Journal of Safety Research 70 (2019) 213–222

Contents lists available at ScienceDirect

Journal of Safety Research

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

Understanding crash potential associated with teen driving: Survey analysis using multivariate graphical method

Subasish Das, [a,*] Lisa Minjares-Kyle, [b] Lingtao Wu, [a] Russell H. Henk [c]

[a] Texas A&M Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX 77843, United States
[b] Texas A&M Transportation Institute, Texas A&M University System, 701 N. Post Oak Rd. Suite 430, Houston, TX 77024, United States
[c] Texas A&M Transportation Institute, Texas A&M University System, 1100 NW Loop 410, Suite 605, San Antonio, TX, United States

ARTICLE INFO

Article history:
Received 1 November 2018
Received in revised form 17 May 2019
Accepted 15 July 2019
Available online 23 July 2019

Keywords:
Teen driver survey
Crash potentials
Licensing
Multivariate graphical method
Taxicab correspondence analysis

ABSTRACT

Introduction: Teen crash involvement is usually higher than other age groups, and they are typically overrepresented in car crashes. To infer teen drivers' understanding of crash potentials (factors that are associated with crash occurrence), two sources of data are generally used: retrospective data and prospective data. Retrospective data sources contain historical crash data, which have limitations in determining teen drivers' knowledge of crash potentials. Prospective data sources, like surveys, have more potential to minimize the research gap. Prior studies have shown that teen drivers are more likely to be involved in crashes during their early driving years. Thus, there is a benefit in examining how teen drivers' understanding of crash potentials change during their transition through licensing stages (i.e., no licensure to unrestricted licensure). *Method*: This study used a large set of teen driver survey data (a dataset from approximately 88,000 respondents) of Texas teens to answer the research question. Researchers provided rankings of the crash potentials by gender and licensure stages using a multivariate graphical method named taxicab correspondence analysis (TCA). *Results*: The findings show that driving behavior and understanding of crash potentials differ among teens based upon various licensing stages. *Practical applications:* Findings from this study can help government authorities to refine policies of teen driver licensing and implement potential countermeasures for safety improvement.

1. Introduction

Throughout the United States, motor-vehicle collisions continue to be a concern across multiple disciplines due to the alarming number of lives lost each year on our nation's roadways. These alarming rates are disturbingly high for America's youth, ages 13–25, as motor-vehicle collisions continue to be the leading cause of injury and death for this age group. The CDC cites riding and driving in a car to be the #1 threat to teen safety (2016). Recent crash data revealed that 1908 young drivers (ages 15–20) lost their lives in 2016, showing almost no improvement from the previous year (NHTSA, 2016). On a per-mile basis, teens drive fewer miles per year than other age groups, yet they experience the highest overall crash rates (NHTSA, 2016). The total population for this age group has decreased from 2007 to 2016 by 3.4%; despite this, there has been an increase in the number of young, licensed drivers (NHTSA, 2016). As of 2016, young drivers account for 5.4% of all licensed drivers in the United States– a 2.1% increase from 2015; but this age group represents 9% of all drivers involved in fatal crashes (NHTSA, 2016). Research into the causes of teen crashes has built a formidable case identifying driver inexperience along with several other variables such as teen passengers, nighttime driving, and lack of seat belt use as the primary contributors to an increased crash risk (Williams, 2003). It has been proposed that an understanding of crash potentials (an alternative term of the word “risk”) influences driving behaviors because younger teen drivers are more likely to underestimate crash potentials and, thus, more likely to engage in perilous driving behaviors (Brown & Groeger, 1988; Jonah & Dawson, 1987; Rhodes & Pivik, 2011). There is a need for a robust study in understanding the key contributing factors and patterns of these factors that trigger teen driver involved crashes.

Conventional statistical modeling techniques pay little attention to data visualization. Correspondence analysis (CA) can help mitigate this gap by analyzing the hypothesis testing in order to identify patterns of association in the data. CA can easily accommodate various scales of dataset sizes. The target of CA is to minimize the loss of information so that the maximum amount of information is retained. There is a new sturdy–robust–resistant variant of CA called Taxicab Correspondence Analysis (TCA). It can smoothly handle an abundance of data and produce satisfactory and meaningful results in the presence of outliers. Application of CA has been widely used in survey analysis. This paper focuses on a significantly large dataset (information regarding approximately 88,000 respondents) of teen survey data collected in Texas to

* Corresponding author.
E-mail addresses: s-das@tti.tamu.edu (S. Das), l-minjares@tti.tamu.edu (L. Minjares-Kyle), l-wu@tti.tamu.edu (L. Wu), r-henk@tamu.edu (R.H. Henk).

https://doi.org/10.1016/j.jsr.2019.07.009

investigate the understanding of crash potentials among teen drivers. The application of TCA on this dataset was deemed appropriate due to the method's suitability in tackling the research problem related to this large data set.

2. Earlier work and research context

Research on young drivers has yielded a vast amount of information regarding the contributing causes to increased crash risk and the time period in which teens are most at risk for a crash. Newly licensed drivers are at the highest risk for crashes within the first six months of obtaining their license (Mayhew, Simpson, & Pak, 2003; McCartt, Shabanova, & Leaf, 2003a). This risk decreases as they gain more driving experience, but conversely, research shows that a teen's transition from being newly licensed to unrestricted licensure can lead to higher levels of risky driving behaviors and increased crash risk (Brown & Groeger, 1988; Curry, Pfeiffer, Durbin, & Elliot, 2015; Qing, Feng, Kim, Klauer, & Simons-Morton, 2018). Voas and Kelley-Baker found that this is because as teens move further from parental influences and into the community with their peers, their exposure to risky behaviors may grow, causing an uptick in riskier behaviors with little control (Voas & Kelly-Baker, 2008).

These stages of licensure occur due to the Graduated Driver Licensing system (GDL) that gradually grants novice drivers' privileges after a period of restricted driving has been completed (Williams, 2017). Some variance of the GDL system can be found within all 50 states and across several other nations. It can include driver education, supervised driving hours, and probationary periods as steps towards gaining an unrestricted license (p. 29, Williams, 2017). The implementation of GDL has proven to be effective in reducing teen traffic safety crashes over the years (Williams, 2017).

Studies have found links between teens' acknowledgment or recognition of crash potentials in connection to their observed and self-reported driving behaviors (Gershon et al., 2018; Harbeck & Glendon, 2013; Hatfield & Fernandes, 2009; Ouimet et al., 2008; Simons-Morton et al., 2011; Trankle, Gelau, & Metker, 1990; Voas & Kelly-Baker, 2008), with some limitations (Simons-Morton et al., 2016). Despite the recognition of risk factors like using a cell phone, driving tired, or riding with their peers, teens reported either doing the activity themselves or witnessing the behavior in other teen drivers (Ehsani et al., 2015; Fernandes, Hatfield, & Soames Job, 2010; McDonald & Sommers, 2015; Mirman, Albert, Jacobsohn, & Winston, 2012; Mirman, Durbin, Lee, & Seifert, 2017; Rundmo & Iversen, 2004). For example, a national survey of teens asked respondents to rank items that affected driver safety and use of a cell phone was ranked high as making "a lot of difference" in driving safely, yet teens reported witnessing their peers talking on a cell-phone "often or always" (Ginsburg et al., 2008). Understanding teens' likelihood in engaging in these behaviors is important, as research has linked them to an increased likelihood of engaging in other dangerous behaviors (O'Malley Olsen, Schults, & Eaton, 2013).

There is a research gap in the current form of teen driving survey studies in that they only provide general descriptive information regarding teen driving. This study mitigates the research gap by applying a robust survey tool. Many recent transportation studies have applied CA (Baireddy, Zhou, & Jalayer, 2018; Das, Avelar, Dixon, & Sun, 2018; Das, Brimley, Lindheimer, & Pant, 2017; Das, Jha, Fitzpatrick, Brewer, & Shimu, 2019; Das & Sun, 2015; Das & Sun, 2016; Factor, Yair, & Mahalel, 2010; Fontaine, 1995; Jalayer, Pour-Rouholamin, & Zhou, 2018), but this study uses TCA due to its capability of higher variance explanation. Many studies focused on retrospective approaches by using state maintained crash databases. One potential limitation of these studies is the lack of sufficient information on the perception of crash potentials, which is unavailable in police-reported crashes. The current study was based on prospective study design and it examined an extensive dataset (over 88,000 survey entries) to identify the critical understanding of the crash potentials among teen drivers.

3. Survey design and participants

3.1. Survey design

The survey instrument used in this study was developed by the Teens in the Driver Seat (TDS) staff to obtain traffic safety knowledge and patterns of driving behaviors prior to intervention, as well as facilitate longitudinal analyses over time. The survey instrument has also been approved by the Texas A&M Institutional Review Board. The survey was four pages long and consisted of 12 main questions (see Fig. 1). Part One (questions 1–5) of the survey obtained demographic information including age, school, gender, and grade level. Part Two (question 6) of the survey asked the teen to write down the top five risk factors that contribute to teens being hurt or killed in car crashes in an open-ended question. Part Three (questions 7–11) of the survey focuses on traffic safety background including:

- License status and driver education history.
- Crash and citation history.
- The on-road driving test that is taken before receiving the license.
- Car crash history for either participant or family member.
- Receipt of a traffic citation.

The last section of the survey, Part Four (question 12 a-m), focuses on respondents' experience in risky driving behaviors and the frequency (Never; Some – 1–5 times; and A Lot – more than 1–5 times) that teens have engaged in those behaviors in the previous month. It is important to note that for some of the questions there are many missing values. For example, around 80% of the entries in the question about age (question 2) are either redundant or missing.

3.2. Survey participants

The study participants consisted of high school students from Texas. Teenagers in Texas are eligible to apply for driver licensing beginning at age 15 and, as of 2013, they are restricted in some form until the age of 18 (Texas Department of Transportation, 2018). The initial selection of the school was random as schools were recruited through a voluntary state education-outreach program. A total of 209 high schools (57% of which were located in urban areas) responded that they would be able to distribute the surveys during a regularly scheduled seminar or homeroom class period. Schools received a cover letter explaining the purpose of the survey, a copy for parents to be informed of their child's potential participation in the study, and a participant information sheet that informed respondents of their rights, including an explanation that participation was optional. Students were free to leave some or all the responses blank if they did not want to answer. No names or identifiers were associated with the participants. The data were collected from 2007 to 2017 and includes high school students estimated to be between the ages of 13 and 19 based on their reported grade levels. Hence, this study can provide insights on differences across teenagers of different ages. The sample consisted of 88,065 surveys returned (45,566 females, 41,437 males, and 1062 unreported gender) and the average number of respondents per school was 420.

4. Exploratory data analysis

4.1. Descriptive statistics

The final dataset contains 88,065 respondent data with 23 questions answered completely or in part. The survey contains responses for 28 questions, or parts of the questions (question 6 has 5 parts, and question 12 has 13 parts). Responses regarding several questions (question 1, 2, 4, and 8) were not included in the final analysis due to redundancy or a larger number of missing values. For example, grade level is highly

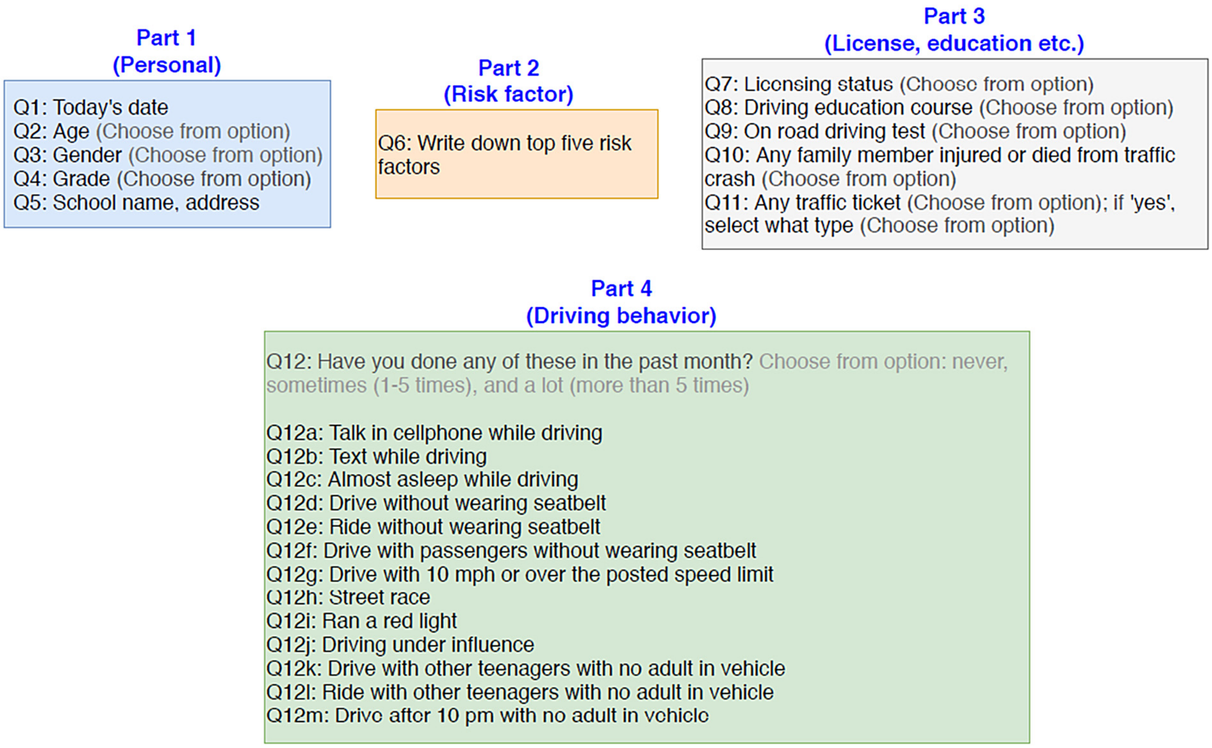


Fig. 1. Teen survey questions.

correlated with licensing status, so it was unnecessary to include both variables. Responses of question 5 were converted into land use variable (rural vs. urban) for analysis. Around 52% of the respondents did not have licenses. However, these respondents were likely obtaining licenses in the near future. Due to the complex nature of the survey questions and different focus areas, there is a need for general descriptive statistics before applying complex survey tool like TCA.

4.2. Ranking of crash potentials by gender

The respondents provided written responses regarding the top five risk factors in driving. Using a slope graph is also an excellent tool to see the ranking differences with the highest slopes (see Fig. 2). This data visualization technique shows the ranking of a certain variable in an ordinal manner. These text-based responses are sorted into 23 key categories. The rankings for each category were based on the frequency that it was identified as a risk factor by the respondents. The top four most-commonly cited risk factors are the same for both female and male respondents: drinking and driving, texting on cell phone, talking on cell phone, and drugs. Use of seat belt is the fifth risk factor according to the female respondents, while it is sixth for the male respondents. Speeding is also ranked in the higher position among the female respondents than the male respondents. Three other risk factors (distractions, too many teen passengers, and sleeping/fatigue) have the same ranks

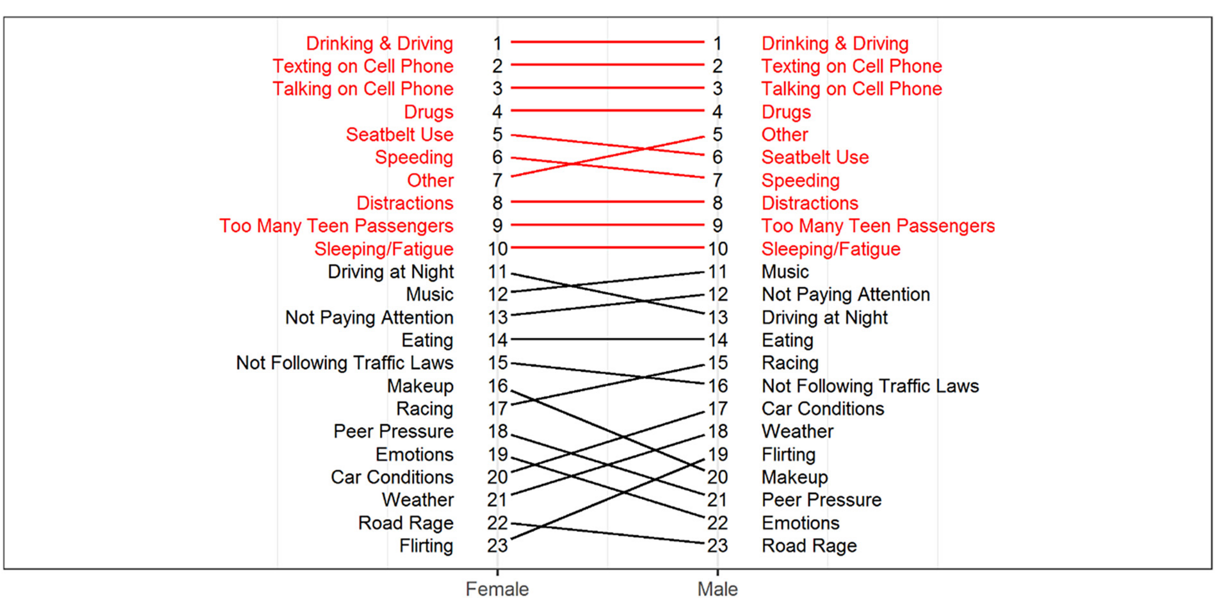


Fig. 2. Ranking of crash potentials by gender.

for both female and male respondents. Several risk factors (music, inattention, racing, car conditions, weather, flirting) are ranked in the higher positions among the male respondents than the female respondents. On the other hand, some risk factors (driving at night, not following traffic laws, makeup, peer pressure, emotions, and road rage) are positioned in the higher ranks among the female respondents as compared to males. Two risk factors have the highest slope differences between the two genders; flirting is ranked in the 19th position among the male respondents while it is ranked in 23rd position among the female respondents, and makeup is ranked in the 16th position among the female respondents compared to 20th for male respondents.

4.3. Experience of crash potentials by licensing status

Fig. 3 shows the stacked bar plot of the experience of crash potentials based on teens' licensing status. Most of the teen respondents with provisional and unrestricted driving license responded that they usually use cell phones while driving; this is quite alarming. Interestingly, around 22% of unrestricted licensed teen drivers drove while tired or almost fell asleep while driving in the past month. This percentage is lower among teen respondents with no license or a learning permit. Teen respondents with a learner permit also show a higher percentage of driving or riding in a car with no seat belt. Furthermore, more than 70% of the teen drivers with an unrestricted driving license responded that they have driven 10 mph above the speed limit. This percentage is also high for teen drivers with a provisional driving license. Around 70–75% of the teen respondents with provisional and unrestricted driving licenses reported never participating in road racing. More than one-third of the teen respondents with an unrestricted driving license reported red light running. As a GDL requirement, adult supervision is mandatory for newly licensed teens who are driving at night or riding with more than one teen passenger (Texas Department of Transportation, 2018). The survey contained three questions related to adult supervision (driving or riding with other teens and driving at night), and most of the unrestricted license holders reported the experience of driving with no adults in the vehicle, while 30% of respondents with a provisional license reported this experience.

5. Methodology

5.1. Multivariate graphical method: taxicab correspondence analysis

5.1.1. Theory

In a series of papers, Choulakian explained the extended theory of TCA (Choulakian, 2006a; Choulakian, 2006b; Choulakian, 2013). This section of the paper provides a brief overview of the basic concepts of TCA. Correspondence analysis (CA) is based on Euclidean distance, whereas Taxicab correspondence analysis (TCA) is based on the Manhattan, City Block or Taxicab distance. Let $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ and a vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ to evaluate these distances:

$$\textit{Euclidean Distance} = ED(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \left[\text{with } L_2\text{Norm} = \|v\|_2 = \sqrt{\sum_{i=1}^{n} (v_i)^2}\right] \tag{1}$$

$$\textit{Taxicab Distance} = TD(X, Y) = \sum_{i=1}^{n} |x_i - y_i| \left[\text{with } L_1\text{Norm} = \|v\|_1 = \sum_{i=1}^{n} |v_i|\right] \tag{2}$$

The concepts of CA and TCA are founded in singular value decomposition (SVD), wherein a real matrix A is decomposed as $M\Lambda^{1/2}N'$, with Λ the diagonal matrix of the real non-negative eigenvalues of AA', M the orthogonal matrix of the corresponding eigenvectors, and N the matrix of eigenvectors of $A'A$ (with constraints $M'M = I \text{ and } N'N = I$). The SVD theory corresponds to the reconstruction formula of a k-rank matrix:

$$a_{ij} = \sum_{i=1}^{k} \sqrt{\lambda_\alpha} m_{i\alpha} n_{i\alpha} \tag{3}$$

To solve the equivalent optimization problem, this approach consists of finding the first vectors m_1 and n_1 principal component of A.

$max\ \|Am\|_2$ subject to $\|m\|_2 = 1$.

$max\ \|A'n\|_2$ subject to $\|n\|_2 = 1$

Taxicab Correspondence Analysis is defined as the Taxicab Singular Value Decomposition of the data table $D = T - rl'$, taking into account the table's profiles, respectively $R = D_r^{-1}D$ for the rows and $L = D_l^{-1}D$ for the columns. Unlike CA, the solution is recursive, considering at each step the residuals from the previous factors. This leads to the reconstruction formula

$$T = p_r p_c' + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} B_\alpha C_\alpha' \tag{4}$$

Elementwise the formula becomes:

$$t_{ij} = t_{i.} t_{.j} + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} B_{i\alpha} C_{j\alpha} \tag{5}$$

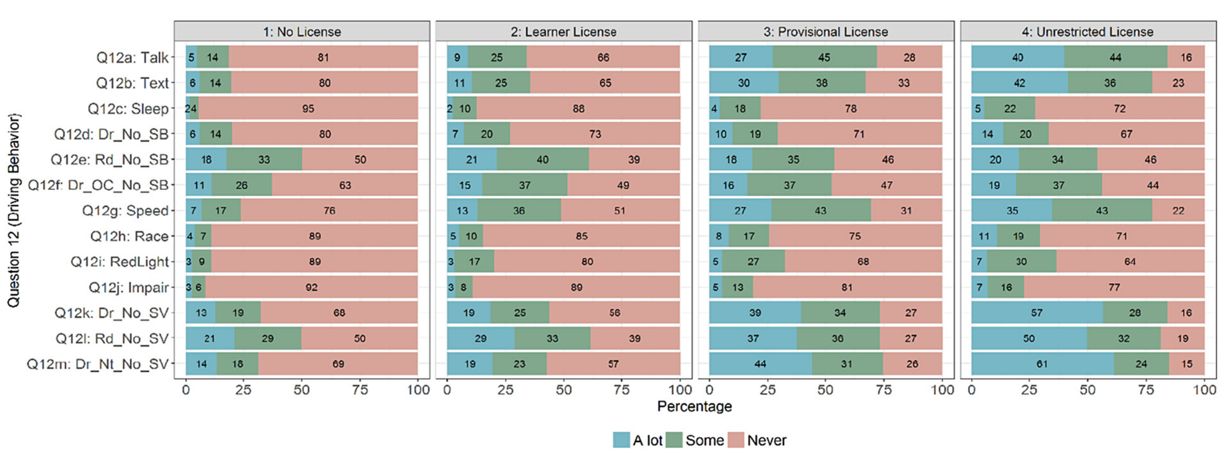


Fig. 3. Experience of crash potentials by licensing status.

After transformation:

$$n_{ij} = nr_i l_j \left(1 + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} b_{i\alpha} c_{j\alpha}\right) \tag{6}$$

6. Results and discussions

Correspondence analysis has recently been gaining popularity among archeologists and it is often applied to archeological abundance data. Sometimes data sets are sparse, where the degree of sparsity is defined as the percentage of zero abundances. For sparse data sets, three types of potential outliers may be identified: rare observations, zero-block structure, and relatively high valued cells. These multivariate statistical methods summarize the knowledge extraction from a complex dataset set by projecting the multivariate data with on a two-dimensional space. The notion is to produce a superimposed map in the form of a biplot that can be used for data visualization and interpretation.

An important property of TCA and CA is that columns (or rows) with identical profiles (conditional probabilities) receive identical factor scores. One important advantage of TCA over CA is that it stays as close as possible to the original data– acting on the correspondence matrix P without calculating a dissimilarity (or similarity) measure between the rows or columns. The CA and TCA maps will be different from each other for the dataset with missing values. TCA is usually more robust than CA when handling missing values. The percentage of variance explained by both axes (see Fig. 4) is around 55% (axis 1 explain 34.37% of variance, and axis 2 explains 20.28% of variance). The locations of the variable categories indicate their association patterns. For example, no license and learner licensing are positioned in the upper quadrants while provisional and unrestricted licensing are in the lower quadrants. The positions of male versus female and urban versus rural contexts are also located in different quadrants. Responses regarding Part Three questions are also located in different quadrants. As the general TCA plot is not suitable for determining all distinct feature patterns, quadrant-specific TCA plots (as illustrated in Figs. 5 and 6) were developed.

Fig. 5a provides an amplified version of Quadrant 1 from Fig. 4. This quadrant represents female teen drivers with no driving license, no citations, and it also represents the response of "never" (e.g., never talk on a phone or texting, never fall asleep while driving a car, never run a red light). The results indicate that female teens tend to engage in less risky driving behaviors than male teens, especially prior to obtaining any sort of licensing. This conclusion is reasonable and consistent with previous studies (Özkan & Lajunen, 2005; Özkan & Lajunen, 2006; Rhodes & Pivik, 2011). For example, Turker and Lajunen had similar findings when they investigated the association between gender and risky driving behaviors, traffic offenses, and accident involvement among young drivers (Özkan & Lajunen, 2005). Furthermore, the researchers analyzed the data using regression methods, and they found that young male drivers reported more ordinary violations than young female drivers. Rhodes and Pivik reported similar results in their study, finding that male teen drivers demonstrated more risky driving than females (Rhodes & Pivik, 2011). In this quadrant, seatbelt and speeding are two main risk factors. Another interesting finding is that makeup was identified as a risk factor; this aligns with a study by Klauer et al. in which they analyzed the naturalistic driving study (NDS) data and identified applying makeup while driving as a key inattention type, especially for female drivers (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). They found that applying makeup increases the likelihood of crash or near-crash involvement by more than two times.

Quadrant 2 is plotted in Fig. 5b and mainly represents the response of "some" (e.g., talking or texting while driving between 1 and 5 times). This quadrant primarily contains teens that have a learner permit and are novice drivers. Typically, this group of teens is supervised by their parents while driving (Simons-Morton & Ouimet, 2006). Parental management helps minimize the risks involved. There might be some cases, however, that the teens drive independently and engage

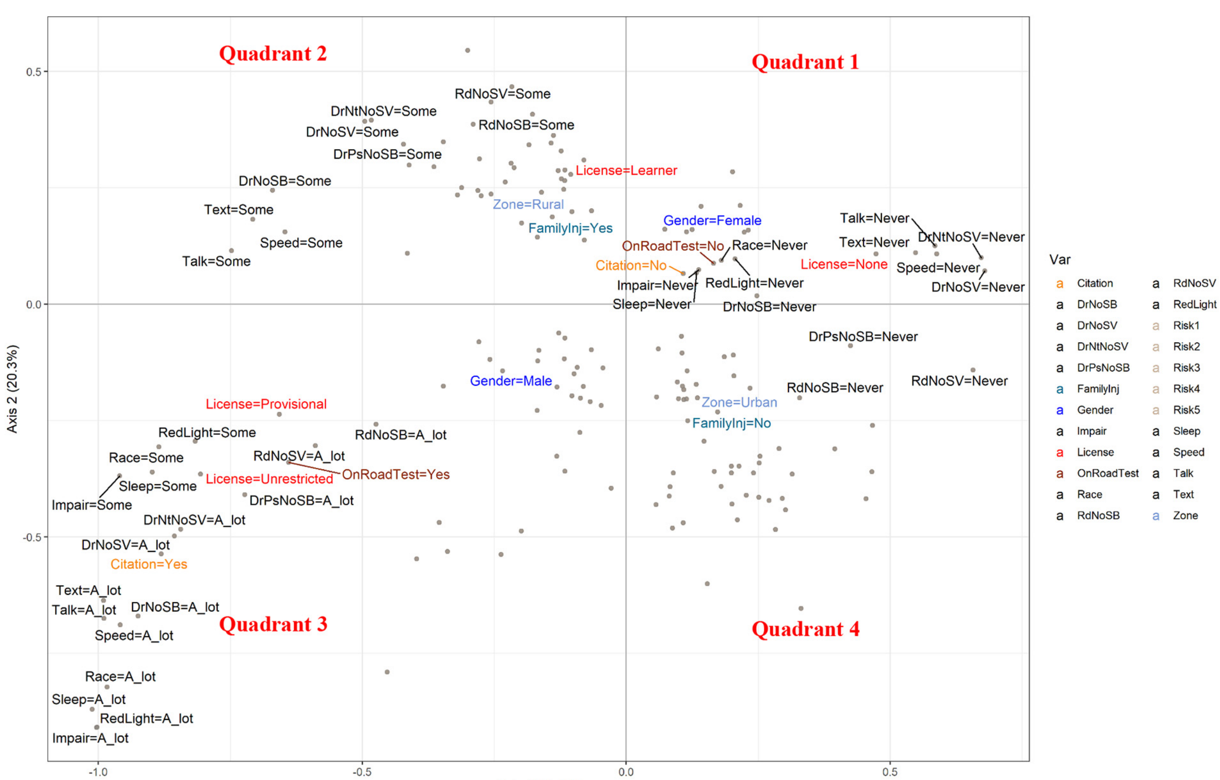


Fig. 4. TCA plot of the attributes (Note: 'Dot will no texts' indicates risk attributes).

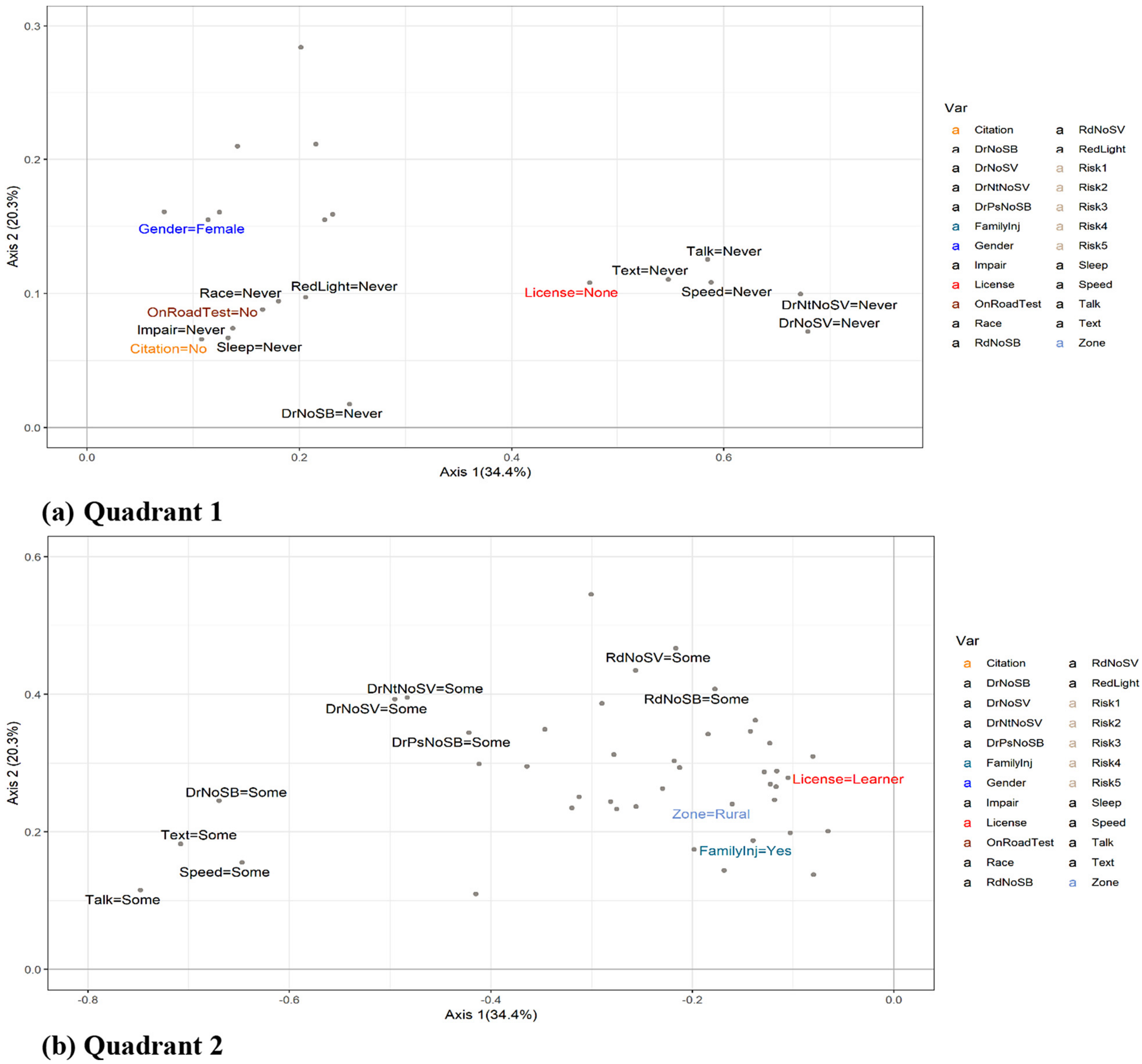
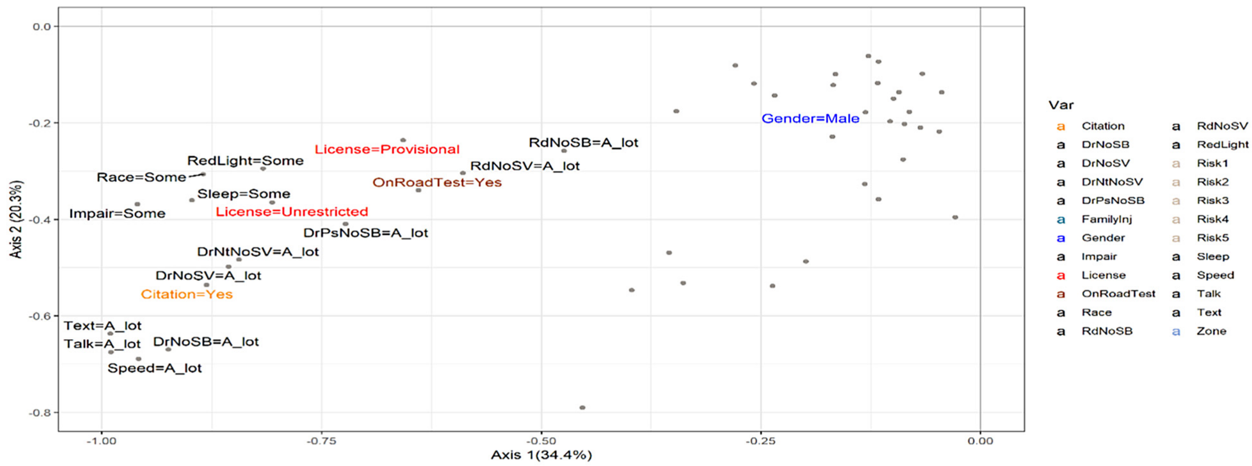


Fig. 5. TCA plot of Quadrant 1 and Quadrant 2.

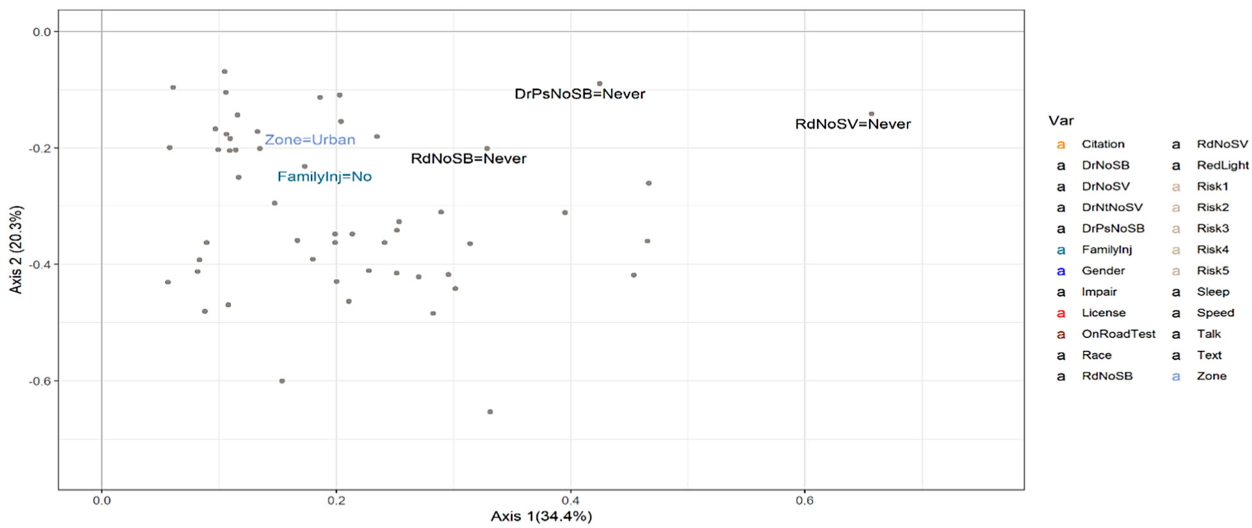
in some risky behaviors (Simons-Morton, 2007). Thus, there are occasions of talking on the phone, texting, and speeding while driving. In Quadrant 2, it can be observed that key risk factors for teen drivers include drinking, drugs, fatigue, and cellphone use while driving. These findings are also supported by previous studies (Curry, Hafetz, Kallan, Winston, & Durbin, 2011; Olsen, Shults, & Eaton, 2013).

Fig. 6a illustrates Quadrant 3 in more detail, mainly representing male teen drivers with provisional or unrestricted licenses. Drivers included in this quadrant seem to exhibit the riskiest driving behavior. As shown in Fig. 6a, there is a high involvement in risky driving performances such as distractions (i.e., talking and texting while driving), speeding, driving without wearing a seatbelt, and driving at night. It is possible that teen drivers are more likely to engage in risky driving without supervision from their parents; this is consistent with some previous studies after getting the driving permit. Williams found that the driving risk of teenagers is particularly high right after licensure (Williams, 2003). The crash rate is the highest in the first month of licensure, and it drops sharply during the next few months. Ebbesen and Haney found that males generally take more risks than females. Additionally, driving late at night and having passengers present are also associated with higher crash rates for teen drivers (Ebbesen & Haney, 1973); this is shown in Quadrant 3. Flirting was also identified as a risk factor in all five questions, so it is likely that teens have more interactions while driving. Simon-Morton et al. found that teenage drivers drove faster than the general traffic and allowed shorter headways, particularly in the presence of a male teenage passenger; and that the presence of male teenage passengers, in general, was associated with risky driving behaviors among teenage drivers (Simons-Morton, Lerner, & Singer, 2005).

Fig. 6b shows Quadrant 4 in more detail, which mainly represents respondents from an urban area. As can be seen, teen drivers in this quadrant are less likely to engage in risky driving behaviors. The responses are "never" to driving without a seatbelt, with passengers not wearing seatbelts, or without someone over the age of 21 in the car. This finding is also reflected in previous studies (McCartt, Shabanova, & Leaf, 2003b; Peek-Asa, Britton, Young, Pawlovich, & Falb, 2010; Simons-Morton et al., 2005). For example, McCartt et al. reported that teen drivers have a higher citation rate in rural areas (McCartt et al., 2003b). Peek-Asa et al. analyzed rates of overall crashes as well as fatal or severe injury crashes among teenage drivers for urban and rural areas and found that rural teen crashes were nearly five times more likely to lead to a fatal or severe injury crash than urban teen crashes (Peek-Asa et al.,



(a) Quadrant 3



(b) Quadrant 4

Fig. 6. TCA plot of Quadrant 3 and Quadrant 4.

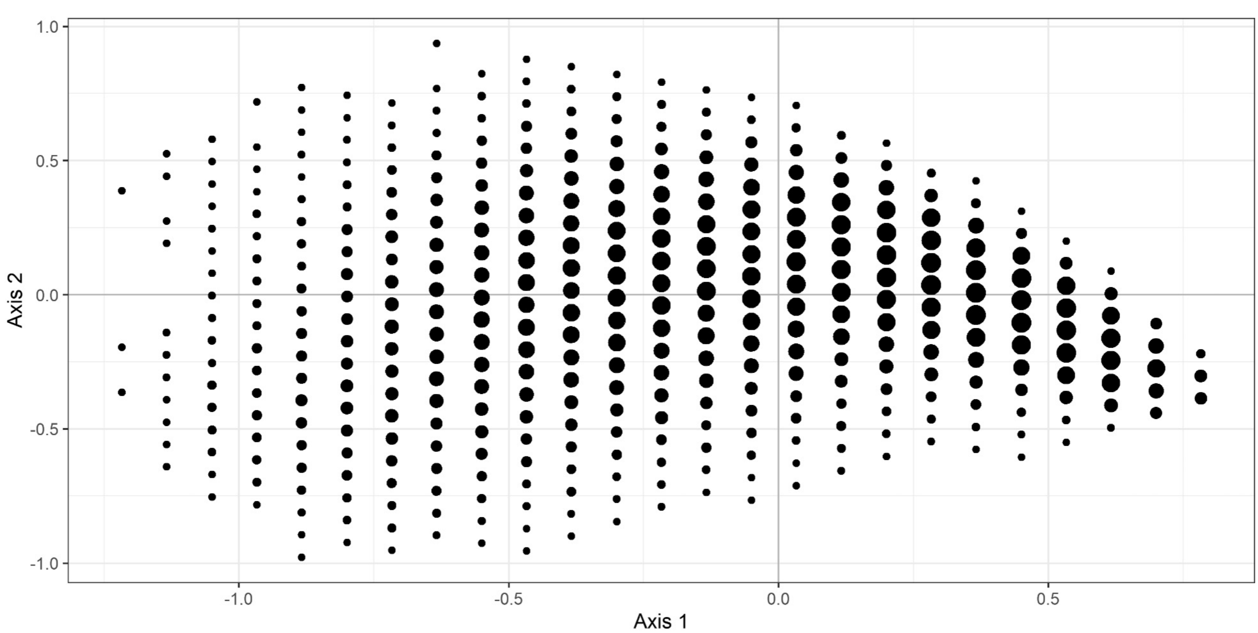


Fig. 7. Clusters from Key Surrogate Measures.

Table 1
Log odds ratios based on key surrogates.

Clusters	Respondent %	Male vs Female	Urban vs Rural	LicenseUnres YES vs NO	LicenseProvi YES vs NO	LicenseLearn YES vs NO	LicenseNone YES vs NO
Cluster01	0.01%	–	–	1.546	1.774	–	–
Cluster02	0.06%	3.397	−1.657	1.323	1.081	0.282	−3.390
Cluster03	0.24%	2.557	−1.250	1.141	1.135	0.433	−3.177
Cluster04	0.57%	2.130	−1.203	1.515	0.920	0.097	−3.069
Cluster05	1.01%	1.906	−1.010	1.525	0.713	−0.039	−2.257
Cluster06	1.60%	1.262	−0.671	1.566	0.696	−0.222	−2.118
Cluster07	2.29%	1.019	−0.680	1.281	0.843	−0.077	−1.894
Cluster08	2.95%	0.786	−0.568	1.280	0.780	−0.196	−1.699
Cluster09	3.90%	0.541	−0.453	1.165	0.736	0.029	−1.631
Cluster10	4.85%	0.340	−0.401	1.002	0.854	−0.040	−1.478
Cluster11	5.28%	0.102	−0.316	0.911	0.707	0.038	−1.249
Cluster12	5.91%	0.030	−0.208	0.759	0.706	0.135	−1.140
Cluster13	6.13%	−0.024	−0.155	0.626	0.686	0.146	−0.991
Cluster14	6.01%	−0.101	−0.125	0.385	0.400	0.303	−0.660
Cluster15	5.78%	−0.155	−0.023	0.053	0.343	0.461	−0.496
Cluster16	6.26%	−0.120	−0.060	−0.371	0.006	0.501	−0.111
Cluster17	6.40%	−0.109	−0.040	−0.766	−0.356	0.475	0.222
Cluster18	7.01%	−0.024	−0.078	−1.360	−0.766	0.278	0.670
Cluster19	7.57%	−0.081	0.016	−1.869	−1.225	0.016	1.089
Cluster20	7.85%	−0.104	0.251	−2.299	−1.521	−0.323	1.474
Cluster21	7.24%	−0.259	0.398	−2.649	−2.088	−0.664	1.893
Cluster22	5.61%	−0.464	0.831	−3.522	−2.627	−1.206	2.498
Cluster23	3.62%	−0.706	1.291	−4.069	−3.434	−1.626	3.002
Cluster24	1.52%	−1.465	2.162	–	−4.765	−2.461	3.995
Cluster25	0.34%	–	–	–	–	–	–

2010). While this may be due to multiple factors such as the type of roadways typically found in rural localities (two lane roads) or the higher potential of remote locations of crashes and delayed emergency service response time, understanding the differences between urban and rural drivers is important to develop a further understanding of teen driver crashes.

This study has performed analysis to determine the key patterns based on the following key surrogate measures:

- Gender of the respondents (Male vs Female)
- Location of the schools (Urban vs Rural)
- Comparison between licensing status of the respondents (four groups: License Unres YES vs NO, License Provi YES vs NO, License Learn YES vs NO, and License None YES vs NO).

Fig. 7 illustrates the distribution of each responses on both axes. Based on axis 1, 25 distinctive clusters (each cluster is represented by a vertical line on the x-axis) are visible. It is found that the clusters are located on both the positive and negative side of the x-axis. The sizes of the points indicate the counts of respondents with same coordinates. Table 1 lists the key patterns of the clusters by computing the log-odds ratio of the surrogate conditions with respect to the marginal distribution. The interpretation of LOR ($X = x$) is:

- LOR($X = x$) = 0.00 indicates that the proportion of category A in cluster x equals the proportion of category B in the sample. For example, the LOR value for the respondents in provisional licensing status (yes vs. no) is almost 0 (0.006) for Cluster16. It indicates that the proportions of provisional licensing status (yes vs. no) for this cluster (6.3% of total respondents) are not different.
- LOR($X = x$) > 0.00 indicates that the proportion of category A in cluster x is greater than the proportion of category B in the sample. For example, the LOR value for "Male vs. Female" respondents is 3.397 for Cluster02. This cluster represents 0.06% of the respondents. The LOR value indicates that this cluster has more male respondents than female respondents. The other LOR values in this cluster indicate the prevalence of other surrogate measures in the form of LOR values.
- LOR($X = x$) < 0.00 indicates that the proportion of category A in cluster x is smaller than the proportion of category B in the sample. For example, the LOR value for "Urban vs. Rural" respondents is −1.657 for Cluster02. This indicates that this cluster represents a higher number of rural respondents than the urban respondents.

The clustering of the respondents highlights some key patterns of the surrogate measures. The key patterns are following:

- Five clusters (Cluster02, Cluster03, Cluster04, Cluster11, and Cluster12) show similar patterns: high number of male respondents, moderately high number of rural respondents, moderately high number of respondents with driving permits. These clusters represent 12.06% of all respondents.
- Five clusters (Cluster05, Cluster06, Cluster07, Cluster08, and Cluster10; representing 12.68% of all respondents) show similar patterns: slight to moderately high number of male respondents, slightly high number of rural respondents, slightly high number of respondents with unrestricted and provisional driving permits.
- Three clusters (Cluster13, Cluster14, and Cluster15) show similar patterns: slightly high number of female respondents, slightly high number of rural respondents, slightly high number of respondents with driving permits. These clusters represent 17.92% of all respondents.
- Four clusters (Cluster21, Cluster22, Cluster23, and Cluster24) show similar patterns: slightly high number of female respondents, slight to moderately high number of urban respondents, high number of respondents with no driving permits. These clusters represent 24.33% of all respondents.

7. Conclusions

The teen survey data collected from Texas showed interesting trends and differences between respondents of the four categories of license statuses. This study found that males with provisional or unrestricted licenses were among the highest risk group, which is consistent with previous research findings. The transition to an unrestricted license has been shown to increase crash risk and, as the data reflects, these drivers reported engaging in higher risk driving behaviors more frequently than novice drivers. Countermeasures to address this issue could include GDL revisions on age restrictions, which have been successfully enacted in New Jersey as well as in other countries (Curry et al., 2015; Williams, 2017). GDL in New Jersey increased the minimum age a teen can obtain a learner's permit to 16 (or 17 with no driver education) and applied GDL restrictions to new drivers all the way up to the age of 21 (Curry et al., 2015). Teens that remained in a restricted license phase for a longer period of time had a crash risk potential that remained steady, while teens transitioning into a full licensure had an increased crash risk (Curry et al., 2015). Ultimately, all states should consider extending the age for the different GDL stages. Additionally, Texas restrictions continue until the age of 18, but an individual can wait to be licensed until 18 and have no restrictions, despite the fact that car crashes remain the leading cause of injury and death for young adults until the age of 25 (CDC, 2016; Texas Department of Transportation, 2018). A recent study conducted by the National Highway Traffic Safety Administration (NHTSA) used a final sample of 14 selected studies representing 13 different States, and three represented GDL programs across most or all of the United States; this study found that GDL programs were associated with statistically reliable reductions in traffic crashes outcomes among teen drivers (Masten, Thomas, Korbelak, Peck, & Blomberg, 2015). It is important to note that the current study is unique as it brings to light the complexity in identifying key crash potentials among the teens in Texas. Due to its large sampling size, the analysis and related conclusions can provide insights into teen driver trends across the country.

In addition, the data in this study showed that rural and novice drivers with learners permit status reported participating in risky behaviors like riding without a seatbelt, using a cell phone while (texting and talking), and speeding at least one to five times within the past month. While the frequency of engaging in these behaviors is lower, due to the types of roadways these types of behaviors are occurring on, there is a greater likelihood of a fatality and is therefore an important issue to address. Countermeasures should include targeted education to rural localities whose population continues to be overrepresented in car crashes compared to urban localities (Rural, 2016). Another study showed that teen driver education programs, stringent driving requirements, strict driving tests, and safe driving education can all help decrease the overall rate of teen crashes (Duddu, Kukkapalli, & Pulugurtha, 2018). Many states and metropolitan planning organizations (MPOs) consider teen driving as an emphasis area for safety improvement. For example, the Alabama Safe Teen Driving Coalition has worked with several agencies and organizations throughout the state to create a "Safe Teen Driving Toolkit" (Children's of Alabama, 2019). Our findings show that distraction is one of the key risk factors for the teen drivers. Birmingham MPO initiated a distracted-driver simulator program to educate young drivers on the dangers of distraction (Regional Planning Commission of Greater Birmingham, 2019). This approach would be beneficial for many other state DOTs and MPOs. To improve teen driving safety, several key initiatives have been taken: (a) implement media campaign to help the public understand GDL requirements, (b) communicate safe driving behaviors to young drivers and prepare preteens for the responsibility of safe driving, and (c) educate parents about the issues of teen driving. Another key risk factor is speeding behavior of the teen drivers. Several countermeasures can be taken to reduce the speeding related teen driving behavior such as safe driving insurance policy for the teen drivers, usage based insurance programs, and in-vehicle driving monitoring.

It is difficult to identify a countermeasure specific to only young male drivers, but countless studies and crash data highlight the continued need to target this demographic. Education and outreach are essential for this population and efforts to reach young males should be encouraged – both prior to driving and after receiving an unrestricted license. This study also conducted analysis to determine the significant patterns based on key surrogate measures. The findings show several patterns of the key surrogates based on gender, location, and licensing permits: rural male respondents are likely to have some driving permits, while urban female respondents are less likely to have any driving permits. The findings from the TCA must be interpreted with the surrogate measures from the cluster analysis. A more in-depth analysis on the surrogate measures by including risk perceptions can be conducted in future.

Acknowledgement

The authors are grateful to an anonymous reviewer for providing useful comments and suggestions.

References

Baireddy, R., Zhou, H., & Jalayer, M. (2018). Multiple correspondence analysis of pedestrian crashes in rural Illinois. *Transportation Research Record: Journal of the Transportation Research Board*.

Brown, I. D., & Groeger, J. A. (1988). Risk perception and decision taking during the transition between novice and experienced driver status. *Ergonomics*, *31*, 585–597.

CDC (2016). Motor Vehicle Safety, Teen Drivers. https://www.cdc.gov/motorvehiclesafety/teen_drivers/index.html Accessed: July 27, 2018.

Children's of Alabama. Alabama safe teen driving toolkit. https://www.childrensal.org/safe-teen-driving-toolkit Accessed: May 17, 2019.

Choulakian, V. (2006a). L1-norm projection pursuit principal component analysis. *Computational Statistics & Data Analysis*, *50*, 1441–1451.

Choulakian, V. (2006b). Taxicab correspondence analysis. *Psychometrika*, *71*, 1–13.

Choulakian, V. (2013). Graph partitioning by correspondence analysis and taxicab correspondence analysis. *Journal of Classification*, *30*, 397–427.

Curry, A. E., Hafetz, J., Kallan, M. J., Winston, F. K., & Durbin, D. R. (2011). Prevalence of teen driver errors leading to serious motor vehicle crashes. *Accident Analysis & Prevention*, *43*(4), 1285–1290.

Curry, A. E., Pfeiffer, M. R., Durbin, D. R., & Elliot, M. R. (2015). Young driver crash rates by licensing age, driving experience and license phase. *Accident Analysis and Prevention*, *80*, 243–250.

Das, S., Avelar, R., Dixon, K., & Sun, X. (2018). Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis & Prevention*, *111*, 43–55.

Das, S., Brimley, B. K., Lindheimer, T., & Pant, A. (2017). *Safety impacts of reduced visibility in inclement weather. Final report: ATLAS-2017-19.* University of Michigan.

Das, S., Jha, K., Fitzpatrick, K., Brewer, M., & Shimu, T. (2019). Pattern identification from elderly cyclist fatal crashes. *Transportation Research Board Annual Meeting, Washington D.C.*

Das, S., & Sun, X. (2015). Factor association with multiple correspondence analysis in vehicle–pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2519*, 95–103.

Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research*, *39*, 146–155.

Duddu, V., Kukkapalli, V., & Pulugurtha, S. (2018). Crash risk factors associated with injury severity of teen drivers. *IATSS Research*.

Ebbesen, E. B., & Haney, M. (1973). Flirting with death: Variables affecting risk taking at intersections. *Journal of Applied Social Psychology*, *3*, 303–324.

Ehsani, J. P., Hayie, D. L., Luthers, C., Perlus, J., Gerber, E., Ouimet, M. C., ... Simons-Morton, B. (2015). Teen drivers' perceptions of their peer passengers. *Transportation Research Board: Journal of the Transportation Research Board*, *2516*.

Factor, R., Yair, G., & Mahalel, D. (2010). Who by accident? The social morphology of car accidents. *Risk Analysis.*, *30*(9).

Fernandes, R., Hatfield, J., & Soames Job, R. F. (2010). A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*, 179–196.

Fontaine, H. (1995). A typological analysis of pedestrian accidents. *Presented at the 7th workshop of ICTCT, Paris, 26–27 October*.

Gershon, P., Ehsani, J., Zhu, C., O'Brien, F., Klauer, S., Dingus, T., & Simons-Morton, B. (2018). Vehicle ownership and other predictors of teenagers risky driving behavior: Evidence from a naturalistic driving study. *Accident Analysis and Prevention*, *118*, 96–101.

Ginsburg, K. R., Winston, F. K., Senserrick, T. M., Garcia-Espana, F., Kinsman, S., Quistberg, A., ... Elliot, M. R. (2008). National young-driver survey: Teen perspective and experience with factors that affect driving safety. *Pediatrics*, *121*, 1391–1403.

Harbeck, E. L., & Glendon, A. I. (2013). How reinforcement sensitivity and perceived risk influence young drivers' reported engagement in risky driving behaviors. *Accident Analysis and Prevention*, *54*, 73–80.

Hatfield, J., & Fernandes, R. (2009). The role of risk-propensity in the risky driving of younger drivers. *Accident Analysis and Prevention*, *41*, 25–35.

Jalayer, M., Pour-Rouholamin, M., & Zhou, H. (2018). Wrong-way driving crashes: A multiple correspondence approach to identify contributing factors. *Traffic Injury Prevention*, *19*, 35–41.

Jonah, B. A., & Dawson, N. E. (1987). Youth and risk: Age differences in risky driving, risk perception, and risk utility. *Alcohol, Drugs, and Driving*, *3*, 13–29.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. Report no. DOT HS 810 594.* NHTSA.

Masten, S., Thomas, F., Korbelak, K., Peck, R., & Blomberg, R. (2015). *Meta-Analysis of Graduated Driver Licensing Laws. Report no. DOT HS 812 211.*
Mayhew, D. R., Simpson, H. M., & Pak, A. (2003). Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis and Prevention*, *35*, 683–691.
McCartt, A., Shabanova, V., & Leaf, W. (2003a). Driving experience, crashes, and traffic citations of teenage beginning drivers. *Accident Analysis and Prevention*, *35*, 311–320.
McCartt, A. T., Shabanova, V. I., & Leaf, W. A. (2003b). Driving experience, crashes and traffic citations of teenage beginning drivers. *Accident Analysis & Prevention*, *35*, 311–320.
McDonald, K., & Sommers, M. S. (2015). Teen drivers' perceptions of inattention and cell phone use while driving. *Traffic Injury Prevention.*, *16*, 52–58.
Mirman, J. H., Albert, D., Jacobsohn, L. S., & Winston, F. K. (2012). Factors associated with adolescents' propensity to drive with multiple passengers and to engage in risky driving behaviors. *Journal of Adolescent Health*, *50*, 634–640.
Mirman, J. H., Durbin, D. R., Lee, Y., & Seifert, S. J. (2017). Adolescent and adult driver's mobile phone use while driving with different interlocutors. *Accident Analysis and Prevention*, *104*, 18–23.
NHTSA (2016). Rural/Urban Comparison of Traffic Safety Fatalities. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812521 Accessed: July 27, 2018.
NHTSA. Traffic Safety Facts: Young Drivers, 2016 Data. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812498 Accessed: July 27, 2018.
Olsen, E. O., Shults, R. A., & Eaton, D. K. (2013). Texting while driving and other risky motor vehicle behaviors among US high school students. *Pediatrics*, *131*, e1708–e1715.
O'Malley Olsen, E., Schults, R. A., & Eaton, D. K. (2013). Texting while driving and other risky motor vehicle behaviors among US high school students. *Pediatrics*, *131*, e1078–e1715.
Ouimet, M. C., Simons-Morton, B. G., Noelcke, E. A., Williams, A. F., Leaf, W. A., Preusser, D. F., & Hartos, J. L. (2008). Perceived risk and other predictors and correlates of teenagers' safety belt use during the first year of driving. *Traffic Injury Prevention*, *9*, 1–16.
Özkan, T., & Lajunen, T. (2005). Why are there sex differences in risky driving? The relationship between sex and gender-role on aggressive driving, traffic offences, and accident involvement among young Turkish drivers. *Aggressive Behavior*, *31*, 547–558.
Özkan, T., & Lajunen, T. (2006). What causes the differences in driving between young men and women? The effects of gender roles and sex on young drivers' driving behaviour and self-assessment of skills. *Transportation Research Part F: Traffic Psychology and Behaviour*, *9*, 269–277.
Peek-Asa, C., Britton, C., Young, T., Pawlovich, M., & Falb, S. (2010). Teenage driver crash incidence and factors influencing crash injury by rurality. *Journal of Safety Research*, *41*, 487–492.
Qing, L., Feng, G., Kim, I., Klauer, S. G., & Simons-Morton, B. G. (2018). A Bayesian finite mixture change-point model for assessing the risk of novice teenage drivers. *Journal of Applied Statistics*, *45*, 604–625.
Regional Planning Commission of Greater Birmingham. 2040 Regional Transportation Plan. http://rpcgb.org/transportation-planning/regional-transportation-plan/ Accessed: May 17, 2019.
Rhodes, N., & Pivik, K. (2011). Age and gender differences in risky driving: The roles of positive affect and risk perception. *Accident Analysis and Prevention*, *43*, 1879–2057.
Rundmo, T., & Iversen, H. (2004). Risk perception and driving behaviour among adolescents in two Norwegian counties before and after a traffic safety campaign. *Safety Science*, *42*, 1–21.
Simons-Morton, B. (2007). Parent involvement in novice teen driving: Rationale, evidence of effects, and potential for enhancing graduated driver licensing effectiveness. *Journal of Safety Research*, *38*, 193–202.
Simons-Morton, B., Lerner, N., & Singer, J. (2005). The observed effects of teenage passengers on the risky driving behavior of teenage drivers. *Accident Analysis & Prevention*, *37*, 973–982.
Simons-Morton, B., Ouimet, M., Zhang, Z., Lee, S., Klauer, S., Wang, J., ... Dingus, T. (2011). Crash and risky driving involvement among novice adolescent drivers and their parents. *American Journal of Public Health*, *101*, 2362–2367.
Simons-Morton, B., & Ouimet, M. C. (2006). Parent involvement in novice teen driving: A review of the literature. *Injury Prevention*, *12*, i30–i37.
Simons-Morton, B. G., Li, K., Ehsani, J., Ouimet, M. C., Perlus, J., & Klauer, S. G. (2016). Are perceptions about driving risk and driving skill prospectively associated with risky driving among teenagers? *Transportation Research Record: Journal of the Transportation Research Board*, *2584*.
Texas Department of Transportation, Graduated Driver Licensing Informational Flyer. https://ftp.dot.state.tx.us/pub/txdot-info/trf/graduated_driver_licensing.pdf. Accessed July 27, 2018.
Trankle, U., Gelau, C., & Metker, T. (1990). Risk perception and age-specific accidents of young drivers. *Accident Analysis and Prevention*, *22*, 119–125.
Voas, R., & Kelly-Baker, T. (2008). Licensing teenagers: Nontraffic risks and beneftis in the transition to driving status. *Traffic Injury and Prevention*, *9*, 89–97.
Williams, A. (2017). Graduated driver licensing (GDL) in the United States in 2016: A literature review and commentary. *Journal of Safety Research*, *63*, 29–41.
Williams, A. F. (2003). Teenage drivers: Patterns of risk. *Journal of Safety Research*, *34*, 5–15.

Subasish Das is an Associate Transportation Researcher at Texas A&M Transportation Institute (TTI). He received his Ph.D. in Civil Engineering from University of Louisiana at Lafayette in 2015 and holds a Master of Science in Civil Engineering from the same university in 2012. He has more than 10 years of national and international experience associated with transportation safety engineering research projects. His primary fields of research interest are roadway safety, roadway design, and associated operational issues. Dr. Das is the author or co-author of over 60 technical papers or research reports.

Lisa Minjares-Kyle is an Associate Transportation Researcher under the Youth Transportation Safety Program at the Texas A&M Transportation Institute (TTI). Her primary area of expertise involves research, development of educational materials, and outreach pertaining to young drivers. Topics of interest include distracted driving, impaired driving, child passenger safety, peer-to-peer outreach and data analysis. She has conducted extensive literature reviews on multiple transportation topics and has co-authored several traffic safety technical papers as well as two traffic safety programs regarding drug and alcohol traffic safety issues. She led in the development of the Continuing Professional Education workshop for Teen Traffic Safety and played an active role in the development of the Peer-to-Peer Training Initiative for young teens.

Lingtao Wu has more than ten years of experience in research on topics related to crash data analysis, highway design, safety effectiveness evaluation, and crash modification factor (CMF) development. He has authored over 30 technical papers and book chapters, and is a widely recognized researcher in traffic safety. Dr. Wu has lead or been involved in more than ten projects. He is the recipient of numerous national and international awards, such as the Transportation Research Board (TRB) Young Researcher Paper Award, the International Road Federation Fellowship, the European Road Safety Award, and the China Highway and Transportation Society Prize.

Russell H. Henk has been a member of TTI's staff for 29 years, a registered Professional Engineer in the State of Texas for 23 years, and currently serves as Program Manager of the Youth Transportation Safety Program in TTI's Center for Transportation Safety. He has supervised and/or participated in over 100 research projects and technical studies. These technical activities have primarily been focused on transportation system operations, management and safety, with most of his recent projects being in the area of young driver and passenger safety.