TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–10

DOI: 10.1177/0361198119841570
journals.sagepub.com/home/trr
SAGE

Elderly Pedestrian Fatal Crash-Related Contributing Factors: Applying Empirical Bayes Geometric Mean Method

Subasish Das[1], Apoorba Bibeka[1], Xiaoduan Sun[2], Hongmin "Tracy" Zhou[3], and Mohammad Jalayer[4]

Abstract
Recent statistics show that around 20% of all pedestrian fatalities (1,002 out of 5,376) in 2015 were pedestrians over the age of 65. There is a need to identify issues associated with elderly pedestrian crashes to develop effective countermeasures. This study aimed to determine the key associations between contributing factors of elderly pedestrian crashes. The authors analyzed three years (2014 to 2016) of elderly pedestrian fatal crashes from the Fatality Analysis Reporting System in the United States by using empirical Bayes (EB) data mining. The findings of this study revealed several association patterns with high crash potential for elderly pedestrians that include backing vehicle-related crashes for female pedestrians (especially those aged 79 and above), segment-related crashes at night for 65 to 69 year-old male pedestrians, crossing an expressway at night for male pedestrians, especially the 65 to 69 year group, failure to yield while crossing at intersections, and crashes occurring in the dark with poor street lighting. The findings of this study could help authorities determine effective countermeasures for this group of vulnerable road users.

The number of people older than 64 years is expected to grow to over 83.7 million in 2050, "almost double its estimated population of 43.1 million" in 2012 (*1*). Pedestrian fatalities increased 27% from 2007 to 2016 in the United States (*2*). Additionally, according to a National Highway Traffic Safety Administration (NHTSA) report "19 percent of all pedestrian fatalities (1,002 out of 5,376) in 2015 were pedestrians over the age of 65" (*3*). Issues associated with an increasing elderly pedestrian population, and the overrepresentation of elderly pedestrians in fatal crashes warrant investigation. Identification of clusters and patterns in large datasets could provide information about a group of factors and their association with crash outcomes. Therefore, research efforts are required that will utilize newer data resources and analytical tools to better understand the patterns of elderly pedestrian crashes.

In contrast to conventional statistical modeling, in which a structural relation is developed, the data mining approach assumes no definite structure but follows different algorithms to determine the key associations from complex datasets. The objective of this research was to identify the association between key crash attributes and different age groups of elderly pedestrians. To accomplish this, the researchers applied an empirical Bayes (EB) data mining approach: a data mining approach because of its capability to identify key patterns from a large pool of potential patterns. Data mining is a popular tool among traffic safety researchers owing to its ability to identify significant associations from complex crash datasets. This study used three years (2014 to 2016) of elderly pedestrian fatal crash data from the Fatality Analysis Reporting System (FARS). The advantage of this database is the level of detail it provides about crashes. The findings from this study could help authorities to determine effective countermeasures for this group of vulnerable road users.

Earlier Work and Research Context

In recent years, research on vulnerable roadway users (pedestrians and bicyclists) has increased significantly.

[1]Texas A&M Transportation Institute, College Station, TX
[2]University of Louisiana at Lafayette, Lafayette, LA
[3]Texas A&M Transportation Institute, Houston, TX
[4]Department of Civil and Environmental Engineering, Rowan University, Glassboro, NJ

Corresponding Author:
Address correspondence to Subasish Das: s-das@tti.tamu.edu

However, elderly pedestrians have received very little research interest. This section summarizes some of the studies in the field of highway safety, aiming to understand the association of significant contributing factors to vulnerable road-user fatal crashes. The studies use different age limits in defining elderly pedestrians. For instance, some studies considered people over 65 as elderly pedestrians (*4*) whereas others defined it as people over 60, or 62 (*5*, *6*). A study by Siram et al. helps determine this limit by showing a statistical difference in the injuries sustained and mortality rates of pedestrians over the age of 65 (*7*). The authors found that pedestrians over 65 years were more likely to have crash-related skeletal fractures. The findings indicate that 65 years is a reasonable cut-off for defining elderly pedestrians.

Several other studies demonstrated that elderly pedestrians are prone to fatal injuries (*4*, *8–10*). Kim et al. attributed this to the physical deterioration caused by age (*8*). Moreover, driving simulator-based studies show that elderly pedestrians tend to have an unsafe margin while crossing one and two-way streets (*5*, *11*, *12*). The situation is worsened when the vehicles are approaching at high speed. Elderly pedestrians find it difficult to factor in the time gap of an approaching vehicle in the far lane: Liu et al. found that pedestrians generally consider distance rather than the available time gap (*13*). In addition, elderly pedestrians are unable to compensate by speeding up after making an unsafe crossing decision. Oxley et al. similarly found considerations of distance over time gap in elderly pedestrians when they conducted a simulator-based study (*14*). In a previous study, they found that elderly pedestrians choose similar safety margins as younger pedestrians in less complicated situations (*15*). Dommes et al. attributed this to motor skill deterioration (*12*). Studies based on crash data show similar results. Zegeer et al. found that a greater number of elderly pedestrians are involved in fatal crashes while crossing wider streets (e.g., four to five lanes) (*4*). Niebuhr et al. found that elderly pedestrians have a higher risk of being severely injured in full-frontal crashes at various collision speeds (*6*). Eluru et al. found that age was one of the critical variables in determining injury severity (*16*). Rosenbloom et al. evaluated the effect of socioeconomic indicators on the safety behavior of elderly pedestrians. They found that elderly pedestrians from a high socioeconomic level within the city made better safety crossing decisions than elderly pedestrians of a low socioeconomic level (*17*). In general, female pedestrians suffer higher injury severity than male pedestrians and are less likely to survive a crash with vehicles and more likely to be involved in a fatal crash than their male counterparts (*18–20*).

The above studies suggest various causations of near- and far-side crashes. Geraghty et al. conducted a study to identify the mobility and cognitive measures responsible for near- and far-side crashes in elderly pedestrians (*21*). These measures included walking speed, startup delay, visual attention, spatial planning, reaction time, spatial working memory, and updating abilities. This study indicated that elderly women were more likely to be involved in near-side crossing errors as they potentially have less driving experience than elderly men. Moreover, balance ability, processing speed, and inhibition were correlated with near-side crossing errors. Elderly pedestrians were also involved in more far-side crossing errors. Declining spatial planning abilities were predictive of far-side crossing errors. Walking speed and self-rated mobility were effective in identifying both near- and far-side errors.

The literature review indicates a research gap in identifying key contributing factors and frequent crash scenarios of elderly pedestrian crashes. This study applied a robust data mining tool, empirical Bayes geometric mean (EBGM), to quantify the significant crash potentials from three years of FARS elderly pedestrian crash data. This database has some unique features that can help identify the crash location and precise crash types that are strongly associated with fatal crash occurrences.

Data Description

Traffic crash files at state level contain insufficient information about types of pedestrian crashes. This practice hampers the development of effective countermeasures to prevent pedestrian crashes. To mitigate this problem, Pedestrian Crash Typing was developed to describe the pre-crash actions of the involved parties to better define the sequence of events and precipitating activities leading to crashes between motor vehicles and pedestrians (*22*). The FARS data used is maintained by the NHTSA. In 2010, NHTSA integrated parts of a stand-alone crash typing application, the Pedestrian and Bicycle Crash Analysis Tool, into FARS. Since 2014, FARS has incorporated crucial information about pedestrian crashes. These variables are:

- PB30 – Crash Type—Pedestrian
- PB31 – Crash Location—Pedestrian
- PB32 – Pedestrian Position
- PB33 – Pedestrian Initial Direction of Travel
- PB34 – Motorist Initial Direction of Travel
- PB35 – Motorist Maneuver
- PB36 – Intersection Leg
- PB37 – Pedestrian Scenario
- PB38 – Crash Group—Pedestrian

Figure 1 depicts the methodology flowchart applied in this study. Pedestrian crash data were extracted first from FARS "PBTYPE" data. To identify the elderly

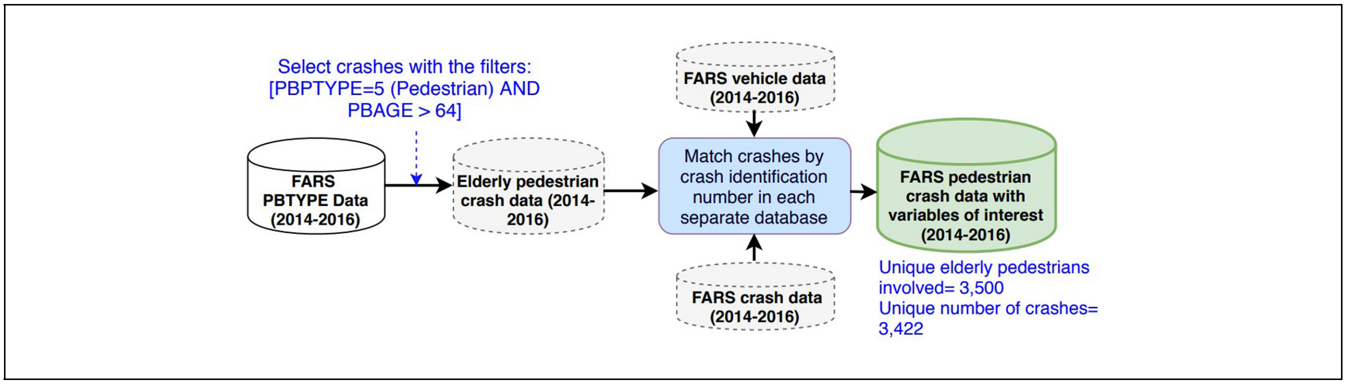

Figure 1. Flowchart of the methodology.

pedestrians, an age threshold (age 65 and older) was applied. Later, both crash and vehicle crash data were integrated with the elderly pedestrian data through matching with the crash identification number. The final database contained 340 features for crash information and 350 features for elderly pedestrian information.

The data preparation is shown in the flowchart of Figure 1. The primary objective of this study was to examine the crash type, pedestrian position, gender, age group, and other critical contributing factors to determine the associations between different categories of variable. The primary dataset contained 3,422 elderly pedestrian crash records with 23 variables. Some of the variables were omitted owing to redundancy. Based on the correlation analysis, five key variables were selected for further analysis including crash type, key crash groups, lighting condition, gender, and age of the pedestrians.

Data Analysis and Discussion

Descriptive Statistics

The FARS database has 16 crash groups and 55 crash types associated with pedestrians. The crash types were categorized into several groups of factors: vehicle, roadway, and pedestrian-related factors, lighting conditions, and gender. As the elderly pedestrians vary in perception-reaction tasks while walking, four major age groups were developed for this analysis: 65 to 69, 70 to 74, 75 to 79 and >79. Table 1 shows the percentage distributions of pedestrian crashes according to key variables by different age groups. All four elderly age groups have the highest percentages of crashes related to crossing roadways when the involved vehicle is not turning. Very elderly pedestrians (age 75 and older) are highly associated with crossing roadway (vehicle turning) crashes. This is evidenced by research from Zegeer et al. (*4*). Pedestrians of this group are also involved more in backing vehicle crashes. Pedestrians aged 70 and above are more likely to be involved in crashes when crossing a roadway than moving along a roadway, a finding that is supported by Kitali et al. (*10*). Pedestrians in the age group 79 years and above are overrepresented in driveway-related crashes. Crashes that occurred on roadways in the dark with street lighting are highly represented in the 65 to 79 age group. The 65 to 69 age group has the highest percentage of crashes on roadways in the dark with street lighting, whereas the 79 and above age group has the highest rate of daytime crashes and higher percentages of dawn and dusk crashes. This may indicate that the relatively younger elderly pedestrians are more likely to walk at night whereas very elderly pedestrians may try to avoid doing so. Male pedestrians show lower proportions than female pedestrians as their ages advance. This may be on account of very elderly women tending to be more active pedestrians than very elderly men, as older women tend to be in better physical condition (*19*).

Figure 2 shows the slope graph of the key crash types by age group. Out of the 55 crash types, the 25 crash types with higher proportions are kept as the described crash types in the FARS dataset; the other 30 crash types were merged to develop the crash type “others.” A slope graph ranks the factors (crash types) responsible for pedestrian fatalities by frequency for each age group. A factor that has the highest frequency for an age group is ranked 1. A slope graph also tracks the rank of different factors across age group. The highest number of pedestrian fatalities across all age groups was caused by the pedestrian’s failure to yield to an oncoming vehicle. Several previous studies also found elderly pedestrians electing to cross roadways in unsafe gaps to be an issue in elderly pedestrian crashes (*5*, *12*, *14*, *15*). Pedestrian fatalities owing to motorists failing to yield are ranked second in the age groups 65 to 69, 75 to 79 and >79, and third in the age group 70 to 74. Thus, motorists’ failure to yield is an important factor in relation to elderly pedestrian fatalities. Crashes involving left-turning

Table 1. Percentage Distribution among Key Variables of Pedestrian Crashes by Age Groups

Variable	Attributes	Age: 65–69	Age: 70–74	Age: 75–79	Age: > 79
Vehicle, roadway, and pedestrian-related factors	Crossing roadway—vehicle not turning	43.22	42.38	46.82	44.94
	Crossing roadway—vehicle turning	7.44	12.87	14.03	14.25
	Other/unknown—insufficient details	11.93	9.37	10.77	9.03
	Walking/running along roadway	9.85	6.99	6.36	4.46
	Dash/dart-out	6.02	6.57	5.87	6.20
	Unusual circumstances	6.78	6.71	4.08	6.31
	Pedestrian in roadway—circumstances unknown	5.58	5.17	4.08	3.37
	Backing vehicle	1.31	1.96	2.94	4.79
	Crossing expressway	2.63	3.22	1.31	1.09
	Unique midblock	1.64	1.54	1.14	1.96
	Driveway access/ driveway access-related	0.88	0.84	0.82	1.52
	Non-traffic way	0.88	1.40	0.82	1.09
	Bus-related	0.98	0.42	0.49	0.33
	Multiple threat/trapped	0.55	0.28	0.49	0.54
	Working or playing in roadway	0.22	0.28	0.00	0.11
	Waiting to cross	0.11	0.00	0.00	0.00
	Sum	100.00	100.00	100.00	100.00
Lighting conditions	Daylight	28.99	36.64	43.23	52.45
	Dark—lighted	39.28	33.99	31.97	25.90
	Dark—not lighted	25.05	21.40	19.09	14.80
	Dawn	2.41	2.80	2.28	3.05
	Dusk	2.52	3.08	1.79	2.94
	Dark—unknown lighting	1.09	1.68	0.98	0.65
	Other/unknown	0.66	0.42	0.65	0.22
	Sum	100.00	100.00	100.00	100.00
Gender	Male	68.05	66.43	62.48	60.28
	Female	31.95	33.57	37.52	39.72
	Sum	100.00	100.00	100.00	100.00

Note: A darker red color indicates a higher percentage of crashes of all age groups.

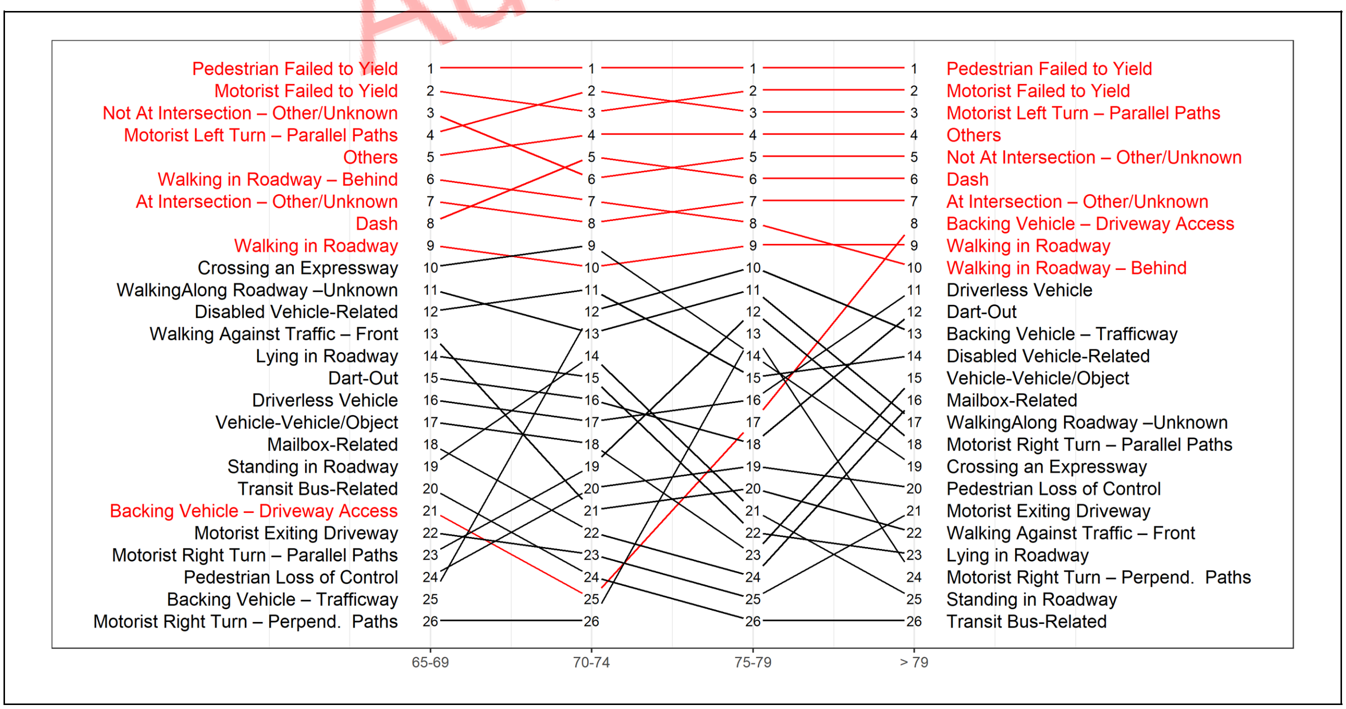

Figure 2. Slope graph of the key crash types by age group.

vehicles also have a high frequency. These types of crashes could be caused by either pedestrian failure to yield or motorist failure to yield. Elderly pedestrian fatalities caused by vehicles backing out of driveways shows an interesting trend. This variable ranked seventeenth or below for elderly pedestrians younger than 79 but ranked eighth for elderly pedestrians over the age of 79. This might be the result of physical and cognitive deterioration or other related problems. Crashes involving pedestrians walking or running along and against the direction of traffic decline with increasing age. This might be because older elderly pedestrians might be less likely to walk or run on roadways owing to physical limitations.

Empirical Bayes Data Mining

Safety analyses of elderly pedestrian issues usually involve categorical data, and contingency tables provide an effective way of representing categorical data. The relative reporting ratio (RR) statistic is often used for evaluating contingency tables. RR is the ratio of the frequency of cell N_{ij} to the expected frequency ($E = N_{i.} \times N_{.j}$) of a cell under a condition of independence. However, RR is not a good estimator when the actual count in a cell is not large. The EBGM method is an association rule-learning method. Rules based methods have been gaining popularity among transportation safety researchers. The most frequently used methods applied to transportation safety engineering studies are association rules mining (*23–32*), EB and EB mining (*33–35*), and correspondence analysis (*36–41*). Data mining is an algorithm-based model that can deal with larger sets of data and can produce results in the form of rules with weighting scores. As there are no prior assumptions, there is no risk of interpretability issues from the perspective of deviation from the assumptions. This makes data mining significantly different from conventional statistical modeling. EBGM overcomes the limitation of RR and considers N_{ij} and E irrespective of the sample size. EBGM adds Bayesian shrinkage corrections to RR to account for high RR values obtained from a small N_{ij} and small E by shrinking the test statistics toward 1. The test fails to reject the null hypothesis of independence for this value. Moreover, the effect of the shrinkage factor reduces as the count increases; it becomes negligible for very large counts. This method provides more stable results than RR and accounts for sampling variation. Thus, this method is well suited for large, sparse contingency tables. The authors used the EBGM method for conducting the analysis as the elderly pedestrian dataset was both large and sparse. The open source statistical software R package “openEBGM” was used for implementing the EBGM method in this study (*42*).

Consider a three-way contingency table in which index variable i and j represent two factors and k represents the stratification. Elements in each cell are represented by N_{ijk}. The expected count assuming independence between factor i and j conditional on stratification variable k would be E_{ij}.

where

$$N_{ij} = N_{ij.} = \sum_k N_{ijk}$$

$$N_{i.j} = \sum_j N_{ijk}$$

$$N_{.jk} = \sum_i N_{ijk}$$

$$N_{..k} = \sum_i \sum_j N_{ijk}$$

$$E_{ij} = \sum_{jk} \frac{N_{i.k} N_{.jk}}{N_{..k}}$$

Consider N_{ij} is Poisson distributed with an unknown mean μ_{ij} and $\lambda_{ij} = \mu_{ij}/E_{ij}$ is the decision statistics for detecting unusually large frequencies or each cell in the contingency table. Then statistic λ is drawn from a mixture of two gamma distributions with mixing proportion p. The following equation shows the distribution of the λ statistic:

$$\pi(\lambda|\alpha_1, \beta_1, \alpha_2, \beta_2, p) = p \times Gamma(\lambda|\alpha_1, \beta_1) + (1-p) \times Gamma(\lambda|\alpha_2, \beta_2)$$

The distribution of the λ statistic depends on five parameters. Assuming $E, \alpha_1, \beta_1, \alpha_2, \beta_2, and\ p$ are known, the marginal distribution of N_{ij} would be the mixture of two negative binomial distributions. The posterior expected value of $\log_2 \lambda_{ij}$ ($E_{\lambda_{ij}|n}[\log_2 \lambda_{ij}]$) would be the Bayesian statistic for RR, which can be derived using the following form:

$$E[\log(\lambda)|N = n] = Q_n[\psi(\alpha_1 + n) - \log(\beta_1 + E)] + (1 - Q_n)[\psi(\alpha_2 + n) - \log(\beta_2 + E)]$$

The digamma function ($dlog(\Gamma(x))/dx$) and Q_n is the posterior probability that λ came from the first component of the mixture. $E_{\lambda_{ij}|n}[\log_2 \lambda_{ij}]$ is close to log(RR) for large counts but it shrinks toward lower values when E and n_{ij}/E are not large. To get $E_{\lambda_{ij}|n}[\log_2 \lambda_{ij}]$ in the same scale as RR and obtain a value that is easily comparable and interpretable, DuMuchel computed EBGM ($EBGM_{ij}$), the geometric mean of $E_{\lambda_{ij}|n}[\log_2 \lambda_{ij}]$. This is given by the equation $EBGM_{ij} = 2^{E_{\lambda_{ij}|n}[\log_2 \lambda_{ij}]}$. In EB methods, the choice of prior parameters is obtained from

Table 2. Top 10 Combination Groups with High PRRs

No.	Variable 1	Variable 2	*E*	RR	PRR	EBGM
1	Backing vehicle—driveway access, backing vehicle, dawn	Female (70–74)	0.076	13.2	13.2	1.04
2	Disabled vehicle-related, unusual circumstances, dusk	Female (70–74)	0.076	13.2	13.2	1.04
3	Driveway access—other/unknown, driveway access/ driveway access-related, dusk	Female (70–74)	0.076	13.2	13.2	1.04
4	Motorist failed to yield, crossing roadway—vehicle not turning, dark—unknown lighting	Female (70–74)	0.076	13.2	13.2	1.04
5	Mailbox-related, unique midblock, dusk	Female (65–69)	0.277	10.8	10.9	1.27
6	Mailbox-related, unique midblock, dark- unknown lighting	Female (65–69)	0.092	10.8	10.9	1.04
7	Motorist right turn—parallel paths, crossing roadway—vehicle turning, dawn	Female (65–69)	0.092	10.8	10.9	1.04
8	Pedestrian on vehicle, unusual circumstances, dawn	Female (65–69)	0.092	10.8	10.9	1.04
9	Backing vehicle—driveway access, backing vehicle, dark—not lighted	Female (>79)	0.231	8.7	8.7	1.12
10	Backing vehicle—trafficway, backing vehicle, dusk	Female (>79)	0.115	8.7	8.7	1.03

Note: *E* = expected counts; RR = relative reporting ratio; PRR = proportional relative ratio; EBGM = empirical Bayes geometric mean.

the data. In this methodology $\alpha_1, \beta_1, \alpha_2, \beta_2, and\, p$ are obtained by maximizing the likelihood of these parameters taken together. This likelihood is the marginal distribution of n_{ij}.

EBGM Modeling and Estimates

This method provides valuable information about the significance of frequency of a given combination of "surrogate" and "outcome" in the contingency table. The algorithm has the ability to handle large contingency tables. After removing rows with missing information, the final dataset contained information on 3,161 person-level crashes with 679 unique combinations of "surrogate–outcome."

The first step was to calculate actual "surrogate–outcome" count combinations (surrogate indicates the combination of the crash conditions: crash type—crash group—lighting, outcome indicates the associated gender–age group), expected counts (E) under the row/column independence assumption, RR, and proportional reporting ratio (PRR). Stratification is beneficial in controlling confounding variables. Stratification will affect E and RR, but not PRR. The Es were calculated by summing the E from every stratum. Ideally, each stratum should contain several unique reports to ensure good estimates of E. The actual count (N) and E are used to estimate the hyperparameters of the prior distribution. A large contingency table will have many cells, resulting in computational difficulties for the optimization routines needed for estimation. The hyperparameters are estimated by minimizing the negative log-likelihood function. The optimized hyperparameters are $(\alpha_1, \beta_1, \alpha_2, \beta_2, P) = (1.03 \times 10^{-8}$, 0.877, 6.284, 6.434, and 0.003). EBGM is a measure of central tendency of the posterior distributions $\lambda_{ij}|N = n$. Scores much larger than 1 indicate surrogate–outcome pairs that are reported at an unusually high rate. Table 2 lists the top 10 combination groups [in the form of a rule, for example, {Backing Vehicle—Driveway Access, Backing Vehicle, Dawn $=>$ Female (70–74)}] with high PRRs. All outcomes in the top 10 rules relate to female pedestrians. Driveway-related (a backing vehicle) presented in three of the surrogates. Another interesting pattern was the events mostly occur in poor lighting conditions. For very elderly female pedestrians, backing vehicle-related crashes were overrepresented. Furthermore, RR and PRR had the same values for surrogate–outcome rules 5 through 8, whereas the EBGM score for Rule 5 [{Mailbox-related, Unique Midblock, Dusk $=>$ Female (65–69)}] was higher than for the rest of the groups. This indicates this combination group, identified by the EBGM algorithm, is more likely to be associated with pedestrian crashes of elderly females aged between 65 and 69 than the other four groups; RR and PRR methods failed to identify this pattern.

EBGM is the antilog of the mean of the *In*-transformed posterior distribution. It can be used as a measure of central tendency of the posterior distribution. The EBGM scores indicate an adjusted estimate for the relative reporting ratio. For example, "Backing Vehicle—Driveway Access, Backing Vehicle, Daylight $=>$ Female (>79)" pair had an EB score of 4.33. The interpretation is that this pair occurs in the data 4.33 times more frequently than expected under the assumption of no association between the surrogate and the outcome. The 5% and 95% quantiles of the posterior distributions can be used to create two-sided 90% credibility intervals for λ_{ij}, given N_{ij}. Because of the Bayesian shrinkage property, the EB scores are much more stable than RR and PRR for small counts.

Table 3 lists the top 25 combination groups with high EBGM and quantiles (the list is sorted in descending order based on quantile 5% values). Among the top 25 rules, female elderly pedestrians were mostly associated with crossing roadway-related crashes (out of 11 elderly

Table 3. Problem Groups with High EBGM Scores and Quantiles

No.	Variable 1	Variable 2	*N*	EBGM	Q05	Q95
1	Backing vehicle—driveway access, backing vehicle, daylight	Female (>79)	17	4.33	4.34	4.47
2	Motorist right turn—perpendicular paths, crossing roadway—vehicle turning, daylight	Female (75–79)	6	1.63	0.89	4.43
3	Walking/running along roadway—direction/position unknown, walking/running along roadway, dark—not lighted	Male (65–69)	10	1.56	0.98	4.41
4	Walking in roadway, pedestrian in roadway—circumstances unknown, dark—lighted	Male (65–69)	15	1.46	1.05	1.99
5	Motorist left turn—parallel paths, crossing roadway—vehicle turning, daylight	Female (>79)	44	1.43	1.15	1.76
6	Motorist failed to yield, crossing roadway—vehicle not turning, daylight	Female (>79)	28	1.41	1.09	1.80
7	Walking/running along roadway with traffic—from behind, walking/running along roadway, dark—not lighted	Male (65–69)	18	1.38	1.03	1.82
8	Pedestrian failed to yield, crossing roadway—vehicle not turning, daylight	Male (>79)	71	1.38	1.15	1.64
9	Motorist exiting driveway, driveway access/ driveway access-related, daylight	Female (>79)	7	1.37	0.88	4.39
10	Motorist failed to yield, crossing roadway—vehicle not turning, daylight	Female (75–79)	18	1.32	0.99	1.75
11	Crossing an expressway, crossing expressway, dark—not lighted	Male (65–69)	14	1.29	0.94	1.74
12	Motorist left turn—parallel paths, crossing roadway—vehicle turning, daylight	Female (70–74)	27	1.29	1.00	1.65
13	Mailbox-related, unique midblock, dusk	Female (65–69)	3	1.27	0.77	4.39
14	Crossing an expressway, crossing expressway, dark—lighted	Male (70–74)	7	1.26	0.85	1.79
15	Pedestrian failed to yield, crossing roadway—vehicle not turning, dark—not lighted	Male (75–79)	44	1.25	1.00	1.55
16	Driverless vehicle, unusual circumstances, daylight	Female (> 79)	8	1.23	0.86	1.72
17	Pedestrian loss of control, unusual circumstances, daylight	Female (>79)	5	1.21	0.80	1.77
18	Pedestrian failed to yield, crossing roadway—vehicle not turning, daylight	Female (>79)	41	1.21	0.97	1.50
19	Dash, dash/dart-out, dark—not lighted	Male (>79)	11	1.19	0.85	1.63
20	Not at intersection—other/unknown, other/unknown—insufficient details, daylight	Male (>79)	16	1.19	0.88	1.59
21	Pedestrian failed to yield, crossing roadway—vehicle not turning, dark—lighted	Male (65–69)	119	1.19	1.03	1.37
22	At intersection—other/unknown, other/unknown—insufficient details, dark—not lighted	Male (65–69)	7	1.18	0.82	1.66
23	Crossing an expressway, crossing expressway, dark—lighted	Male (65–69)	7	1.18	0.82	1.66
24	Pedestrian failed to yield, crossing roadway—vehicle not turning, dawn	Male (65–69)	10	1.18	0.84	1.62
25	Pedestrian failed to yield, crossing roadway—vehicle not turning, other/unknown	Female (65–69)	3	1.17	0.75	1.80

Note: *N* = count; EBGM = empricial Bayes geometric mean score; Q05 = five percentile of EBGM; Q95 = ninety five percentile of EBGM.

female groups, six groups were related to crossing roadway). This is supported by findings from Zegeer et al. who found that more elderly pedestrians are involved in crossing roadway crashes and from Clifton et al. that a higher percentage of crashes involving female pedestrians occurred in areas with high pedestrian activity (*4*, *43*). Rules associated with very elderly females have higher EBGM scores than other female pedestrian age groups. This is consistent with the trend presented in Table 1 showing that more very elderly female pedestrians were involved in crashes, and is supported by studies conducted by Niebuhr et al. (*6*) and Eluru et al. (*16*). Given that fewer female than male pedestrians were represented in the dataset (see Table 1), combination groups for elderly female pedestrian crashes ranked higher than those for male crashes, thereby indicating an overrepresented association. Rules with segment-related crashes of male pedestrians aged between 65 and 69 years had higher EBGM values than other age groups of male pedestrians. This is consistent with the trend (Table 1) that more male pedestrians in this age group were involved in crashes than other age groups. All segment-related male pedestrian crashes occurred at night. This aligns with the crash pattern identified by Prato et al. of male pedestrians being hit by vehicles when crossing a roadway at night (*44*). Three rules had a crash surrogate of crossing expressways at nighttime. These rules were all associated with male pedestrians younger than 75 years old; in two, males aged 65 to 69 years were involved. This indicates that younger elderly males are more likely to risk crossing high-speed roadways than very elderly males, but the high-speed traffic plus the darkness factor increases the risk of a fatal crash, as found by Kitali et al. (*10*). In fact, elderly pedestrian crashes in dark conditions were presented in 10 rules. Out of these 10 rules, 6 were associated with the dark and no street lighting. This further confirms the pattern that lighting is an essential factor that affects pedestrian safety (*10*, *44–46*).

Conclusions

According to the Bureau of Labor Statistics, the elder adult population in the workforce aged 65 and older in the United States has risen from 10.8% in 1985 to 19.2% in 2017. There is a clear notion that the average age of the U.S. population is increasing every year, justifying prioritizing the elderly population in roadway safety planning. This study identified patterns of fatal crashes among elderly pedestrians to better comprehend the underlying causes. The EBGM method was applied to analyze three years of FARS fatal crash data on elderly pedestrians. The findings of this study follow:

- Backing vehicle-related crashes for female pedestrians (especially those aged 79 and above) were significantly high, according to EBGM scores. For roadway-related issues such as crossing a road or an intersection, walking along a segment roadway or a driveway, high pedestrian demand areas need to be planned with better pedestrian accessibility by providing pedestrian facilities such as sidewalks, crosswalks, median islands, overpasses, and underpasses. Well-designed curbs and sidewalks could help mitigate such crashes.
- Segment-related crashes at night for 65 to 69 year-old male pedestrians were high. Crashes occurring at night with poor street lighting conditions were also high in relation to EBGM scores. Enhanced nighttime street lighting should be provided at locations experiencing high numbers of nighttime pedestrian crashes, especially at high-speed locations. Note that countermeasures to improve pedestrian safety should be implemented systematically to achieve the expected benefit. Well-designed curbs and sidewalks are needed to separate pedestrians and vehicles. Speed controls in areas where the risk to pedestrians is high are believed to be effective in reducing the severity of segment-related pedestrian crashes (*47*). Signs and markings should be well designed and maintained with improved visibility to inform both drivers and pedestrians about oncoming roadway users.
- Crossing the expressway at night for male pedestrians, especially the 65 to 69 year group, showed high EBGM values. However, these crashes are not the most frequent. The EBGM scores indicate that such crashes are more likely to be associated with 65 to 69 year-old elderly pedestrians.
- Failure to yield while crossing at intersections, and crossing intersections while a vehicle is turning were associated with elderly pedestrians. Better pedestrian accessibility through providing appropriate pedestrian facilities are required in areas of high pedestrian demand.

Urban transportation professionals need to understand how being elderly can potentially affect safe mobility; they need to consider the safety and mobility of the elderly population at every step of the planning and design processes. The American Association of Retired Persons Public Policy Institute published a comprehensive public report, “Planning Complete Streets for an Elderly America” (*48*). The report discusses a broad approach for integrating policy changes in transportation design sectors to account for elderly road users. It also provides a four-step implementation strategy to ensure successful planning focusing on agency staffs’ skillsets, policy standards and procedures, and collection of data. The FHWA’s *Handbook for Designing Roadways for the Elderly Population* provides practitioners with a practical information source that links elderly road-user performance to highway design, operational, and traffic engineering features (*49*). The patterns of factors found in this study could be incorporated into these guidelines for better enforcement and decision making on policies and strategies.

To address behavioral factors, public campaigns might be considered. These campaigns should focus on educating elderly pedestrians about safe travel behaviors to avoid high-risk situations, such as crossing roadways only at crosswalks and educating them about the factors they need to consider while crossing an intersection. Moreover, driver education could also be implemented to raise awareness of the safety issues surrounding elderly pedestrians. Advanced technology in connected/automated vehicles and smartphones could also be useful in effecting behavioral changes to improve elderly pedestrian safety. Vehicle to pedestrian (V2P) communication could be established based on dedicated short-range communications using onboard equipment in vehicles and smartphones (*50*). The V2P system could provide warnings to both the driver and pedestrian about potential risks so they can take actions to avoid collisions.

The current study is not without limitations. This study focused on roadway geometry and pedestrians' position-related information to identify patterns of association between variables in elderly pedestrian crashes. Additional variables (for example, behavioral and demographic) from FARS and census data could be incorporated to determine more robust findings.

Acknowledgments

The authors thank four anonymous reviewers for their valuable comments that helped to improve the manuscript.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: SD; data collection: SD; analysis and interpretation of results: SD; draft manuscript preparation: SD, AB, XS, HZ, MJ. All authors reviewed the results and approved the final version of the manuscript.

References

1. Ortman, J. M., V. A. Velkoff, and H. Hogan. *An Elderly Nation: The Elderly Population in the United States*. United States Census Bureau, Economics and Statistics Administration, US Department of Commerce, Washington, D.C., 2014.
2. Retting, R. *Pedestrian Traffic Fatalities by State*. Governors Highway Safety Association, Washington, D.C., 2017.
3. *Traffic Safety Facts 2015 Data – Pedestrians*. Publication DOT-HS-812-375. US Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C., 2017.
4. Zegeer, C. V., J. C. Stutts, H. Huang, M. Zhou, and E. Rodgman. Analysis of Elderly Pedestrian Accidents and Recommended Countermeasures. *Transportation Research Record: Journal of the Transportation Research Board*, 1996. 1405: 56–63.
5. Dommes, A., V. Cavallo, J. B. Dubuisson, I. Tournier, and F. Vienne. Crossing a Two-Way Street: Comparison of Young and Old Pedestrians. *Journal of Safety Research*, Vol. 50, 2014, pp. 27–34.
6. Niebuhr, T., M. Junge, and E. Rosén. Pedestrian Injury Risk and the Effect of Age. *Accident Analysis & Prevention*, Vol. 86, 2016, pp. 121–128.
7. Siram, S. M., V. Sonaike, O. B. Bolorunduro, W. R. Greene, S. Z. Gerald, D. C. Chang, E. E. Cornwell, and T. A. Oyetunji. Does the Pattern of Injury in Elderly Pedestrian Trauma Mirror That of the Younger Pedestrian? *Journal of Surgical Research*, Vol. 167, 2011, pp. 14–18.
8. Kim, J. K., G. F. Ulfarsson, V. N. Shankar, and S. Kim. Age and Pedestrian Injury Severity in Motor-Vehicle Crashes: A Heteroskedastic Logit Analysis. *Accident Analysis & Prevention*, Vol. 40, 2008, pp. 1695–1702.
9. Tournier, I., A. Dommes, and V. Cavallo. Review of Safety and Mobility Issues among Elderly Pedestrians. *Accident Analysis & Prevention*, Vol. 91, 2016, pp. 24–35.
10. Kitali, A. E., T. Sando, E. E. Ozguven, and R. Moses. Understanding the Factors Associated with Severity of Elderly Population-Involved Pedestrian Crashes in Florida. *Advances in Transportation Studies*, Vol. 42, 2017, pp. 85–98.
11. Dommes, A., V. Cavallo, F. Vienne, and I. Aillerie. Age-Related Differences in Street-Crossing Safety before and after Training of Elderly Pedestrians. *Accident Analysis & Prevention*, Vol. 44, 2012, pp. 42–47.
12. Dommes, A., T. L. Lay, F. Vienne, N. T. Dang, A. P. Beaudoin, and M. C. Do. Towards an Explanation of Age-Related Difficulties in Crossing a Two-Way Street. *Accident Analysis & Prevention*, Vol. 85, 2015, pp. 229–238.
13. Liu, Y. C., and Y. C. Tung. Risk Analysis of Pedestrians' Road-Crossing Decisions: Effects of Age, Time Gap, Time of Day, and Vehicle Speed. *Safety Science*, Vol. 63, 2014, pp. 77–82.
14. Oxley, J. A., E. Ihsen, B. N. Fildes, J. L. Charlton, and R. H. Day. Crossing Roads Safely: An Experimental Study of Age Differences in Gap Selection by Pedestrians. *Accident Analysis & Prevention*, Vol. 37, 2005, pp. 962–971.
15. Oxley, J. A., B. Fildes, E. Ihsen, J. Charlton, and R. Day. Differences in Traffic Judgements between Young and Old Adult Pedestrians. *Accident Analysis & Prevention*, Vol. 29, 1997, pp. 839–847.
16. Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis & Prevention*, Vol. 40, 2008, pp. 1033–1054.
17. Rosenbloom, T., Y. Sapir-Lavid, and A. Perlman. Risk Factors in Road Crossing among Elderly Pedestrians and Readiness to Adopt Safe Behavior in Socio-Economic Comparison. *Accident Analysis & Prevention*, Vol. 93, 2016, pp. 23–31.
18. Lee, C., and M. Abdel-Arty. Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida, *Accident Analysis and Prevention*, Vol. 37, No. 4, 2005, pp. 775–786.
19. Ginter, E., and V. Simko. Women Live Longer Than Men. *Bratislavske Lekarske Listy*, Vol. 114 No. 2, 2013, pp. 45–49.
20. Ward, H., J. Cave, A. Morrison, R. Allsop, A. Evans, C. Kuiper, and L. Willumsen. *Pedestrian Activity and Accident Risk*. Automobile Association Foundation for Road Safety Research, London, UK, 1994.
21. Geraghty, J., C. Holland, and K. Rochelle. Examining Links between Cognitive Markers, Movement Initiation and Change, and Pedestrian Safety in Elderly Adults. *Accident Analysis & Prevention*, Vol. 89, 2016, pp. 151–159.
22. *2016 FARS/CRSS Pedestrian Bicyclist Crash Typing Manual: A Guide for Coders using the FARS/CRSS Ped/Bike Typing Tool*. NHTSA, Washington, D.C., 2018.
23. Pande, A., and M. AbdelAty. Market Basket Analysis of Crash Data from Large Jurisdictions and Its Potential as a Decision Support Tool. *Safety Science*, Vol. 47, No. 1, 2009, pp. 145–154.
24. Montella, A. Identifying Crash Contributory Factors at Urban Roundabouts and Using Association Rules to Explore Their Relationships to Different Crash Types. *Accident Analysis and Prevention*, Vol. 43, No. 4, 2011, pp. 1451–1463.

25. Montella, A., M. Aria, A. D'Ambrosio, and F. Mauriello. Analysis of Powered Two-Wheeler Crashes in Italy by Classification Trees and Rules Discovery. *Accident Analysis and Prevention*, Vol. 49, 2012, pp. 58–72.
26. Weng, J., J. Zhu, X. Yan, and Z. Liu. Investigation of Work Zone Crash Casualty Patterns Using Association Rules. *Accident Analysis and Prevention*, Vol. 92, 2016, pp. 43–52.
27. Das, S., L. Minjares-Kyle, R. E. Avelar, K. K. Dixon, and B. Bommanayakanahalli. Improper Passing Related Crashes on Rural Roadways: Using Association Rules Negative Binomial Miner. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
28. Ait-Mlouk, A., F. Gharnati, and T. Agouti. An Improved Approach for Association Rule Mining Using a Multi-Criteria Decision Support System: A Case Study in Road Safety. *European Transport Research Review*, Vol. 9, No.3, 2017, p. 13.
29. Das, S., A. Dutta, R. Avelar, K. Dixon, X. Sun, and M. Jalayer. Supervised Association Rules Mining on Pedestrian Crashes in Urban Areas: Identifying Patterns for Appropriate Countermeasures. *International Journal of Urban Science*, Vol. 23, No. 1, 2019, pp. 30–48.
30. Das, S., A. Dutta, M. Jalayer, A. Bibeka, and L. Wu. Factors Influencing the Patterns of Wrong-Way Driving Crashes on Freeway Exit Ramps and Median Crossovers: Exploration Using 'Eclat' Association Rules to Promote Safety. *International Journal of Transportation Science and Technology*, Vol. 7, No. 2, 2018, pp. 114–123.
31. Das, S., A. Dutta, R. Avelar, K. Dixon, X. Sun, and M. Jalayer. Supervised Association Rules Mining on Pedestrian Crashes in Urban Areas: Identifying Patterns for Appropriate Countermeasures. *International Journal of Urban Sciences*, Vol. 23, No. 1, 2018, pp. 1–19.
32. Das, S., A. Mudgal, A. Dutta, and S. R. Geedipally. Vehicle Consumer Complaint Reports Involving Severe Incidents: Mining Large Contingency Tables. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 72–82.
33. Das, S., A. Dutta, S. R. Geedipally, C. Ma, and Z. Elgart. Effect of Vehicular Defects on Crash Severity: A Bayesian Data Mining Approach. Presented at 98th Annual Meeting of the Transportation Research Board, Washington, D.C., 2019.
34. Zou, Y., J. E. Ash, B. J. Park, D. Lord, and L. Wu. Empirical Bayes Estimates of Finite Mixture of Negative Binomial Regression Models and its Application to Highway Safety. *Journal of Applied Statistics*, Vol. 45, No. 9, 2018, pp. 1652–1669.
35. Sun, X., S. Das, Z. Zhang, F. Wang, and C. Leboeuf. Investigating Safety Impact of Edgelines on Narrow, Rural Two-Lane Highways by Empirical Bayes Method. *Transportation Research Record: Journal of the Transportation Research Board*, 2014. 2433: 121–128.
36. Das, S., K. Jha, K. Fitzpatrick, M. Brewer, and T. H. Shimu. Pattern Identification from Older Bicyclist Fatal Crashes. *Transportation Research Record: Journal of the Transportation Research Board* (forthcoming).
37. Jalayer, M., H. Zhou, and S. Das. *Exploratory Analysis of Run-off-Road Crash Patterns*. Data Analytics for Smart Cities, CRC Press, Boca Raton, FL, 2018.
38. Fitzpatrick, K., and S. Das. *Influences on Bicyclists and Motor Vehicles Operating Speed within a Corridor*. Safe-D: Safety through Disruption Research Project. https://www.vtti.vt.edu/utc/safe-d/index.php/projects/influences-on-bicyclists-and-motor-vehicles-operating-speed-within-a-corridor/
39. Das, S., R. Avelar, K. Dixon, and X. Sun. Investigation on the Wrong Way Driving Crash Patterns using Multiple Correspondence Analysis. *Accident Analysis & Prevention*, Vol. 111, 2018, pp. 43–55.
40. Das, S., B. K. Brimley, T. Lindheimer, and A. Pant. *Safety Impacts of Reduced Visibility in Inclement Weather*. Atlas-TTI Research Report. Advancing Transportation Leadership and Safety (ATLAS) Center, Ann Arbor, MI, 2017.
41. Das, S., and X. Sun. Association Knowledge for Fatal Run-off-Road Crashes by Multiple Correspondence Analysis. *IATSS Research*, Vol. 39, No. 2, 2016, pp. 146–155.
42. Canida, T., and J. Ihrie. OpenEBGM: An R Implementation of the Gamma-Poisson Shrinker Data Mining Model. *The R Journal*, Vol. 9, No. 2, 2017, pp. 499–519.
43. Clifton, K. J., C. Burnier, and K. Kreamer Fults. *Women's Involvement in Pedestrian–Vehicle Crashes*. *Influence of Personal and Environmental Factors*. Transportation Research Board – Research on Women's Issues in Transportation, Washington, D.C., 2004.
44. Prato, C. G., V. Gitelman, and S. Bkhor. Mapping Patterns of Pedestrian Fatal Accidents in Israel. *Accident Analysis & Prevention*, Vol. 44, No. 1, 2012, pp. 56–62.
45. Turner, S., I. Sener, M. Martin, S. Das, E. Shipp, R. Hampshire, K. Fitzpatrick, L. Molnar, R. Wijesundera, M. Colety, and S. Robinson. *Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure to Risk at Areawide Levels and on Specific Transportation Facilities*. Report No. FHWA-SA-17-041. U.S. Department of Transportation, Federal Highway Administration, Washington, D.C., 2017.
46. Das, S., A. Dutta, G. Medina, L. Minjares-Kyle, and Z. Elgart. Extracting Patterns from Twitter to Promote Biking. *IATSS Research*, 2018. https://doi.org/10.1016/j.iatssr.2018.09.002
47. Grundy, C., R. Steinbach, and P. Edwards. Effect of 20 mph Traffic Speed Zones on Road Injuries in London, 1986–2006: Controlled Interrupted Time Series Analysis. *BMJ*, Vol. 339, 2009.
48. Lynott, J., J. Haase, K. Nelson, ASLA, A. Taylor, B. McCann, and E. R. Stollof. *Planning Complete Streets for an Elderly America*. AARP Public Policy Institute, Washington, D.C., 2009.
49. Brewer, M., D. Murillo, and A. Pate. *Handbook for Designing Roadways for the Elderly Population*. Report No. FHWA-SA-14-015, Federal Highway Administration, Washington, D.C., 2014.
50. Wu, X., R. Miucic, S. Yang, S. Al-Stouhi, J. Misener, S. Bai, and W.h. Chan. Cars Talk to Phones: A DSRC Based Vehicle-Pedestrian Safety System. *Proc., IEEE Vehicular Technology Conference (VTC Fall)*. IEEE, Vancouver, BC, Canada, 2014, pp. 1–7.

The Standing Committee on Safe Mobility of Older Persons (ANB60) peer-reviewed this paper (19-03411).