Accident Analysis and Prevention 50 (2013) 1073–1081

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender

Joon-Ki Kim[a,1], Gudmundur F. Ulfarsson[b,*], Sungyop Kim[c,2], Venkataraman N. Shankar[d,3]

[a] Korea Research Institute for Human Settlements, National Infrastructure & GIS Research Division, 224 Simin-ro, Dongan-gu, Anyang-si, Gyeonggi-do 431-712, Republic of Korea
[b] University of Iceland, Faculty of Civil and Environmental Engineering, Hjardarhagi 6, IS-107 Reykjavik, Iceland
[c] University of Missouri – Kansas City, Department of Architecture, Urban Planning and Design, 5100 Rockhill Road, Kansas, MO 64110-2499, USA
[d] Pennsylvania State University, Department of Civil and Environmental Engineering, 226C Sackett Building, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 7 September 2011
Received in revised form 18 July 2012
Accepted 8 August 2012

Keywords:
Age
Gender
Single-vehicle
Heterogeneity
Mixed logit
Injury severity

ABSTRACT

This research develops a mixed logit model of driver-injury severity in single-vehicle crashes in California. The research especially considers the heterogeneous effects of age and gender. Older drivers (65+ years old) were found to have a random parameter with about half the population having a higher probability of a fatal injury given a crash than the comparison group of 25–64 year olds with all other factors than age kept constant. The other half of the 65+ population had a lower probability of fatal injury. Heterogeneity was also noted in vehicle age, but related to the gender of the driver, with males linked to, on average, a higher probability of fatal injury in a newer vehicle compared with females, all other factors kept constant. These effects lend support to the use of mixed logit models in injury severity research and show age and gender based population heterogeneity.

Several other factors were found to significantly increase the probability of fatal injury for drivers in single-vehicle crashes, most notably: male driver, drunk driving, unsafe speed, older driver (65+) driving an older vehicle, and darkness without streetlights.

1. Introduction

Single-vehicle crashes pose significant road safety risk. For example, occupant fatalities in single-vehicle crashes accounted for 46% of all motor vehicle fatalities in the United States in 2008 (NHTSA, 2010). Thus, there have been a number of studies on single-vehicle crashes to examine the crash characteristics (for example, Chang and Yeh, 2006; Chen and Chen, 2011; Jehle et al., 2007; Islam and Mannering, 2006; Schneider et al., 2009; Xie et al., 2012; Zhu et al., 2010). It has been reported that fatal single-vehicle crashes often occurred on rural roads (e.g. Xie et al., 2012) and rollover crashes result in severe injuries (e.g. Jehle et al., 2007). Chen and Chen (2011) found that, in adverse driving conditions, such as inclement weather and/or complex terrain, trucks are often involved in single-vehicle crashes in addition to multi-vehicle crashes.

There also have been studies on driver and environmental factors associated with single-vehicle crashes. Age is clearly an important factor when it comes to injury severity. Older drivers are more likely to be killed or seriously injured in traffic crashes than median age drivers (Zhang et al., 2000). Therefore, the impact of societal aging may have an adverse effect on traffic safety (Kent et al., 2003). Gender has also been reported as a significant injury factor in crashes. For example, Ulfarsson and Mannering (2004) reported that males and females as groups have significant differences with respect to injury severity. Such differences can be important; for example, crashes involving male drivers are more likely to be fatal than those of female drivers (Massie et al., 1995). The importance of gender heterogeneity cannot be ignored even though an Australian study (Leitgeb et al., 2011) found that gender had no significant influence on mortality associated with traumatic brain injury involving vehicle crash after controlling for age.

A recent study by Morgan and Mannering (2011) investigated the effects that age and gender factors have on injury severities by considering single-vehicle crashes that occurred on dry, wet, and snow/ice-covered roadway surfaces. The study found that there are substantial differences across age/gender groups under different roadway-surface conditions and argued that drivers perceive and react to pavement surface conditions in different ways based on gender and as they age. Another recent study (Obeng, 2011) reported that driver condition, type of crash, type of vehicle driven

* Corresponding author. Tel.: +354 525 4907; fax: +354 525 4632.
E-mail addresses: kimjoonki@krihs.re.kr (J.-K. Kim), gfu@hi.is (G.F. Ulfarsson), kims@umkc.edu (S. Kim), shankarv@engr.psu.edu (V.N. Shankar).
[1] Tel.: +82 310 380 0285; fax: +82 310 380 0484.
[2] Tel.: +1 816 235 1725; fax: +1 816 235 5226.
[3] Tel.: +1 814 865 9434.

0001-4575/$ – see front matter
http://dx.doi.org/10.1016/j.aap.2012.08.011

and vehicle safety features have different effects on females' and males' injury severity risks in signalized intersection crashes.

Previously, single-vehicle crashes were analyzed without injury categories being simultaneously analyzed (e.g. Dissanayake and Lu, 2002; Zhu et al., 2010). Also, other studies tend to focus on a particular type of single-vehicle crash: fixed object single-vehicle crashes (Holdridge et al., 2005), rollover crashes (Krull et al., 2000), signalized intersection crashes (Obeng, 2011), and crashes by roadway surface conditions (Morgan and Mannering, 2011) and vehicle type (Yau, 2004).

However, there have been a limited number of studies that identified and compared the factors that have different roles on injury severity of single-vehicle crashes between age groups in a comprehensive manner. Islam and Mannering (2006) and Morgan and Mannering (2011) critically analyzed gender and age related differences in injury in single-vehicle crashes. However, in the former study, individuals were assumed to be identical with regard to their factors due to one parameter being estimated for males and the effect of the parameter therefore being identical for all male drivers (Islam and Mannering, 2006). In the Morgan-Mannering study (2011), the focus was given to road surface conditions employing a mixed logit model.

The present paper intends to contribute to a better understanding of single-vehicle crashes by employing a mixed logit model with a person-specific heterogeneous mean specified as a function of age and gender, and by testing the statistical significance of such a flexibility in the model. Mixed logit models have been successfully applied to study injury severities in single-vehicle crashes as shown in Morgan and Mannering (2011) as well as other crash types (for example, Kim et al., 2010). The factors found important for injury severity in previous research allow the development of a hypotheses-based multivariate model. Previous research gives insight into the general characteristics affecting injury severity and this work expands on that research by developing a model that allows random parameters on variables and the possibility of a heterogeneous mean as a function of age and gender for those random parameter distributions.

Identifying factors affecting safety and more in-depth understanding of the differences in age and gender of drivers will help to improve transportation safety and more critically understand the effects of age and gender. This is especially necessary given the increasing number of older drivers in an aging society. This paper further analyzes the effect of heterogeneity due to age through the use of interaction terms between age and the variables in the model. There can be unobserved factors having to do with the road, vehicle, and environment that could lead to heterogeneity and random parameters. However, the effect of those is minimized as the model is conditioned on a crash already having occurred.

A primary reason to focus on single-vehicle crashes in this study is that the characteristics of multi-vehicle crashes are potentially different and require a separate study that isolates the role of driver age in the interaction between multiple vehicles in a crash. This study aims to more accurately identify and specify driver, vehicle, roadway, temporal, and environmental factors in single-vehicle crashes that have impact or play different roles in affecting injury severity while accounting for possible person-specific heterogeneity due to age and gender.

2. Data description

To perform this investigation, data on single-vehicle crashes (omitting non-private vehicles, heavy truck involved crashes, and pedestrian crashes) is drawn from all reported crashes in the State of California during the years 2003–2004 with data supplied by the California Highway Patrol. This yields 18,183 observations. To explore the data descriptively it is separated into age groups. Younger drivers are defined as 16–24 years old, working age drivers are defined as those 25–64 years old, and older drivers are defined as those 65 years or older. Table 1 shows descriptive statistics of selected key variables classified into these three age groups by injury severity outcome.

Table 1 shows the distributions of the single-vehicle crashes across injuries for younger drivers, working age drivers, and older drivers, respectively: fatal injury (2.1%, 2.5%, and 4.8%), severe injury (6.5%, 7.0%, and 5.3%), visible injury (50.1%, 47.3%, and 50.5%), and complaint of pain or no injury (41.2%, 43.2%, and 39.3%). These results show that the older drivers are represented to a greater degree in the fatal injury category relative to the other age groups of drivers. In terms of gender, male drivers are represented to a greater extent in single-vehicle crashes than female drivers, with the difference more pronounced for the younger drivers (58.5% male and 41.5% female) compared with the other age groups (working age: 53.6% male and 46.4% female; older age: 53% male and 47% female).

When considering the behavioral factors, the working age group is overrepresented in most cases. Younger drivers are overrepresented in cell phone in use and unsafe speed crashes compared to the other age groups. Older drivers have been found less likely to, for example, drink and drive or speed, and more likely to wear seat belts than the other age groups. These distributions fit prior research which has noted risk taking behavior among younger drivers and risk averse behavior among older drivers (Kostyniuk and Molnar, 2008; Horwood and Fergusson, 2000). Regarding vehicle choice, the older drivers are also more likely to be driving older vehicles than the other age groups.

Crashes occurred rarely under adverse weather conditions; they mainly occurred on clear days, which account for 75.8%, 73.9%, and 78.1% of the crashes among younger drivers, working age drivers, and older drivers, respectively. When considering lighting conditions, most crashes occurred during daylight, 42.8%, 53.1%, and 73.3% for younger drivers, working age drivers, and older drivers, respectively. Roughly more than half of the crashes reported for the younger driver group occurred during darkness (including dusk and dawn), in contrast with the much lesser fraction of older drivers (26.7%).

3. Methodology

This research uses the mixed logit model (Bhat, 1998; Train, 2003; Hensher and Greene, 2003) to explore effects of various observable characteristics on driver injury severities conditional on a single-vehicle crash having occurred. This model is chosen due to the mixed logit model's ability to capture heterogeneity through the use of random parameters. Not only that, but the mixed logit model allows explanatory variables to affect the mean of the distribution of the random parameters.

The ordered logit and probit models are constrained to find only one coefficient on each variable and it is in one direction, either towards higher severity or towards lower severity. It is not impossible to use random parameter ordered models but the ordering constraint remains. This is a constraint because it is not inconceivable that a variable can both increase the probability of low and high severity. Also, a variable can tend towards middle severities and away from the low and high severities.

The multinomial logit model relaxes this ordinal constraint on the effect of variables but brings the constraint of independence of irrelevant alternatives which results in the so-called red bus/blue bus problem (McFadden, 1974) and it does not have random parameters. The mixed logit model can relax both these constraints. The random parameter is especially important here and has seen useful applications in severity analysis (see for example,

Table 1
Descriptive statistics for key variables classified by age and injury severity.

	Driver age	Fatal injury		Severe injury		Visible injury		Complaint of pain/no injury		Total	
Driver characteristics											
Driver injury	16–24	166	(2.1)	507	(6.5)	3900	(50.1)	3206	(41.2)	7779	(42.8)
	25–64	237	(2.5)	657	(7.0)	4443	(47.3)	4054	(43.2)	9391	(51.6)
	65+	49	(4.8)	54	(5.3)	512	(50.5)	398	(39.3)	1013	(5.6)
Gender											
Male	16–24	132	(2.9)	352	(7.7)	2480	(54.5)	1583	(34.8)	4547	(58.5)
Female		34	(1.1)	155	(4.8)	1420	(43.9)	1623	(50.2)	3232	(41.5)
Male	25–64	173	(3.4)	416	(8.3)	2603	(51.7)	1840	(36.6)	5032	(53.6)
Female		64	(1.5)	241	(5.5)	1840	(42.2)	2214	(50.8)	4359	(46.4)
Male	65+	31	(5.8)	24	(4.5)	300	(55.9)	182	(33.9)	537	(53.0)
Female		18	(3.8)	30	(6.3)	212	(44.5)	216	(45.4)	476	(47.0)
Accident characteristics											
Seat belt											
Yes	16–24	32	(0.9)	190	(5.1)	1817	(49.1)	1664	(44.9)	3703	(47.6)
No		134	(3.3)	317	(7.8)	2083	(51.1)	1542	(37.8)	4076	(52.4)
Yes	25–64	62	(1.4)	203	(4.7)	1953	(45.1)	2108	(48.7)	4326	(46.1)
No		175	(3.5)	454	(9.0)	2490	(49.2)	1946	(38.4)	5065	(53.9)
Yes	65+	21	(3.9)	24	(4.4)	267	(49.2)	231	(42.5)	543	(53.6)
No		28	(6.0)	30	(6.4)	245	(52.1)	167	(35.5)	470	(46.4)
Driving drunk											
Yes	16–24	86	(4.6)	200	(10.8)	1137	(61.4)	430	(23.2)	1853	(23.8)
No		80	(1.3)	307	(5.2)	2763	(46.6)	2776	(46.8)	5926	(76.2)
Yes	25–64	100	(4.1)	264	(10.9)	1427	(58.7)	638	(26.3)	2429	(25.9)
No		137	(2.0)	393	(5.6)	3016	(43.3)	3416	(49.1)	6962	(74.1)
Yes	65+	10	(10.8)	8	(8.6)	49	(52.7)	26	(28.0)	93	(9.2)
No		39	(4.2)	46	(5.0)	463	(50.3)	372	(40.4)	920	(90.8)
Cell phone in use											
Yes	16–24	2	(1.2)	3	(1.7)	91	(52.9)	76	(44.2)	172	(2.2)
No		164	(2.2)	504	(6.6)	3809	(50.1)	3130	(41.1)	7607	(97.8)
Yes	25–64	3	(1.6)	10	(5.2)	102	(53.1)	77	(40.1)	192	(2.0)
No		234	(2.5)	647	(7.0)	4341	(47.2)	3977	(43.2)	9199	(98.0)
Yes	65+	0	(0.0)	1	(16.7)	4	(66.7)	1	(16.7)	6	(0.6)
No		49	(4.9)	53	(5.3)	508	(50.4)	397	(39.4)	1007	(99.4)
Wrong side of road											
Yes	16–24	1	(1.2)	8	(9.6)	46	(55.4)	28	(33.7)	83	(1.1)
No		165	(2.1)	499	(6.5)	3854	(50.1)	3178	(41.3)	7696	(98.9)
Yes	25–64	2	(1.9)	18	(16.7)	52	(48.1)	36	(33.3)	108	(1.2)
No		235	(2.5)	639	(6.9)	4391	(47.3)	4018	(43.3)	9283	(98.8)
Yes	65+	0	(0.0)	0	(0.0)	6	(66.7)	3	(33.3)	9	(0.9)
No		49	(4.9)	54	(5.4)	506	(50.4)	395	(39.3)	1004	(99.1)
Improper turning											
Yes	16–24	36	(3.4)	112	(10.7)	643	(61.4)	256	(24.5)	1047	(13.5)
No		130	(1.9)	395	(5.9)	3257	(48.4)	2950	(43.8)	6732	(86.5)
Yes	25–64	38	(2.7)	162	(11.5)	819	(58.1)	391	(27.7)	1410	(15.0)
No		199	(2.5)	495	(6.2)	3624	(45.4)	3663	(45.9)	7981	(85.0)
Yes	65+	5	(9.6)	4	(7.7)	27	(51.9)	16	(30.8)	52	(5.1)
No		44	(4.6)	50	(5.2)	485	(50.5)	382	(39.8)	961	(94.9)
Unsafe speed											
Yes	16–24	40	(4.9)	83	(10.2)	395	(48.5)	297	(36.4)	815	(10.5)
No		126	(1.8)	424	(6.1)	3505	(50.3)	2909	(41.8)	6964	(89.5)
Yes	25–64	46	(6.0)	77	(10.0)	406	(52.9)	238	(31.0)	767	(8.2)
No		191	(2.2)	580	(6.7)	4037	(46.8)	3816	(44.2)	8624	(91.8)
Yes	65+	5	(12.2)	2	(4.9)	23	(56.1)	11	(26.8)	41	(4.0)
No		44	(4.5)	52	(5.3)	489	(50.3)	387	(39.8)	972	(96.0)
Vehicle characteristics											
Vehicle age (years)											
0–5	16–24	51	(2.1)	133	(5.4)	1180	(47.8)	1107	(44.8)	2471	(31.8)
6–10		52	(2.1)	162	(6.5)	1226	(49.3)	1045	(42.1)	2485	(31.9)
11+		63	(2.2)	212	(7.5)	1494	(52.9)	1054	(37.3)	2823	(36.3)
0–5	25–64	69	(2.2)	186	(5.9)	1398	(44.4)	1495	(47.5)	3148	(33.5)
6–10		78	(2.7)	164	(5.7)	1401	(48.3)	1257	(43.3)	2900	(30.9)
11+		90	(2.7)	307	(9.2)	1644	(49.2)	1302	(38.9)	3343	(35.6)
0–5	65+	7	(2.6)	13	(4.8)	122	(45.0)	129	(47.6)	271	(26.8)
6–10		12	(4.0)	13	(4.3)	169	(56.1)	107	(35.5)	301	(29.7)
11+		30	(6.8)	28	(6.3)	221	(50.1)	162	(36.7)	441	(43.5)
Environmental characteristics											
Weather											
Clear	16–24	128	(2.2)	394	(6.7)	3023	(51.3)	2350	(39.9)	5895	(75.8)

Table 1 (*Continued*)

	Driver age	Fatal injury		Severe injury		Visible injury		Complaint of pain/no injury		Total	
Cloudy		31	(2.3)	89	(6.7)	632	(47.4)	582	(43.6)	1334	(17.1)
Clear	25-64	176	(2.5)	527	(7.6)	3368	(48.5)	2873	(41.4)	6944	(73.9)
Cloudy		44	(2.6)	98	(5.8)	763	(44.8)	799	(46.9)	1704	(18.1)
Clear	65+	34	(4.3)	42	(5.3)	391	(49.4)	324	(41.0)	791	(78.1)
Cloudy		13	(7.9)	8	(4.9)	91	(55.5)	52	(31.7)	164	(16.2)
Lighting											
Daylight	16–24	37	(1.1)	163	(4.9)	1620	(49.1)	1482	(44.9)	3302	(42.8)
Dusk–dawn		4	(1.5)	19	(7.2)	131	(49.6)	110	(41.7)	264	(3.4)
Dark with streetlights		71	(2.5)	226	(7.8)	1453	(50.4)	1133	(39.3)	2883	(37.3)
Dark without streetlights		53	(4.2)	97	(7.6)	665	(52.3)	456	(35.9)	1271	(16.5)
Daylight	25–64	85	(1.7)	257	(5.2)	2265	(45.8)	2341	(47.3)	4948	(53.1)
Dusk–dawn		9	(2.6)	24	(6.8)	161	(45.7)	158	(44.9)	352	(3.8)
Dark with streetlights		91	(3.4)	247	(9.2)	1345	(50.2)	998	(37.2)	2681	(28.8)
Dark without streetlights		51	(3.8)	126	(9.4)	638	(47.7)	523	(39.1)	1338	(14.4)
Daylight	65+	32	(4.3)	42	(5.7)	377	(50.9)	289	(39.1)	740	(73.3)
Dusk–dawn		2	(7.1)	1	(3.6)	15	(53.6)	10	(35.7)	28	(2.8)
Dark with streetlights		10	(5.4)	9	(4.9)	91	(49.2)	75	(40.5)	185	(18.3)
Dark without streetlights		4	(7.0)	2	(3.5)	29	(50.9)	22	(38.6)	57	(5.6)

Percentages are in parentheses.

Milton et al., 2008; Anastasopoulos and Mannering, 2011; Savolainen et al., 2011). Importantly, as a point of additional insight in terms of heterogeneity contributions to the occurrence of severity, as discussed in the introduction, age is a key factor. As we age, we become more fragile, and lose some of our abilities; however this happens in varying degrees across the population. Age itself as a number is not necessarily meaningful as a modeling construct. As a group we can note that 40 year olds are more uniform than 80 year olds, with some 80 year olds able to drive and function but with others greatly medically impaired (Meuser et al., 2009). This increasing heterogeneity with age and its importance in injury severity modeling has been shown for pedestrians (Kim et al., 2010).

The analysis in this paper is therefore based on the mixed logit model. Its formulation begins by specifying a function that determines driver-injury severity. It is written as

$$U_{ni} = \boldsymbol{\beta}_{ni} x_n + \varepsilon_{ni}, \tag{1}$$

where U_{ni} is the propensity function that determines the probability of discrete injury severity outcome i for individual driver n, x_n are explanatory variables, the vector $\boldsymbol{\beta}_{ni}$ contains the estimable parameters, and the ε_{ni} denotes a stochastic error term which is identically and independently distributed extreme value. To allow individual heterogeneity and random parameters, the parameter is written as a linear combination of a fixed parameter, a heterogeneity term, and a random term. The parameter $\boldsymbol{\beta}_{ni}$ in (1) is written as

$$\boldsymbol{\beta}_{ni} = \mathbf{m}_i + \mathbf{M}\mathbf{s}_n + \boldsymbol{\Gamma}_i \boldsymbol{\eta}_{ni} \tag{2}$$

where $\mathbf{m}_i$ is the fixed parameter that is identical for all individuals, $\mathbf{s}_n$ is a matrix of individual-specific characteristics that generate individual heterogeneity (in this work these are age and gender) in $\boldsymbol{\beta}_{ni}$, $\mathbf{M}$ is a relevant matrix of estimable parameters on the heterogeneity variables (here age and gender), $\boldsymbol{\eta}_{ni}$ is a vector of uncorrelated random variables from a chosen distribution (for a more detailed explanation, please refer to Hensher and Greene, 2003; Train, 2003). $\boldsymbol{\Gamma}_i$ is a lower triangular matrix allowing an estimable variance of $\boldsymbol{\beta}_{ni}$ and a correlation of the parameters. It is possible that one or both parameters on age and gender in $\mathbf{M}$ are not statistically significant. That indicates the lack of heterogeneity due to either age or gender or both. The mean will then be constrained to be fixed for all observations. The standard deviation of the random parameter with a fixed mean will be tested. If the standard deviation is also not found statistically significantly different from zero, the parameter will not be found random and not suffering from effects of unobserved heterogeneity.

The probability of individual n suffering injury severity i conditional on a crash having occurred and unconditioned on $\boldsymbol{\eta}_{ni}$ (the random term) is

$$P_{ni}(\mathbf{x}_n|\boldsymbol{\beta}_{ni}) = \int \frac{\exp(\boldsymbol{\beta}_{ni}\mathbf{x}_n)}{\sum_{j=1}^{I}\exp(\boldsymbol{\beta}_{nj}\mathbf{x}_n)} f(\boldsymbol{\eta}_{ni}|\boldsymbol{\Omega}_n)\, d\boldsymbol{\eta}_{ni}, \tag{3}$$

where $f(\boldsymbol{\eta}_{ni}|\boldsymbol{\Omega}_n)$ is the joint density function of $\boldsymbol{\eta}_{ni}$. Eq. (3) can be estimated with the maximum simulated likelihood estimation (MSLE) method (Walker and Ben-Akiva, 2002).

The mixed logit model was estimated with NLOGIT 4.0 (Hensher and Greene, 2003) using 500 Halton draws (see Train, 2003, for explanation of Halton draws) and the typical simulation run took about 4 h to complete on a Dell Latitude D500 Dual Core computer. Each test of one coefficient as a random parameter required such a run. During model development numerous runs are therefore required, on the order of a hundred runs. The total computation time therefore is on the order of four months. This is a real limitation on the number of random parameters that can be tested or explored. Exploring more heterogeneity causes, such as environmental conditions, the road, etc., would increase the computation time considerably.

Of practical concern is how it is determined if a variable has a random parameter or not, and then the distribution of that random parameter. Here it is done in a stepwise fashion. Each variable is tested in the model as either fixed or random. Statistical testing of the improvement in likelihood is used to determine the best fit. This cannot be done completely in a linear fashion, i.e. testing all variables in turn, since the order of those tests can matter. An iterative approach to testing parameters for randomness and the distribution is therefore used, cycling through the stepwise procedure until the model is stable to changes.

The present model has four possible injury severity outcomes. The model parameters themselves cannot be readily interpreted in that case since a positive parameter may in fact lead to a reduction in probability. This can be proven mathematically (Greene, 2011). Furthermore, due to the additions of random parameters with heterogeneous means, the effect of a variable that enters both into such a mean and is standalone in the model will be hard to interpret by directly viewing only the estimated parameters. For this reason it is helpful to consider the marginal effect of each variable on the probabilities of the different injury severity outcomes.

In this study, the observed variables are all 0 and 1 indicator variables. To uncover the marginal effect of the kth indicator variable we examine the change in estimated probability of injury severity due to a change in the Boolean status of the variable in question (that is a switch from a value of zero to a value of one):

$$E_{x_{nk}}^{P_{ni}} = \frac{P_{ni}[\text{given}\, x_{nk} = 1] - P_{ni}[\text{given}\, x_{nk} = 0]}{P_{ni}[\text{given}\, x_{nk} = 0]}. \quad (4)$$

This is called the direct pseudo-elasticity of the probability with respect to the explanatory variable (see Ulfarsson and Mannering, 2004). It is a point value, i.e. it is calculated for each observation. To represent the entire dataset, the average direct pseudo-elasticity for the data is calculated and reported. The interpretation of these values is straightforward; for example a 0.1 average direct pseudo-elasticity of a variable on a probability means that the probability increases by 10% on average when the variable is switched from 0 to 1.

4. Results

The estimation results for the mixed logit model are presented in Table 2. When a parameter on a variable was not found statistically significantly different from zero at the 0.1 level it was constrained out of the model. Furthermore, if a set of parameters on a variable were not different across injury severities at the 0.1 level, those parameters were constrained to be the same across injury severities.

To investigate the statistical significance of the mixed logit model, a likelihood ratio test was performed to compare the results with the same specification estimated with the multinomial logit model (McFadden, 1974). The result was a likelihood ratio statistic of 17.26 with four degrees of freedom (the multinomial logit model yielded a log-likelihood at convergence of −17,026 with 47 parameters). This result shows that the mixed logit model is superior with 99.5% confidence. In exploring differences between the multinomial logit model and the mixed logit model, it is evident that most parameters are quite similar between the two. The most notable differences were for the fatal injury equation. However, we conclude that the difference is not large and that the primary benefit of the mixed logit model in this setting is the greater interpretive power given to the researcher through the random parameters.

In addition to exploring the heterogeneity in parameters due to age and gender via the random parameter framework in the model, the variables were tested for age related heterogeneity through the use of interaction terms. The three age groups discussed in the data description were used as interactions with the other variables (driver age groups: 16–24, 25–64, 65+). An interaction term is created by multiplying the age group indicator with each of the variables. The results show that the use of such interaction variables tends to make the random parameter heterogeneity insignificant. That is, interaction variables serve to also capture heterogeneity and can be used as a way of representing heterogeneity within a multinomial logit model and thereby avoiding the complexity of the random parameters in the mixed logit model. However, interaction terms also present a complexity of their own, as filling the model with interaction terms quickly yields a model that can be difficult to interpret. It is our conclusion that the mixed logit model produces a cleaner model than a multinomial logit model with interaction terms on all variables, yet supplies a rich interpretation of the effects under study.

The overall result in terms of the heterogeneity is that the parameters on three variables are found to be affected by heterogeneity. The older driver (65+) variable receives a random parameter although its mean is fixed. The newer vehicle (0–5 years old vehicle) variable receives a random parameter with a heterogeneous mean that is a function of gender. There is one interaction term that captures heterogeneity, the older driver in an older vehicle. These effects are interpreted in the following sections. This shows that heterogeneity is statistically significant but the effect is not as large as was expected, with only three parameters affected by heterogeneity.

As noted in Section 3, the parameters of the mixed logit model are not readily interpreted as direct effects on the probability, i.e. a positive parameter may reduce a probability. The parameters in Table 2 are therefore used in Eq. (4) to calculate the average direct pseudo-elasticity of each variable on each probability. These results are displayed in Table 3 and aid the interpretation of the model results, discussed in the following sections. The interpretation of an average direct pseudo-elasticity, e.g. 20%, for a variable on a probability means that when the variable is switched from zero to one, the probability increases by 20% on average.

4.1. Driver characteristics

Drivers of ages 25–64 were used as a comparison group. The model results show that drivers aged 65 and over are more likely to be fatally injured when they are in single-vehicle crashes than the other age groups of drivers (Table 3). Also, the parameter for older driver is found to be random (with a uniform distribution). The parameters of the distribution are estimated to have a mean of 0 and standard deviation of 2.655 for fatal injury, which indicates that individual drivers among older drivers aged 65+ have different parameters and are thereby heterogeneous.

In the multinomial logit model (suppressed due to space limitation), the parameter of older drivers aged 65+ is 0.829. That means, all older drivers are more likely to be fatally injured than other age groups of drivers when in a crash. But in the mixed logit model (Table 2) the older driver group is heterogeneous. Because the mean of the random distribution of the heterogeneity is at zero, half of the distribution results in positive parameters and half results in negative parameters. The mixed logit model thereby indicates that about half of older drivers have a higher probability to be fatally injured than the working age group, but the other half has a lower probability, all other variables being kept constant. This indicates greater variance in the older driver group, and this greater variance lends statistical support to the hypothesis that older drivers are more heterogeneous than the other age groups. This result is reasonable in part because older drivers can have varying degrees of interactions under crash conditions with injury defenses, e.g. seatbelts, airbags, vehicle impact resistance, which reduce dependence on a driver's body to withstand shock. The driver's body is increasingly fragile with age but in varying degrees across the population. Fragility plays a significant role in exacerbating injury severity. This effect is not found significant for the younger drivers. Similar results were discovered in a study of pedestrian-injury severity (Kim et al., 2010).

Male drivers are associated with increased probability of fatal injury in single-vehicle crashes (107% in Table 3). This is a common result and it is generally found that males are overrepresented in the fatal injury category (e.g. Ulfarsson and Mannering, 2004).

4.2. Crash characteristics

Similar to what has been found in other studies (e.g. Ulfarsson and Mannering, 2004), seat belt usage reduced the probability of serious injury in crashes (fatal injury: −60%, severe injury: −42%, visible injury: −5%, and complain of pain or no injury: 22%, in Table 3).

Table 2
Mixed logit model of single-vehicle crash injury severities.

	Fatal injury			Severe injury			Visible injury		
Alternative specific constant	−3.249	(0.140)	‡	−2.616	(0.102)	‡	−0.192	(0.044)	‡
Driver characteristics									
Younger driver, age 16–24	−0.199	(0.108)					0.107	(0.032)	‡
Older driver, age 65+							0.313	(0.074)	‡
Std. dev. of distribution of an older driver	2.655	(0.476)	‡						
Male	0.690	(0.137)	‡	0.426	(0.068)	‡	0.419	(0.033)	‡
Crash characteristics									
Seat belt	−1.131	(0.117)	‡	−0.752	(0.067)	‡	−0.254	(0.033)	‡
Drinking	1.014	(0.116)	‡	0.743	(0.086)	‡	0.736	(0.049)	‡
Cell phone in use				−0.636	(0.276)	†			
Physical impairment				0.477	(0.284)				
Wrong side of road				0.653	(0.216)	‡			
Improper turning				0.520	(0.098)	‡	0.219	(0.057)	‡
Unsafe speed	0.754	(0.127)	‡	0.437	(0.099)	‡			
Vehicle characteristics									
0–5 year old vehicle	−0.806	(0.229)	‡	−0.157	(0.039)	‡			
Heterogeneity in the mean: male	0.706	(0.260)	‡						
11+ year old vehicle				0.403	(0.067)	‡	0.087	(0.039)	†
Older driver (age 65+) & 11+ year old vehicle	0.624	(0.359)							
Roadway characteristics									
Intersection	−0.480	(0.224)	†				−0.148	(0.062)	†
US highway	−0.429	(0.214)	†						
State route				0.197	(0.083)	†			
Wet road surface	−0.604	(0.179)	‡	−0.540	(0.108)	‡	−0.319	(0.049)	‡
Environmental characteristics									
Cloudy	0.223	(0.132)							
Dark with streetlights	0.234	(0.128)		0.282	(0.080)	‡	−0.083	(0.037)	†
Dark without streetlights	0.702	(0.141)	‡	0.371	(0.094)	‡			
Temporal characteristics									
6:00 AM–9:59 AM				0.216	(0.100)	†			
Weekend	0.288	(0.123)	†	0.326	(0.077)	‡	0.139	(0.041)	‡
Spring (March–June)				0.214	(0.092)	†	0.098	(0.038)	†
Summer (June–August)				0.375	(0.091)	‡	0.138	(0.039)	‡
Winter (December–February)				0.299	(0.092)	‡			
Number of observations: 18,086									
Log-likelihood at zero: −25072.52, Log-likelihood at convergence: −17017.29									
Adjusted ρ^2: 0.32									

All variables are 1/0 indicators. For example, the "younger driver" is one if the motorist was in between 16 and 24 years old and zero if not. Standard errors of parameter estimates are in parentheses. All variables are significant at the 0.1 level: ‡ at the 0.01 level and † at the 0.05 level. Parameters that were not significant at the 0.1 level were restricted to zero and omitted from table. The "complain of pain or no injury" alternative is the base case with its parameters set at zero.

When drinking was involved in crashes, the probability of a more severe injury increased (fatal injury: 73%, severe injury: 32%, visible injury: 31%, and complain of pain or no injury: −37%, in Table 3). Note that driver intoxication could be endogenous because drinking is related with risk taking behavior. So, the effects of drinking could be overestimated.

Cell phone use slightly increases the probability of fatal injury. Although studies indicate that in terms of crash risk, cell phone usage has detrimental effects (e.g. Laberge-Nadeau et al., 2003), cell phone usage is here found to be minimally related to injury severity in single-vehicle crashes.

The speed at crash is an important variable when considering injury severity. Unfortunately, an accurate estimate of the speed of the crash is rarely entered in crash databases. Such a speed variable is not available here. However, other variables work to capture this effect, namely road classifications and the variable unsafe speed, which remains statistically significant in the model. In crashes where unsafe speed was noted, there was an increase in the probability of fatal and severe injury, 100% and 46%, respectively (in Table 3), showing the considerable severity implications of unsafe speed.

The literature (e.g. McGwin and Brown, 1999) has noted improper turning as a difficulty among some older drivers in terms of crash frequency. This study investigated the relationship of improper turning and age using interaction variables for improper turning with drivers' age. All such effects turned out to be not significantly different from zero at the 0.1 level. Still, improper turning is overall associated with increased probability of severe injuries in single-vehicle crashes (in Table 2).

4.3. Vehicle characteristics

It was found that driver gender accounted for heterogeneity around the mean of the estimated parameter for the newer vehicle (0–5 year old vehicle) indicator variable. The positive parameter of heterogeneity around the mean for newer vehicles indicates that although newer vehicles can reduce injury severity in single-vehicle crashes (fatal injury: −23% and severe injury: −6%, in Table 3), this effect is reduced among male drivers (Table 2). For example, if a female driver riding newer vehicle is involved in a single-vehicle crash, the estimated parameter of "0–5 year old vehicle" is −0.806, a male driver, however, is −0.1 ($-0.1 = -0.806 + 0.706$). This result indicates that males in new vehicles may drive more aggressively than females, so the added safety benefit of a newer vehicle is not fully realized for male drivers.

When older drivers who drive older vehicles (vehicle 11 years old and over) are involved in single-vehicle crashes, they are more

Table 3
Average direct pseudo-elasticities of the variables.

Variables	Fatal injury	Severe injury	Visible injury	Complain of pain or no injury
Driver characteristics				
Younger driver, age 16–24	−22%	−5%	6%	−5%
Older driver, age 65+	105%	−17%	13%	−17%
Male	107%	19%	18%	−22%
Crash characteristics				
Seat belt	−60%	−42%	−5%	22%
Drinking	73%	32%	31%	−37%
Cell phone in use	3%	−45%	3%	3%
Physical impairment	−4%	55%	−4%	−4%
Wrong side of road	−6%	81%	−6%	−6%
Improper turning	−14%	45%	7%	−14%
Unsafe speed	100%	46%	−6%	−6%
Vehicle characteristics				
0–5 year old vehicle	−23%	−6%	−6%	10%
11+ year old vehicle	−7%	39%	2%	−7%
Older driver (age 65+) & 11+ year old vehicle	83%	−2%	−2%	−2%
Roadway characteristics				
Intersection	−33%	8%	−7%	8%
US highway	−34%	1%	1%	1%
State route	−1%	20%	−1%	−1%
Wet road surface	−33%	−29%	−12%	22%
Environmental characteristics				
Cloudy	24%	−1%	−1%	−1%
Dark with streetlights	28%	34%	−7%	1%
Dark without streetlights	92%	38%	−5%	−5%
Temporal characteristics				
6:00 AM–9:59 AM	−2%	22%	−2%	−2%
Weekend	21%	26%	4%	−9%
Spring (March–June)	−6%	16%	4%	−6%
Summer (June–August)	−9%	32%	4%	−9%
Winter (December–February)	−2%	32%	−2%	−2%

Shading indicates the largest changes in probability due to a marginal change in a variable.

likely suffer from fatal injury (83% in Table 3). This result could be reasoned by an older driver's fragility and lesser safety features of an older vehicle.

4.4. Roadway characteristics

A wet road surface was significant for injury severity. This variable shows a linear distribution where it increased the probability of injury toward complain of pain or no injury (fatal injury: −33%, severe injury: −29%, visible injury: −12%, and complain of pain or no injury: 22%, in Table 3). A wet road surface is likely to impact crash risk more so than injury severity, except through the effect of a wet road on braking or maneuvering which can affect injury severity. This result indicates that drivers may be appropriately compensating for a wet road surface, e.g. by driving slower than on a dry road.

4.5. Environmental characteristics

Older drivers have a tendency to avoid driving during adverse conditions, such as driving at night, and Table 1 shows that tendency compared to the younger group and the working age group. Lighting conditions will primarily affect crash risk but can also affect injury severity since reduced visibility can lead to drivers braking later or taking less effective avoidance maneuvers, leading to a more severe crash. Darkness with or without streetlights showed up as significantly increasing the probability of fatal injury and severe injury (28% and 92%, respectively, in Table 3).

Cloudy weather conditions increase the probability of fatal injury (24% in Table 3). Drivers seem more affected by the relatively little reduction in visibility due to clouds, compared with clear skies. We can postulate that cloudy weather may reduce color contrast and that leads to the increase of probability of fatal injury. This effect is one of the reasons many nations have a mandatory daylight running lights law. The lights on the vehicle will make it stand out more even in daylight and that is a safety benefit. Such vision related factors will mostly affect crash risk, which is not under study in this paper, but as explained for darkness, vision factors can also lead to a less effective response by the driver, i.e. braking or avoidance maneuvers, which in turn can lead to a more severe crash. The significance of this variable indicates that drivers are not compensating for cloudy conditions in their driving, as they compensate for other conditions such as rain.

4.6. Temporal characteristics

Weekends were associated with an increase of the probability of fatal injury and severe injury (21% and 26%, respectively, in Table 3). This result may be caused by driver behavior and type of activities on weekends compared with weekdays, which include more commuting on busier roads.

5. Discussion and conclusions

This study examines characteristics associated with injury severity of drivers conditional on a single-vehicle crash having occurred. The characteristics describe driver, crash, vehicle, roadway, environmental, and temporal effects. A motivating hypothesis was that age and gender lead to unobserved heterogeneity, especially age. The results indicate this to a degree, with three parameters being affected by such heterogeneity.

A random parameter with a uniform distribution is found statistically significant for older drivers (aged 65 and over) and the fatal injury outcome. The result indicates that for about half the older population, the probability of a fatal injury is higher than for the middle age group, and for the other half the probability is lower. This indicates that conditional on a crash occurring, older age is associated with both greater and lesser chance of fatality compared with the working age group. This fits the duality that older drivers as a group may be the safest drivers, yet some among them are quite fragile, even medically impaired and at greater risk of a crash (Meuser et al., 2009).

A random parameter was also significant for the variable representing new to 5 year old vehicles (Table 2). Here it was not age, but rather gender that was linked to the heterogeneity identified in this variable. A newer vehicle is associated with lesser injury severity in general but that effect on fatal injury is somewhat reduced for male drivers.

The identified offset effect (compared with females) on injury severity for male drivers driving a new to 5 year old vehicle is another argument in favor of using the mixed logit model to explore injury severity, since such effects are difficult to find with the multinomial logit without using numerous interaction terms. This offset effect can also be supported by the finding of Winston et al. (2006). They found that drivers' risk compensation could render the benefits of safety devices insignificant.

The safety benefits of the population driving newer vehicles with better safety mechanisms can also inform policy makers. A policy that enables drivers to remove older, less safe cars from the road and upgrade to safer vehicles can have overall safety benefits. These types of policies have been instituted, e.g. by Germany and the UK but in response to economic crisis as a means to fuel the economy (Massey, 2009). However, these programs can also be expected to benefit traffic safety. However, as the mixed logit model is able to show, there exists a certain compensation effect for males, which suggests the need for direct marketing and education efforts to reduce male driver tendency to compensate for added safety by riskier behavior.

The results show, as does much other research, the benefits of seatbelts, the adverse effects of drunk driving, and of speeding. The research therefore lends support to the continued efforts to enforce seatbelt use, curtail drunk driving, and reduce speeding.

Darkness, especially without streetlights, is a factor that is associated with increased severity of crashes. This is not a crash risk interpretation as this study is conditioned on a crash having occurred. This indicates that from a policy and education standpoint, drivers need to be educated to compensate properly for darkness. If they did compensate, e.g. by reducing speed as necessary, darkness would not appear significantly different from daylight. Also, the results indicate a beneficial safety effect from streetlights compared to darkness without streetlights. Importantly, this holds when controlling for roadway type, so this is not merely a road type effect.

Taken together, the research has shown that population heterogeneity is caused by age and gender, especially older age (65+) when it comes to single-vehicle crash severities. This lends support for the use of the mixed logit model in severity modeling. Key topics for future policy are to: account for increasing population variance with demographic change; acknowledge and account for gender differences, support population move towards newer and safer vehicles; continue efforts to increase seatbelt use, reduce drunk driving and speeding, and improve street lighting.

Acknowledgments

The authors thank the Department of California Highway Patrol for providing the data used in this study. The authors thank the Korea Research Institute for Human Settlements for providing research support and the institute's National Infrastructure & GIS Research Division for providing facilities.

References

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.

Bhat, C., 1998. Accommodating variations in responsiveness to level-of-service variables in travel mode choice models. Transportation Research A 32 (7), 455–507.

Chang, H.-L., 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.

Chen, F., Chen, S., 2011. Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. Accident Analysis and Prevention 43 (5), 1677–1688.

Dissanayake, S., Lu, J.J., 2002. Factors influential in making an injury severity difference to older drivers involved in fixed object–passenger car crashes. Accident Analysis and Prevention 34, 609–618.

Greene, W.H., 2011. Econometric Analysis, 7th ed. Prentice Hall, New Jersey, 1232 pp.

Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30 (2), 133–176.

Holdridge, J.M., Shankar, V.N., Ulfarsson, G.F., 2005. The crash severity impacts of fixed roadside objects. Journal of Safety Research 36 (2), 139–147.

Horwood, L.J., Fergusson, D.M., 2000. Drink driving and traffic accidents in young people. Accident Analysis and Prevention 32 (6), 805–814.

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, 267–276.

Jehle, D., Kuebler, J., Auinger, P., 2007. Risk of injury and fatality in single vehicle rollover crashes: danger for the front seat occupant in the outside arc. Academic Emergency Medicine 14 (10), 899–902.

Kent, R., Funk, J., Crandall, J., 2003. How future trends in societal aging, air bag availability, seat belt use, and fleet composition will affect serious injury risk and occurrence in the United States. Traffic Injury Prevention 4 (1), 24–32.

Kim, J.-K., Ulfarsson, G.F., Shankar, V.N., Mannering, F.L., 2010. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. Accident Analysis and Prevention 42 (6), 1751–1758.

Kostyniuk, L.P., Molnar, L.J., 2008. Self-regulatory driving practices among older adults: health, age and sex effects. Accident Analysis and Prevention 40, 1576–1580.

Krull, K., Khattak, A.J., Council, F.M., 2000. Injury effects of rollovers and events sequence in single-vehicle crashes. Transportation Research Record 1717, 46–54.

Laberge-Nadeau, C., Maag, U., Bellavance, F., Lapierre, S.D., Desjardins, D., Messier, S., Saïdi, A., 2003. Wireless telephones and the risk of road crashes. Accident Analysis and Prevention 35, 649–660.

Leitgeb, J., Mauritz, W., Brazinova, A., Janciak, I., Majdan, M., Wilbacher, I., Rusnak, M., 2011. Effects of gender on outcomes after traumatic brain injury. Journal of Trauma – Injury, Infection and Critical Care 71 (6), 1620–1626.

Massey, R., 2009. A £2,500 handout to scrap your old car: ministers back scheme to kickstart motor industry. Mail Online, The Daily Mail. March 11. http://www.dailymail.co.uk/news/article-1161089/2-500-scrap-old-car-Ministers-scheme-kickstart-motor-industry.html# (accessed 06.09.11).

Massie, D.L., Campbell, K.L., Williams, A.F., 1995. Traffic accident involvement rates by driver age and gender. Accident Analysis and Prevention 27 (1), 73–87.

McFadden, D.L., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

McGwin, G., Brown, D.B., 1999. Characteristics of traffic crashes among young, middle-aged, and older drivers. Accident Analysis and Prevention 31, 181–198.

Meuser, T.M., Carr, D.B., Ulfarsson, G.F., 2009. Motor-vehicle crash history and licensing outcomes for older drivers reported as medically impaired in Missouri. Accident Analysis and Prevention 41, 246–252.

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.

Morgan, A., Mannering, F.L., 2011. The effects of road-surface conditions, age, and gender on driver-injury severities. Accident Analysis and Prevention 43 (5), 1852–1863.

NHTSA, 2010. Traffic Safety Facts: 2008 Data Overview. DOT HS 811 162. National Center for Statistics and Analysis, National Highway Traffic Safety Administration, Washington, DC. http://www-nrd.nhtsa.dot.gov/Pubs/811162.pdf.

Obeng, K., 2011. Gender differences in injury severity risks in crashes at signalized intersections. Accident Analysis and Prevention 43 (4), 1521–1531.

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.H., Savolainen, P.T., 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.

Xie, Y., Zhao, K., Huynh, N., 2012. Analysis of driver injury severity in rural single-vehicle crashes. Accident Analysis and Prevention 47, 36–44.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.

Ulfarsson, G.F., Mannering, F.L., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Analysis and Prevention 36, 135–147.

Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Mathematical Social Science 43 (3), 303–343.

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.

Yau, K.K.W., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accident Analysis and Prevention 36, 333–340.

Zhang, J., Lindsay, J., Clarke, K., Robbins, G., Mao, Y., 2000. Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. Accident Analysis and Prevention 32, 117–125.

Zhu, H., Dixon, K.K., Washington, S., Jared, D.M., 2010. Single-vehicle fatal crash prediction for two-lane rural highways in the Southern United States. Transportation Research Record 2147, 88–96.