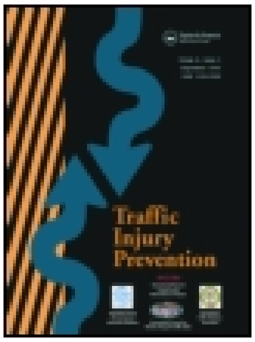

Taylor & Francis
Taylor & Francis Group

Traffic Injury Prevention

ISSN: 1538-9588 (Print) 1538-957X (Online) Journal homepage: https://www.tandfonline.com/loi/gcpi20

A semi-automated tool for identifying agricultural roadway crashes in crash narratives

Amber Brooke Trueblood, Ashesh Pant, Jisung Kim, Hye-Chung Kum, Marcelina Perez, Subasish Das & Eva Monique Shipp

To cite this article: Amber Brooke Trueblood, Ashesh Pant, Jisung Kim, Hye-Chung Kum, Marcelina Perez, Subasish Das & Eva Monique Shipp (2019): A semi-automated tool for identifying agricultural roadway crashes in crash narratives, Traffic Injury Prevention

To link to this article: https://doi.org/10.1080/15389588.2019.1599873

TRAFFIC INJURY PREVENTION
https://doi.org/10.1080/15389588.2019.1599873

A semi-automated tool for identifying agricultural roadway crashes in crash narratives

Amber Brooke Trueblood[a], Ashesh Pant[a], Jisung Kim[b], Hye-Chung Kum[c], Marcelina Perez[a], Subasish Das[d], and Eva Monique Shipp[a]

[a]Center for Transportation Safety, Texas A&M Transportation Institute, College Station, Texas; [b]Transportation Planning, Texas A&M Transportation Institute, College Station, Texas; [c]Department of Health Policy & Management, Texas A&M University, College Station, Texas; [d]Traffic Operations & Roadway Safety Division, Texas A&M Transportation Institute, College Station, Texas

ABSTRACT

Objective: Crash reports contain precoded structured data fields and a crash narrative that can be a source of rich information not included in the structured data. The narrative can be useful for identifying vulnerable roadway users, such as agricultural workers. However, using the narratives often requires manual reviews that are time consuming and costly. The objective of this research was to develop a simple and relatively inexpensive, semi-automated tool for screening crash narratives and expediting the process of identifying crashes with specific characteristics, such as agricultural crashes.

Methods: Crash records for Louisiana from 2010 to 2015 were obtained from the Louisiana Department of Transportation (LaDOTD). Records with narratives were extracted and stratified by vehicle type. The majority of analyses focused on a vehicle type of farm equipment (Type T). Two keyword lists, an inclusion list and an exclusion list, were created based on the published literature, subject-matter experts, and findings from a pilot project. Next, a semi-automated tool was developed in Microsoft Excel to identify agricultural crashes. Lastly, the tool's performance was assessed using a gold standard set of agricultural narratives identified through manual review.

Results: The tool reduced the search space (e.g., number of narratives that need manual review) for narratives requiring manual review from 6.7 to 59.4% depending on the research question. Sensitivity was high, with 96.1% of agricultural crash narratives being correctly classified. Of the gold standard agricultural narratives, 58.3% included an equipment keyword and 72.8% included a farm equipment brand.

Conclusion: This article provides information on how crash narratives can supplement structured crash data. It also provides an easy-to-implement method to facilitate incorporating narratives into safety research along with keyword lists for identifying agricultural crashes.

ARTICLE HISTORY
Received 6 November 2018
Accepted 23 March 2019

KEYWORDS
Agriculture; injury; motor vehicle crashes; occupational health; text mining; crash narratives

Introduction

In 2016, the Agriculture, Forestry, and Fishing, and Hunting (AFF) sector accounted for 2,351,500 employees in the United States (U.S. Department of Labor 2017b). This number may be an undercount since the U.S. Department of Agriculture and the National Institute for Occupational Safety and Health (NIOSH) estimates range from 1.4 to 3.2 million farmers alone (NIOSH 2018; U.S. Department of Agriculture 2018). Although the AFF sector is not the largest in the United States, it is one of the most hazardous in regards to fatal and nonfatal work injuries (NIOSH 2018). In 2015, the AFF sector had the third highest number of fatal work injuries and the highest fatal work injury rate (22.8 per 100,000 full-time equivalents; U.S. Department of Labor 2017a). Roadway crashes make a notable contribution to these events with a higher crash related fatal injury rate than the other sectors (Green et al. 2011). From 2003 to 2008, the rate of occupational transportation fatal injuries in the AFF sector was 5.59 per 100,000 employed workers compared to 0.94 per 100,000 employed workers for all sectors combined (Green et al. 2011).

Several studies utilized crash data to characterize farm equipment crashes and identify associated factors (Costello et al. 2009; Gkritza et al. 2010; Harland et al. 2014, 2018; see Bibliography in Appendix, online supplement). Farm equipment vehicles are used to transport farm products or supplies to and from farming activities (e.g., harvesting) and often are slow moving (e.g., tractors). Variables associated with farm equipment crashes or their severity include nonfamily hired drivers, younger drivers, passenger status, injury history, impairment, lack of restraint use, vehicle age, roadway geometry, and crash factors (e.g., rear end collision,

CONTACT Amber Brooke Trueblood A-Trueblood@tti.tamu.edu Center for Transportation Safety, Texas A&M Transportation Institute, 2929 Research Parkway, Office 317, College Station, TX 77843.
Associate Editor H. Clay Gabler oversaw the review of this article.
Supplemental data for this article can be accessed on the publisher's website.

single-vehicle crash, farm vehicle crossing centerline or median, obstructed vision, rural area, dry road, lighting conditions, speed limit) (Costello et al. 2009; Gkritza et al. 2010; Harland et al. 2018).

Structured data fields contained within the crash report are the basis of most traffic safety studies. The crash reports are completed by law enforcement officers and are then submitted and housed by state departments of transportation. The fields in the crash report and their precoded categories vary by state. Most states do not have structured data fields for detailed information on agricultural or other types of occupational crashes or equipment (e.g., construction, mowing). In addition, officers complete a free-text crash narrative to supplement the information in structured crash data. The crash narrative can serve as a rich source for helping to identify crashes involving special equipment or populations, establishing crash causation including the vehicle at fault, and obtaining other information that could support injury prevention efforts. Literature shows that analysis of crash narratives typically relied on multiple reviewers to manually code narratives (Iragavarapu et al. 2015; Pollack et al. 2013). In addition, more advanced data mining methodologies and predictive models have been applied to crash narratives (Das et al. 2018; Fitzpatrick et al. 2017; Nayak et al. 2009). For example, Fitzpatrick and colleagues (2017) utilized logistic regression to determine whether crashes were speeding related based on developed crash typologies. The researchers then validated the results of the logistic regression through manually reviewing a sample of crash narratives. Limited studies utilize crash narratives to obtain information on farm equipment crashes. Most recently, Gkritza et al. (2010) used structured crash data, as well as narratives, to identify farm vehicle crashes in Iowa for an analysis of injury severity and crash characteristics. Text mining has also been conducted on death certificates (e.g., cause of death, job classification) and fatality surveillance systems (e.g., Kentucky Fatality and Control Evaluation) to characterize farm tractor injuries, as well as newspaper clippings to identify agricultural injury fatalities (Bernhardt and Langley 1999; Bunn et al. 2008; Goodman et al. 1985; Hayden et al. 1995). Most of these publications are more than 10 years old and only a few utilized crash narratives. There is no literature that utilizes text mining methodologies to identify agricultural crashes based on narratives.

Although incorporating crash narratives into safety analyses can be tremendously beneficial because they contain additional information, a significant limitation to using narratives or open text fields is the time and human resources required for manual review. This limitation can be minimized through complex computer science methodologies or other software tools. However, computer science methodologies often require advanced programming skills.

The objective of this article was to describe the development of a simple semi-automated methodology to identify agricultural crashes from crash report narratives that can be widely used with little technical training, is relatively inexpensive, and can be applied to other fields. To our knowledge, this tool would be one of the first to meet this objective. Our goal was to identify agricultural crashes defined as crashes involving farm equipment vehicles used for farming purposes (e.g., cotton picking), other equipment used for farming purposes (e.g., bunchers), transportation of farming goods and products (e.g., crops), or persons who work in farming. A secondary goal was to develop a gold standard data set that can be used to develop and validate sophisticated machine learning methods. To this end, this article describes an easy-to-use and adaptable semi-automated tool to identify agricultural crashes using narratives from Louisiana.

Methods

Data collection

We obtained structured crash data and narratives for the period from 2010 to 2015 from the Louisiana Department of Transportation and Development (LaDOTD). This research was approved by the Texas A&M Institutional Review Board (study # 2016-0592D). The State of Louisiana defines farm equipment as "a vehicle designed and used primarily as a farm implement for drawing plows, moving machines, and other implements of husbandry" (LaDOTD 2005, p. 29). In the *Uniform Motor Vehicle Traffic Crash Report*, farm equipment vehicles were designated as Type T, and crashes involving Type T vehicles were designated as Type T crashes (LaDOTD 2005). The definition used for Type T vehicles did not meet our project's definition (i.e., a crash involving an agricultural worker, vehicle, or equipment), so we set out to develop a semi-automated method to identify agricultural crashes using information in the Type T narratives. We also wanted to assess the prevalence of agricultural crashes mixed into similar but different vehicle types, including (1) Type D: Passenger car, light truck, van, or an SUV with a trailer; (2) Type L: Single-unit truck with 2 axles; and (3) Type Z: Other. These vehicle types were the most likely to also have agricultural crashes present.

Methodology development

This article focuses on developing a semi-automated method to identify agricultural crashes as well as assess the performance of the method compared to manual review (e.g., an individual or team reviewing and classifying each record). Semi-automated methods are a hybrid human–computer process where the goal is to use automated tools to make the manual review process more efficient and feasible rather than replace them entirely. The next 2 sections describe the 2 steps taken to create the methodology, including the development of (1) a gold standard data set and (2) keyword lists and the Excel methodology.

Gold standard data set development

Type T (farm equipment) vehicles were reviewed first because these were the most likely to include agricultural vehicle types. There were a total of 573 crashes involving Type T vehicles, but only 448 crashes contained narratives

Table 1. Gold standard data set by manual review classification and vehicle body style.

	Type T (farm equipment)	Type D (passenger car, light truck, van, or an SUV with a trailer)	Type L (single-unit truck with 2 axles)	Type Z (other)
Agricultural	103	13	3	1
Nonagricultural	150	129	81	59
Ambiguous	195	108	166	190
Total	448	250	250	250

with at least 100 characters. All of the narratives were combined into a single text file for easy processing in Excel. To facilitate the development and evaluation of methods to flag agricultural crashes, we conducted a rigorous manual review to construct a gold standard data set. Two reviewers read and classified all Type T (farm equipment) crash narratives as either agricultural, nonagricultural, or ambiguous. Ambiguous narratives were unclear and had no information to conclusively determine whether they were agricultural or not. For example, if a narrative simply stated that vehicle 1 and vehicle 2 were involved in a crash, it was marked as ambiguous. Next, if a narrative mentioned a tractor without a purpose or activity, during manual review the reviewers classified it as ambiguous. The LaDOTD (2005) classifies grass cutters and mowers as farm equipment, but these were not classified as agricultural based on the project definition. To understand how many grass cutters and mowers are coded as Type T (farm equipment), there was an additional flag as part of the manual review.

Next, Types D (passenger car, light truck, van, or an SUV with a trailer), L (single-unit truck with 2 axles), and Z (other) vehicles were reviewed because of a higher probability of agricultural vehicles being included based on vehicle body style. A sample of 750 crashes was taken for these additional vehicle types to explore the prevalence of agricultural crashes in these vehicle types. All of the narratives were combined into a text file per vehicle type for easy processing in Excel. Next, the same steps taken for Type T manual review were applied to a stratified random sample of approximately 250 crashes (100 narratives with inclusion keywords, indicative of agricultural crashes, and 150 without these keywords) for the 3 other vehicle types D, L, and Z. The manual review resulted in a gold standard data set of 1,198 crashes (Table 1).

Keyword and methodology development

Two keyword lists were developed through a review of literature, discussions with subject-matter experts, and findings from a student pilot project. The inclusion keyword list included the following categories: Farmworkers (e.g., farmer), crops or farm animals (e.g., cotton), farm equipment (e.g., planter), farm equipment brands (e.g., John Deere), and other farm-related vocabulary (e.g., "farm"; see Table A1, online supplement). The exclusion keyword list was developed which included nonagricultural terms, including construction and mowing terms (see Table A2, online supplement).

Next, the keyword lists were used to develop an Excel-based algorithm. The steps undertaken in Excel to analyze and flag the narratives are as follows (in-depth methodology and the Excel tool can be found at: http://groups.tti.tamu.edu/cts/files/2018/10/Type-T-Excel-Methodology_Final.pdf):

1. An Excel workbook was created and a worksheet was added with the narratives, supplemental text if provided, and crash identifiers.
2. The inclusion and exclusion keywords were added to a second worksheet in the Excel workbook. The keywords were formatted by adding a space before and after the keyword to ensure that the algorithm matched on full words only (exact match). For example, if there is a keyword "rig," the algorithm should only flag the narrative if a word "rig" exists. The algorithm should not flag the narrative if it contains the word "brigade," because the second, third, and fourth letters from "brigade" form the substring "rig.
3. The narratives were then formatted to remove punctuation and alphanumeric characters (e.g., commas, semicolons, hyphens, periods). In addition, a blank space was added at the start and end of each narrative to allow a keyword to be flagged if it was the first or last word in the narrative.
4. A SEARCH function was then applied to each column to check whether any inclusion or exclusion terms from a keyword group were present in the narrative. The SEARCH function was applied separately. The SEARCH function returned a "1" in separate columns if the narrative in that row contained any inclusion or exclusion keywords; otherwise, it was marked with "0." It is important to note that the SEARCH function ignores capitalization.
5. Narratives were then classified as ambiguous if both an inclusion and exclusion keyword were present.

Analysis

Lastly, to assess performance of the Excel tool, standard epidemiologic measures (e.g., sensitivity, also known as *recall* in computer science, specificity, false positives, false negatives, positive predictive value, and negative predictive value) of how well a tool or test classifies a characteristic or condition and standard measures of tool performance in computer science (e.g., reduction of search space or the number of narratives that need manual review and sensitivity) were computed (Gordis 2014) for crashes involving Type T (farm equipment) crashes. STATA 14 SE was utilized (StataCorp LLC, College Station, TX). Because very few agricultural crashes were identified through manual review in the other vehicle types, the majority of the analysis in this article focuses on crashes involving Type T (farm equipment) vehicles.

For the calculation of the epidemiologic measures, only Type T narratives that were found to be unambiguous by the manual review were included ($n = 253$). In addition, crashes classified as ambiguous by the tool were included with the nonagricultural crashes because the exclusion keyword list that was applied after the inclusion keyword list indicated that these crashes were most likely nonagricultural. Consequently, a crash narrative with both an inclusion and an exclusion keyword would ultimately be classified as nonagricultural.

Results

Of the 448 manually reviewed Type T narratives, 253 (56.5%) were deemed unambiguous based on the manual review as shown in Table 2, which means these narratives could be classified as either agricultural or nonagricultural. The manual review determined 103 (40.7%) of these unambiguous narratives were classified as agricultural versus 150 (59.3%) that were classified as nonagricultural.

Based on the application of the semi-automated tool on the 253 unambiguous gold standard crashes, sensitivity and specificity were both high and false positives and false negatives were low. For sensitivity, 96.1% (95% confidence interval [CI], 90.4–98.9) of gold standard agricultural crashes were correctly classified as agricultural by the semi-automated tool ($n = 99$). For specificity, 92.7% (95% CI, 87.3–96.3) of gold standard nonagricultural crashes were correctly classified as nonagricultural ($n = 139$). The tool misclassified 3.9% (95% CI, 1.1–9.62) of gold standard agricultural crashes as nonagricultural (false negative; $n = 4$) and 7.3% (95% CI, 3.7–12.7) of gold standard nonagricultural crashes were misclassified as agricultural (false positive; $n = 11$). The positive predictive value was 90.0% (95% CI, 82.8–94.9), and the negative predictive value was 97.2% (95% CI, 93.0–99.2). Of the 4 false negatives, one narrative was correctly identified as agricultural with the inclusion keyword list but still excluded after applying the exclusion keyword list. The 3 remaining narratives that the algorithm could not detect were all tractors traveling to a worksite that did not match any of the other inclusion keywords. There were 11 narratives flagged as agricultural, but manual review found them to be nonagricultural. These included 2 parish or county tractors, 2 grass cutters without a tractor specified, 1 grass cutter attached to a parish tractor, 2 tractors involved with an electrical company or work, 3 tractors with no purpose specified, and 1 recreational vehicle. These 11 narratives did not contain sufficient information and did not use any exclusion list terms to be flagged by the exclusion keyword list. For example, phrases such as "cutting the grass" were not on the exclusion list. Ambiguous narratives presented a challenge for classification by both the gold standard manual review and the Excel tool. Although they had at least 100 characters, the ambiguous narratives simply had too few words to be identified as agricultural or nonagricultural by a human or a computer; there was not good agreement in classification of ambiguous narratives. Only 30 (6.7%) of the crash narratives were classified by the tool as ambiguous. The tool classified 1 agricultural crash and 27 nonagricultural crashes as ambiguous. Based on manual review, 2 of the gold standard ambiguous crashes were also classified by the tool as ambiguous. Sixty percent ($n = 18$) of the 30 excluded crashes were identified as being a grass cutter. There were 195 ambiguous crashes identified by manual review. Of these, 42 (21.5%) were classified by the Excel tool as agricultural, 151 (77.4%) as nonagricultural, and 2 (1.0%) as ambiguous. The large number classified as nonagricultural by the tool is because crash narratives without an inclusion keyword default to this category.

To explore differences in character lengths, the means and medians were calculated for ambiguous narratives and agricultural and nonagricultural narratives (data not shown). The mean character length was 1,223 characters for ambiguous narratives and the average character length was 1,522 characters for agricultural and nonagricultural narratives. In comparison, the median character length was 909 for ambiguous narratives and 1,212 for agricultural and nonagricultural narratives.

The semi-automated tool significantly reduced the number of records that would need a manual review to a more feasible size. The size of the search space (e.g., number of narratives that need manual review) is dependent on the research question and tolerance for accepting false positives or false negatives as shown in Table 3. The proportion of narratives for Type T (farm equipment) vehicles that needed manual review ranged from 6.7 to 59.4% based on the research question as shown in Table 3.

An additional analysis was conducted to identify the likelihood of agricultural crashes involving vehicle types that were similar but different from Type T (farm equipment), including Type D (passenger car, light truck, van, or an

Table 2. Automated keyword tool and manual review results for Type T (farm equipment).

		Manual review			
		Agricultural	Nonagricultural	Ambiguous	Total
Tool	Agricultural	99	11	42	152
	Nonagricultural	3	112	151	266
	Ambiguous	1	27	2	30
	Total	103	150	195	448

Table 3. Search space reductions based on hypothetical research questions.

	Type T (farm equipment) ($n = 448$)		Types D, L, and Z ($n = 750$)	
Hypothetical research question	Narratives that require manual review	Percentage of narratives	Narratives that require manual review	Percentage of narratives
Validate agricultural crashes	152	33.90	239	31.90
Examine ambiguous crashes	30	6.70	65	8.70
Review crashes identified as nonagricultural to reduce false negatives	266	59.40	446	59.50

Table 4. Categories of manually verified farm equipment narratives.[a]

	Agricultural ($n = 103$)	Percentage of agricultural crash narratives ($n = 103$)
Person	3	2.90
Crop	51	49.00
Equipment	60	58.30
Brands	75	72.80
Other	59	57.30

[a]A narrative may include more than 1 keyword; thus, the sum is greater than 100%.

SUV with a trailer), L (single-unit truck with 2 axles), and Z (other). From 0.4 to 5.2% of the Type D (passenger car, light truck, van, or an SUV with a trailer), L (single-unit truck with 2 axles), and Z (other) sample narratives were agricultural based on manual review (see Table 1). Type D (passenger car, light truck, van, or an SUV with a trailer) had the highest percentage of agricultural crashes at 5.2% ($n = 13$; see Table 1). The search space to examine the likelihood of agricultural crashes involving these vehicle types was also reduced significantly from 8.7 to 59.5% as shown in Table 3.

Of the 103 agricultural crashes identified in the gold standard data set, 60 (58.3%) contained an equipment inclusion keyword and 75 (72.8%) contained a farm equipment brand inclusion keyword, indicating the importance of these keyword categories in identifying agricultural crashes (see Table 4).

Discussion

Although agricultural crashes are less common than other crash types, they still impact the health and safety of workers in the AFF sector, their families and coworkers, as well as other roadway users. The lack of information makes it difficult to develop appropriate countermeasures and programs (Gkritza et al. 2010). This article provides information regarding how crash narratives can be used to identify specific types of roadway users and to quantify potential misclassification for analyses based solely on the type of vehicle. For example, it can be difficult to confirm an agricultural crash through vehicle type alone. Thus, identifying crashes through crash narratives can help to improve crash identification and characterization.

One of the main barriers to using narratives effectively in a crash study is the manual review and coding of hundreds to thousands of records. The findings presented in this article illustrate how a semi-automated method can screen narratives for specific characteristics. For example, the algorithm correctly flagged 96.1% of narratives classified as agricultural in the gold standard data set. Although the process still required a manual review, the automated steps substantially narrowed the search space and the time needed for manual review. An important lesson learned is that the development of the applicable keyword list is an iterative process done through literature review, discussions with experts, and review of sample narratives. It is also important to account for special characters, punctuation, and spacing.

In regards to agricultural crashes, this article provides evidence that the use of the vehicle type or style from the structured data can lead to misclassification. In this case, the misclassification involved falsely identifying vehicles as agricultural (e.g., lawn mowers, construction vehicles). The use of exclusion terms was effective in identifying misclassification through flagging these narratives for review and exclusion. This misclassification would not have been a concern if the analysis focused only on crashes involving tractors rather than equipment and vehicles involved in agricultural activities.

The article found that 0.4 to 5.2% of the other vehicle types sampled (Types D, L, and Z) included agricultural crashes, which indicates that if only Type T (farm equipment) vehicles are included, agricultural crashes may be missed. Thus, for surveillance, it could be important to include additional vehicle types to capture all agricultural crashes.

A limitation of using crash narratives is that a portion of narratives may not contain sufficient information to be meaningful. In this analysis, 43.5% of Type T (farm equipment) narratives did not contain sufficient information to be classified as agricultural or not. Type T (farm equipment) narratives cannot be relied upon to identify only agricultural crashes. Rather, they can be useful for excluding nonagricultural crashes from an analysis and for estimating certain forms of misclassification. The issue of ambiguous narratives due to insufficient text or information sorely needs to be addressed. Much could be gained if more narratives were completed with information recorded in a systematic fashion. The impact of this issue could be minimized through training officers on how to complete narratives and the value of narratives for occupational health and injury prevention, specifically roadway safety. The quality and accuracy of any data improves as it is used more. Thus, if narratives are used to inform research and policy on a larger scale, comprehensive training and completion of the narrative could be a positive consequence.

Despite these limitations, a strength of the method presented in this article is that it can be widely used, quickly, and with little technical training across topic areas. This overcomes a major barrier to the use of narratives in crash research. Keyword lists can be developed and applied to additional populations and crash types. Future research could examine the performance of this approach for other specialized vehicle types or other state crash databases. The gold standard data combined with the lessons learned from this analysis can be used to inform the development and evaluation of complex machine learning based models.

In conclusion, this is one of the first articles to present an easy-to-implement Excel tool to facilitate incorporating crash narratives into an analysis. Existing literature has focused on more advanced data mining methodologies, as well as predictive modeling (e.g., logistic regression) that require expertise in these fields (Das et al. 2018; Fitzpatrick et al. 2017; Nayak et al. 2009). This tool requires a moderate Excel skill level but overall is easy to implement with little technical training and is relatively inexpensive. Lastly, the tool can be modified and utilized with any narrative or text field. The tool and more detailed methodology are publicly available to aid other agencies and researchers utilizing the Excel methodology.

Acknowledgment

The authors would like to thank the LaDOTD for assistance with providing the data and assisting with clarification questions.

Funding

This article was supported by CDC/NIOSH under Cooperative Agreement No. U50 OH07541 to the Southwest Center for Agricultural Health, Injury Prevention, and Education at the University of Texas Health Science Center at Tyler. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC/NIOSH.

Data availability statement

The data sets generated during the current study are not publicly available due to data use agreements. Data can be requested through the LaDOTD.

ORCID

Amber Brooke Trueblood http://orcid.org/0000-0002-1404-347X
Hye-Chung Kum http://orcid.org/0000-0002-6882-8053
Subasish Das http://orcid.org/0000-0002-1671-2753
Eva Monique Shipp http://orcid.org/0000-0002-4034-8031

References

Bernhardt JH, Langley RL. Analysis of tractor-related deaths in North Carolina from 1979 to 1988. *J Rural Health*. 1999;15:285–295.

Bunn TL, Slavova S, Hall L. Narrative text analysis of Kentucky tractor fatality reports. *Accid Anal Prev*. 2008;40:419–425.

Costello TM, Schulman MD, Mitchell RE. Risk factors for a farm vehicle public road crash. *Accid Anal Prev*.2009;41:42–47.

Das S, Dutta A, Avelar R, Dixon K, Sun X, Jalayer M. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*. 2018;23(1):30–48.

Fitzpatrick CD, Rakasi S, Knodler MA Jr. An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression. *Accid Anal Prev*. 2017;98:57–63.

Gkritza K, Kinzenbaw CR, Hallmark S, Hawkins N. An empirical analysis of farm vehicle crash injury severities on Iowa's public road system. *Accid Anal Prev*. 2010;42:1392–1397.

Goodman RA, Smith JD, Sikes RK, Rogers DL, Mickey JL. Fatalities associated with farm tractor injuries: an epidemiologic study. *Public Health Rep*. 1985;100:329–333.

Gordis L. *Epidemiology*. 5th ed. Philadelphia, PA: Elsevier Saunders; 2014.

Green MK, Harrison R, Leinenkugel K, et al. Occupational highway transportation deaths—United States, 2003–2008. *MMWR Morb Mortal Wkly Rep*. 2011;60(16):497–502.

Harland KK, Bedford R, Wu H, Ramirez M. Prevalence of alcohol impairment and odds of a driver injury or fatality in on-road farm equipment crashes. *Traffic Inj Prev*. 2018;19(3):230–234.

Harland KK, Greenan M, Ramirez M. Not just a rural occurrence: differences in agricultural equipment crash characteristics by rural–urban crash site and proximity to town. *Accid Anal Prev*. 2014;70:8–13.

Hayden GJ, Gerberich SG, Maldonado G. Fatal farm injuries: a five-year study utilizing a unique surveillance approach to investigate the concordance of reporting between two data sources. *J Occup Environ Med*. 1995;37:571–577.

Iragavarapu V, Khazraee SH, Lord D, Fitzpatrick K. Pedestrian fatal crashes on freeways in Texas. Paper presented at: Transportation Research Board 94th Annual Meeting; January 11–15, 2015; Washington, DC. Available at: https://trid.trb.org/view/1337822. Accessed March 11, 2019.

Louisiana Department of Transportation & Development. State of Louisiana Uniform Motor Vehicle Traffic Crash Report. LA: 2005. Available at: http://wwwsp.dotd.la.gov/Inside_LaDOTD/Divisions/Multimodal/Highway_Safety/Misc%20Documents/2005_LA_Crash_Report_Guide01.pdf. Accessed March 12, 2019.

National Institute for Occupational Safety and Health. *Agricultural Safety*. 2018. Available at: https://www.cdc.gov/niosh/topics/aginjury/default.html. Accessed October 25, 2018.

Nayak R, Piyatrapoomi N, Weligamage J. Application of text mining in analysing road crashes for road asset management. Paper presented at: 4th World Congress on Engineering Asset Management; September 28–30, 2009; Athens, Greece.

Pollack KM, Yee N, Canham-Chervak M, Rossen L, Bachynski KE, Baker SP. Narrative text analysis to identify technologies to prevent motor vehicle crashes: examples from military vehicles. *J Safety Res*. 2013;44:45–49.

U.S. Department of Agriculture. *Farm Demographics—U.S. Farmers by Gender, Age, Race, Ethnicity, and More*. 2018. Available at: https://www.agcensus.usda.gov/Publications/2012/Online_Resources/Highlights/Farm_Demographics/. Accessed October 25, 2018.

U.S. Department of Labor, Bureau of Labor Statistics. *Census of Fatal Occupational Injuries*. 2017a. Available at: https://www.bls.gov/news.release/cfoi.toc.htm. Accessed October 25, 2018.

U.S. Department of Labor, Bureau of Labor Statistics. *Employment by Major Industry Sector*. 2017b. Available at: https://www.bls.gov/emp/tables/employment-by-major-industry-sector.htm. Accessed October 25, 2018.