident injuries for signalized intersections in rural areas. KSCE Journal of Civil Engineering 10 (3), 219–225.
Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway–highway interfaces. Accident Analysis and Prevention 38 (2), 346–356.
Ouyang, Y., Shankar, V., Yamamoto, T., 2002. Modeling the simultaneity in injury causation in multi-vehicle collisions. Transportation Research Record 1784, 143–152.
Pai, C., 2008. Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: focusing on the effects of driver's failure to yield and junction control measures. Accident Analysis and Prevention 40 (2), 479–486.
Pai, C., 2009. Motorcyclist injury severity in angle crashes at T-junctions: identifying significant factors and analyzing what made motorists fail to yield to motorcycles. Safety Science 47 (8), 1097–1106.
Pai, C., Saleh, W., 2007. An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK. Safety Science 45 (8), 832–847.
Paleti, R., Eluru, N., Bhat, C., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accident Analysis and Prevention 42 (6), 1839–1854.
Park, E.-S., Carlson, P., Porter, R., Anderson, C., 2012. Safety effects of wider edge lines on rural, two-lane highways. Accident Analysis and Prevention 48, 317–325.
Park, E.-S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019, 1–6.
Park, E-U, Park, J., Lomax, T., 2010a. A fully Bayesian multivariate approach to before–after safety evaluation. Accident Analysis and Prevention 42 (4), 1118–1127.
Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention 41 (4), 683–691.
Park, B.J., Lord, D., Hart, J., 2010b. Bias properties of Bayesian statistics in finite mixture of negative regression models for crash data analysis. Accident Analysis and Prevention 45 (2), 741–749.
Patil, S., Geedipally, S., Lord, D., 2012. Analysis of crash severities using nested logit model—accounting for the underreporting of crashes. Accident Analysis and Prevention 45 (1), 646–653.
Peek-Asa, C., Britton, C., Young, T., Pawlovich, M., Falb, S., 2010. Teenage driver crash incidence and factors influencing crash injury by rurality. Journal of Safety Research 41 (6), 487–492.
Pei, X., Wong, S.C., Sze, N., 2011. A joint-probability approach to crash prediction models. Accident Analysis and Prevention 43 (3), 1160–1166.
Peng, Y., Lord, D., 2011. Application of latent class growth model to longitudinal analysis of traffic crashes. Transportation Research Record 2236, 102–109.
Persaud, B., Nguyen, T., 1998. Disaggregate safety performance models for signalized intersections on Ontario provincial roads. Transportation Research Record 1635, 113–120.
Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2013. Evaluating the road safety effects of a fuel cost increase measure by means of zonal crash prediction modeling. Accident Analysis and Prevention 50, 186–195.
Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering 122 (2), 105–113.
Puhani, P., 2000. The Heckman correction for sample selection and its critique. Journal of Economic Surveys 14 (1), 53–68.

Qin, X., Ivan, J., Ravishankar, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis and Prevention 36 (2), 183–191.
Quddus, M., 2008. Time series count data models: an empirical application to traffic accidents. Accident Analysis and Prevention 40 (5), 1732–1741.
Quddus, M., Noland, R., Chin, H., 2002. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research 33 (4), 445–462.
Quddus, M., Wang, C., Ison, S., 2010. Road traffic congestion and crash severity: an econometric analysis using ordered response models. Journal of Transportation Engineering 136 (5), 424–435.
Rana, T., Sikder, S., Pinjari, A., 2010. Copula-based method for addressing endogeneity in models of severity of traffic crash injuries. Transportation Research Record 2147, 75–87.
Renski, H., Khattak, A., Council, F., 1999. Effect of speed limit increases on crash injury severity: analysis of single-vehicle crashes on North Carolina interstate highways. Transportation Research Record 1665, 100–108.
Rifaat, S., Tay, R., 2009. Effects of street patterns on injury risks in two-vehicle crashes. Transportation Research Record 2102, 61–67.
Rifaat, S., Tay, R., de Barros, A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accident Analysis and Prevention 43 (1), 276–283.
Riviere, C., Lauret, P., Ramsamy, J., Page, Y., 2006. A Bayesian neural network approach to estimating the energy equivalent speed. Accident Analysis and Prevention 38 (2), 248–259.
Russo, B., Savolainen, P., Schneider, W., 2013. Comparison of factors affecting injury severity in angle collisions by fault status using a bivariate ordered probit model. Analytic Methods in Accident Research. (in preparation).
Saccomanno, F., Nassar, S., Shortreed, J., 1996. Reliability of statistical road accident injury severity models. Transportation Research Record 1542, 14–23.
Santolino, M., Bolance, C., Alcaniz, M., 2012. Factors affecting hospital admission and recovery stay duration of in-patient motor victims in Spain. Accident Analysis and Prevention 49, 512–519.
Savolainen, P., Ghosh, I., 2008. Examination of factors affecting driver injury severity in Michigan's single-vehicle-deer crashes. Transportation Research Record 2078, 17–25.
Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Analysis and Prevention 39 (5), 955–963.
Savolainen, P., Mannering, F., Lord, D., Quddus, M., 2011. The statistical analysis of crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis and Prevention 43 (5), 1666–1676.
Schneider, W., Savolainen, P., 2011. Comparison of motorcyclist injury severity among various crash types. Transportation Research Record 2265, 70–80.
Schneider, W., Savolainen, P., Zimmerman, K., 2009. Driver injury severity resulting from single-vehicle crashes along horizontal curves on rural two-lane highways. Transportation Research Record 2102, 85–92.
Sellers, K., Shmueli, G., 2010. A flexible regression model for count data. Annals of Applied Statistics 4 (2), 943–961.
Shaheed, M., Gritza, N., Bilionis, D., 2013. A latent class logit analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research. (in preparation).
Shankar, V., Albin, R., Milton, J., Mannering, F., 1998. Evaluating median cross-over likelihoods with clustered accident counts: an empirical inquiry using random effects negative binomial model. Transportation Research Record 1635, 44–48.
Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural accident frequencies. Accident Analysis and Prevention 27 (3), 371–389.
Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of Safety Research 27 (3), 183–194.
Shankar, V., Mannering, F., Barfield, W., 1996. Statistical analysis of accident severity on rural freeways. Accident Analysis and Prevention 28 (3), 391–401.
Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequency as zero-altered probability processes: an empirical inquiry. Accident Analysis and Prevention 29 (6), 829–837.
Shankar, V., Ulfarsson, G., Pendyala, R., Nebergall, M., 2003. Modeling crashes involving pedestrians and motorized traffic. Safety Science 41 (7), 627–640.
Shibata, A., Fukuda, K., 1994. Risk factors of fatality in motor vehicle traffic accidents. Accident Analysis and Prevention 26 (3), 391–397.
Shimamura, M., Yamazaki, M., Fujita, G., 2005. Method to evaluate the effect of safety belt use by rear seat passengers on the injury severity of front seat occupants. Accident Analysis and Prevention 37 (1), 5–17.
Shin, K., Washington, S., van Schalkwyk, I., 2009. Evaluation of the Scottsdale Loop 101 automated speed enforcement demonstration program. Accident Analysis and Prevention 41 (3), 393–403.
Siddiqui, N., Chu, X., Guttenplan, M., 2006. Crossing locations, light conditions, and pedestrian injury severity. Transportation Research Record 1982, 141–149.
Song, J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis 97 (1), 246–273.
Srinivasan, K., 2002. Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. Transportation Research Record 1784, 132–142.
Sze, N., Wong, S.C., 2007. Diagnostic analysis of the logistic model for pedestrian injury. Accident Analysis and Prevention 39 (6), 1267–1278.
Terza, J., Basu, A., Rathouz, P., 2008. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. Journal of Health Economics 27 (3), 531–543.
Toy, E., Hammitt, J., 2003. Safety impacts of SUVs, vans, and pickup trucks in two vehicle crashes. Risk Analysis 23 (4), 641–650.
Tsui, K.L., So, F.L., Sze, N.N., Wong, S.C., Leung, T.F., 2009. Misclassification of injury severity among road casualties in police reports. Accident Analysis and Prevention 41 (1), 84–89.
Turner, S., Nicholson, A., 1998. Using accident prediction models in area wide crash reduction studies. In: Proceedings of the 9th Road Engineering Association of Asia and Australasia Conference. Wellington, NZ.
Ulfarsson, G., Mannering, F., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Analysis and Prevention 36 (2), 135–147.
Ulfarsson, G., Shankar, V., 2003. An accident count model based on multi-year cross-sectional roadway data with serial correlation. Transportation Research Record 1840, 193–197.
Ukkusuri, S., Hasan, S., Aziz, H., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transportation Research Record 2237, 98–106.
Usman, T., Fu, L., Miranda-Moreno, L., 2012. A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level. Accident Analysis and Prevention 48, 368–378.
Venkataraman, N., Ulfarsson, G., Shankar, V., Oh, J., Park, M., 2011. Model of relationship between interstate crash occurrence and geometrics: exploratory insights from random parameter negative binomial approach. Transportation Research Record 2236, 41–48.
Venkataraman, N., Ulfarsson, G., Shankar, V., 2013. Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accident Analysis and Prevention 59, 309–318.
Venkataraman, N., Shankar, V., Ulfarsson, G., Deptuch, D., 2013. Modeling the effects of interchange configuration on heterogeneous influences of interstate geometrics on crash frequencies. Analytic Methods in Accident Research. (in preparation).
Verbeke, G., Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 91 (433), 217–221.
Viera Gomes, S., Geedipally, S., Lord, D., 2012. Estimating the safety performance of urban intersections in Lisbon. Portugal Safety Science 50 (9), 1732–1739.

Wang, C., Quddus, M, Ison, S., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accident Analysis and Prevention 43 (6), 1979–1990.
Wang, X., Kockelman, K., 2005. Use of heteroscedastic ordered logit model to study severity of occupant injury: distinguishing effects of vehicle weight and type. Transportation Research Record 1908, 195–204.
Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38 (6), 1137–1150.
Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis and Prevention 40 (5), 1674–1682.
Wang, Z., Chen, H., Lu, J., 2009. Exploring impacts of factors contributing to injury severity at freeway diverge areas. Transportation Research Record 2102, 43–52.
Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis, second edition Chapman and Hall/ CRC, Boca Raton, FL.
Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes. Journal of Risk and Uncertainty 32 (2), 83–99.
World Health Organization, 2009. Global Status Report on Road Safety: Time for Action. World Health Organization, Geneva, Switzerland.
World Health Organization, 2013. Global Status Report on Road Safety 2013: Supporting a Decade of Action. World Health Organization, Geneva, Switzerland.
Wu, Z., Sharma, A., Mannering, F., Wang, S., 2013. Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. Accident Analysis and Prevention 54, 90–98.
Xie, K., Wang, X., Ozbay, K., Yang, H., 2013. Crash frequency modeling for signalized intersections in a high-density urban road network. In: Proceedings of the Fourth International Conference on Road Safety and Simulation, Rome, Italy.
Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural networks: an empirical analysis. Accident Analysis and Prevention 39 (5), 922–933.
Xie, Y., Zhang, Y., 2008. Crash frequency analysis with generalized additive models. Transportation Research Record 2061, 39–45.
Xie, Y., Zhang, Y., Liang, F., 2009. Crash injury severity analysis using Bayesian ordered probit models. Journal of Transportation Engineering 135 (1), 18–25.
Xie, Y., Zhao, K., Huynh, N., 2012. Analysis of driver injury severity in rural single-vehicle crashes. Accident Analysis and Prevention 47, 36–44.
Xiong, Y., Mannering, F., 2013. The heteroscedastic effects of guardian supervision on adolescent driver-injury severities: a finite mixture-random parameters approach. Transportation Research Part B 49, 39–54.
Xiong, Y., Tobias, J. Mannering, F., 2013. Unobserved Heterogeneity in Vehicle Crash-Severity Modeling: A Markov Switching Random Parameters Approach. Working Paper.
Xu, C., Tarko, A., Wang, W., Liu, P., 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accident Analysis and Prevention 57, 30–39.
Yamamoto, T., Hashiji, J., Shankar, V., 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accident Analysis and Prevention 40 (4), 1320–1329.
Yamamoto, T., Shankar, V., 2004. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects. Accident Analysis and Prevention 36 (5), 869–876.
Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis and Prevention 59, 506–521.
Yasmin, S., Eluru, N., Ukkusuri, S., 2013. Alternative ordered response frameworks for examining pedestrian injury severity in New York City. Journal of Transportation Safety and Security, http://dx.doi.org/10.1080/19439962.2013.839590.
Yasmin, S., Eluru, N., Bhat, C., Tay, R., 2014. A latent segmentation generalized ordered logit model to examine factors influencing driver injury severity. Analytic Methods in Accident Research. (this issue).
Yasmin, S., Eluru, N., Pinjari, A., Tay, R., 2013. Examining Driver Injury Severity in Two Vehicle Crashes—A Copula Based Approach. Technical Paper. Department of Civil Engineering and Applied Mechanics, McGill University.
Ye, F., Lord, D., 2011. Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit. Transportation Research Record 2241, 51–58.
Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit, and mixed logit. Analytic Methods in Accident Research. (this issue).
Ye, X., Pendyala, R., Washington, S., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safety Science 47 (3), 443–452.
Ye, X., Pendyala, R., Shankar, V., Konduri, K., 2013. A simultaneous model of crash frequency by severity level for freeway sections. Accident Analysis and Prevention 57, 140–149.
Yu, R., Abdel-Aty, M., 2013a. Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions. Accident Analysis and Prevention 56, 51–58.
Yu, R., Abdel-Aty, M., 2013b. Multi-level Bayesian analysis for single- and multi-vehicle freeway crashes. Accident Analysis and Prevention 58, 97–105.
Yu, R., Abdel-Aty, M., 2013c. Utilizing support vector machine in real-time crash evaluation. Accident Analysis and Prevention 51, 252–259.
Yu, R., Abdel-Aty, M., Ahmed, M., 2013. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis and Prevention 50. (371–376, 1051).
Zajac, S., Ivan, J., 2003. Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut. Accident Analysis and Prevention 35 (3), 369–379.
Zhu, X., Srinivasan, S., 2011. A comprehensive analysis of factors influencing the injury-severity of large-truck crashes. Accident Analysis and Prevention 43 (1), 49–57.
Zoi, C., Cohen, S., Karlaftis, M., 2010. Vehicle occupant injury severity on highways: an empirical investigation. Accident Analysis and Prevention 42 (6), 1606–1620.
Zou, Y., Zhang, Y., Lord, D., 2013. Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis. Accident Analysis and Prevention 50, 1042–1051.
Zou, Y., Zhang, Y., Lord, D., 2014. Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models. Analytic Methods in Accident Research. (this issue).