

2017

Analysis of workers' compensation claims data for improving safety outcomes in agribusiness industries

Sai Kumar Ramaswamy
Iowa State University

Follow this and additional works at: <http://lib.dr.iastate.edu/etd>

 Part of the [Agriculture Commons](#), [Industrial Engineering Commons](#), and the [Occupational Health and Industrial Hygiene Commons](#)

Recommended Citation

Ramaswamy, Sai Kumar, "Analysis of workers' compensation claims data for improving safety outcomes in agribusiness industries" (2017). *Graduate Theses and Dissertations*. 15403.
<http://lib.dr.iastate.edu/etd/15403>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

**Analysis of workers' compensation claims data for improving safety outcomes in
agribusiness industries**

by

Sai Kumar Ramaswamy

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Industrial and Agricultural Technology

Program of Study Committee:
Gretchen A. Mosher, Major Professor
Steven A. Freeman
Charles R. Hurburgh, Jr.
Keri L. Jacobs
Charles V. Schwab
Mack C. Shelley

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

Copyright © Sai Kumar Ramaswamy, 2017. All rights reserved.

This dissertation is dedicated to
Amma, Appa, Anupama and Avantika.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
CHAPTER 1. INTRODUCTION TO RESEARCH.....	1
Purpose of Research.....	4
Research Questions	6
Measurement and Methodology	7
Organization of Dissertation	13
References.....	13
CHAPTER 2. USING WORKERS' COMPENSATION CLAIMS DATA TO CHARACTERIZE OCCUPATIONAL INJURIES IN THE BIOFUELS INDUSTRY	24
Abstract	24
Introduction.....	25
Background.....	28
Methods and Data	33
Results and Discussion	36
Conclusion	50
References.....	51
CHAPTER 3. USING WORKERS' COMPENSATION CLAIMS DATA TO CHARACTERIZE OCCUPATIONAL INJURIES IN THE COMMERCIAL GRAIN ELEVATOR INDUSTRY	65
Abstract	65
Introduction.....	66
Methods and Data	71
Results and Discussion	74
Conclusions.....	83
References.....	85
CHAPTER 4. PREDICTING DAYS AWAY FROM WORK: COMPARING MODEL PERFORMANCE.....	96
Abstract	96
Introduction.....	97
Background.....	99
Methods and Data	101
Results and Discussion	106
Conclusions.....	112
References.....	113

CHAPTER 5. GENERAL CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK.....	122
Conclusions.....	122
Limitations	125
Recommendations for future research	126
APPENDIX.....	129
DECISION TREE MODEL.....	129

ACKNOWLEDGMENTS

First, I thank my major professor Dr. Gretchen Mosher for providing me the opportunity to earn my doctorate degree and for all her help, guidance, support, and encouragement during this study.

I thank Dr. Steve Freeman, Dr. Charles Hurburgh, Dr. Keri Jacobs, Dr. Charles Schwab and Dr. Mack Shelley for serving on my program of study committee and for their help and guidance throughout this project. I thank the mid-west based insurance company for their help with the data used in this study.

I thank my fellow graduate students, Saxon Ryan, Mathew Harvey, John Haughrey, Nurhaizan Mohd Zainudin, Wesley Chang, Fatemeh Davoudi, and Dr. Steve Simpson for their support and encouragement and for creating a fun work environment. I thank Jeremy Hadler for his help with the statistical analysis and Bret Ramirez for his help reviewing sections of this dissertation.

Finally, I thank my parents Ramamani Ramaswamy, and Ramaswamy Chakravarthy, my wife Anupama Vembar, and daughter Avantika Chakravarthy, for the love, support and constant encouragement all through the ups and downs of this endeavor without which none of this would have been possible.

CHAPTER 1. INTRODUCTION TO RESEARCH

Prevention of occupational injuries is a major issue for non-farm agricultural workplaces (Field et al., 2014). Furthermore, improving workplace safety outcomes requires learning from past incidents, identifying the most significant causes and implementing targeted prevention strategies (Abdolhamidzadeh, Abbasi, Rashtchian, & Abbasi, 2011; Anderson, 2009; Gotcheva et al., 2016; Pasman, 2009). However, obtaining detailed records of past incidents is a challenge acknowledged by investigators across several industrial sectors including agribusiness such as grain elevators and biofuel production (Calvo Olivares, Rivera, & Nunez McLeod, 2014, 2015; Dong, Largay, Wang, & Windau, 2014; Keren, 2010; Meel et al., 2007; Riedel & Field, 2011). Workers' compensation claims have been offered by previous researchers as an excellent data source to address existing informational gaps about safety incidents and injuries in the workplace (Utterback, Meyers, & Wurzelbacher, 2014; Wurzelbacher et al., 2016).

Workers' compensation claims data provide case-level injury information such as cause of injury, nature of injury, and type of injury. The claims also provide demographic information such as age, gender and cost information such as medical and indemnity payments (Nestoriak & Pierce, 2009; Reville, Bhattacharya, & Weinstein, 2001; Utterback et al., 2012; Wuellner, Adams, & Bonauto, 2016). Typically, workers' compensation claims dataset contain several thousand rows of data, a key factor in their applicability for injury surveillance (Oleinick & Zaidman, 2004; Wurzelbacher et al., 2016). Furthermore, the majority of employers in the U.S, including those in the grain elevator and biofuels industry, purchase workers' compensation insurance to provide benefits to their employee who suffers

a work-related injury (Sengupta, Reno, Burton Jr, & Baldwin, 2012). Therefore, analysis of workers' compensation claims data of biofuels and grain elevator companies can provide key insights on contributing factors of occupational injuries needed for enhancing safety outcomes in these high- hazard agricultural-based work environments. (Doupbrate, Rosecrance, Reynolds, Stallones, & Gilkey, 2009; Doupbrate, Rosecrance, Stallones, Reynolds, & Gilkey, 2009; Utterback & Schnorr, 2010).

Workers' compensation claims data have been used in the past to study work-related injuries in agriculture, but most of these have focused on farm-level analyses. For example, Spector et al. (2016) investigated heat exposure and injury risk of farm workers using Washington State Fund workers' compensation claims. Similarly, Doupbrate et al. (2009) examined livestock handling injuries of dairy operations in Colorado using workers compensation claims. Even though workers at grain elevators and biofuels are exposed to a wide variety of safety hazards and incur a higher rate of occupational injuries than do workers in other industries, very little research has examined the use of workers' compensation claims to characterize injuries in these environments.

Extracting useful information from large datasets such as workers' compensation claims requires a scientific and systematic technique (Gnanadesikan, 2011; Springmeyer, Blattner, & Max, 1992). Statistical methods such as correlation and regression have traditionally been used to explore data trends, link cause and effect, and develop prediction models (Chen, Jiang, Wang, & Tang, 2016; Nisbet, Miner, & Elder Iv, 2009). For example, Gorucu, Murphy, and Kassab (2015) used a chi-square test to gain a better understanding of the relationship between age, gender, cause of injury and number of fatalities in data from Pennsylvania State University's farm and agricultural injury database. Similarly, Reiner,

Gerberich, Ryan, and Mandel (2016), used logistic regression to investigate injuries from large agricultural machinery with data drawn from a rural injury database.

In the last decade, data mining is rapidly evolving as the preferred method of investigators for analyzing large amounts of data such as workers' compensation claims (Liao, Chu, & Hsiao, 2012; Sanmiquel, Rossell, & Vintro, 2015). Data mining is a multidisciplinary field that encompasses classical statistical techniques and new computational techniques such as decision trees and association rules (Anand et al., 2006). Over the last few years, data mining techniques such as decision trees are being used to uncover hidden patterns in large dataset across various disciplines (Cheng, Leu, Cheng, Wu, & Lin, 2012). However, the application of these techniques to characterize injury data is not a common use of the technique (Nenonen, 2013).

In this study a large workers' compensation claims dataset obtained from a leading private insurance company was investigated using statistical techniques such as chi-square tests, regression analysis, and data mining techniques such as decision trees. This dataset consisted of claims submitted by non-farm agricultural businesses such as grain elevators and biofuel producers. Obtaining injury data for these businesses is a challenge. Since, workers' compensation claims has been suggested as a good alternative source for injury and incident analysis, the objective of this study is to analyze these claims, identify injury causes, risks, and problem areas so supervisors and safety professionals can make decisions needed to improve safety outcomes in the workplace. Furthermore, safety incidents that cause injuries and fatalities have a widespread impact (Leigh & Marcin, 2012), and therefore mitigating these incidents using a proactive data-driven approach rather than just compliance can benefit

the worker, the organization, and society-at-large (Mosher, 2011; Simmons, Matos, & Simpson, 2016).

Purpose of Research

Occupational injuries in the U.S. continue to be a major concern in several industries including agribusiness such as grain elevators and biofuel production (Boden, O'Leary, Applebaum, & Tripodis, 2016; Lander, Nielsen, & Lauritsen, 2016; Mabila, Gracia, Cohen, Almberg, & Friedman, 2015; Wurzelbacher et al., 2016). According to the U.S. Bureau of Labor Statistics (BLS), in 2015 the incident rate of non-fatal occupational injuries for the agricultural industry including production agriculture was 71% higher than the national average (BLS, 2016).

Work-related injuries not only affect the injured worker and their family adversely but also influence the company's medical, liability and insurance premium costs (Hajakbari & Minaei-Bidgoli, 2014). In addition to direct costs such as medical and indemnity payments, there are indirect costs associated with workplace injuries. These include equipment damage, equipment repair, incident investigation time, the cost of hiring and training an injured worker's replacement, loss of reputation, loss of employee morale, loss of confidence and negative media attention (Gavious, Mizrahi, Shani, & Minchuk, 2009; Griend, 2011; OSHA, 2016). According to Manuele (2013), for every dollar in direct costs, there are an estimated \$4 of indirect costs associated with work-related injuries.

Based on the most recent data available in research literature, the total cost of occupational injuries and illnesses in the U.S. is estimated to be \$250 billion (Leigh, 2011). According to Marucci-Wellman et al. (2015), in the U.S., nearly \$1 billion is spent each week just to cover the direct costs of severe work-related injuries. In their study of

occupational injury costs in production agriculture, Leigh, McCurdy, and Schenker (2001) suggested that injuries in the agricultural industry cost 30% more than the national average on a per person basis. In a more recent study, Costich (2010) estimated the average cost of hospitalization for agricultural injuries be \$12,056. Assuming the direct cost of injuries in on-farm and non-farm agricultural industries to be equal and using the ratio suggested by Manuele (2013), the mean indirect costs of occupational injuries in the agricultural industry is estimated at \$48,224. Since the agricultural industry (farm as well as non-farm) is currently under financial stress due to low commodity prices (Ehmke, 2016), such high injury costs represent a further threat to their profitability. Furthermore, unsafe work environments influence employee perceptions and their actions thus affecting the success of organizational initiatives such as quality management practices (Mosher, Keren, & Hurburgh Jr, 2013). For this reason, an improved understanding of injuries and fatalities in the agricultural industry is necessary to prevent work-related injury risks before they occur.

The purpose of this study was to characterize the direct cost of occupational injury using the information obtained from the workers' compensation claims including variables such as age, tenure of employee, and nature, cause and type of injury. A secondary purpose of the study was to identify and classify at-risk groups within the grain elevator and biofuels production industry. To improve the safety of work conditions in these work environments, a better understanding of injuries and an enhanced process for identifying at-risk groups is necessary. The grain and biofuels industries spend a great deal of time and resources to mitigate safety hazards, so additional information on the injuries and at-risk groups has potential to maximizing the return on safety investments.

Research Questions

The overarching goal of this study was to characterize the injury direct costs and number of days away from work based on the injured employee demographic and injury attributes. This dissertation research consists of three individual studies and the research objectives of each of those studies are listed below.

In the first study the relationship between the direct costs of injury and the injured employee's demographic and injury attributes were examined for biofuel producers. The claim amount was used as a measure of the direct costs and the research questions examined are:

- Is the claim amount in the biofuels industries independent of the age of the injured employee?
- Is the claim amount in the biofuels industries independent of the tenure of the injured employee in the present organization?
- Is the claim amount in the biofuels industries independent of the type of claim?
- Is the claim amount in the biofuels industries independent of the nature of injury of the injured worker?
- Is the claim amount in the biofuels industries independent of the cause of injury of the injured worker?
- Is the claim amount in the biofuels industries independent of the body part affected of the injured worker?

In the second study the relationship between the direct costs of injury and the injured employee's demographic and injury attributes were examined for commercial grain handling

facilities. Just as in the first study, the claim amount was used as a measure of the direct costs and the research questions examined are:

- Is the claim amount in the commercial grain handling facilities independent of the age of the injured employee?
- Is the claim amount in the commercial grain handling facilities independent of the tenure of the injured employee in the present organization?
- Is the claim amount in the commercial grain handling facilities independent of the nature of injury of the injured worker?
- Is the claim amount in the commercial grain handling facilities independent of the body part affected of the injured worker?

In the third study the number of days away from work of the injured employee was investigated based on the employee's demographic, injury and organizational attributes. The previous two studies that focused on a particular agriculture-related industry such as biofuel and commercial grain elevators. This study examined the number of days away for various agriculture-related industries such as bulk commodity handling, food manufacturing, grocery and retail stores. The research objective of this study was:

- Is a linear regression model better than a decision tree model at predicting the number of days away from work based on the injured employee's age, tenure, nature of injury, cause of injury, body part injured, type of organization and the nature of their job?

Measurement and Methodology

The dataset used in this study was obtained from a private insurance company specializing in insurance products for agricultural businesses. The dataset was sent to

researchers in an electronic format and consisted of individual claims filed by the injured employee or their employer with the insurance company. These claims were reported to the insurance company from 2008 to 2016. The oldest claim in the dataset had an injury date of January 1, 2008, and the newest claim injury date was March 11, 2016.

Nearly all workers in the U.S. are covered by workers' compensation insurance (Utterback et al., 2014). Employers generally provide this benefit to their employees by either purchasing insurance from an insurance carrier or through self-insurance (Reville, Polich, Seabury, & Giddens, 2001). When an employee is injured on the job, the program pays the employee's medical and indemnity costs. To provide information and to facilitate the payment, employers must create a report of the worker's injury to (Utterback et al., 2012). Data collected during the claims process are provided by employees, employers, insurance companies and other involved parties (Utterback et al., 2014). All information recorded in the dataset were vetted and verified by insurance company personnel. The collection of information from multiple stakeholders makes workers' compensation claims records an excellent data source for occupational injuries (Dement et al., 2004; Janicak, 2010; Kim, Dropkin, Spaeth, Smith, & Moline, 2012; Reville et al., 2001).

The data used in this study consisted of 35,686 rows and 18 columns of data. The complete list of the 18 columns is provided in Table 1. The dataset did not contain any personal information that could be used to identify either the injured employee or their employer. Using the demographic information provided in the dataset, age of the employee was calculated as the difference between the date of birth and the injury date. Similarly, the tenure of the employee was calculated as the difference between the date of hire and injury date. The number of days away from work (DAFW) was calculated as the difference between

the injury date and the date returned to work. The claim amount variable used as a proxy for injury severity and was categorized as “less than \$3,000”, “\$3,001-\$9,999”, and “10,000 or greater”.

Table 1: *List of variables*

#	Column Name	Description
1	Claim	Unique identifier for each claim record
2	Effective Year	Filing year of the claim
3	Account	Unique identifier to differentiate claims for each customer
4	Market	Type of business (biodiesel or ethanol)
5	Gender	Gender of injured worker
6	Accident State	Name of state where injury occurred
7	Jurisdiction State	Name of state where the headquarters of the employer is located
8	Date of Birth	Date of birth of injured worker
9	Date of Hire	Date on which the present company hired the injured worker
10	Injury Date	Date on which the injury occurred
11	Date returned to work	Date on which the injured employee returned back to work
12	Claim Description	One-line narration of incident resulting in injury; Example: "Employee was cleaning equipment and employee opened up a line and acid sprayed in his face and mouth."
13	Claim Status	If the claim is still open or closed
14	Type of claim	Indicates if the claim was "medical only", "permanent disability", "death".
15	Body Part	Body part(s) injured
16	Cause of Injury	Main cause of injury. For example: "Dust, gasses or fumes inhalation", "Foreign matter in eyes", "Chemical exposure" etc.
17	Nature of Injury	Describes the type of injury such as Fracture, Strain, Contusion etc.
18	Claim amount	Total amount paid out in medical, indemnity and other miscellaneous payments. Used as a proxy for injury severity in this study.

Variables analyzed in this study were mostly categorical variables, therefore, statistical analysis began with the construction of frequency counts, percentages, and contingency tables. To characterize and analyze the claim amount against employee and injury variables, contingency tables were used to classify the variable pairs and chi-square

tests were used to test for an independent relationship. To identify the at-risk groups for posthoc tests, residual analysis was used to determine the nature of relationship between the row variable and the column variable of the contingency tables (Agresti & Finlay, 2008). The residual is the difference between the observed value of a specific variable pair and its expected value (Agresti & Finlay, 2008). A positive residual implies that the observed value was greater than the expected value, while a negative residual implies the observed value was less than expected value. Descriptive analysis was performed using Microsoft Excel and inferential analyses were performed with statistical software SAS version 9.4. According to Agresti and Finlay (2008), an adjusted residual of ± 2 is evidence of dependence between the row and column variables while an adjusted residual of ± 3 is evidence of strong dependence. Comparing the adjusted residuals of each cell in a contingency table helped identify the cells where the degree of dependence between the two variables was the strongest (Sharpe, 2015).

The contingency table and chi-square test method of statistical analysis were used previously to investigate injuries in agriculture-related industries such as crop and dairy farming. For example, Javadi and Rostami (2007) used contingency tables and chi-square test to categorize and identify causes of on-farm machinery injuries. Similarly, Sprince et al. (2003) used chi-square tests to determine the strength of association between farm injuries of Iowa farmers and variables such as farming work hours, education beyond high school and age. Finally, Karttunen and Rautiainen (2011) evaluated the risk factors for declined work ability among full-time dairy farmers using the chi-square statistical tests.

The exponential growth of data over the last two decades has led to the development of new set of tools to analyze large datasets (Anand et al., 2006). These tools are part of the

field of study called data mining. Data Mining involves retrieval and analysis of large datasets to successfully uncover hidden patterns among the data variables (Cheng et al., 2012). Data mining includes statistical tools such as regression and modern computing tools such as decision trees (Anand et al., 2006). According to Sanmiquel, Rossell, and Vintro (2015), decision tree models are extremely useful for investigating injury data and identifying the key factors contributing to these injuries.

Decision trees are one of the most widely used data mining technique (Liao et al., 2012). This technique was first proposed by Breiman et al. (1984) to recursively partition a set of data into homogeneous groups and displayed graphically in an inverted tree-like structure. This representation of information in an intuitive and easy to visualize format is a reason for the popularity of decision trees in data analysis (Elith, Leathwick, & Hastie, 2008). Decision trees can be used to characterize both numeric as well as categorical dependent variables (Loh, 2011). If the dependent variable is categorical, then the model is called classification tree, and if the dependent variable is numeric then decision tree is known as regression tree model (Razi & Athappilly, 2005).

There are several advantages of using decision trees for investigating large datasets such as workers compensation claims, used in this study (Savolainen, Mannering, Lord, & Quddus, 2011). Unlike statistical techniques that are parameter based and require certain assumptions for their model parameters, decision trees are non-parametric and hence do not require any such pre-defined assumptions (Breiman, 2001; Chang & Chen, 2005). Another advantage of decision tree is that they can be used with any type of predictor variables such as numeric, binary or categorical for characterizing a dependent variable (Elith et al., 2008; Harb, Yan, Radwan, & Su, 2009). Since decision tree have a hierarchical structure where

response of one input variable depends on the values of inputs higher in the tree, interactions between predictor variables are automatically modeled (Elith et al., 2008). Finally, decision tree models while being straightforward to interpret, also have prediction accuracy comparable to those of statistical techniques such as regression (Chang & Wang, 2006; Lim, Loh, & Shih, 2000; Meleddu & Pulina, 2016).

Building decision trees involve three major phases i) tree building ii) tree pruning and iii) testing (Breiman et al., 1984). The decision tree model is constructed starting with the complete data, and partitioning the data using a set of rules and one predictor variable at a time to create two or more mutually exclusive groups (Cheng et al., 2012). At each partition, the process is repeated, and a large tree is grown until each of partitioned group are as homogenous as possible so the outcomes can be predicted accurately (Nenonen, 2013). Once the tree building is complete, the tree is then pruned by removing branches that do not contribute significantly in characterizing the dependent variable (Strobl, Malley, & Tutz, 2009). Tree pruning ensures the decision tree model is kept as simple as possible without significant loss of prediction accuracy (De'ath & Fabricius, 2000). Finally, the pruned decision tree model is tested either using a subset of the existing data or a whole new set of data (Loh, 2011). This process of tree building, tree pruning, and testing is repeated until the model with the least prediction error is found (Breiman et al., 1984; Loh, 2011; Strobl et al., 2009).

In this study the decision model was used to characterize the dependent variable “days away from work” (DAFW) based on predictor variables, age, tenure, nature of injury, cause of injury and body part injured. The decision model was developed using the SAS Enterprise Miner version 13.1, application software. The data were prepared for data mining

analysis using MS- Excel. The decision tree model was compared with a multiple linear regression model to identify the method that best predicts the DAFW.

Organization of Dissertation

This dissertation is written in the alternative manuscript format as defined by Iowa State University's Graduate College. Chapter one is the general introduction which outlines the basic ideas behind the research and summarizes the goals and objectives. Chapters two, three, and four, are three manuscripts formatted for submission to specified journals. Chapter five is a general summary and interpretation of findings, recommendations for further research, and conclusions.

References

- Abdolhamidzadeh, B., Abbasi, T., Rashtchian, D., & Abbasi, S. A. (2011). Domino effect in process-industry accidents - An inventory of past events and identification of some patterns. *Journal of Loss Prevention in the Process Industries*, 24(5), 575-593.
doi:10.1016/j.jlp.2010.06.013
- Agresti, A., & Finlay, B. (2008). *Statistical Methods for the Social Sciences* (4th ed.). NJ: Prentice Hall.
- Anand, S., Keren, N., Tretter, M. J., Wang, Y. J., O'Connor, T. M., & Mannan, M. S. (2006). Harnessing data mining to explore incident databases. *Journal of Hazardous Materials*, 130(1-2), 33-41. doi:10.1016/j.jhazmat.2005.07.046
- Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis and Prevention*, 41(3), 359-364.
doi:10.1016/j.aap.2008.12.014

BLS. (2016). Employer-Reported Workplace Injury and Illness Summary [Press release].

Retrieved from <http://www.bls.gov/news.release/osh.nr0.htm>

Boden, L. I., O'Leary, P. K., Applebaum, K. M., & Tripodis, Y. (2016). The impact of non-fatal workplace injuries and illnesses on mortality. *American journal of industrial medicine*, 59(12), 1061-1069.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199-215. doi:10.1214/ss/1009213726

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.

Calvo Olivares, R. D., Rivera, S. S., & Nunez Mc Leod, J. E. (2014). Database for accidents and incidents in the biodiesel industry. *Journal of Loss Prevention in the Process Industries*, 29, 245-261. doi:10.1016/j.jlp.2014.03.010

Calvo Olivares, R. D., Rivera, S. S., & Nunez Mc Leod, J. E. (2015). Database for accidents and incidents in the fuel ethanol industry. *Journal of Loss Prevention in the Process Industries*, 38, 276-297. doi:10.1016/j.jlp.2015.10.008

Chang, L.-Y., & Wang, H.-W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis and Prevention*, 38(5), 1019-1027. doi:10.1016/j.aap.2006.04.009

Chang, L. Y., & Chen, W. C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 36(4), 365-375.
doi:10.1016/j.jsr.2005.06.013

- Chen, J., Jiang, Q., Wang, Y., & Tang, J. (2016). *Study of data analysis model based on big data technology*. Paper presented at the Big Data Analysis (ICBDA), 2016 IEEE International Conference on.
- Cheng, C. W., Leu, S. S., Cheng, Y. M., Wu, T. C., & Lin, C. C. (2012). Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. *Accident Analysis and Prevention*, *48*, 214-222.
doi:10.1016/j.aap.2011.04.014
- Costich, J. (2010). Who Pays for Agricultural Injury Care? *Journal of Agromedicine*, *15*(1), 54-57. doi:10.1080/10599240903389714
- De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, *81*(11), 3178-3192.
doi:10.1890/0012-9658(2000)081[3178:cartap]2.0.co;2
- Dement, J. M., Pompeii, L. A., Ostbye, T., Epling, C., Lipscomb, H. J., James, T., . . . Thomann, W. (2004). An integrated comprehensive occupational surveillance system for health care workers. *American Journal of Industrial Medicine*, *45*(6), 528-538.
doi:10.1002/ajim.20017
- Dong, X., Largay, J. A., Wang, X., & Windau, J. A. (2014). Fatalities in the construction industry: findings from a revision of the BLS Occupational Injury and Illness Classification System. *Monthly Lab. Rev.*, *137*, 1.
- Douphrate, D. I., Rosecrance, J. C., Reynolds, S. J., Stallones, L., & Gilkey, D. P. (2009). Tractor-Related Injuries: An Analysis of Workers' Compensation Data. *Journal of Agromedicine*, *14*(2), 198-205. doi:10.1080/10599240902773215

- Douphrate, D. I., Rosecrance, J. C., Stallones, L., Reynolds, S. J., & Gilkey, D. P. (2009). Livestock-Handling Injuries in Agriculture: An Analysis of Colorado Workers' Compensation Data. *American Journal of Industrial Medicine*, 52(5), 391-407. doi:10.1002/ajim.20686
- Ehmke, T. (2016). Grain Elevators Braced for a Challenging 2016. Retrieved from <http://www.cobank.com/Knowledge-Exchange/Centers-of-Excellence/~media/Files/Searchable%20PDF%20Files/Knowledge%20Exchange/2016/Front%20Page%20%20Grain%20Elevators%20Report%20%20Apr%202016.pdf>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813. doi:10.1111/j.1365-2656.2008.01390.x
- Field, W. E., Heber, D. J., Riedel, S. M., Wettschurack, S. W., Roberts, M. J., & Grafft, L. J. (2014). Worker hazards associated with the use of grain vacuum systems. *Journal of agricultural safety and health*, 20(3), 147-163.
- Gavious, A., Mizrahi, S., Shani, Y., & Minchuk, Y. (2009). The costs of industrial accidents for the organization: Developing methods and tools for evaluation and cost-benefit analysis of investment in safety. *Journal of Loss Prevention in the Process Industries*, 22(4), 434-438. doi:10.1016/j.jlp.2009.02.008
- Gnanadesikan, R. (2011). *Methods for statistical data analysis of multivariate observations* (Vol. 321): John Wiley & Sons.
- Gorucu, S., Murphy, D. J., & Kassab, C. (2015). A multi-year analysis of fatal farm and agricultural injuries in Pennsylvania. *Journal of agricultural safety and health*, 21(4), 281-298.

- Gotcheva, N., Oedewald, P., Wahlstrom, M., Macchi, L., Osvalder, A. L., & Alm, H. (2016). Cultural features of design and shared learning for safety: A Nordic nuclear industry perspective. *Safety Science*, *81*, 90-98. doi:10.1016/j.ssci.2015.04.014
- Griend, N. V. (2011). Creating a Culture of Safety at an Ethanol Plant. *Ethanol Producer Magazine*. Retrieved from <http://www.ethanolproducer.com/articles/7490/creating-a-culture-of-safety-at-an-ethanol-plant>
- Hajakbari, M. S., & Minaei-Bidgoli, B. (2014). A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data. *Journal of Loss Prevention in the Process Industries*, *32*, 443-453. doi:10.1016/j.jlp.2014.10.013
- Harb, R., Yan, X. D., Radwan, E., & Su, X. G. (2009). Exploring precrash maneuvers using classification trees and random forests. *Accident Analysis and Prevention*, *41*(1), 98-107. doi:10.1016/j.aap.2008.09.009
- Janicak, C. (2010). *Safety Metrics: Tools and Techniques for Measuring Safety Performance*. Government Institutes. Lanham, MD.
- Javadi, A., & Rostami, M. A. (2007). Safety assessments of agricultural machinery in Iran. *Journal of agricultural safety and health*, *13*(3), 275-284.
- Karttunen, J. P., & Rautiainen, R. H. (2011). Risk factors and prevalence of declined work ability among dairy farmers. *Journal of agricultural safety and health*, *17*(3), 243-257.
- Keren, N. (2010). Incident Database-based Framework for Establishing Industrial Safety Performance Assessments. *Journal of Industrial Technology*, *26*(2), 1-12.

- Kim, H., Dropkin, J., Spaeth, K., Smith, F., & Moline, J. (2012). Patient handling and musculoskeletal disorders among hospital workers: Analysis of 7 years of institutional workers' compensation claims data. *American Journal of Industrial Medicine*, 55(8), 683-690. doi:10.1002/ajim.22006
- Lander, F., Nielsen, K. J., & Lauritsen, J. (2016). Work injury trends during the last three decades in the construction industry. *Safety Science*, 85, 60-66. doi:10.1016/j.ssci.2015.10.013
- Leigh, J. P. (2011). Economic Burden of Occupational Injury and Illness in the United States. *Milbank Quarterly*, 89(4), 728-772. doi:10.1111/j.1468-0009.2011.00648.x
- Leigh, J. P., & Marcin, J. P. (2012). Workers' Compensation Benefits and Shifting Costs for Occupational Injury and Illness. *Journal of Occupational and Environmental Medicine*, 54(4), 445-450. doi:10.1097/JOM.0b013e3182451e54
- Leigh, J. P., McCurdy, S. A., & Schenker, M. B. (2001). Costs of occupational injuries in agriculture. *Public Health Reports*, 116(3), 235-248.
- Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303-11311. doi:10.1016/j.eswa.2012.02.063
- Lim, T. S., Loh, W. Y., & Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40(3), 203-228. doi:10.1023/a:1007608224229
- Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, 1(1), 14-23. doi:10.1002/widm.8

- Mabila, S. L., Gracia, G., Cohen, R., Almberg, K., & Friedman, L. S. (2015). Injury and illness data for Illinois mining industry employees, 1990 to 2012. *Journal of Occupational and Environmental Medicine*, 57(12), 1305-1310.
- Manuele, F. A. (2013). *On the practice of safety*: John Wiley & Sons.
- Marucci-Wellman, H. R., Courtney, T. K., Corns, H. L., Sorock, G. S., Webster, B. S., Wasiak, R., . . . Leamon, T. B. (2015). The direct cost burden of 13 years of disabling workplace injuries in the US (1998-2010): Findings from the Liberty Mutual Workplace Safety Index. *Journal of Safety Research*, 55, 53-62.
doi:10.1016/j.jsr.2015.07.002
- Meel, A., O'Neill, L. M., Levin, J. H., Seider, W. D., Oktem, U., & Keren, N. (2007). Operational risk assessment of chemical industries by exploiting accident databases. *Journal of Loss Prevention in the Process Industries*, 20(2), 113-127.
doi:10.1016/j.jlp.2006.10.003
- Meleddu, M., & Pulina, M. (2016). Evaluation of individuals' intention to pay a premium price for ecotourism: An exploratory study. *Journal of Behavioral and Experimental Economics*, 65, 67-78. doi:10.1016/j.socec.2016.08.006
- Mosher, G. A. (2011). Measurement and analysis of the relationship between employee perceptions and safety and quality decision-making in the country grain elevator.
- Mosher, G. A., Keren, N., & Hurburgh Jr, C. R. (2013). Development of a Quality Decision-Making Scenario to Measure How Employees Handle Out-of-Condition Grain. *Applied Engineering in Agriculture*, 29(5), 807.
- Nenonen, N. (2013). Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and

- diseases statistics database. *Applied Ergonomics*, 44(2), 215-224.
doi:10.1016/j.apergo.2012.07.001
- Nestoriak, N., & Pierce, B. (2009). Comparing Workers' Compensation claims with establishments' responses to the SOII. *Monthly Labor Review*, 132(5), 57-64.
- Nisbet, R., Miner, G., & Elder Iv, J. (2009). *Handbook of statistical analysis and data mining applications*: Academic Press.
- Oleinick, A., & Zaidman, B. (2004). Methodologic issues in the use of workers' compensation databases for the study of work injuries with days away from work. I. Sensitivity of case ascertainment. *American Journal of Industrial Medicine*, 45(3), 260-274. doi:10.1002/ajim.10333
- OSHA. (2016). Safety Pays. Retrieved from
<https://www.osha.gov/Region7/fallprotection/safetypays.html>
- Pasman, H. J. (2009). Learning from the past and knowledge management: Are we making progress? *Journal of Loss Prevention in the Process Industries*, 22(6), 672-679.
doi:10.1016/j.jlp.2008.07.010
- Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, 29(1), 65-74. doi:10.1016/j.eswa.2005.01.006
- Reiner, A. M., Gerberich, S. G., Ryan, A. D., & Mandel, J. (2016). Large Machinery-Related Agricultural Injuries Across a Five-State Region in the Midwest. *Journal of Occupational and Environmental Medicine*, 58(2), 154-161.
doi:10.1097/jom.0000000000000584

- Reville, R. T., Bhattacharya, J., & Weinstein, L. R. S. (2001). New methods and data sources for measuring economic consequences of workplace injuries. *American Journal of Industrial Medicine*, 40(4), 452-463. doi:10.1002/ajim.1115
- Reville, R. T., Polich, S., Seabury, S. A., & Giddens, E. (2001). *Permanent disability at private, self-insured firms*: Rand Corporation.
- Riedel, S., & Field, W. (2011). *Summary of over 800 grain storage and handling-related entrapments and suffocations documented in the US between 1970 and 2010*. Paper presented at the Presentation at XXXIV CIOST-CIGR V Conference, Vienna, Austria.
- Sanmiquel, L., Rossell, J. M., & Vintro, C. (2015). Study of Spanish mining accidents using data mining techniques. *Safety Science*, 75, 49-55. doi:10.1016/j.ssci.2015.01.016
- Savolainen, P. T., Mannering, F., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention*, 43(5), 1666-1676. doi:10.1016/j.aap.2011.03.025
- Sengupta, I., Reno, V. P., Burton Jr, J. F., & Baldwin, M. L. (2012). Workers' compensation: Benefits, coverage, and costs, 2010. *National Academy of Social Insurance*.
- Sharpe, D. (2015). Your Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research & Evaluation*, 20.
- Simmons, H. E., Matos, B., & Simpson, S. A. (2016). Analysis of injury data to improve safety and training. *Journal of Chemical Health and Safety*.

- Spector, J. T., Bonauto, D. K., Sheppard, L., Busch-Isaksen, T., Calkins, M., Adams, D., . . . Fenske, R. A. (2016). A Case-Crossover Study of Heat Exposure and Injury Risk in Outdoor Agricultural Workers. *Plos One*, *11*(10). doi:10.1371/journal.pone.0164498
- Sprince, N. L., Zwerling, C., Lynch, C. F., Whitten, P. S., Thu, K., Logsdan–Sackett, N., . . . Alavanja, M. C. R. (2003). Risk factors for agricultural injury: A case–control analysis of Iowa farmers in the Agricultural Health Study. *Journal of agricultural safety and health*, *9*(1), 5.
- Springmeyer, R. R., Blattner, M. M., & Max, N. L. (1992). A CHARACTERIZATION OF THE SCIENTIFIC-DATA ANALYSIS PROCESS. *Visualization 92 : Proceedings*, 235-242.
- Strobl, C., Malley, J., & Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods*, *14*(4), 323-348. doi:10.1037/a0016973
- Utterback, D., Meyers, A., & Wurzelbacher, S. (2014). Workers' Compensation Insurance: A Primer for Public Health. *Department of Health and Human Services: National Institute for Occupational Safety and Health*.
- Utterback, D. F., & Schnorr, T. M. (2010). Use of workers' compensation data for occupational injury & illness prevention. *National Institute for occupational safety and health, US Department of Labour*, 2010-2152.
- Utterback, D. F., Schnorr, T. M., Silverstein, B. A., Spieler, E. A., Leamon, T. B., & Amick, B. C., III. (2012). Occupational Health and Safety Surveillance and Research Using

- Workers' Compensation Data. *Journal of Occupational and Environmental Medicine*, 54(2), 171-176. doi:10.1097/JOM.0b013e31823c14cb
- Wuellner, S. E., Adams, D. A., & Bonauto, D. K. (2016). Unreported workers' compensation claims to the BLS Survey of Occupational Injuries and Illnesses: Establishment factors. *American Journal of Industrial Medicine*, 59(4), 274-289. doi:10.1002/ajim.22563
- Wurzelbacher, S. J., Al-Tarawneh, I. S., Meyers, A. R., Bushnell, P. T., Lampl, M. P., Robins, D. C., . . . Raudabaugh, J. A. (2016). Development of methods for using workers' compensation data for surveillance and prevention of occupational injuries among State-insured private employers in Ohio. *American journal of industrial medicine*, 59(12), 1087-1104.

CHAPTER 2. USING WORKERS' COMPENSATION CLAIMS DATA TO CHARACTERIZE OCCUPATIONAL INJURIES IN THE BIOFUELS INDUSTRY

Manuscript submitted to the *Safety Science*

Sai K. Ramaswamy^a, Gretchen A. Mosher^a

^aDepartment of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA

Abstract

Biofuels production is a fast growing and emerging industry. Occupational injuries are a serious problem due to their human, financial and social costs, yet little research has been published on injuries in the biofuels industry. Learning from past injuries are essential for preventing future occurrences, but the lack of injury information hinders this effort in the biofuels industry. The present study addresses this knowledge gap by utilizing data from over 900 workers' compensation claims reported from 2008 to 2016 by ethanol and biodiesel facilities in the U.S. to characterize injury costs and severity. The total amount paid for each claim was used as a measure of injury severity, and the effects of age, tenure, type of claim, body part injured, nature and, cause of injury on the cost of work-related injuries were investigated. Contingency tables were used to classify the variable pairs, chi-square test and chi-square residuals were employed to evaluate the relation between the variable pairs and identify the at-risk groups. Results showed age and tenure of employee, type of claim, body part injured, nature of injury and, cause of injury have a significant influence in determining the claim amount. Age group 46-50, tenure group 1-2 years, strain and fractures injuries, slips, falls, or trips and, injuries to lower extremities were some of the at-risk groups identified. The findings from the study will assist biofuel producers to develop precisely targeted safety interventions that are effective in preventing worker injuries and also mitigating the financial and social losses from occupational injuries.

Keywords: *Occupational injury, Workers Compensation, Injury severity, Contingency Table*

Introduction

The production of biofuels such as ethanol and biodiesel is a fast-growing business witnessing constant changes and improvements to its production processes (Dias et al., 2012; Gubicza et al., 2016; Moreno & Cozzani, 2015; Priambodo et al., 2015; Scovronick & Wilkinson, 2014). The manufacturing of biofuels involves processing, handling, and storing of grains such as corn, sorghum, wheat, and oilseeds as well as hazardous chemicals such as ammonia and sulfuric acid. The combination of grain handling and chemical hazards present a dangerous work environment (OSHA, 2016a, 2016b). Handling and storing flammable liquids, working with heavy equipment, dealing with combustible dust and confined spaces, grain engulfment, working at heights, slips, falls, and trips are just some of the occupational safety hazards in biofuels production (OSHA, 2016b).

The presence of occupational safety hazards is a precursor for incidents and injuries in the workplace (Bevilacqua, Ciarapica, & Mazzuto, 2012; Khanzode, Maiti, & Ray, 2012; Vredenburg, 2002). Data gathered from trade journals, Occupational Safety and Health Administration (OSHA) records, Environmental Protection Agency (EPA) reports, academic, and newspaper articles suggest an increase in frequency of incidents and injuries, which in turn has resulted in higher levels of injuries and fatalities in biofuels producing facilities (Calvo Olivares, Rivera, & Nunez Mc Leod, 2014, 2015; Rivera, Olivares, Baziuk, & Mc Leod, 2015). Despite the increased risk of worker injuries, very little scientific work has explored health and safety in the biofuels industry (Harper, Etchells, Summerfield, & Cockton, 2008; Rivera et al., 2015; Riviere & Marlair, 2010).

Work-related injuries not only affect the injured worker and their family adversely but also impact the company in the form of increased medical, liability and insurance premium costs (Hajakbari & Minaei-Bidgoli, 2014). In addition to direct costs such as medical and indemnity payments, there are several indirect costs associated with workplace injuries. These indirect costs include equipment damage, equipment repair, accident investigation time, the cost of hiring and training an injured worker's replacement, loss of reputation, loss of employee morale, loss of confidence and negative media attention (Gavious, Mizrahi, Shani, & Minchuk, 2009; Griend, 2011; Manuele, 2013). According to Bird, Germain, and Veritas (1996), for every dollar in direct costs, there are \$5 to \$50 in property damage costs and \$1 to \$3 in other indirect costs associated with work-related injuries. Furthermore, Manuele (2013) suggested that the ratio of direct to indirect costs used by safety practitioners to estimate total injury costs is 1: 4.

The average direct cost estimate for a work-related injury in the biofuels industry is \$7,150 (Griend, 2011). Using the 1:4 ratio, the estimate for indirect costs per injury equals approximately \$28,600. Since biofuels production is a highly cost-sensitive business (Festel, 2008; Haarlemmer, Boissonnet, Peduzzi, & Setier, 2014), such high injury costs represent a threat to the profitability of a biofuels operation. Hence, an improved understanding of injuries and fatalities in the biofuels industry is necessary to prevent work-related injury risks before they occur.

Learning from past safety events is a critical component of improving worker safety and preventing work-related injuries (Kletz, 2008; Pasma, 2009). Examining injuries and identifying associated causes provides valuable information that can help prevent recurrence of similar injuries (Ferjencik & Jalovy, 2010). Analysis of incident and injury data can also

help identify at-risk groups (Anderson, 2009; Pirdavani, Brijs, & Wets, 2010), so targeted injury prevention strategies can be developed, thus improving the return on safety investments (Abdolhamidzadeh, Abbasi, Rashtchian, & Abbasi, 2011; Khanzode, Maiti, & Ray, 2011; B. K. Kim et al., 2012). In one example, Chettouh, Hamzi, and Benaroua (2016) examined incidents in an oil refinery and uncovered evidence that employees lacked safety awareness. One result of the research was suggested improvements to the hiring process and increased investments in safety education and professional competency. Likewise, Marhavidas, Koulouriotis, and Mitrakas (2011) analyzed occupational injury data for an electric power provider and found that workers under the age of 45 years had the greatest risk of fractures, bruises and sprains injuries, caused due to slips, falls, and impacts with stationary objects.

Obtaining detailed historical records of safety events for data analysis is a challenge in the process industry (Meel et al., 2007; Pasman, 2009). In the U.S., organizations such as Occupational Safety and Health Administration, U.S Environmental Protection Agency, National Fire Protection Association, and the National Response Center track and collect data on industrial incidents (Keren, 2010). However, these organizations differ in their interests, procedures, and scope of data collection, and it is difficult to use their data for studying past incidents in a specific industry (Morrison, Fecke, & Martens, 2011; Tauseef, Abbasi, & Abbasi, 2011). While some investigation has resulted in the development of an incident database for the biofuels industry (Calvo Olivares et al., 2014, 2015), this database does not contain detailed historical records of work-related injuries.

The majority of employers in the U.S, including those in the biofuels industry, purchase workers' compensation insurance to provide medical and indemnity benefits to an

employee who suffers a work-related injury (Sengupta, Reno, Burton Jr, & Baldwin, 2012). For an employer, workers' compensation insurance covers direct costs of a work-related injury, including medical expenses and wage replacement incurred by the injured employee (Bird et al., 1996; Griend, 2011; Manuele, 2013). Workers' compensation data contains information that can contribute to injury prevention activities (Utterback et al., 2012). Several researchers have used workers' compensation claims data to study occupational injuries in various industries (Coleman & Kerkering, 2007; Frank Neuhauser, Mathur, & Pines, 2013; Sears, Blonar, Bowman, Adams, & Silverstein, 2013; Smith, Hogg-Johnson, Mustard, Chen, & Tompa, 2012). To date, little research has explored the application of workers' compensation claims data to characterize occupational injuries in the biofuels industry.

This study examined occupational injuries using workers' compensation claims data provided by a leading Midwest-based insurance company from biofuel production facilities. The purpose of this study was to characterize the direct cost of occupational injury using the information obtained from the workers' compensation claims including variables such as age, tenure of employee, and nature, cause and type of injury. A secondary purpose of the study was to identify and classify at-risk groups within the biofuels production industry.

Background

For the last ten years, the biofuels industry in the United States has been one of the fastest growing areas of the agribusinesses sector (Olivares, Rivera, Baziuk, & Mc Leod, 2014; OSHA, 2016a). Between 2006 and 2012, biofuel production in the U.S increased more than three-fold, making the U.S the number one producer of biofuel products in the world (Energy Information Administration, 2016). The rapid growth in biofuels production has

been accompanied by an increasing number of occupational injuries in the industry (Moreno & Cozzani, 2015; Olivares et al., 2014; Rivera, Olivares, Baziuk, & McLeod, 2015).

A typical large-scale commercial biofuels facility utilizes approximately 50 tons of raw biological material daily (Vimmerstedt, Bush, & Peterson, 2013). Most commercial biofuels facilities store large quantities of grain along with chemical additives necessary for biofuel production including sulfuric acid, sodium hydroxide, methanol, and glycerol within their facility (Hardy, Holz-Clause, Shepherd, & Hurburgh, 2006; Marchetti, Miguel, & Errazu, 2008; Vlysidis, Binns, Webb, & Theodoropoulos, 2011). From an occupational safety perspective, biofuel production combines the hazards of both grain handling and chemical processing facilities, which increases the importance of safety interventions for the industry (OSHA, 2016a).

The production of ethanol and biodiesel involves a substantial amount of routine processes (Nigam & Singh, 2011). The routine nature of the process encourages a tendency to ignore safety aspects (Rivera et al., 2015), despite the known documented hazards of grain handling and chemical facilities (Freeman, Kelley, Maier, & Field, 1998; Niskanen, 2012; Reniers, 2009; Riedel & Field, 2011; Roberts & Field, 2010). As the biofuels industry continues to evolve with new and improved production technologies, workers in this industry will likely be exposed to unknown and known risks including explosions, fire, electrical, confined spaces, contact with chemicals, and slips, trips and falls (Spellman, 2013).

The existence of occupational safety hazards amplifies the risk of worker injuries (Bevilacqua et al., 2012; Khanzode et al., 2012). Workplace safety incidents are classified as major or simple (Jorgensen, 2011, 2015, 2016). Major safety events result in a high number of injuries and/or fatalities and cause widespread damage to property and the environment

(Jorgensen, 2016). Major safety events such as an explosion, fire, or containment of chemicals are rare events involving complex event sequences, and they make up a small percentage of the total number of workplace incidents (Jorgensen, 2015, 2016). Smaller safety events impact the immediate occupational area and result in the injury or fatality to a single employee (Jorgensen, 2016). Smaller safety events such as slips, falls, trips, contusions, and lacerations, occur more frequently and result in a higher rate of worker injuries and fatalities than do major safety incidents (Jorgensen, 2011, 2015, 2016). According to the International Association of Oil and Gas Producers, more than half of all oil and gas incidents are small incidents such as slips, trips and falls (Attwood, Khan, & Veitch, 2006). These small workplace safety incidents are seldom investigated (Jorgensen, 2011) because they are perceived to be workers' fault rather than engineering or environment issues (Vredenburgh, 2002).

The lack of a unified source of safety incident data in the biofuels industry has been a major challenge to identifying probable contributing factors of incidents that can help develop appropriate intervention strategies (Calvo Olivares et al., 2014, 2015; Mulloy et al., 2013; Sumner & Layde, 2009). Furthermore, the report of incidents and worker injuries occurring in the biofuels industry are difficult to obtain as they are spread across multiple sources including OSHA investigation summaries, media reports, and trade association publications (Calvo Olivares et al., 2014, 2015; OSHA, 2016a). Obtaining information on simple safety incidents is still a major challenge in the agribusiness industries including biofuels, as there is no single source or entity collecting data on these incidents (Doughrate, Rosecrance, & Wahl, 2006; Issa et al., 2016; Riedel & Field, 2011).

Workers' compensation insurance claims records can partially address the informational gap about occupational injuries in the biofuels industry, enhancing the ability to develop effective safety intervention (Utterback, Meyers, & Wurzelbacher, 2014).

Workers' compensation insurance provides an injured worker medical benefits, a portion of the employee's wage, and a lump sum payment when the employee suffers a permanent impairment (Sengupta et al., 2012). Most employers in all states in the U.S. except Texas are required to provide their employees with workers' compensation insurance. U.S. employers in all industries spend approximately \$85 billion each year on workers' compensation insurance costs for their employees (Sengupta et al., 2012; Utterback & Schnorr, 2010). In addition to providing benefits to the injured worker, workers' compensation insurance also protects employers from lawsuits and monetary losses resulting from occupational injuries.

Workers' compensation data contains valuable information commonly used in injury characterization (Utterback et al., 2012). In addition to information on the direct costs of the injury such as medical, indemnity, and disability payments, data on the industry, occupation, nature of injury, cause of injury and demographic information of the injured worker are also captured in workers' compensation claims data (Nestoriak & Pierce, 2009; Utterback et al., 2012). The size and volume of workers' compensation datasets provide a comprehensive understanding of injury patterns, which can then be used to analyze causal factors leading to an injury (Oleinick & Zaidman, 2004).

Previous research has shown that workers' compensation claims data can be used to characterize the risk, scope, and nature of workplace injuries across multiple industries. Frank Neuhauser, Mathur, and Pines (2013) utilized workers' compensation data to compare injury incidence by gender and age while controlling for the occupation and type of industry

of the injured worker. Sears, Blonar, Bowman, Adams, and Silverstein (2013) used workers' compensation data from the Washington State Department of Labor and Industries to predict occupational disability and medical cost outcomes. Smith, Hogg-Johnson, Mustard, Chen, and Tompa (2012) compared risk factors associated with severe versus less severe occupational injuries using workers' compensation data in industries such as agriculture, mining, and manufacturing. Coleman and Kerkering (2007) studied occupational injuries in coal mines and used workers' compensation data to distinguish between lower and higher risk operations and time periods. Schwatka, Butler, and Rosecrance (2013) studied the relationship between age and injury type on claim amount in the construction industry using workers' compensation claims from 1998 to 2008. Finally, Bookman (2012), Douphrate, Rosecrance, and Wahl (2006) studied workplaces injuries in production agriculture with a focus on injury prevention.

Despite the benefits of using workers' compensation claims data to characterize and predict injuries, very little research has effectively utilized the information source to characterize work-related injuries in the biofuels industry. Accordingly, the purpose of this study was to examine workers' compensation claims data to investigate how variables such as employee age, tenure, nature and cause of injury and type of claim influence the claim amount. Understanding these relationships helps in the development of focused mitigation strategies to prevent worker injuries. Also, the findings of this study may be used to model the direct costs associated with occupational injuries so biofuels producers and workers' compensation insurance providers can better understand the risks and losses from occupational injuries.

Methods and Data

Nearly all workers in the U.S. are covered by workers' compensation insurance provided by their employer (Utterback et al., 2014). Employers provide this benefit to their employees by either purchasing insurance from an insurance carrier or through self-insurance (Reville, Polich, Seabury, & Giddens, 2001). When an employee is injured on the job, the insurance carrier or the self-insured employer pays the medical and indemnity costs. To provide information and to facilitate the payment, employers must create a report of the worker's injury to inform their insurance provider (Utterback et al., 2012). Data collected during the claims process are provided by employees, employers, insurance companies and other involved parties (Utterback et al., 2014). The collection of information from multiple stakeholders makes claims records an excellent data source for work-related injuries (Dement et al., 2004; Janicak, 2010; Kim, Dropkin, Spaeth, Smith, & Moline, 2012; Reville, Bhattacharya, & Weinstein, 2001).

The dataset used in this study were obtained from a private insurance company headquartered in a Midwest state. The company specializes in insurance products for agribusinesses including biofuels facilities. The dataset consisted of 921 claims reported from 2008 to 2016. Of the 921 claims, 145 claims were from biodiesel producing facilities, while the remaining 776 were from ethanol producing facilities. The oldest claim in the dataset had an injury date of January 2nd 2008, and the newest claim injury date was February 24th 2016.

Table 1: *List of variables*

#	Column Name	Description
1	Claim	Unique identifier for each claim record
2	Effective Year	Filing year of the claim
3	Account	Unique identifier to differentiate claims for each customer
4	Market	Type of business (biodiesel or ethanol)
5	Gender	Gender of injured worker
6	State	Name of state where injury occurred
7	Date of Birth	Date of birth of injured worker
8	Date of Hire	Date on which the present company hired the injured worker
9	Injury Date	Date on which the injury occurred
10	Claim Description	One-line narration of incident resulting in injury; Example: "Employee was cleaning equipment and employee opened up a line and acid sprayed in his face and mouth."
11	Claim Status	If the claim is still open or closed
12	Type of claim	Indicates if the claim was "medical only", "permanent disability", "death".
13	Body Part	Body part(s) injured
14	Cause of Injury	Main cause of injury. For example: "Dust, gasses or fumes inhalation", "Foreign matter in eyes", "Chemical exposure" etc.
15	Nature of Injury	Describes the type of injury such as Fracture, Strain, Contusion etc.
16	Claim amount	Total amount paid out in medical, indemnity and other miscellaneous payments. Used as a proxy for injury severity in this study.

The list of variables obtained from the dataset used in this research are shown in Table 1. All information recorded in the dataset were vetted and verified by insurance company personnel. Using the demographic information provided in the dataset, age of the employee was calculated as the difference between the date of birth and the injury date. Similarly, the tenure of the employee was calculated as the difference between the date of hire and injury date. The claim amount variable used as a proxy for injury severity was categorized as “<\$3,000”, “\$3,000-\$9,999”, and “10,000+”.

Since the variables analyzed in this study were all categorical, the statistical analysis began with the construction of frequency counts, percentages, and contingency tables. To characterize the claim amount using the employee and injury variables, contingency tables were used to classify the variable pairs and chi-square tests were used to validate the relationship. For the posthoc tests to identify the at-risk groups, residual analysis was used to determine the nature of relationship between the row variable and the column variable of the contingency tables (Agresti & Finlay, 2008). The residual is the difference between the observed value of a specific variable pair and its expected value (Agresti & Finlay, 2008). A positive residual implies that the observed value was greater than the expected value, while a negative residual implies the observed value was less than expected value. Descriptive analysis was performed using Microsoft Excel and the inferential analyses were performed with statistical software SAS version 9.4.

In SAS, residuals are standardised and calculated as:

$$\text{Standardized residual} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}(1-p_{i.})((1-p_{.j}))}}$$

Where n_{ij} is the observed value, e_{ij} is the expected value for the i^{th} row j^{th} column cell. $p_{i.}$ is row total for i^{th} row and $p_{.j}$ is the column total for the j^{th} column. According to Agresti and Finlay (2008), an adjusted residual of +/- 2 is evidence of dependence between the row and column variables while an adjusted residual of +/- 3 is evidence of strong dependence. Comparing the adjusted residuals of each cell in a contingency table helped identify the cells where the degree of dependence between the two variables was the strongest (Sharpe, 2015).

Since, the purpose of this study was to characterize the direct cost of occupational injury using the demographic (age and tenure) and injury characteristics (nature, type, and

cause of injury, body part injured). The claim amount was used as a proxy for direct injury cost, and the broad research question that guided this study was: Is the claim amount in selected biofuels operations dependent on the employee demographics and injury characteristics? Specifically, the following sub research questions listed below were analyzed:

Is the claim amount independent of:

- i. Age of employee
- ii. Tenure of employee
- iii. Type of claim
- iv. Nature and Cause of injury
- v. Body part injured

Results and Discussion

Characterizing claim amount based on employee age

The first research question investigated if the claim amount and the age of the injured employee were independent. The claim amount is the sum of all payments made by the workers' compensation insurance provider to the injured employee. This amount includes medical, indemnity, and other miscellaneous payments made to the injured employee as compensation for their work related injury. For this reason, severe injuries, such as those resulting in disability or death have a higher claim amount than less severe injuries requiring less medical treatment (Sears, Blonar, & Bowman, 2014; Sears et al., 2013).

Table 2: Age of employee and claim amount

Age (years)	Claim amount			Total
	<\$3000	\$3000- \$9999	\$10,000 & above	
<25	71 (2.7*)	3 (-1.1)	2 (-2.4*)	76
26-30	94 (2.6*)	4 (-1.3)	5 (-2.1*)	103
31-35	99 (1.8)	4 (-1.5)	9 (-1.0)	112
36-40	89 (-1.2)	11 (1.2)	14 (0.5)	114
41-45	107 (0.1)	11 (0.7)	12 (-0.6)	130
46-50	92 (-3.2**)	12 (1.1)	24 (3.1**)	128
51-55	100 (-0.6)	8 (-0.3)	17 (1.1)	125
56-60	68 (-1.9)	11 (2.0*)	12 (0.7)	91
60+	36 (0.6)	1 (-1.2)	5 (0.2)	42
Total	756	65	100	921

χ^2 35.09; df= 16; p-value= 0.0039 and α = 0.05; N=921

Residuals in brackets; * evidence of relationship; ** evidence of strong relationship

To explore the relationship between employee age and the claim amount, a contingency table with age as the row variable and claim amount as the column variable was tabulated as shown in Table 2. Each cell in the contingency table is a count of claims corresponding to the respective age group and claim amount category. The last row in Table 2 shows the total number of claims in each claim amount category, while the last column indicates the number of claims corresponding to each age group.

The distribution of the number of claims based on the claim amount, as shown in the last row of Table 2 indicates that nearly 90% of the claims were less than \$10,000. This finding implies that the majority of work-related injuries in the dataset used in this study

were minor injuries, requiring less medical attention. In other words, only one out of the ten injuries recorded, resulted in a claim amount higher than \$10,000. The distribution of the number of claims, based on age group as shown in the last column of Table 2, shows 77% of the claims involved an employee between 25 and 55 years of age. According to the Bureau of Labor Statistics, 80% of the total workforce in the U.S are in the age group of 25-55 years (BLS, 2016) . Hence, this study's findings that majority of the work related injuries in the biofuel facilities were in the 25-55-year group is consistent with the overall distribution of workforce in the United States.

A chi-square test was conducted to evaluate if the claim amount was independent of the age of the employee. The test results showed a p-value of less than 0.05 providing evidence that the claim amount and the age of the employee were not independent. This finding implies that the age of the employee is a significant factor that can be used to determine the claim amount. Since injury severity in this study was measured using the claim amount, the findings of this study are consistent with previous research that suggested the link between the age of the employee and severity of work-related injuries (Laflamme, 1996; Rogers & Wiatrowski, 2005; Salminen, 2004; Takahashi & Miura, 2016). According to Rogers and Wiatrowski (2005) and Salminen (2004), young workers below the age of 25 years have a higher risk of injuries than older workers. However, the injuries to older workers are likely to be more severe when compared with younger workers. More severe injuries require increased medical attention and could also result in lost workdays, resulting in a higher claim amount.

To identify at-risk age groups, the residual values from the chi-square test shown in brackets in Table 2 were examined. The residual values indicate the strongest relationship

between age and claim amount are in the 46-50 age group category. Also, evidence of a relationship between age and claim amount was found in the 26-30, less than 25, and 56-60 age categories. It is noteworthy that in the 46-50 age group category, the residual value indicates a greater than expected number of claims in the \$10,000 and above category and fewer than expected number of claims in the below \$3,000 category. This implies, that employees in the 46-50 age category are likely to have more severe injuries, which requires higher levels of workers compensation payments. The residual values for the 26-30 and below 25 age categories, however, indicate exactly opposite to that of the 46-50 age group category. In both the 26-30 and below 25 age categories, the number of claims in the below \$3,000 category are more than expected value while the number of claims in the \$10,000 and above category are fewer than the expected value.

Also noteworthy is the change in the sign of the residuals as the age of employee increases. For the below \$3,000 category, the sign of residual changes from positive to negative as the age of employee increases except for the age groups 41-45 and 60 and above. Similarly, in the above \$10,000 category, the sign of residual changes from negative to positive as the age of the employee increases except for the age group 41-45. This change in the sign of residuals implies, that the number of claims for minor injuries decreases as the age of the employee increases while the number of claims for severe injuries increases with the increase in employee age. This finding further corroborates the results of previous studies (Laflamme, 1996; Rogers & Wiatrowski, 2005; Salminen, 2004; Takahashi & Miura, 2016) who suggested that older workers are likely to have more severe injuries than younger workers.

*Characterizing claim amount based on employee tenure***Table 3: Tenure of employee and claim amount**

Tenure (years)	Claim amount			Total
	<\$3,000	\$3000- \$9999	\$10,000+	
<1	104 (0.0)	11 (0.7)	12 (-0.6)	127
1-2	267 (2.8*)	14 (-2.1*)	26 (-1.7)	307
3-5	185 (-0.1)	16 (0.0)	25 (0.1)	226
6-10	116 (-1.8)	17 (2.2*)	18 (0.4)	151
11-20	51 (0.1)	3 (-0.7)	8 (0.5)	62
20+	28 (-2.9*)	4 (0.6)	11 (3.2**)	43
Total	751	65	100	916

$\chi^2 = 21.34$; $df = 10$; $p\text{-value} = 0.0189$ and $\alpha = 0.05$; $N = 916$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

The second research question investigated if the tenure of the injured employee and the claim amount were independent of each other. The contingency table used to investigate this research question is shown in Table 3. Only 916 records out of the 921 available records were analyzed because the remaining records did not contain the tenure information. The distribution of the number of claims by employee tenure indicates that nearly half of all injury-causing safety incidents involved an employee with two or fewer years of work experience.

Similarly, close to three-quarters of all injury-causing safety incidents involved an employee with five or fewer years of experience. This finding is different from the results of Chen, Yao and Wu (2013) who studied employees in the petrochemical production industry. Chen, Yao and Wu (2013) analyzed occupational injuries recorded in the Council of Labor

Affairs in Taiwan, and found that the majority of occupational incident victims were employees with more than three years of work experience. According to Vinodhkumar and Bhasi (2009), the length of service of an employee in a company influences their skills and attitudes towards safety. Therefore, this difference could be attributed to employees in biofuels production taking less time to learn about occupational hazards as compared to the employees in the petrochemicals industry. Since, the production of biofuels involves a routine process (Nigam & Singh, 2011), employees in this industry could have a shorter learning cycle when compared to employees in other petrochemical industries. It is worth noting that the study of Chen, Yao and Wu (2013) analyzed only those safety incidents that resulted in injuries to three or more employees while this study analyzed safety incidents that resulted in injuries to one or more employees.

The chi-square test of independence between the tenure of the injured employee and the claim amount resulted in a p-value of less than the 0.05 significance level. This result implies, the tenure of the injured employee is also a significant factor that can be used to determine the claim amount. This result of a significant relationship between the tenure of the injured employee and claim are expected. This expectation was because previous studies in the petrochemical industries (Cheng et al., 2013; Nouri, Azadeh, & Fam, 2008) as well as non-petrochemical industries such as construction (Lopez Arquillos, Rubio Romero, & Gibb, 2012; Suarez-Cebador, Carlos Rubio-Romero, & Lopez-Arquillos, 2014) have suggested a relationship between tenure and injury severity. Studies by Lopez Arquillos, Rubio Romero, and Gibb (2012) and Suarez-Cebador, Carlos Rubio-Romero, & Lopez-Arquillos (2014) found that employees with three or less years of work experience have the highest number of

occupational injuries, while Nouri, Azadeh, and Fam (2008) suggested that more experienced employees tend to have higher number of incidents.

The residual values from the chi-square test used to investigate the at-risk groups showed that employees in the 1-2 years and 20 years and above tenure category show evidence of a significant relationship with the claim amount. In addition, one cell in the 6-10 years tenure category also demonstrated evidence of a significant relationship with the claim amount. This result implies that both newer employees and highly experienced employees in the biofuel facilities investigated in this study are the most at-risk groups for occupational injury. This finding aligns with the findings of Khanzode et al. (2012). According to Khanzode et al. (2012), the risk of injury is high during the initial years on the job, then decreases when an employee acquires sufficient work experience but increases again for the highly experienced employee.

Observing the sign of the residuals shows that in the below \$3,000 category the sign changes from positive to negative as the tenure increases, but in the \$10,000 and above category the sign changes from negative to positive as tenure increases. This finding implies that with the increase in employee tenure, there are greater numbers of claims in the severe injury category than in the minor injury category. This shift in injury severity could be attributed to the age of the employee since tenure and age are highly correlated (Vinodkumar & Bhasi, 2009). As tenure of the employee increase so does their age, reducing their ability to tolerate impact from safety incidents (Brorsson, 1989) thereby increasing the likelihood of severe injuries.

Characterizing claim amount based on the type of claim

Table 4: Type of claim and claim amount

Type of claim	Claim amount			Total
	<\$3,000	\$3,000- \$9,999	\$10,000+	
Medical only	747 (23.6**)	38 (-6.8**)	10 (-23.5**)	795
Temporary disability	3 (-8.5**)	11 (8.0**)	8 (3.9**)	22
Death or Permanent disability	6 (-21.5**)	16 (3.5**)	82 (23.7**)	103
Total	756	65	100	921

$\chi^2 = 696.97$; $df = 4$; $p\text{-value} = <0.001$ and $\alpha = 0.05$; $N=921$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

The third research question investigated the relationship between the claim amount and the type of claim. The variable type of claim, had three categories, “medical only”, “temporary disability”, and “death or permanent disability”. “Medical-only” relates to occupational safety incidents where the employee’s injury required medical treatment only and did not result in any temporary or permanent incapacitation. This implies the claims in the medical-only category had a low injury severity and required minimal medical attention. The next claim category – temporary disability – is defined as injuries where employees are not able to work fully or to their expected capacity. In these cases, injured employees are paid two-thirds of their monthly salary in addition to the medical payments (Manning, 2012). The final claim category – death or permanent disability – has the highest injury severity. A claim is classified in the death or permanent disability category when the injured employee is expected to have either long-term physical challenges or permanent loss of functionality with some part of the body (Corso, Finkelstein, Miller, Fiebelkorn, & Zaloshnja, 2006). The

contingency table used to classify the claim amount using the type of claim is shown in Table 4.

The distribution of claims based on the type of claim showed that 86% of the claims were medical only, 2% of the claims involved temporary disability, and 11% of the claims involved the death or disability of an employee. Out of the 103 claims in the death and permanent disability category, only one claim involved the death of the employee. The finding implies, that the majority of injuries in the biofuels facilities investigated in this research were minor injuries requiring only medical attention and no lost workdays. The chi-square test to investigate the relationship between the type of claim and claim amount had a p-value of less than 0.05 significance level, implying that type of claim is a significant factor that can be used to determine the claim amount. The test results for this research question shows a large chi-square statistics value; this implies a strong relationship between the type of claim and the claim amount.

This evidence of a strong relationship between the two variables explains why both claim amount and type of claim can be used as a proxy for injury severity as suggested by previous research (Beery et al., 2014). Medical-only type of claims are the least severe and have the lowest claim amount when compared with the other two types of claim: death or permanent disability, and temporary disability. The residuals from the chi-square test shown in brackets in Table 4 further corroborate the findings of Beery et al.(2014). Furthermore, with the change in the type of injury from medical only category in Table 4 to the death or disability category, the number of claims in the less than \$3,000 category decreases significantly while the number of claims in the above \$10,000 category increases significantly. The finding implies that minor injuries in the biofuels facilities are most likely

to have a claim amount of less than \$3,000 , while severe injuries are more likely to have a claim amount more than \$10,000. It is noteworthy that for each cell in Table 4 the value of residual is greater than ± 3 . The large residual values provide further evidence indicating a strong statistical relationship between the type of injury and claim amount.

Characterizing claim amount based on the nature and cause of injury

The fourth research question investigated if the claim amount was independent of the nature and cause of injury. The contingency tables used to classify the claim amount by nature of injury are shown in Tables 5.

Table 5: Nature of injury and claim amount

Nature of injury	Claim amount			Total
	<\$3,000	\$3,000- \$9,999	\$10,000+	
Strain & sprain	262 (-2.9*)	28 (1.1)	49 (2.7*)	339
Others	110 (-1.0)	11 (0.4)	18 (0.9)	139
Laceration	115 (3.1*)	6 (-1.1)	4 (1.5)	125
Burn	101 (1.3)	6 (1.7)	12 (-3.0**)	119
Contusion	105 (1.9)	7 (-1.0)	7 (-1.7)	119
Foreign body	39 (2.6*)	1 (-1.2)	0 (-2.3*)	40
Fracture	24 (-3.7**)	6 (2.0)	10 (2.9*)	40
Total	756	65	100	921

$\chi^2 = 40.04$; $df = 12$; $p\text{-value} = <0.001$ and $\alpha = 0.05$; $N=921$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

The distribution of claims in Table 5 revealed that strain and sprain injuries were the most common type of injuries followed by “others” and lacerations. Fractures and foreign body were the least common types of injuries. Nearly 37% of claims were for strain and

sprain injuries while only 4% of the claims were for fractures or foreign body. The Chi-square test suggested that the claim amount was not independent of the nature of the injury. This finding means the nature of the injury is a significant factor that can be used to determine the claim amount.

Examining the chi-square residuals indicates the nature of injury categories showing an evidence of a strong relationship with claim amount were: strain and sprain, laceration, burn, foreign body, and fracture. However, the sign of residuals indicates that, in the case of strain and sprain and fractures injuries, the number of claims in the \$3,000 or below category was less than the expected value, but the number of claims in the \$10,000 and above category was more than the expected value. This finding suggests that strain and sprain and fracture injuries are likely to be more expensive when compared to all other types of injuries. Likewise, lacerations and foreign body injuries tend to have lower costs than all other types of injuries and are likely to be less expensive.

A review of the literature suggests strain and sprain injuries are the most common type of injury across many industries (Nur, Dawal, & Dahari, 2014; Schwatka et al., 2013; van Tulder, Malmivaara, & Koes, 2007). In the U.S., strain injuries alone cost \$6.5 billion in workers' compensation costs with the average claim ranging from \$5000 to \$8000 (Baldwin & Butler, 2006; van Tulder et al., 2007). However, studies by Attwood, Khan, and Veitch (2006) and Cheng, Yao, and Wu (2013) investigating occupational injuries in the oil and gas and petrochemical industries did not list strain injuries in their list of injury types. The fact that strain and sprain injuries were the most common type of injuries in this study of biofuels facilities suggests that there may be tasks unique to the biofuels industry that may be contributing to such higher levels of strain injuries as compared to other petrochemical

industries. Fracture, burn, foreign body, laceration and contusion type of injuries have all been listed as types of injuries in petrochemical industries by prior studies, yet these type of injuries are not as prevalent in biofuels production (Bertolini, Bevilacqua, Ciarapica, & Giacchetta, 2009; Owens & Hazeldean, 1995; Wu, 2004).

Consistent with the nature of injury, the classification of claim amount based on the cause of injury as shown in Table 6 indicated that “strain or injured by” is the most common cause of injury followed by slips, falls, and trips. Nearly 47% of claims involved a strain injury or a slip, trip, or fall injury. The chi-square test showed the claim amount was not independent of the cause of the injury thus suggesting the cause of injury is also significant factor useful in determining the claim amount.

Table 6: Cause of injury and claim amount

Cause of injury	Claim amount			Total
	<\$3,000	\$3,000-\$9,999	\$10,000+	
Strain or injured by	197 (-1.9)	19 (0.4)	36 (2.1*)	252
Slip, fall or trip	131 (-3.6**)	15 (0.7)	34 (3.9**)	180
Others	130 (2.8*)	7 (-1.1)	7 (-2.5)	144
Heat or cold Exposures	115 (1.6)	5 (-1.6)	12 (-0.7)	132
Struck or injured by	89 (0.8)	11 (1.4)	5 (-2.1*)	105
Cut, puncture, scrape	77 (2.1*)	4 (-0.9)	4 (-1.9)	85
Caught in, or under, or between	17 (-1.0)	4 (1.9)	2 (-0.3)	23
Total	756	65	100	921

$\chi^2 = 39.6$; $df = 12$; $p\text{-value} = <0.001$ and $\alpha = 0.05$; $N=921$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

Residuals from the chi-square test indicate slips, trips, and falls have the strongest relationship with claim amount. Evidence of a significant relationship with claim amount was also observed in one cell corresponding to the strain or injured by, others, struck or injured by and cut puncture scrape categories. However, the sign of residuals suggests that for categories strain or injured by and slips falls or trips, the number of claims in the \$10,000 and above category was more than the expected value while the number of claims in the \$3,000 or below category was less than the expected value. This finding suggests that injuries due to slips, trips, and falls or injuries involving a strain or injured by are likely to be more expensive than the other injuries. For the remaining categories including others, struck or injured by and cut puncture scrape, the residual values indicate the claim amount are more likely to be in the below \$3,000 category. Based on these analyses, targeting slips, trips, and falls and strain related injuries through safety interventions could potentially reduce the claim amount significantly.

Characterizing claim amount based on injured body part

The final research question investigated if the claim amount was independent of the injured body part. A contingency table constructed to address this research question is shown in Table 7.

The distribution of data suggests that upper extremities such as hands and fingers were the most frequently injured body part, followed by the trunk and lower extremities such as toes and feet. In eight out of 10 claims filed by the biofuel facilities, the injured body part was either an upper extremity, trunk or the lower extremity. This finding is consistent with the results of Marhavidas, Koulouriotis, and Mitrakas (2011) who also reported that the most injured body parts were arms, legs, head and neck of the employee . The chi-square test

results indicated the claim amount was not independent of the body part injured. This finding suggests the body part injured is a significant factor that can be used to determine the claim amount.

Table 7: Body part injured and claim amount

Body part	Claim amount			Total
	<\$3,000	\$3,000-\$9,999	\$10,000+	
Upper extremities	273 (1.0)	19 (-1.1)	34 (-0.3)	326
Trunk	183 (0.1)	21 (1.6)	18 (-1.5)	222
Lower extremities	127 (-3.3**)	14 (0.6)	32 (3.6**)	173
Head & neck	149 (3.6**)	8 (-1.2)	5 (-3.5**)	162
Multiple	24 (-3.1**)	3 (0.2)	11 (3.7**)	38
Total	756	65	100	921

$\chi^2 = 40.5$; $df = 8$; $p\text{-value} = <0.001$ and $\alpha = 0.05$; $N=921$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

Examining the chi-square residuals indicates that lower extremities, head and neck, and multiple body part injuries show evidence of strongest relationship with the claim amount. For both lower extremities and multiple body part injuries, the residuals indicate that the number of claims in the below \$3,000 category was less than the expected value while the number of claims in the greater than \$10,000 category was more than the expected value. This finding suggests that injuries to lower extremities and multiple body parts tend to be more expensive compared to the rest of the categories. However, for the head and neck injuries, the residuals indicate these injuries are more likely to be inexpensive compared with other injuries.

With further investigation of the claim records relating to injuries on lower extremities and multiple body parts showed, that slips, trips, and falls were the leading cause of these injuries. Heat or cold exposures were found to have the next highest number of claims corresponding to multiple and head and neck injuries. Reviewing the claims description showed that grain dust, hot and cold liquid entering the eye were most frequent description mentioned for the head and neck and multiple body part injuries. These findings suggest that implementing safety interventions to prevent slips, trips, and falls and providing improved protective equipment to prevent hot and cold exposure to employees' eyes and face has potential to significantly help reduce the claim amount.

Conclusion

Occupational injuries in the biofuels industry have received little attention in the research literature. Lack of a centralized source of data to investigate these incidents continues to be a challenge. This study proposes the use of workers' compensation claims data as a useful resource for investigating workplace injuries in the biofuels industry. The objective of this study was to characterize the relationship of the claim amount with employee age and tenure, nature and cause of injury, type of claim and the body part injured. These data are all collected as part of the workers' compensation claim process. This study found that the employee age, tenure, type of claim, injured body part, nature and cause of injury have a significantly influence on the claim amount. Since the claim amount is a proxy for injury severity this study shows that employee age, tenure, type of claim, injured body part, nature and cause of injury have a significantly influence on the severity of occupational injuries. Of all the variables analyzed, the type of claim was found to have the most significant influence on the claim amount. Furthermore, employees in the age group 46-50,

1-2 years' tenure groups have been identified as the most at-risk groups. Strain and sprain, laceration, foreign body, and fractures have a significant influence on the claim amount, while strain or injured by, slips, fall, or trips were found to be the most significant causes. Finally, injuries to lower extremities, head and neck, and injuries to multiple body parts have the influence on claim amount.

While workers' compensation data are extremely useful in injury prevention studies, the recording of information during the workers' compensation claims process could be prone to human errors. Also, the scope of analysis is narrowed by the information available in the dataset. However, analysis of a large number of claims, recorded over an extended period of time, characterizes the strength and rigorousness of this study. The findings of this study will enhance the understanding of the risks of injury as it highlights areas where safety efforts can be focused. Future work could involve analyzing the relationships between the non-cost related variables and also simultaneously investigating all the variables and their interaction effects. A multivariate model of the claim amount can be constructed so biofuels producers, as well as the worker's compensation insurance providers, can better analyze the risks from occupational accidents.

References

- Abdolhamidzadeh, B., Abbasi, T., Rashtchian, D., & Abbasi, S. A. (2011). Domino effect in process-industry accidents - An inventory of past events and identification of some patterns. *Journal of Loss Prevention in the Process Industries*, 24(5), 575-593.
doi:10.1016/j.jlp.2010.06.013
- Agresti, A., & Finlay, B. (2008). *Statistical Methods for the Social Sciences* (4th ed.). NJ: Prentice Hall.

- Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis and Prevention*, *41*(3), 359-364.
doi:10.1016/j.aap.2008.12.014
- Attwood, D., Khan, F., & Veitch, B. (2006). Occupational accident models - Where have we been and where are we going? *Journal of Loss Prevention in the Process Industries*, *19*(6), 664-682. doi:10.1016/j.jlp.2006.02.001
- Baldwin, M. L., & Butler, R. J. (2006). Upper extremity disorders in the workplace: Costs and outcomes beyond the first return to work. *Journal of Occupational Rehabilitation*, *16*(3), 303-323. doi:10.1007/s10926-006-9043-2
- Beery, L., Harris, J. R., Collins, J. W., Current, R. S., Amendola, A. A., Meyers, A. R., . . . Bertke, S. J. (2014). Occupational Injuries in Ohio Wood Product Manufacturing: A Descriptive Analysis With Emphasis on Saw-Related Injuries and Associated Causes. *American Journal of Industrial Medicine*, *57*(11), 1265-1275.
doi:10.1002/ajim.22360
- Bertolini, M., Bevilacqua, M., Ciarapica, F. E., & Giacchetta, G. (2009). Development of Risk-Based Inspection and Maintenance procedures for an oil refinery. *Journal of Loss Prevention in the Process Industries*, *22*(2), 244-253.
doi:10.1016/j.jlp.2009.01.003
- Bevilacqua, M., Ciarapica, F. E., & Mazzuto, G. (2012). Analysis of injury events with fuzzy cognitive maps. *Journal of Loss Prevention in the Process Industries*, *25*(4), 677-685.
doi:10.1016/j.jlp.2012.02.004
- Bird, F. E., Germain, G. L., & Veritas, N. (1996). *Practical loss control leadership*: Det Norske Veritas (USA).

- BLS. (2016). Employment status of the civilian noninstitutional population by age, sex, and race. Retrieved from <http://www.bls.gov/cps/cpsaat03.htm>
- Bookman, J. A. (2012). *Describing Agricultural Injury in Ohio Using the Ohio Bureau of Workers' Compensation Database*. The Ohio State University.
- Brorsson, B. (1989). Age and injury severity. *Scandinavian Journal of Social Medicine*, 17(4), 287-290.
- Calvo Olivares, R. D., Rivera, S. S., & Nunez Mc Leod, J. E. (2014). Database for accidents and incidents in the biodiesel industry. *Journal of Loss Prevention in the Process Industries*, 29, 245-261. doi:10.1016/j.jlp.2014.03.010
- Calvo Olivares, R. D., Rivera, S. S., & Nunez Mc Leod, J. E. (2015). Database for accidents and incidents in the fuel ethanol industry. *Journal of Loss Prevention in the Process Industries*, 38, 276-297. doi:10.1016/j.jlp.2015.10.008
- Cheng, C.-W., Yao, H.-Q., & Wu, T.-C. (2013). Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. *Journal of Loss Prevention in the Process Industries*, 26(6), 1269-1278.
doi:10.1016/j.jlp.2013.07.002
- Chettouh, S., Hamzi, R., & Benaroua, K. (2016). Examination of fire and related accidents in Skikda Oil Refinery for the period 2002-2013. *Journal of Loss Prevention in the Process Industries*, 41, 186-193.
- Coleman, P. J., & Kerkerling, J. C. (2007). Measuring mining safety with injury statistics: Lost workdays as indicators of risk. *Journal of Safety Research*, 38(5), 523-533.
doi:10.1016/j.jsr.2007.06.005

- Corso, P., Finkelstein, E., Miller, T., Fiebelkorn, I., & Zaloshnja, E. (2006). Incidence and lifetime costs of injuries in the United States. *Injury Prevention, 12*(4), 212-218. doi:10.1136/ip.2005.010983
- Dement, J. M., Pompeii, L. A., Ostbye, T., Epling, C., Lipscomb, H. J., James, T., . . . Thomann, W. (2004). An integrated comprehensive occupational surveillance system for health care workers. *American Journal of Industrial Medicine, 45*(6), 528-538. doi:10.1002/ajim.20017
- Dias, M. O. S., Junqueira, T. L., Cavalett, O., Cunha, M. P., Jesus, C. D. F., Rossell, C. E. V., . . . Bonomi, A. (2012). Integrated versus stand-alone second generation ethanol production from sugarcane bagasse and trash. *Bioresource Technology, 103*(1), 152-161. doi:10.1016/j.biortech.2011.09.120
- Douphrate, D. I., Rosecrance, J. C., & Wahl, G. (2006). Workers' compensation experience of Colorado agriculture workers, 2000-2004. *American Journal of Industrial Medicine, 49*(11), 900-910. doi:10.1002/ajim.20387
- EIA. (2016). *International Energy Statistics*. Washington, DC: U.S. Energy Information Administration Retrieved from <https://www.eia.gov/cfapps/ipdbproject/iedindex3.cfm?tid=79&pid=79&aid=1&cid=r1,&syid=2006&eyid=2012&unit=TBPD>.
- Ferjencik, M., & Jalovy, Z. (2010). What can be learned from incidents in chemistry labs. *Journal of Loss Prevention in the Process Industries, 23*(5), 630-636. doi:10.1016/j.jlp.2010.06.009
- Festel, G. W. (2008). Biofuels - Economic aspects. *Chemical Engineering & Technology, 31*(5), 715-720. doi:10.1002/ceat.200700335

- Frank Neuhauser, M., Mathur, A. K., & Pines, J. (2013). *Gender, Age, and Risk of Injury in the Workplace*. Paper presented at the Use of Workers' Compensation Data for Occupational Safety and Health: Proceedings from June 2012 Workshop.
- Freeman, S. A., Kelley, K. W., Maier, D. E., & Field, W. E. (1998). Review of entrapments in bulk agricultural materials at commercial grain facilities. *Journal of Safety Research*, 29(2), 123-134. doi:10.1016/s0022-4375(98)00008-5
- Gavious, A., Mizrahi, S., Shani, Y., & Minchuk, Y. (2009). The costs of industrial accidents for the organization: Developing methods and tools for evaluation and cost-benefit analysis of investment in safety. *Journal of Loss Prevention in the Process Industries*, 22(4), 434-438. doi:10.1016/j.jlp.2009.02.008
- Griend, N. V. (2011). Creating a Culture of Safety at an Ethanol Plant. *Ethanol Producer Magazine*. Retrieved from <http://www.ethanolproducer.com/articles/7490/creating-a-culture-of-safety-at-an-ethanol-plant>
- Gubicza, K., Nieves, I. U., Sagues, W. J., Barta, Z., Shanmugam, K. T., & Ingram, L. O. (2016). Techno-economic analysis of ethanol production from sugarcane bagasse using a Liquefaction plus Simultaneous Saccharification and co-Fermentation process. *Bioresource Technology*, 208, 42-48. doi:10.1016/j.biortech.2016.01.093
- Haarlemmer, G., Boissonnet, G., Peduzzi, E., & Setier, P.-A. (2014). Investment and production costs of synthetic fuels - A literature survey. *Energy*, 66, 667-676. doi:10.1016/j.energy.2014.01.093
- Hajakbari, M. S., & Minaei-Bidgoli, B. (2014). A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's

- Ministry of Labor data. *Journal of Loss Prevention in the Process Industries*, 32, 443-453. doi:10.1016/j.jlp.2014.10.013
- Hardy, C. L., Holz-Clause, M. S., Shepherd, H. E., & Hurburgh, C. R. (2006). Sourcing corn for ethanol: impacts of local processing. *November, a project of the Iowa Grain Quality Initiative, Iowa State University Extension, Ames, IA.*
- Harper, S., Etchells, J., Summerfield, A., & Cockton, A. (2008). *Health and safety in biodiesel manufacture*. Paper presented at the Institution of Chemical Engineers symposium series.
- Issa, S. F., Field, W. E., Hamm, K. E., Cheng, Y.-H., Roberts, M. J., & Riedel, S. M. (2016). Summarization of Injury and Fatality Factors Involving Children and Youth in Grain Storage and Handling Incidents.
- Janicak, C. (2010). *Safety Metrics: Tools and Techniques for Measuring Safety Performance* Government Institutes. *Lanham, MD.*
- Jorgensen, K. (2011). A tool for safety officers investigating "simple" accidents. *Safety Science*, 49(1), 32-38.
- Jorgensen, K. (2015). Serious work accidents and their causes - an analysis of data from EUROSTAT. *Safety Science Monitor*, 19(2).
- Jorgensen, K. (2016). Prevention of "simple accidents at work" with major consequences. *Safety Science*, 81, 46-58. doi:10.1016/j.ssci.2015.01.017
- Keren, N. (2010). Incident Database-based Framework for Establishing Industrial Safety Performance Assessments. *Journal of Industrial Technology*, 26(2), 1-12.

- Khazode, V. V., Maiti, J., & Ray, P. K. (2011). Injury count model for quantification of risk of occupational injury. *International Journal of Injury Control and Safety Promotion*, 18(2), 151-162. doi:10.1080/17457300.2010.540332
- Khazode, V. V., Maiti, J., & Ray, P. K. (2012). Occupational injury and accident research: A comprehensive review. *Safety Science*, 50(5), 1355-1367. doi:10.1016/j.ssci.2011.12.015
- Kim, B. K., Krams, J., Krug, E., Leaseburge, M., Lemley, J., Alkhalwaldeh, A., . . . Mannan, M. S. (2012). Case study analysis of the financial impact of catastrophic safety events. *Journal of Loss Prevention in the Process Industries*, 25(5), 780-787. doi:10.1016/j.jlp.2012.05.006
- Kim, H., Dropkin, J., Spaeth, K., Smith, F., & Moline, J. (2012). Patient handling and musculoskeletal disorders among hospital workers: Analysis of 7 years of institutional workers' compensation claims data. *American Journal of Industrial Medicine*, 55(8), 683-690. doi:10.1002/ajim.22006
- Kletz, T. A. (2008). Searchlights from the past. *Journal of Hazardous Materials*, 159(1), 130-134. doi:10.1016/j.jhazmat.2007.09.119
- Laflamme, L. (1996). Age-related accident risks among assembly workers: A longitudinal study of male workers employed in the Swedish automobile industry. *Journal of Safety Research*, 27(4), 259-268. doi:10.1016/s0022-4375(96)00024-2
- Lopez Arquillos, A., Rubio Romero, J. C., & Gibb, A. (2012). Analysis of construction accidents in Spain, 2003-2008. *Journal of Safety Research*, 43(5-6), 381-388. doi:10.1016/j.jsr.2012.07.005

- Manning, P. (2012). Do claimants stay on Worker's Compensation longer during tough economic times. *Wyoming Labor Force Trends*, 49(5).
- Manuele, F. A. (2013). *On the practice of safety*: John Wiley & Sons.
- Marchetti, J. M., Miguel, V. U., & Errazu, A. F. (2008). Techno-economic study of different alternatives for biodiesel production. *Fuel Processing Technology*, 89(8), 740-748.
doi:10.1016/j.fuproc.2008.01.007
- Marhavilas, P. K., Koulouriotis, D. E., & Mitrakas, C. (2011). On the development of a new hybrid risk assessment process using occupational accidents' data: Application on the Greek Public Electric Power Provider. *Journal of Loss Prevention in the Process Industries*, 24(5), 671-687. doi:10.1016/j.jlp.2011.05.010
- Meel, A., O'Neill, L. M., Levin, J. H., Seider, W. D., Oktem, U., & Keren, N. (2007). Operational risk assessment of chemical industries by exploiting accident databases. *Journal of Loss Prevention in the Process Industries*, 20(2), 113-127.
doi:10.1016/j.jlp.2006.10.003
- Moreno, V. C., & Cozzani, V. (2015). Major accident hazard in bioenergy production. *Journal of Loss Prevention in the Process Industries*, 35, 135-144.
doi:10.1016/j.jlp.2015.04.004
- Morrison, D. T., Fecke, M., & Martens, J. (2011). Migrating an incident reporting system to a CCPS process safety metrics model. *Journal of Loss Prevention in the Process Industries*, 24(6), 819-826. doi:10.1016/j.jlp.2011.06.008
- Mulloy, K. B., Sumner, S. A., Rose, C., Conway, G. A., Reynolds, S. J., Davidson, M. E., . . . Layde, P. M. (2013). Renewable Energy and Occupational Health and Safety Research Directions: A White Paper from the Energy Summit, Denver Colorado,

- April 11-13, 2011. *American Journal of Industrial Medicine*, 56(11), 1359-1370.
doi:10.1002/ajim.22184
- Nestoriak, N., & Pierce, B. (2009). Comparing Workers' Compensation claims with establishments' responses to the SOII. *Monthly Labor Review*, 132(5), 57-64.
- Nigam, P. S., & Singh, A. (2011). Production of liquid biofuels from renewable resources. *Progress in Energy and Combustion Science*, 37(1), 52-68.
doi:10.1016/j.pecs.2010.01.003
- Niskanen, T. (2012). Results of Finnish national survey in the chemical industry on EU legislation concerned with risk assessment and safety compliance climate. *Journal of Loss Prevention in the Process Industries*, 25(3), 535-543.
doi:10.1016/j.jlp.2011.12.010
- Nouri, J., Azadeh, A., & Fam, I. M. (2008). The evaluation of safety behaviors in a gas treatment company in Iran. *Journal of Loss Prevention in the Process Industries*, 21(3), 319-325. doi:10.1016/j.jlp.2007.11.006
- Nur, N. M., Dawal, S. M., & Dahari, M. (2014). *The Prevalence of Work Related Musculoskeletal Disorders among Workers Performing Industrial Repetitive Tasks in the Automotive Manufacturing Companies*. Paper presented at the International Conference on Industrial Engineering and Operations Management, Bali, Indonesia. Abstract retrieved from <http://ieom.org/ieom2014/pdfs/303.pdf>.
- Oleinick, A., & Zaidman, B. (2004). Methodologic issues in the use of workers' compensation databases for the study of work injuries with days away from work. I. Sensitivity of case ascertainment. *American Journal of Industrial Medicine*, 45(3), 260-274. doi:10.1002/ajim.10333

- OSHA. (2016a). *Ethanol Processing*. Occupational Safety & Health Administration
Retrieved from https://www.osha.gov/dts/osta/otm/otm_iv/chapters_all.pdf.
- OSHA. (2016b). *Green Job Hazards: Biofuels*. Occupational Safety & Health Administration
Retrieved from <https://www.osha.gov/dep/greenjobs/biofuels.html>.
- Owens, K. A., & Hazeldean, J. A. (1995). Fires, explosions and related incidents at work in 1992-1993. *Journal of Loss Prevention in the Process Industries*, 8(5), 291-297.
doi:10.1016/0950-4230(95)00035-y
- Pasman, H. J. (2009). Learning from the past and knowledge management: Are we making progress? *Journal of Loss Prevention in the Process Industries*, 22(6), 672-679.
doi:10.1016/j.jlp.2008.07.010
- Pirdavani, A., Brijs, T., & Wets, G. (2010). A Multiple Criteria Decision-Making Approach for Prioritizing Accident Hotspots in the Absence of Crash Data. *Transport Reviews*, 30(1), 97-113. doi:10.1080/01441640903279345
- Priambodo, R., Chen, T.-C., Lu, M.-C., Gedanken, A., Liao, J.-D., & Huang, Y.-H. (2015). Novel Technology for Bio-diesel Production from Cooking and Waste Cooking Oil by Microwave Irradiation. *Clean, Efficient and Affordable Energy for a Sustainable Future*, 75, 84-91. doi:10.1016/j.egypro.2015.07.143
- Reniers, G. (2009). An optimizing hazard/risk analysis review planning (HARP) framework for complex chemical plants. *Journal of Loss Prevention in the Process Industries*, 22(2), 133-139. doi:10.1016/j.jlp.2008.10.005
- Reville, R. T., Bhattacharya, J., & Weinstein, L. R. S. (2001). New methods and data sources for measuring economic consequences of workplace injuries. *American Journal of Industrial Medicine*, 40(4), 452-463. doi:10.1002/ajim.1115

- Reville, R. T., Polich, S., Seabury, S. A., & Giddens, E. (2001). *Permanent disability at private, self-insured firms*: Rand Corporation.
- Riedel, S., & Field, W. (2011). *Summary of over 800 grain storage and handling-related entrapments and suffocations documented in the US between 1970 and 2010*. Paper presented at the Presentation at XXXIV CIOST-CIGR V Conference, Vienna, Austria.
- Rivera, S. S., Olivares, R. D. C., Baziuk, P. A., & McLeod, J. E. N. (2015). *Assessment of Biofuel Accident Risk: A Preliminary Study*. Paper presented at the Proceedings of the World Congress on Engineering.
- Riviere, C., & Marlair, G. (2010). The use of multiple correspondence analysis and hierarchical clustering to identify incident typologies pertaining to the biofuel industry. *Biofuels Bioproducts & Biorefining-Biofpr*, 4(1), 53-65.
doi:10.1002/bbb.187
- Roberts, M., & Field, B. (2010). A disturbing trend-US Grain Entrapments on the Increase. *Resource Magazine*, 17(4), 10-11.
- Rogers, E., & Wiatrowski, W. J. (2005). Injuries, illnesses, and fatalities among older workers. *Monthly Labor Review*, 128(10), 24-30.
- Salminen, S. (2004). Have young workers more injuries than older ones? An international literature review. *Journal of Safety Research*, 35(5), 513-521.
doi:10.1016/j.jsr.2004.08.005
- Schwatka, N. V., Butler, L. M., & Rosecrance, J. C. (2013). Age in relation to worker compensation costs in the construction industry. *American Journal of Industrial Medicine*, 56(3), 356-366. doi:10.1002/ajim.22093

- Scovronick, N., & Wilkinson, P. (2014). Health impacts of liquid biofuel production and use: A review. *Global Environmental Change-Human and Policy Dimensions*, 24, 155-164. doi:10.1016/j.gloenvcha.2013.09.011
- Sears, J. M., Blanar, L., & Bowman, S. M. (2014). Predicting work-related disability and medical cost outcomes: A comparison of injury severity scoring methods. *Injury-International Journal of the Care of the Injured*, 45(1), 16-22. doi:10.1016/j.injury.2012.12.024
- Sears, J. M., Blanar, L., Bowman, S. M., Adams, D., & Silverstein, B. A. (2013). Predicting Work-Related Disability and Medical Cost Outcomes: Estimating Injury Severity Scores from Workers' Compensation Data. *Journal of Occupational Rehabilitation*, 23(1), 19-31. doi:10.1007/s10926-012-9377-x
- Sengupta, I., Reno, V. P., Burton Jr, J. F., & Baldwin, M. L. (2012). Workers' compensation: Benefits, coverage, and costs, 2010. *National Academy of Social Insurance*.
- Sharpe, D. (2015). Your Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research & Evaluation*, 20.
- Smith, P., Hogg-Johnson, S., Mustard, C., Chen, C., & Tompa, E. (2012). Comparing the risk factors associated with serious versus and less serious work-related injuries in ontario between 1991 and 2006. *American Journal of Industrial Medicine*, 55(1), 84-91. doi:10.1002/ajim.21000
- Spellman, F. R. (2013). *Safe Work Practices for Green Energy Jobs*: DEStech Publications, Inc.

- Suarez-Cebador, M., Carlos Rubio-Romero, J., & Lopez-Arquillos, A. (2014). Severity of electrical accidents in the construction industry in Spain. *Journal of Safety Research*, 48, 63-70. doi:10.1016/j.jsr.2013.12.002
- Sumner, S. A., & Layde, P. M. (2009). Expansion of Renewable Energy Industries and Implications for Occupational Health. *Jama-Journal of the American Medical Association*, 302(7), 787-789.
- Takahashi, A., & Miura, T. (2016). At What Age Is the Occupational Accident Risk High? Analysis of the Occurrence Rate of Occupational Accidents by Age *Advances in Safety Management and Human Factors* (pp. 3-10): Springer.
- Tauseef, S. M., Abbasi, T., & Abbasi, S. A. (2011). Development of a new chemical process-industry accident database to assist in past accident analysis. *Journal of Loss Prevention in the Process Industries*, 24(4), 426-431. doi:10.1016/j.jlp.2011.03.005
- Utterback, D., Meyers, A., & Wurzelbacher, S. (2014). Workers' Compensation Insurance: A Primer for Public Health. *Department of Health and Human Services: National Institute for Occupational Safety and Health*.
- Utterback, D. F., & Schnorr, T. M. (2010). Use of workers' compensation data for occupational injury & illness prevention. *National Institute for occupational safety and health, US Department of Labour*, 2010-2152.
- Utterback, D. F., Schnorr, T. M., Silverstein, B. A., Spieler, E. A., Leamon, T. B., & Amick, B. C., III. (2012). Occupational Health and Safety Surveillance and Research Using Workers' Compensation Data. *Journal of Occupational and Environmental Medicine*, 54(2), 171-176. doi:10.1097/JOM.0b013e31823c14cb

- van Tulder, M., Malmivaara, A., & Koes, B. (2007). Repetitive strain injury. *Lancet*, 369(9575), 1815-1822. doi:10.1016/s0140-6736(07)60820-4
- Vimmerstedt, L. J., Bush, B. W., & Peterson, S. (2013). Effects of Deployment Investment on the Growth of the Biofuels Industry. *Golden, CO: National Renewable Energy Laboratory*. <http://dx.doi.org/10.2172/1118095>.
- Vinodkumar, M. N., & Bhasi, M. (2009). Safety climate factors and its relationship with accidents and personal attributes in the chemical industry. *Safety Science*, 47(5), 659-667. doi:10.1016/j.ssci.2008.09.004
- Vlysidis, A., Binns, M., Webb, C., & Theodoropoulos, C. (2011). A techno-economic analysis of biodiesel biorefineries: Assessment of integrated designs for the co-production of fuels and chemicals. *Energy*, 36(8), 4671-4683. doi:10.1016/j.energy.2011.04.046
- Vredenburg, A. G. (2002). Organizational safety: Which management practices are most effective in reducing employee injury rates? *Journal of Safety Research*, 33(2), 259-276. doi:10.1016/s0022-4375(02)00016-6
- Wu, H. C. (2004). Explosion accident analysis of diaminodiphenolether process. *Journal of Loss Prevention in the Process Industries*, 17(5), 373-376. doi:10.1016/j.jlp.2004.05.006

CHAPTER 3. USING WORKERS' COMPENSATION CLAIMS DATA TO CHARACTERIZE OCCUPATIONAL INJURIES IN THE COMMERCIAL GRAIN ELEVATOR INDUSTRY

Manuscript submitted to the *Journal of Agricultural Safety and Health*

Sai K. Ramaswamy^a, Gretchen A. Mosher^a

^aDepartment of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA

Abstract

Workplace injuries in the grain handling industry are common, yet little research has characterized worker injuries in grain elevators across all hazard types. Learning from past injuries is essential for preventing future occurrences, but the lack of injury information for the grain handling industry hinders this effort. The present study addresses this knowledge gap by utilizing data from over 7000 workers' compensation claims reported from 2008 to 2016 by commercial grain handling facilities in the U.S. to characterize injury costs and severity. The total amount paid for each claim was used as a measure of injury severity. The effect of worker age and tenure, cause of injury and body part injured on the cost of work-related injuries were investigated. Contingency tables were used to classify the variable pairs. The chi-square test and chi-square residuals were employed to evaluate the relationship between the variable pairs and identify the at-risk groups. Results showed that age and tenure of employee, cause of injury and body part injured have a significant influence on the cost paid for the claim. Several at-risk groups were identified as a result of the analyses. Findings from the study will assist commercial grain elevators in the development of targeted safety interventions and assist grain elevator safety managers in mitigation of financial and social losses from occupational injuries.

Keywords: grain elevator, grain handling, occupational injuries, at-risk groups

Introduction

The grain handling industry in the U.S. is a hazardous work environment with workers in these facilities constantly at risk of severe and life-threatening occupational injuries (Issa et al., 2016). Common sources of occupational hazards in the grain handling facilities include grain dust, grain engulfment, entrapment in confined spaces, slips, falls, trips, equipment-related hazards, exposure to harmful chemicals and gasses (OSHA, 2016; Snyder & Bobick, 1995). Identification and characterization of past safety incidents can drive potential intervention strategies intended to mitigate injury risks (Cohen, Clark, Silverstein, Sjostrom, & Spielholz, 2006; Kines, Spangenberg, & Dyreborg, 2007; Menckel & Carter, 1985; Verma, Das Khan, Maiti, & Krishna, 2014). However, the majority of past studies investigating injuries and fatalities in grain facilities have focused only on a few safety hazards. This is true even though hazards in a grain handling facility are plentiful. Some of the safety risks affecting the workers in the grain handling industry are, exposure to chemicals and gasses, electrical hazards, noise hazard due to fast moving machinery like conveyors, motors, and augers, slips, trips and fall and finally suffocation and engulfment hazards (Van Fleet, Frank, & Rosenbeck, 2013).

Despite the existence of numerous workplace hazards in the grain handling industry, very few comprehensive studies have thus far examined worker safety across all the hazard categories. Previous research has examined specific hazards and resulting injuries in commercial grain handling environments. For example, Freeman, Kelley, Maier and Field (1998), examined entrapments in various bulk commodities at commercial grain facilities. Similarly, the study by Field, Heber, Riedel, Wettschurack, Roberts, and Grafft (2014), examined hazards associated with grain vacuum systems at commercial grain storage

facilities. No recent comprehensive characterization of injuries in the commercial grain handling setting has been completed. A National Institute for Occupational Safety and Health (NIOSH) study (NIOSH, 1983) was the last time a detailed analysis of injuries in grain elevators was conducted. A review of research literature showed very few follow-up studies to the 1983 NIOSH study were completed, even though the grain handling industry has seen several changes over the last few years (Rosentrater & Williams, 2004).

The grain handling industry performs an important role in U.S. agriculture by handling, storing, distributing and processing a variety of agricultural commodities (Williams & Rosentrater, 2004). According to the National Agricultural Statistics Service (NASS), in 2015, there were 8,638 commercial grain facilities in the United States, storing and handling 11 billion bushels of grains such as corn, wheat, soybean, and oats. In the last five years, the grain storage capacity in the U.S. has increased by 15% while the number of grain storage facilities has decreased by 4% (NASS, 2011, 2016). Furthermore, the average grain stored at each facility has increased by 22% from 2010 to 2015. These numbers suggest U.S. grain handling facilities are getting larger, handling and moving higher volumes of grain as compared to previous years. This expansion of the grain handling industry has resulted in a high rate of occupational injuries and fatalities as compared to prior years (S. Riedel & Field, 2011). According to the National Institute for Occupational Safety and Health (NIOSH), grain-handling machinery is the second largest factor in farm machinery-related deaths and disabilities (Snyder & Bobick, 1995).

The availability of injury data, especially non-fatal injury data is a continuing challenge, limiting potential development of research-based safety intervention in grain handling facilities (Issa et al., 2016; Patel, Watanabe-Galloway, Rautiainen, Haynatzki, &

Gofin, 2016; Zhou & Roseman, 1994). Although, the Occupational Safety and Health Administration (OSHA) standard 29 CFR 1910.272 regulates the grain handling industry, OSHA record-keeping does not always record all injuries and fatalities that occur in the grain industry (Issa, Cheng, & Field, 2016). Furthermore, many grain handling facilities are exempt from OSHA record-keeping requirements because they have fewer than 11 employees (Doupbrate, Rosecrance, & Wahl, 2006; Zhou & Roseman, 1994). Even in larger facilities, because of budgetary, administrative, and logistical constraints, OSHA collects data only from employers deemed as high hazard, and most often, only from companies with more than 40 employees (Leeth, 2012). Additionally, data gathered from grain handling facilities is frequently mixed with other farm-level data, so drawing conclusions on workplace conditions of grain elevators becomes difficult (Doupbrate, Rosecrance, Stallones, Reynolds, & Gilkey, 2009).

The most widely used source for investigating occupational injuries and fatalities across various industries are the Census of Fatal Occupational Injuries (CFOI) and the Survey of Occupational Injuries and Illness (SOII) published by the U.S. Bureau of Labor Statistics (BLS) (Biddle & Marsh, 2002; Nanda, Grattan, Chu, Davis, & Lehto, 2016). Waehrer, Dong, Miller, Haile, and Men (2007), investigated the cost of occupational injuries in the high-hazard construction industry using the SOII data from BLS. Similarly, Asfaw, Pana-Cryan, and Rosa (2011), investigated workplace injuries across five industry sectors using the non-fatal SOII injury data. While the BLS data are a useful source for injury investigations, researchers have also highlighted its limitation for studying workplace injuries in the agricultural industry (Doupbrate, Rosecrance, Stallones, et al., 2009; Landsteiner, McGovern, Alexander, Lindgren, & Williams, 2015; Patel et al., 2016; Riedel & Field,

2013). According to Riedel and Field (2013), BLS data include only annual totals for injuries and fatalities and do not provide detailed information such as causative factors, considered essential for studying workplace hazards. Furthermore, evidence from scientifically published literature also suggests that BLS data significantly underreport work-related injury data, missing between 61 and 88 percent of non-fatal injuries (Boden & Ozonoff, 2008; Leigh, Du, & McCurdy, 2014; Leigh, Marcin, & Miller, 2004; Rosenman et al., 2006). For this reason, a need exists for an alternative injury data source to investigate workplace injuries in the grain handling industry.

Workers' compensation insurance claims records can partially address the informational gap about occupational injuries in the grain handling industry, enhancing the ability of firms to develop effective safety interventions (Utterback, Meyers, & Wurzelbacher, 2014). Workers' compensation insurance provides an injured worker medical benefits, a portion of the employee's wage, and a lump sum payment when the employee suffers a permanent impairment (Sengupta, Reno, Burton Jr, & Baldwin, 2012). Employers in all states in the U.S. except Texas are required to provide their employees with workers' compensation insurance. Each year, companies in the U.S. across all industries spend approximately \$85 billion on workers' compensation insurance costs (Sengupta et al., 2012; Utterback & Schnorr, 2010). In addition to providing benefits to the injured worker, workers' compensation insurance also protects employers from lawsuits resulting from occupational injuries.

Workers' compensation data contains valuable information commonly used in injury characterization (Utterback et al., 2012). In addition to information on the direct costs of the injury such as medical, indemnity, and disability payments, data on the industry, occupation,

nature of injury, cause of injury, and demographic information of the injured worker are also captured in workers' compensation claims data (Nestoriak & Pierce, 2009; Utterback et al., 2012). Several studies have highlighted the use of workers' compensation claims data as an excellent data source that provides information on workplace injuries and their contributing factors (Dement et al., 2004; Foley, Rauser, Rappin, & Bonauto, 2013; Meyers et al., 2013). The size and volume of workers' compensation datasets provide a comprehensive understanding of injury patterns, which can then be used to analyze causal factors leading to an injury (Oleinick & Zaidman, 2004).

Previous research has shown that workers' compensation claims data can be used to characterize the risk, scope, and nature of workplace injuries across multiple industries. Neuhauser, Mathur, and Pines (2013) utilized workers' compensation data to compare injury incidence by gender and age while controlling for the occupation and type of industry of the injured worker. Sears, Blonar, Bowman, Adams, and Silverstein (2013) used workers' compensation data to predict occupational disability and medical cost outcomes. Smith, Hogg-Johnson, Mustard, Chen, and Tompa (2012) compared risk factors associated with severe versus less severe occupational injuries using workers' compensation data in industries such as agriculture, mining, and manufacturing. Coleman and Kerkering (2007) studied occupational injuries in coal mines and used workers' compensation data to distinguish between lower and higher risk operations and time periods. Schwatka, Butler, and Rosecrance (2013) studied the relationship between age and injury type on claim amount in the construction industry using workers' compensation claims from 1998 to 2008.

Review of research literature also showed that workers' compensation claims data have been used previously to characterize occupational injuries in some agricultural-based

industries. For example, Bell and Helmkamp (2003) examined workers' compensation injury claims to investigate injury patterns and rates of non-fatal logging injuries. Douphrate, Rosecrance, Reynolds, Stallones, and Gilkey (2009) studied tractor-related injuries by analyzing workers' compensation data. Similarly, Bookman (2012) used workers' compensation data to investigate occupational injuries of Ohio agricultural workers over a ten-year period (1999-2008). Despite the validated benefits of utilizing workers' compensation claims data for studying agricultural injuries, limited research has expanded the use of these data to study occupational injuries in the grain handling industry.

This study investigated occupational injuries in grain elevators using workers' compensation claims data provided by a leading Midwest-based insurance company. The purpose of this study was to characterize the direct cost of occupational injury using the information obtained from the workers' compensation claims including variables such as body part injured, cause of injury, employee age, and tenure. A secondary purpose of the study was to identify and classify at-risk groups within the grain handling industry to enable development of targeted intervention strategies for mitigating the risk of occupational injuries.

Methods and Data

The dataset used in this study obtained from a private insurance company headquartered in a Midwest state. The claim data were dated from January 2008 to March 2016. The variables used in this research are shown in Table 1, and were used from the dataset, with the exception of employee age and tenure. The age of the employee was calculated as the difference between the date of birth and the injury date. Similarly, the tenure of the employee was calculated as the difference between the date of hire and injury

date. To simplify analysis, the claim amount was categorized as “<\$3,000”, “\$3,000-\$9,999”, and “10,000+”.

Table 1: *List of variables in dataset*

#	Column Name	Description
1	Claim	Unique identifier for each claim record
2	Effective Year	Filing year of the claim
3	Account	Unique identifier to differentiate claims for each customer
4	Market	Type of business (grain elevator)
5	Gender	Gender of injured worker
6	State	Name of state where injury occurred
7	Date of Birth	Date of birth of injured worker
8	Date of Hire	Date on which the present company hired the injured worker
9	Injury Date	Date on which the injury occurred
10	Claim Description	One-line narration of incident resulting in injury; Example: "Employee was cleaning equipment and employee opened up a line and acid sprayed in his face and mouth."
11	Claim Status	If the claim is still open or closed
12	Body Part	Body part(s) injured
13	Cause of Injury	Main cause of injury. For example: "Cut, puncture or scrape", "heat or cold exposures", "Fall, slip or trip" etc.
14	Nature of Injury	Describes the type of injury such as Fracture, Strain, Contusion etc.
15	Claim amount	Total amount paid out in medical, indemnity and other miscellaneous payments. Used as a proxy for injury severity in this study.

Grain elevators are classified as off-farm commercial enterprises and are required to provide workers compensation insurance to their employees (Ag web, 2015). Employers provide this benefit to their employees by either purchasing insurance from an insurance carrier or through self-insurance (Reville, Polich, Seabury, & Giddens, 2001). When an employee is injured on the job, the insurance carrier or the self-insured employer pays the medical and indemnity costs. To provide information and to facilitate the payment, employers must create a report of the worker’s injury to inform their insurance provider (Utterback et al., 2012). Data collected during the claims process are provided by employees,

employers, insurance companies and other involved parties (Utterback et al., 2014). The collection of information from multiple stakeholders makes claims records an excellent data source for work-related injuries (Dement et al., 2004; Janicak, 2010; Kim, Dropkin, Spaeth, Smith, & Moline, 2012; Reville, Bhattacharya, & Weinstein, 2001).

The variables used in this study were categorical. For this reason, the statistical analysis began with the construction of frequency counts, percentages, and contingency tables. The chi-square test was used to validate the hypothesis of independence of the claim amount from the demographic (age, tenure) and injury variables (nature of injury, body part injured). This statistical methodology was also used by previous studies investigating injuries in agriculture (Javadi & Rostami, 2007; Karttunen & Rautiainen, 2011; Sprince et al., 2003). Standardized residuals were calculated to identify the source of dependence between the two variables or at-risk groups. The standardized residual is the difference between the observed value of a particular variable and its expected value (Agresti & Finlay, 2008). A positive residual implies that the observed value was greater than the expected value, while a negative residual indicates the observed value was less than expected value. The value and sign (positive/negative) of residuals were used to determine the nature of the relationship between the row and the column variable of the contingency tables (Lopez, Ritzel, Gonzalez, & Alcantara, 2011).

All descriptive and the inferential analyses were performed with statistical software SAS version 9.4. In SAS, residuals are standardized and calculated as:

$$\text{Standardized residual} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}(1-p_{i.})((1-p_{.j}))}}$$

Where n_{ij} is the observed value, e_{ij} is the expected value for the i^{th} row j^{th} column cell. p_i is row total for i^{th} row and p_j is the column total for the j^{th} column. According to Agresti and Finlay (2008), an adjusted residual of +/- 2 is evidence of dependence between the row and column variables while an adjusted residual of +/- 3 is evidence of strong dependence. Examining the adjusted residuals of each cell in a contingency table helped identify the at-risk groups where the degree of dependence between the two variables was the strongest (Sharpe, 2015).

Since the purpose of this study was to characterize the direct cost of occupational injury using the demographic and injury characteristics the claim amount was used as a proxy for direct injury cost. The broad research question that guided this study was: Is the claim amount of injuries in the grain elevators independent of the employee demographics and injury characteristics? Specifically, the following sub research questions listed below were analyzed:

Is the claim amount independent of:

- vi. Age of employee
- vii. Tenure of employee
- viii. Nature of injury and
- ix. Body part injured

Results and Discussion

Characterizing claim amount based on employee age

The first research question investigated if the claim amount and the age of the injured employee were independent. The claim amount is the sum of all payments made by the workers' compensation insurance provider to the injured employee. This amount includes

medical, indemnity, and other miscellaneous payments made to the injured employee as compensation for their work related injury. For this reason, severe injuries, such as those resulting in disability or death have a higher claim amount than less severe injuries requiring less medical treatment (Sears, Blonar, & Bowman, 2014; Sears et al., 2013).

Table 2: Relationship between age of employee and claim amount

Age (years)	Claim amount			Total
	<\$3000	\$3000- \$9999	\$10,000 & above	
<25	979 (8.8**)	64 (-2.3*)	82 (-8.6**)	1125
26-30	636 (5.6**)	45 (-1.4)	67 (-5.5**)	748
31-35	549 (2.3*)	50 (-2.7*)	84 (0.0)	683
36-40	503 (0.2)	40 (0.6)	109 (-1.2)	652
41-45	481 (-2.3*)	51 (0.5)	126 (2.4*)	658
46-50	608 (-3.0**)	64 (0.4)	165 (3.2**)	837
51-55	688 (-2.5*)	68 (-0.6)	180 (3.0**)	936
56-60	652 (-4.0**)	72 (0.7)	188 (4.2**)	912
60+	585 (-5.7**)	87 (3.5**)	176 (4.1**)	848
Total	5681	541	1177	7399

χ^2 180.14; df= 16; p-value= <0.0001 and $\alpha = 0.05$; N=7399

Residuals in brackets; * evidence of relationship; ** evidence of strong relationship

To explore the relationship between employee age and the claim amount, a contingency table with age as the row variable and claim amount as the column variable was tabulated as shown in Table 2. Each cell in the contingency table is a count of claims corresponding to the respective age group and claim amount category. The last row in Table

2 shows the total number of claims in each claim amount category, while the last column indicates the number of claims corresponding to each age group.

The distribution of the number of claims based on the claim amount indicates that nearly 84% of the claims were less than \$10,000, suggesting that small injury claims are the largest workers' compensation expense for grain handling organizations. The distribution of the number of claims based on age group, shows 50% of the claims involved an employee less than 45 years of age and 76% of the claims involved an employee less than 55 years old. One out of every four claims involved an employee older than 55 years. This result is different from the studies by Douphrate et al. (2009), Bookman (2012), and Reiner, Gerberich, Ryan, and Mandel (2016), where injuries to employees below the age of 55 years constituted over 90% of the injuries. One reason for this difference in distribution between the current study and previous studies could be the type of occupation investigated in these studies. For example, Bookman (2012) analyzed workers compensation claims of employees in various agricultural occupations such as poultry and egg producers, logging or tree removal, and fisheries and hatcheries. Bookman's (2012) study did not include grain elevators. Similarly, Reiner et al. (2016) study investigated only farm injuries caused by large machinery such as augers, balers, and harvesting machinery.

A chi-square test was conducted to evaluate if the claim amount varied based on the age of the employee. The test results showed a p-value of less than 0.05 providing evidence that the claim amount and the age of the employee were not independent. This finding implies that the claim amount varied based on the age of the employee and that the age of the employee is a significant factor that can be used to determine the claim amount. The finding is consistent with previous studies that also found a significant relationship between age and

injury severity (Laflamme, 1996; Rogers & Wiatrowski, 2005; Salminen, 2004; Takahashi & Miura, 2016). According to Rogers and Wiatrowski (2005) and Salminen (2004), young workers below the age of 25 years have a higher risk of injuries than older workers.

However, the injuries to older workers are likely to be more severe when compared with younger workers. More severe injuries require increased medical attention and could also result in lost workdays, resulting in a higher claim amount. To identify at-risk age groups, residuals were examined. The residuals indicate a strong relationship between age and claim amount across all age groups except employees in the 36 to 40 years' age group category. It is noteworthy that as the age of the employee increases from < 25 years to 40 years the sign of residual indicates a greater than expected number of claims in the below \$3,000 category and fewer than expected number of claims in the \$10,000 and above category. This finding implies that employees up to 40 years of age are likely to have less severe injuries, which requires minimum levels of workers' compensation payments. The residual values for the 41 and above age group categories indicate exactly opposite to that of the below 40 year age group categories. In all the 41 and above age group categories, the number of claims in the below \$3,000 category are less than expected value while the number of claims in the \$10,000 and above category are more than the expected value. This finding implies that injuries to grain elevator employees who are older than 40 years are likely to be more severe and expensive as compared to employees who are 40 years and younger. In other words, as the age of the employee increases, the number of claims for minor injuries tends to decrease. At the same time, as the age of the employee increases, the number of claims for major injuries tends to increase, suggesting that older employees in this environment should heighten their focus on safe work practices. It must be noted that in this study the minor

claims with younger employees become more “major” as the employee ages. This finding further corroborates the results of previous studies (Laflamme, 1996; Rogers & Wiatrowski, 2005; Salminen, 2004; Takahashi & Miura, 2016) who suggested that older workers are likely to have more severe injuries than younger workers.

Characterizing claim amount based on employee tenure

Table 3: Relationship between tenure of employee and claim amount

Tenure (years)	Claim amount			Total
	<\$3,000	\$3000-\$9999	\$10,000+	
<1	1138 (2.5*)	98 (-0.8)	199 (-2.4*)	1435
1-2	1641 (3.2**)	147 (-0.4)	281 (-3.4**)	2069
3-5	1017 (-0.3)	94 (-0.4)	219 (0.6)	1330
6-10	764 (-2.8*)	76 (0.0)	201 (3.2**)	1041
11-20	648 (-2.9*)	77 (1.6)	164 (2.2*)	889
20+	470 (-1.5)	49 (0.4)	113 (1.4)	632
Total	5678	541	1177	7396

$\chi^2 = 33.56$; $df = 10$; $p\text{-value} = 0.0002$ and $\alpha = 0.05$; $N = 7396$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

The second research question investigated if the tenure of the injured employee and the claim amount were independent of each other. The contingency table used to investigate this research question is shown in Table 3. The distribution of the number of claims by employee tenure indicates that nearly half of all injury-causing safety incidents involved an employee with two or fewer years of work experience. Similarly, 65% of all injury-causing safety incidents involved an employee with five or fewer years of experience and 80% of all incidents involved an employee with ten or less years of work experience. Furthermore, the

general trend of the number of claims by employee tenure indicates that as tenure increases the number of claims decreases. According to Vinodhkumar and Bhasi (2009), the length of service of an employee in a company influences their skills and attitudes towards safety. Mariger et al. (2009) in their comprehensive study of farm injuries observed that experienced workers in agriculture tend to have fewer injuries than less experienced workers.

The chi-square test to evaluate if the claim amount varied based on the tenure of the employee showed a p-value of less than the 0.05, thus suggesting the tenure of the injured employee is also a significant factor that can be used to determine the claim amount. The relationship between employee tenure and injury severity has not been investigated in the agricultural industry. Evidence from studies conducted in other high hazard industries such as construction and petrochemicals suggest a significant relationship between between tenure and injury severity (Cheng, Yao, & Wu, 2013; Lopez Arquillos, Rubio Romero, & Gibb, 2012; Nouri, Azadeh, & Fam, 2008; Suarez-Cebador, Carlos Rubio-Romero, & Lopez-Arquillos, 2014). The residual values from the chi-square test show evidence of a relationship between tenure and claim amount across most age categories. With claims below \$3,000, residuals change from positive to negative as the employee tenure increases, suggesting that the longer the tenure of the employee, the fewer small dollar claims they incur. The opposite is true with claims above \$10,000, with residual values shifting from negative to positive. In this case, the longer the tenure of an employee, the more likely he or she will incur a more expensive workers' compensation claim. This shift in injury severity could be attributed to the age of the employee since tenure and age are highly correlated (Vinodhkumar & Bhasi, 2009). As tenure of the employee increase so does their age, reducing

their ability to tolerate injuries (Brorsson, 1989) thereby increasing the likelihood of severe injuries.

Characterizing claim amount based on cause of injury

The third research question investigated if the claim amount was statistically independent of the cause of injury. The contingency table used to classify the claim amount by cause of injury is shown in Tables 5.

Table 5: Relationship between cause of injury and claim amount

Cause of injury	Claim amount			Total
	<\$3,000	\$3,000-\$9,999	\$10,000+	
Strain or injured by	1552 (-4.3**)	173 (1.8)	387 (3.6**)	2112
Slip, fall or trip	1281 (-11.6**)	178 (3.9**)	448 (10.5**)	1907
Struck or injured by	926 (7.4**)	62 (-2.2*)	94 (-7.0**)	1082
Others	666 (7.5**)	31 (-3.6**)	63 (-6.1**)	760
Cut, puncture, scrape	650 (10.1**)	26 (-3.9**)	30 (-8.9**)	706
Vehicle	190 (-5.3**)	35 (3.0**)	72 (4.0**)	297
Heat or cold Exposures	246 (3.0**)	13 (-1.9)	34 (-2.1*)	293
Caught in, under, or between	175 (-2.6*)	23 (1.2)	49 (1.7)	247
Total	5686	541	1177	7404

$\chi^2 = 351.6$; $df = 14$; $p\text{-value} = <0.001$ and $\alpha = 0.05$; $N=7404$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

The distribution of claims in Table 5 revealed that “strain or injured by” were the most common causes of injuries followed by “slips, falls and trip” and “struck or injured by”.

Nearly 29% of claims were for strain and sprain injuries while “heat or cold exposures” and “caught in, under, or between” accounted for only 4% of the claims. Furthermore, more than half (54%) of the injuries recorded were either due to strain or due to slips and fall injury.

The Chi-square test suggested that the claim amount was not independent of the cause of the injury. This finding means the claim amount is not the same for all type of injuries and that cause of the injury is a significant factor that can be used to determine the claim amount.

Examining the chi-square residuals indicates that certain injuries are more likely to have higher claim amounts. For example, injuries in the strain, slip, fall and trip categories are likely to be more expensive when compared to “struck or injured by” and “others” categories. A review of the agricultural safety and health literature shows that strain, slips and falls are common across many industries, including agribusiness (Bobick & Myers, 1994; Davis & Kotowski, 2007; Douphrate, 2008; Fathallah, Miller, & Miles, 2008) . In the U.S., strain injuries alone cost \$6.5 billion in workers’ compensation costs with the average claim ranging from \$5000 to \$8000 (Baldwin & Butler, 2006; van Tulder, Malmivaara, & Koes, 2007).

Residuals from the chi-square test also indicates that the categories of “slip, trip, or fall”, “struck or injured by”, “others”, “cut, puncture, scrape”, and “vehicle” have the strongest relationship with claim amount. Evidence of a significant relationship with claim amount was also observed in one cell corresponding to the category “caught in, under, or between” and two cells corresponding to the categories strain or injured by and heat or cold exposures. However, the sign of residuals suggests that for categories “strain or injured by” , “slips, falls or trips”, and “vehicles” the claim amount are more likely to be \$10,000 and above and less likely to be below \$3,000. This finding suggests that injuries due to slips, trips, and falls or

injuries involving a strain or those due to vehicles are likely to be more expensive than the other injuries. For the remaining categories including “others”, struck or injured by, and cut puncture scrape, the residual values indicate the claim amount is likely to be lower. Based on these analyses, targeting slips, trips, and falls, strain related injuries and vehicle-related injuries through safety interventions could potentially reduce the claim amount significantly.

Characterizing claim amount based on body part injured

The final research question investigated if the claim amount was statistically independent of the injured body part. A contingency table constructed to address this research question is shown in Table 7.

Table 7: Relationship between body part injured and claim amount

Body part	Claim amount			Total
	<\$3,000	\$3,000-\$9,999	\$10,000+	
Upper extremities	1909 (-2.6*)	182 (-0.4)	454 (3.3**)	2545
Lower extremities	1178 (-5.2**)	126 (0.7)	331 (5.4**)	1635
Trunk	1203 (-0.5)	142 (2.9*)	231 (-1.5)	1576
Head & neck	1071 (11.2**)	58 (-3.6**)	71 (-10.3**)	1200
Others	325 (-2.2*)	33 (0.0)	90 (2.5*)	448
Total	5686	541	1177	7404

$\chi^2 = 155.1$; $df = 8$; $p\text{-value} = <0.001$ and $\alpha = 0.05$; $N=7404$

Residuals in brackets; * evidence of dependence; ** evidence of strong dependence

The distribution of data suggests that upper extremities (such as hands and fingers) were the most frequently injured body part, followed by lower extremities (such as toes and feet) and trunk. In nearly 80% of claims filed, the injured body part was either an upper

extremity, trunk or the lower extremity. This finding is different from the results of the NIOSH study (1983) that reported back, finger and eyes as the three most injured body parts. The chi-square test results indicated the claim amount was not independent of the body part injured. This finding suggests the claim amount varied based on the body part injured and that the body part injured is a significant factor that can be used to determine the claim amount.

Examining the chi-square residuals indicates that upper extremities, lower extremities, head and neck, and other body part categories show evidence of relationship with the claim amount. For upper, lower extremities, and other injury categories, the residuals indicate that the number of claims is more likely to be in the \$10,000 and above category and less likely to be in the below \$3,000 category. This finding suggests that injuries to upper, lower extremities, tend to be more expensive compared to “Head & trunk” and “Others” categories. Similarly, for the trunk, head and neck injuries, the residuals indicate these injuries are more likely to be inexpensive compared with the remaining categories.

Conclusions

Occupational injuries across all hazard categories of the grain handling industry have received little attention in the research literature. One reason for this is a lack of a centralized source of data to quantify the incidents. This study utilized workers’ compensation claims data to investigate patterns of workplace injuries in the grain handling industry. The first research examined if the claim amount was independent of the age of the employee. This study found that the employee age has a significant influence on the claim amount. Furthermore, employees who are 40 years of age and above have a higher likelihood of

severe injury than employees who are below 40 years of age. The second research question investigated if the claim amount was independent of the tenure of the employee. The results showed that the tenure of the employee has a significant influence on the claim amount. Also, employees with less than five years of work experience were found to be the most at risk group since the majority of injuries involved the employees in this category. The third research question examined if the claim amount were independent of the cause of injury. Results showed that the cause of injury has a significant influence on the claim amount thus suggesting that the injury costs varies based on the cause of injury. This study found that the injuries caused due to strain, slips, fall, or trips were the most significant causes of injuries. The final research question investigated if the claim amount were independent of the body part injured. Data showed that the claim amount were significantly related to the body part injured. Injuries to upper and lower extremities, trunk, and injuries to other body parts have the most influence on claim amount.

While workers' compensation data are extremely useful in injury prevention studies, the recording of information during the workers' compensation claims process could be prone to human errors, as they are collected by field agents. Also, the scope of analysis is narrowed by the information available in the dataset. However, the analysis of a large number of claims, recorded over an extended period of time, characterizes the strength and rigorousness of this study. The findings of this study will enhance the understanding of recommended areas of preventative intervention in the grain handling environment. Future work could involve analyzing the relationships between the non-cost related variables and also simultaneously investigating all the variables and their interaction effects. A multivariate model of the claim amount could also be constructed so commercial grain elevators, as well

as the worker's compensation insurance providers, can better analyze the risks contributing to occupational accidents.

References

- Agresti, A., & Finlay, B. (2008). *Statistical Methods for the Social Sciences* (4th ed.). NJ: Prentice Hall.
- Agweb. (2015). Grain Stocks: On- and Off-Farm Grain Storage Capacity. Retrieved from <http://www.agweb.com/article/jan-12-grain-stocks-on--and-off-farm-grain-storage-capacity-agwebcom-editors/>
- Asfaw, A., Pana-Cryan, R., & Rosa, R. (2011). The business cycle and the incidence of workplace injuries: Evidence from the USA. *Journal of Safety Research*, 42(1), 1-8. doi:10.1016/j.jsr.2010.10.008
- Baldwin, M. L., & Butler, R. J. (2006). Upper extremity disorders in the workplace: Costs and outcomes beyond the first return to work. *Journal of Occupational Rehabilitation*, 16(3), 303-323. doi:10.1007/s10926-006-9043-2
- Bell, J. L., & Helmkamp, J. C. (2003). Non-fatal injuries in the west Virginia logging industry: Using workers' compensation claims to assess risk from 1995 through 2001. *American Journal of Industrial Medicine*, 44(5), 502-509. doi:10.1002/ajim.10307
- Biddle, E. A., & Marsh, S. M. (2002). Comparison of two fatal occupational injury surveillance systems in the United States. *Journal of Safety Research*, 33(3), 337-354. doi:10.1016/s0022-4375(02)00030-0
- Bobick, T. G., & Myers, J. R. (1994). AGRICULTURE-RELATED SPRAIN AND STRAIN INJURIES, 1985-1987. *International Journal of Industrial Ergonomics*, 14(3), 223-232. doi:10.1016/0169-8141(94)90098-1

- Boden, L. I., & Ozonoff, A. (2008). Capture-recapture estimates of nonfatal workplace injuries and illnesses. *Annals of Epidemiology*, 18(6), 500-506.
doi:10.1016/j.annepidem.2007.11.003
- Bookman, J. A. (2012). *Describing Agricultural Injury in Ohio Using the Ohio Bureau of Workers' Compensation Database*. The Ohio State University.
- Brorsson, B. (1989). AGE AND INJURY SEVERITY. *Scandinavian Journal of Social Medicine*, 17(4), 287-290.
- Cheng, C.-W., Yao, H.-Q., & Wu, T.-C. (2013). Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. *Journal of Loss Prevention in the Process Industries*, 26(6), 1269-1278.
doi:10.1016/j.jlp.2013.07.002
- Cohen, M. A., Clark, R. E., Silverstein, B., Sjostrom, T., & Spielholz, P. (2006). Work-related deaths in Washington State, 1998-2002. *Journal of Safety Research*, 37(3), 307-319. doi:10.1016/j.jsr.2006.02.007
- Coleman, P. J., & Kerkering, J. C. (2007). Measuring mining safety with injury statistics: Lost workdays as indicators of risk. *Journal of Safety Research*, 38(5), 523-533.
doi:10.1016/j.jsr.2007.06.005
- Davis, K. G., & Kotowski, S. E. (2007). Understanding the ergonomic risk for musculoskeletal disorders in the United States agricultural sector. *American Journal of Industrial Medicine*, 50(7), 501-511. doi:10.1002/ajim.20479
- Dement, J. M., Pompeii, L. A., Ostbye, T., Epling, C., Lipscomb, H. J., James, T., . . . Thomann, W. (2004). An integrated comprehensive occupational surveillance system

- for health care workers. *American Journal of Industrial Medicine*, 45(6), 528-538.
doi:10.1002/ajim.20017
- Douphrate, D. I. (2008). *Ph.D.*, COLORADO STATE UNIVERSITY. Retrieved from
<http://gradworks.umi.com/33/21/3321273.html>
- Douphrate, D. I., Rosecrance, J. C., Reynolds, S. J., Stallones, L., & Gilkey, D. P. (2009).
Tractor-Related Injuries: An Analysis of Workers' Compensation Data. *Journal of
Agromedicine*, 14(2), 198-205. doi:10.1080/10599240902773215
- Douphrate, D. I., Rosecrance, J. C., Stallones, L., Reynolds, S. J., & Gilkey, D. P. (2009).
Livestock-Handling Injuries in Agriculture: An Analysis of Colorado Workers'
Compensation Data. *American Journal of Industrial Medicine*, 52(5), 391-407.
doi:10.1002/ajim.20686
- Douphrate, D. I., Rosecrance, J. C., & Wahl, G. (2006). Workers' compensation experience
of Colorado agriculture workers, 2000-2004. *American Journal of Industrial
Medicine*, 49(11), 900-910. doi:10.1002/ajim.20387
- Fathallah, F. A., Miller, B. J., & Miles, J. A. (2008). Low back disorders in agriculture and
the role of stooped work: scope, potential interventions, and research needs. *Journal
of agricultural safety and health*, 14(2), 221-245.
- Field, W. E., Heber, D. J., Riedel, S. M., Wettschurack, S. W., Roberts, M. J., & Grafft, L. J.
(2014). Worker hazards associated with the use of grain vacuum systems. *Journal of
agricultural safety and health*, 20(3), 147-163.
- Foley, M., Rauser, E., Rappin, C., & Bonauto, D. (2013). *Using Workers' Compensation
Data to Conduct OHS Surveillance of Temporary Workers in Washington State*.
Paper presented at the Use of Workers' Compensation Data for Occupational Safety
and Health: Proceedings from June 2012 Workshop.

- Frank Neuhauser, M., Mathur, A. K., & Pines, J. (2013). *Gender, Age, and Risk of Injury in the Workplace*. Paper presented at the Use of Workers' Compensation Data for Occupational Safety and Health: Proceedings from June 2012 Workshop.
- Freeman, S. A., Kelley, K. W., Maier, D. E., & Field, W. E. (1998). Review of entrapments in bulk agricultural materials at commercial grain facilities. *Journal of Safety Research, 29*(2), 123-134. doi:10.1016/s0022-4375(98)00008-5
- Issa, S. F., Cheng, Y.-H., & Field, W. (2016). Summary of Agricultural Confined-Space Related Cases: 1964-2013. *22*(1), 33-45. doi:10.13031/jash.22.10955
- Issa, S. F., Field, W. E., Hamm, K. E., Cheng, Y.-H., Roberts, M. J., & Riedel, S. M. (2016). Summarization of Injury and Fatality Factors Involving Children and Youth in Grain Storage and Handling Incidents.
- Janicak, C. (2010). *Safety Metrics: Tools and Techniques for Measuring Safety Performance*. Government Institutes. Lanham, MD.
- Javadi, A., & Rostami, M. A. (2007). Safety assessments of agricultural machinery in Iran. *Journal of agricultural safety and health, 13*(3), 275-284.
- Karttunen, J. P., & Rautiainen, R. H. (2011). Risk factors and prevalence of declined work ability among dairy farmers. *Journal of agricultural safety and health, 17*(3), 243-257.
- Kim, H., Dropkin, J., Spaeth, K., Smith, F., & Moline, J. (2012). Patient handling and musculoskeletal disorders among hospital workers: Analysis of 7 years of institutional workers' compensation claims data. *American Journal of Industrial Medicine, 55*(8), 683-690. doi:10.1002/ajim.22006

- Kines, P., Spangenberg, S., & Dyreborg, J. (2007). Prioritizing occupational injury prevention in the construction industry: Injury severity or absence? *Journal of Safety Research*, 38(1), 53-58. doi:10.1016/j.jsr.2006.09.002
- Laflamme, L. (1996). Age-related accident risks among assembly workers: A longitudinal study of male workers employed in the Swedish automobile industry. *Journal of Safety Research*, 27(4), 259-268. doi:10.1016/s0022-4375(96)00024-2
- Landsteiner, A. M. K., McGovern, P. M., Alexander, B. H., Lindgren, P. G., & Williams, A. N. (2015). Incidence Rates and Trend of Serious Farm-Related Injury in Minnesota, 2000–2011. *Journal of Agromedicine*, 20(4), 419-426. doi:10.1080/1059924X.2015.1075449
- Leeth, J. D. (2012). 'OSHA's role in promoting occupational safety and health.' *Foundations and Trends in Microeconomics* 7(4), 267-353.
- Leigh, J. P., Du, J., & McCurdy, S. A. (2014). An estimate of the US government's undercount of nonfatal occupational injuries and illnesses in agriculture. *Annals of Epidemiology*, 24(4), 254-259. doi:10.1016/j.annepidem.2014.01.006
- Leigh, J. P., Marcin, J. P., & Miller, T. R. (2004). An estimate of the US government's undercount of nonfatal occupational injuries. *Journal of Occupational and Environmental Medicine*, 46(1), 10-18. doi:10.1097/01.jom.0000105909.66435.53
- Lopez Arquillos, A., Rubio Romero, J. C., & Gibb, A. (2012). Analysis of construction accidents in Spain, 2003-2008. *Journal of Safety Research*, 43(5-6), 381-388. doi:10.1016/j.jsr.2012.07.005

- Lopez, M. A. C., Ritzel, D. O., Gonzalez, I. F., & Alcantara, O. J. G. (2011). Occupational accidents with ladders in Spain: Risk factors. *Journal of Safety Research*, 42(5), 391-398. doi:10.1016/j.jsr.2011.08.003
- Mariger, S. C., Grisso, R. D., Perumpral, J. V., Sorenson, A. W., Christensen, N. K., & Miller, R. L. (2009). Virginia agricultural health and safety survey. *Journal of agricultural safety and health*, 15(1), 37-47.
- Menckel, E., & Carter, N. (1985). THE DEVELOPMENT AND EVALUATION OF ACCIDENT PREVENTION ROUTINES - A CASE-STUDY. *Journal of Safety Research*, 16(2), 73-82. doi:10.1016/0022-4375(85)90009-x
- Meyers, A., Wurzelbacher, S., Bertke, S., Lampl, P. M., Robins, D., & Bell, J. (2013). *Using Workers' Compensation Data for Surveillance of Occupational Injuries and Illnesses—Ohio, 2005–2009*. Paper presented at the Use of Workers' Compensation Data for Occupational Safety and Health: Proceedings from June 2012 Workshop.
- Nanda, G., Grattan, K. M., Chu, M. T., Davis, L. K., & Lehto, M. R. (2016). Bayesian decision support for coding occupational injury data. *Journal of Safety Research*, 57, 71-82. doi:10.1016/j.jsr.2016.03.001
- NASS. (2011). *Grain Stocks*. Washington D.C.: National Agricultural Statistics Service (NASS) Retrieved from <http://usda.mannlib.cornell.edu/usda/nass/GraiStoc//2010s/2011/GraiStoc-01-12-2011.pdf>.
- NASS. (2016). *Grain Stocks*. Washington D.C.: National Agricultural Statistics Service (NASS) Retrieved from <http://usda.mannlib.cornell.edu/usda/nass/GraiStoc//2010s/2016/GraiStoc-01-12-2016.pdf>.

NIOSH. (1983). *Occupational Safety in Grain Elevators and Feed Mills*. Morgantown, WV:

National Institute for Occupational Safety and Health Retrieved from

<http://www.cdc.gov/niosh/docs/83-126/83-126.pdf>.

Nouri, J., Azadeh, A., & Fam, I. M. (2008). The evaluation of safety behaviors in a gas treatment company in Iran. *Journal of Loss Prevention in the Process Industries*, 21(3), 319-325. doi:10.1016/j.jlp.2007.11.006

Oleinick, A., & Zaidman, B. (2004). Methodologic issues in the use of workers' compensation databases for the study of work injuries with days away from work. I. Sensitivity of case ascertainment. *American Journal of Industrial Medicine*, 45(3), 260-274. doi:10.1002/ajim.10333

OSHA. (2016). Safety and Health Topics | Grain Handling. Retrieved from

<https://www.osha.gov/SLTC/grainhandling/>

Patel, K., Watanabe-Galloway, S., Rautiainen, R., Haynatzki, G., & Gofin, R. (2016).

Surveillance of non-fatal agricultural injuries among farm operators in the Central States region of the United States. University of Nebraska Medical Center. Retrieved from

<http://digitalcommons.unmc.edu/cgi/viewcontent.cgi?article=1141&context=etd>Retrieved

from <http://digitalcommons.unmc.edu/etd/141>

Reiner, A. M., Gerberich, S. G., Ryan, A. D., & Mandel, J. (2016). Large Machinery-Related Agricultural Injuries Across a Five-State Region in the Midwest. *Journal of*

Occupational and Environmental Medicine, 58(2), 154-161.

doi:10.1097/jom.0000000000000584

- Reville, R. T., Bhattacharya, J., & Weinstein, L. R. S. (2001). New methods and data sources for measuring economic consequences of workplace injuries. *American Journal of Industrial Medicine*, 40(4), 452-463. doi:10.1002/ajim.1115
- Reville, R. T., Polich, S., Seabury, S. A., & Giddens, E. (2001). *Permanent disability at private, self-insured firms*: Rand Corporation.
- Riedel, S., & Field, W. (2011). *Summary of over 800 grain storage and handling-related entrapments and suffocations documented in the US between 1970 and 2010*. Paper presented at the Presentation at XXXIV CIOST-CIGR V Conference, Vienna, Austria.
- Riedel, S. M., & Field, W. E. (2013). Summation of the frequency, severity, and primary causative factors associated with injuries and fatalities involving confined spaces in agriculture. *Journal of agricultural safety and health*, 19(2), 83-100.
- Rogers, E., & Wiatrowski, W. J. (2005). Injuries, illnesses, and fatalities among older workers. *Monthly Labor Review*, 128(10), 24-30.
- Rosenman, K. D., Kalush, A., Reilly, M. J., Gardiner, J. C., Reeves, M., & Luo, Z. W. (2006). How much work-related injury and illness is missed by the current national surveillance system? *Journal of Occupational and Environmental Medicine*, 48(4), 357-365. doi:10.1097/01.jom.0000205864.81970.63
- Rosentrater, K. A., & Williams, G. D. (2004). *Design Considerations for the Construction and Operation of Grain Elevator Facilities. Part II: Process Engineering Considerations*. Paper presented at the 2004 ASAE Annual Meeting.

- Salminen, S. (2004). Have young workers more injuries than older ones? An international literature review. *Journal of Safety Research*, 35(5), 513-521.
doi:10.1016/j.jsr.2004.08.005
- Schwatka, N. V., Butler, L. M., & Rosecrance, J. C. (2013). Age in relation to worker compensation costs in the construction industry. *American Journal of Industrial Medicine*, 56(3), 356-366. doi:10.1002/ajim.22093
- Sears, J. M., Blonar, L., & Bowman, S. M. (2014). Predicting work-related disability and medical cost outcomes: A comparison of injury severity scoring methods. *Injury-International Journal of the Care of the Injured*, 45(1), 16-22.
doi:10.1016/j.injury.2012.12.024
- Sears, J. M., Blonar, L., Bowman, S. M., Adams, D., & Silverstein, B. A. (2013). Predicting Work-Related Disability and Medical Cost Outcomes: Estimating Injury Severity Scores from Workers' Compensation Data. *Journal of Occupational Rehabilitation*, 23(1), 19-31. doi:10.1007/s10926-012-9377-x
- Sengupta, I., Reno, V. P., Burton Jr, J. F., & Baldwin, M. L. (2012). Workers' compensation: Benefits, coverage, and costs, 2010. *National Academy of Social Insurance*.
- Sharpe, D. (2015). Your Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research & Evaluation*, 20.
- Smith, P., Hogg-Johnson, S., Mustard, C., Chen, C., & Tompa, E. (2012). Comparing the risk factors associated with serious versus and less serious work-related injuries in ontario between 1991 and 2006. *American Journal of Industrial Medicine*, 55(1), 84-91.
doi:10.1002/ajim.21000

Snyder, K. A., & Bobick, T. G. (1995). Safe grain and silage handling. *NIOSH-Issued Publications*.

Sprince, N. L., Zwerling, C., Lynch, C. F., Whitten, P. S., Thu, K., Logsdan-Sackett, N., . . .

Alavanja, M. C. R. (2003). Risk factors for agricultural injury: A case-control analysis of Iowa farmers in the Agricultural Health Study. *Journal of agricultural safety and health*, 9(1), 5.

Suarez-Cebador, M., Carlos Rubio-Romero, J., & Lopez-Arquillos, A. (2014). Severity of electrical accidents in the construction industry in Spain. *Journal of Safety Research*, 48, 63-70. doi:10.1016/j.jsr.2013.12.002

Takahashi, A., & Miura, T. (2016). At What Age Is the Occupational Accident Risk High? Analysis of the Occurrence Rate of Occupational Accidents by Age. In *Advances in Safety Management and Human Factors* (pp. 3-10): Springer.

Utterback, D., Meyers, A., & Wurzelbacher, S. (2014). Workers' Compensation Insurance: A Primer for Public Health. *Department of Health and Human Services: National Institute for Occupational Safety and Health*.

Utterback, D. F., & Schnorr, T. M. (2010). Use of workers' compensation data for occupational injury & illness prevention. *National Institute for occupational safety and health, US Department of Labour*, 2010-2152.

Utterback, D. F., Schnorr, T. M., Silverstein, B. A., Spieler, E. A., Leamon, T. B., & Amick, B. C., III. (2012). Occupational Health and Safety Surveillance and Research Using Workers' Compensation Data. *Journal of Occupational and Environmental Medicine*, 54(2), 171-176. doi:10.1097/JOM.0b013e31823c14cb

- Van Fleet , E., Frank, O., & Rosenbeck, J. (2013). *A Guide to Safety and Health in Feed and Grain Mills*. Raleigh, NC: North Carolina Department of Labor Retrieved from <http://www.nclabor.com/osha/etta/indguide/ig29.pdf>.
- van Tulder, M., Malmivaara, A., & Koes, B. (2007). Repetitive strain injury. *Lancet*, 369(9575), 1815-1822. doi:10.1016/s0140-6736(07)60820-4
- Verma, A., Das Khan, S., Maiti, J., & Krishna, O. B. (2014). Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports. *Safety Science*, 70, 89-98. doi:10.1016/j.ssci.2014.05.0070925-7535
- Vinodkumar, M. N., & Bhasi, M. (2009). Safety climate factors and its relationship with accidents and personal attributes in the chemical industry. *Safety Science*, 47(5), 659-667. doi:10.1016/j.ssci.2008.09.004
- Waehrer, G. M., Dong, X. W. S., Miller, T., Haile, E., & Men, Y. (2007). Costs of occupational injuries in construction in the United States. *Accident Analysis and Prevention*, 39(6), 1258-1266. doi:10.1016/j.aap.2007.03.012
- Williams, G. D., & Rosentrater, K. A. (2004). *Design Considerations for the Construction and Operation of Grain Elevator Facilities. Part I: Planning, Structural, and Life Safety Considerations*. Paper presented at the 2004 ASAE Annual Meeting.
- Zhou, C., & Roseman, J. M. (1994). AGRICULTURAL INJURIES AMONG A POPULATION-BASED SAMPLE OF FARM OPERATORS IN ALABAMA. *American Journal of Industrial Medicine*, 25(3), 385-402. doi:10.1002/ajim.4700250307

CHAPTER 4. PREDICTING DAYS AWAY FROM WORK: COMPARING MODEL PERFORMANCE

Manuscript to be submitted to the *Journal of Safety Research*

Sai K. Ramaswamy^a, Gretchen A. Mosher^a

^aDepartment of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA

Abstract

The number of days away from work (DAFW) is an important indicator of the severity of a work-related injury. Severity is important because injuries that are severe have higher economic and social costs than those that are less severe. Therefore, predicting factors that influence the DAFW can help supervisors and safety managers evaluate injury risks, identify problem areas, and make decisions to mitigate the impacts of severe work-related injuries. Very little published research has modeled the DAFW using information recorded through workers' compensation reporting. This study built and compared the predictive performance of two models – linear regression and decision tree - using data from a workers' compensation claims dataset (N=10,802). Linear regression and decision tree models were built in the SAS Enterprise Miner application. The Root Mean Squared Error (RMSE), a measure of precision, was used to compare the two model. Results showed that the linear regression model with two-way interactions had a lower RMSE or higher precision than the decision tree model. Based on these findings and the fact that regression models are straightforward and most widely used, this study suggests the linear regression is a viable method for predicting the DAFW. The findings of this study will help safety professionals and insurance companies better understand the effects of employee and injury characteristics such as age, tenure, nature of injury and cause of injury on injury severity. A better

understanding of severity of injuries can facilitate proper planning and development of measures for mitigating occupational injuries.

Keywords: Days away from work, decision tree, linear regression, root mean square error, occupational injuries

Introduction

The number of days away from work (DAFW), also referred as “days lost from work” or “lost work days” is an important indicator of severity of occupational injuries and illness (Azadeh-Fard, Schuh, Rashedi, & Camelio, 2015; BLS, 2016c; Khanzode, Maiti, & Ray, 2012; Marucci-Wellman et al., 2015). A measure of severity is important because, workers suffering severe injuries require more DAFW to recuperate than do workers with less severe injuries (Fordyce et al., 2016; Micheli & Cagno, 2010). Therefore, determining the number of DAFW can estimate the economic impacts of occupational injuries and assist organizations in determining spending priorities for injury intervention efforts (Galizzi & Boden, 2003; Krause, Frank, Dasinger, Sullivan, & Sinclair, 2001; Sears, Bowman, & Hogg-Johnson, 2014). The purpose of this study was to construct two models to predict DAFW using the data recorded during the workers’ compensation claims process.

Previous studies that investigated DAFW using a workers’ compensation data have examined the direct relationship of DAFW with variables such as injured employee’s age, tenure, nature of injury and cause of injury. For example, Fordyce et al. (2016) used the DAFW information from the occupational health and safety database to identify the most at-risk job groups and causes of severe injuries in the electric power industry. Similarly, Onder (2013) analyzed past incidents database from a Turkish coal mine and identified the job group, age group, and the body part associated with the highest probability of exposure to

severe occupational injuries. Margolis (2010) investigated the Mine Safety and Health Administration's database on injuries and illnesses to examine the influence of age and experience of injured employees to the severity of injury using the DAFW.

In observational studies such as these, evaluating the relationship between two variables by controlling the effect of other variables is not possible as in the case of experimental or intervention studies (Rosenbaum, 2002; von Elm et al., 2007). Therefore, multivariate models are required to allow investigators to take into account not just the direct relationship but also interactions among all variables in the study (Cohen, Cohen, West, & Aiken, 2013; Ramsey & Schafer, 2012). A review of the literature showed very little published research had utilized multivariate models to characterize DAFW by leveraging information recorded in injury databases such as the workers' compensation claims.

The focus of this study was to build and compare two multivariate models for modeling the DAFW resulting from occupational injuries. A decision tree model and a multiple linear regression model were built using the information in a workers' compensation claims database obtained from a Midwest-based insurance company. The DAFW was the dependent variable, and other information recorded in the workers' compensation process such as employee age, employee tenure, nature of injury, cause of injury, and body part injured were used as the predictor variables. The root means square error (RMSE) was calculated to compare the two models. The RMSE has been identified as a primary factor in measuring the precision of the model. Therefore, the predicted DAFW values of a model with lowest RMSE value are closer to the actual value than the values predicted by a model with higher RMSE value (Chang, Laird, Mausbach, & Hurburgh, 2001; Honn, Satterfield,

McCauley, Caldwell, & Dongen, 2016; Lee, Chung, & Kang, 2016; Singh, Sachdeva, & Pal, 2016).

Background

Occupational injuries in the U.S. continues to be a major concern in several industries, specifically the process industries (Boden, O'Leary, Applebaum, & Tripodis, 2016; Lander, Nielsen, & Lauritsen, 2016; Mabila, Gracia, Cohen, Almberg, & Friedman, 2015; Wurzelbacher et al., 2016). According to the U.S. Bureau of Labor Statistics (BLS, 2016b), the 2015 incident rate of non-fatal occupational injuries for the manufacturing industry was 26% higher than the national average (BLS, 2016b). Similarly, in food manufacturing, the incident rate was 56% higher, and in wood products and animal production, the incident rate of non-fatal occupational injuries was more than twice the national average (BLS, 2016a).

The cost of these injuries are a burden not just on the injured employee, but their families, and society (Leigh & Marcin, 2012). In the U.S., the cost of occupational injuries and illnesses is estimated to be over \$200 billion annually (Leigh, 2011; National Safety Council, 2016). According to Marucci-Wellman et al. (2015) in the U.S., nearly \$1 billion is spent each week to cover the direct costs of severe work-related injuries, including medical costs and wage replacement for injured employees. Such high costs can pose a serious threat to the profitability of any business enterprise. Therefore, an improved understanding of injury severity, measured in this study using the DAFW, and the factors affecting the injury severity can help estimate the risk of injuries needed for developing measures for preventing reoccurrence of occupational injuries.

Preventing occupational injuries requires learning from past incidents (Abdolhamidzadeh, Abbasi, Rashtchian, & Abbasi, 2011; Pasman, 2009). However, obtaining detailed records of past incidents is a challenge acknowledged by investigators across several industrial sectors including agribusiness (Calvo Olivares, Rivera, & Nunez McLeod, 2014, 2015; Dong, Largay, Wang, & Windau, 2014; Riedel & Field, 2011). Workers' compensation (WC) claims records can partially address these informational gaps needed to characterize and prevent occupational injuries (Utterback, Meyers, & Wurzelbacher, 2014; Wurzelbacher et al., 2016). WC claims data contain information such as medical and indemnity payments, industry type, occupation, nature and cause of injury, and employee demographic data (Nestoriak & Pierce, 2009; Utterback et al., 2012; Wurzelbacher et al., 2016). WC claims data are preferred in injury surveillance and prevention because they contain a large amount of detail in the records (Oleinick & Zaidman, 2004).

To distil such large amounts of data into a few parameters characterizing the phenomenon under study requires a systematic technique (Gnanadesikan, 2011; Springmeyer, Blattner, & Max, 1992). A technique most widely used for analyzing data in injury and incident investigation studies is linear regression. (Tso & Yau, 2007). Therefore, to investigate the large workers' compensation claims data used in this study and characterize the DAFW, a multiple linear regression model was employed. An important limitation of regression analysis is that they require assumptions such as linear relationship between dependent and independent variables, and normal distribution of residuals to be satisfied to adequately model the underlying data (Fragiadakis, Tsoukalas, & Papazoglou, 2014).

Data mining techniques such as Classification and Regression Trees (CART) can overcome the limitations of linear regression and are widely employed to build predictive

models (Chang & Wang, 2006). These non-parametric techniques are rapidly evolving as an effective method of analysis to investigate large datasets such as workers' compensation claims used in this study (Fayyad, PiatetskyShapiro, & Smyth, 1996; Mistikoglu et al., 2015). Data mining is a multidisciplinary field that encompasses classical statistical techniques and new computational techniques, such as decision trees and association rules (Anand et al., 2006). The CART or decision trees are the most widely used data mining technique to uncover hidden patterns in large data sets across various disciplines (Cheng, Leu, Cheng, Wu, & Lin, 2012). However, the application of decision tree techniques has not been commonly used with occupational injury data (Mistikoglu et al., 2015; Nenonen, 2013).

Previous studies that have investigated the days away from work (DAFW) of injured employees focused on a single industry. For example, Onder (2013) evaluated occupational injuries with lost days among workers in the coal mining industry, while Blanch, Torrelles, Aluja, and Salinas (2009) examined age and DAFW for employees at a plastic film manufacturing facility. Fordyce et al. (2016) analyzed fatal and non-fatal injury severity factors using DFWW for employees in the electric power industry. The present study examined DAFW for occupational injuries in bulk commodity handling, food manufacturing, grocery and retail stores. Data mining of historical injury incidents helps identify potential problem areas and provide insights into the causes of occupational injuries (Coleman & Kerkerling, 2007; Nenonen, 2013; Tsioras, Rottensteiner, & Stampfer, 2014).

Methods and Data

The dataset used in this study was obtained from a Midwest-based insurance company that focuses on agribusiness clients. The dataset consisted of approximately 10,800 claims reported from January 2008 to March 2016. The insurance company collected the data

as part of the workers' compensation claim process with inputs provided by the injured employees, their employers, and other parties involved in the claims process. The insurance company personnel vetted all information recorded in the dataset. This collection and review of information from multiple stakeholders is a key reason why workers' compensation claims record are an excellent source of information for investigating occupational injury patterns (Dement et al., 2004; Janicak, 2010; Kim, Dropkin, Spaeth, Smith, & Moline, 2012; Reville, Bhattacharya, & Weinstein, 2001).

Table 1 lists all the variables used in the study. The age of the employee was calculated as the difference between the date of birth and the injury date. Similarly, the tenure of the employee was calculated as the difference between the date of hire and injury date. The DAFW was calculated as the difference between the injury date, and the date returned to work. In this study, DAFW was the dependent variable. Age of the injured employee, employee tenure, employee gender, nature of injury, cause of injury, injured body part, market, and class description were the independent variables used in the model.

Linear regression models are the most commonly used predictive modeling techniques in incident and injury research (Khanzode, Maiti, & Ray, 2012; Lord & Mannering, 2010; Mannering & Bhat, 2014). The rapid growth in the volume of data in recent years has led to the popularity of data mining techniques, and decision trees are one of the most widely used methods for predictive modeling (Chang & Chen, 2005; Liao, Chu, & Hsiao, 2012). In this study, linear regression and decision tree modeling techniques were used to build predictive models. Both the models were developed and evaluated using the SAS Enterprise Miner software application. The use of SAS Enterprise Miner for building and evaluating predictive models has been documented in previous research by Tso and Yau

(2007) when predicting electricity consumption and by Yap, Ong, and Husain (2011) in evaluating creditworthiness. These studies suggest that the algorithms and statistical techniques used by SAS Enterprise Miner for building predictive models have been vetted and consequently, have high reliability.

Table 1: *List of variables*

#	Column Name	Description
1	Claim	Unique identifier for each claim record
2	Effective Year	Filing year of the claim
3	Market	Type of business (such as bulk commodity, food manufacturing)
4	Class description	Job class description
5	Date of Birth	Date of birth of injured worker
6	Date of Hire	Date on which the present company hired the injured worker
7	Gender	Injured employee's gender
8	Injury Date	Date on which the injury occurred
9	Date returned to work	Date on which the injured employee returned to work
11	Body Part	Part of body injured
12	Cause of Injury	Main cause of injury. For example: "Dust, gasses or fumes inhalation", "Foreign matter in eyes", "Chemical exposure".
13	Nature of Injury	Describes the type of injury such as Fracture, Strain, Contusion

In SAS Enterprise Miner, a linear model was build using the stepwise option. In the stepwise method, the model is built by adding one independent variable at a time, satisfying the entry significance level and eliminating any variable that does not meet the stay significance level. The entry and stay significance level for a variable in the model were measured using an F-statistic and were set to 0.05 (Neerchal, Morel, Huang, & Moluh, 2014). The two-factor interactions option in the SAS Enterprise Miner was selected so both the main effects and the two-way interaction effect of the variables could be modeled.

Fitting a linear regression model to a dataset requires several assumptions about the underlying data (Agresti & Finlay, 2008). These assumptions include homoscedasticity or even variability between variables (Long & Ervin, 2000), low correlation among the predictor variables or lack of multicollinearity (Friedman & Wall, 2005) and the absence of outliers in the dataset (Mark & Workman, 2007). When the assumptions of regression are violated the predictive model will result in a poor fit or produce large error values (Chang & Chen, 2005). Therefore in this study, a decision tree model that does not require underlying assumptions about the dataset (Chang & Wang, 2006), was used to compare the error values generated by the regression model in predicting the DAFW.

Decision trees are one of the most widely used data mining techniques (Liao et al., 2012). This technique was first proposed by Breiman et al. (1984) to recursively partition a set of data into homogeneous groups and displayed graphically in an inverted tree-like structure. This representation of information in an intuitive and easy to visualize format is a reason for the popularity of decision trees in data analysis (Elith, Leathwick, & Hastie, 2008). Decision trees can be used to characterize both numeric as well as categorical dependent variables (Loh, 2011). If the dependent variable is categorical, then the model is called a classification tree, and if the dependent variable is numeric, then decision tree is known as a regression tree model (Razi & Athappilly, 2005). Since the DAFW was numeric and continuous variable, regression tree was used in this study.

There are several advantages of using decision trees for investigating large datasets such as the workers' compensation claims used in this study (Savolainen, Mannering, Lord, & Quddus, 2011). Decision trees are a non-parametric method to use with workers' compensation claims, data that often violates the assumptions of regression (Breiman, 2001;

Chang & Chen, 2005). Another advantage of decision trees is that they can use any type of predictor variables, including numeric, binary or categorical to characterize a dependent variable (Elith et al., 2008; Harb, Yan, Radwan, & Su, 2009). Since decision trees have a hierarchical structure where the response of one input variable depends on the values of inputs higher in the tree, interactions between predictor variables are automatically modeled (Elith et al., 2008). Finally, decision tree models, while being straightforward to interpret, also have prediction accuracy comparable to those of statistical techniques such as regression (Chang & Wang, 2006; Lim, Loh, & Shih, 2000; Meleddu & Pulina, 2016).

Building decision trees involve three major phases: i) tree building, ii) tree pruning, and iii) testing (L. Breiman, 2001; Leo Breiman et al., 1984). The decision tree model is constructed starting with the complete data, and partitioning the data using a set of rules and one predictor variable at a time to create two or more mutually exclusive groups (Cheng et al., 2012). At each partition, the process is repeated, and a large tree is grown until each of partitioned groups are as homogenous as possible so the outcomes can be predicted accurately (Nenonen, 2013). Once the tree building is complete, the tree is then pruned by removing branches that do not contribute significantly in characterizing the dependent variable (Strobl, Malley, & Tutz, 2009). Tree pruning ensures that the decision tree model is kept as simple as possible without significant loss of prediction accuracy (De'ath & Fabricius, 2000). Finally, the pruned decision tree model is tested either using a subset of the existing data or a whole new set of data (Loh, 2011). This process of tree building, tree pruning, and testing is repeated until the model with the least prediction error is found (Leo Breiman et al., 1984; Loh, 2011; Strobl et al., 2009).

The SAS Enterprise Miner provides two options for the method of selecting and evaluating splitting rules for a continuous dependent variable. In this study, the F-test option was selected with a significance level of 0.2 a default in the SAS Enterprise Miner. The depth of the tree that specifies the number of levels in the model was set to 5 and the minimum number of observations required for a split search was set to 50 to simplify. All other parameters were left unchanged or “default” option in the SAS Enterprise Miner.

The root mean squared error (RMSE) was used as a measure to compare the error generated by the two models. The RMSE is the measure of the difference between the predicted value generated by the model and the actual value. The general representation of the RMSE is given by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

where \hat{y}_i is the predicted value obtained from the model, y_i is the actual value and n is the number of observations. A small value of RMSE means the predicted values are close to the actual values of the dependent variable indicating less error thereby demonstrating the performance of a model (Lee et al., 2016). The RMSE has been a widely used criteria to compare predictive models in safety studies (Amnieh, Mozdianfard, & Siamaki, 2010; Karacan & Goodman, 2012; Singh, Sachdeva, & Pal, 2016) involving copper mines, coalmines, and road incidents.

Results and Discussion

In this study, workers’ compensation claims from 2008 to 2016 were used to model the DAFW. The range of DAFW values in the dataset was 0 to 365 days. Descriptive

analysis of the DAFW for the 10,800 claims showed that approximately 24% of the claims had no DAFW. Nearly 51% of the claims had 1 to 10 DAFW, 18% of the claims had 11-100 DAFW while 7% of the claims had 101-365 DAFW. The variables i) age, ii) tenure, iii) gender, iv) market, v) class description, vi) nature of injury, vii) cause of injury and viii) body part injured collected during the claim process were used to model the DAFW using linear regression and decision tree modeling techniques.

In the stepwise process, the regression model is built by adding one variable at a time. Starting with a model with just the intercept, one variable is added or removed at a time based on the entry and stay significance level criteria of 0.05. The results of the regression modeling of DAFW are shown in Table 2. In the final model generated by the stepwise process, only the age of employee and nature of injury variables were significant out of the eight independent variables entered into the model. Likewise, as shown in Table 2, only four interaction variables were significant out of the 28 possible two-way interactions. The regression model developed in this study for predicting the DAFW was significant at a 95% confidence level. This finding suggests that a linear model can characterize the relationship between the DAFW and the eight independent variables. The coefficient of determination (R^2) that describes the proportion of the variance in the DAFW explained by the regression model was calculated as 0.19. Six independent variables were not significant to the model, including tenure, gender, market, class description, cause of injury and body part group.

Table 2: Stepwise regression model results

Effect	Type of effect	Degrees of freedom	Sum of Squares	F-Value	p-value
Age	Main	1	256662.17	124.18	<0.005
Nature of injury	Main	12	101176.29	4.08	<0.005
Body part * Cause of injury	Interaction	96	333803.30	1.68	<0.005
Body part * Nature of injury	Interaction	68	546172.23	3.89	<0.005
Class description*Market	Interaction	275	948874.77	1.67	<0.005
Cause of injury* Nature of injury	Interaction	164	488407.35	1.44	<0.005

$R^2 = 0.19$; $N = 10802$

The regression coefficient for the variable age from the regression model indicated that the injured employee's age has a positive and significant relationship with the DAFW. This finding indicates that older employees are likely to have a greater number of DAFW than younger employees for the same type of injury. This finding is consistent with previous studies in the research literature. Salminen (2004) and Takahashi & Miura (2016) also documented a significant relationship between age and injury risks, finding that older workers had more severe injuries, resulting in more days off work. Furthermore, Onder (2013) and Margolis (2010), who investigated DAFW in mining industry, also found that the age of the worker has a positive and statistically significant relationship with the DAFW.

In addition to the age of the injured employee, the nature of injury was the only other variable that was significant in the regression model. The nature of injury variable describes the type of occupational injury, for example, strain, burn, laceration or contusion. The analysis of the likelihood estimates showed that fracture and dislocation injuries are likely to have a higher DAFW than any other type of injury. Similarly, laceration, inflammation and burn injuries are more likely to have a lower number of

DAFW. This finding is consistent with the studies by Fordyce et al. (2016), Baril, Berthelette, and Massicotte (2003), and Krause et al (2001), who also reported a significant relationship between nature of injury and DAFW.

It is noteworthy that in this study the effect of the cause of injury and the body part injured on the DAFW were not significant in the regression model. Previous research by Baril et al.(2003), Fordyce et al.(2016) and Tsioras, Rottensteiner, & Stampfer (2014) have reported that the cause of injury and the body part injured have a significant relationship with the DAFW. This difference between the findings of this study and previous studies maybe due to fact that in the agribusiness industry, the type of injury is not independent of the cause of injury and body part injured. For example, injuries due to slips, trips, and falls may have varying DAFW depending on the type of injury (fracture, sprain or inflammation). Likewise injuries to the head and neck due to slips, trips, and falls may have a varying DAFW than the same type of injury to lower extremities. The significant interaction effect of body part injured with the nature of injury and the significant interaction effect of cause of injury with the body part injured provide evidence that the nature of injury, cause of injury and body part injured are not independent in the agribusiness industries. Finally, the regression model showed the interaction of class description and market were significant. This finding suggests that some jobs in a particular industry could have a significantly different number of DAFW than the same work in another industry.

According to the decision tree model, the overall average DAFW for all claims investigated in this study was calculated as 20.2 days. This number is much higher than

the average values of 8 and 9 DAFW as reported by the BLS in 2014 and 2015, respectively (BLS, 2015). This difference in average DAFW between agribusiness industries and other industries supports the argument made by previous researchers that injuries in the agribusinesses such as grain elevators tend to be more severe than those in other industries when measured using DAFW (Reiner, Gerberich, Ryan, & Mandel, 2016).

In the decision tree model, the effect of age of the injured employee and the nature of the injury on the DAFW were significant, just as in the regression model. Likewise, the variables tenure, cause of injury, class description, market and gender do not have a significant impact on the DAFW. A notable observation from the tree model was the DAFW for strain injuries was significantly higher than the DAFW for burn, lacerations, puncture and foreign body injuries. Furthermore, fracture injuries tended to have some of the highest DAFW while lacerations and foreign body tended to have the lowest DAFW. As in the regression model, the interaction effect of nature of injury and body part were significant in the decision tree model. For example, strain injuries involving either an upper extremity or a lower extremity tended to have a higher DAFW than strain injuries to any other body part. Finally, the DAFW for younger employees were significantly lower than older employees. For example, employees 44.5 years and older suffering from strain injury had an average DAFW of 59 days while the DAFW for employees below 44.5 years age suffering from the same injury type was only 25 days. Similarly, the average DAFW for employees below 34.5 of age suffering from burns and

puncture injuries was 5 days, while the DAFW was 13 days for employees 34.5 years or older suffering from the same type of injury.

The two models were compared using the root mean square error (RMSE) values. Models with small RMSE have a higher accuracy than models with large RMSE values (Willmott et al., 1985). The values of the RMSE for the two models are shown in Table 3. Results showed the multiple linear regression model built using the step-wise method yielded a lower RMSE value (44.14) than the decision tree model (46.19). Since regression models are straightforward, robust and the most widely used prediction models, this study suggests that the linear regression are an appropriate approach to model the DAFW.

Table 3: RMSE values of regression and tree models for all claims

Model type	Sum of squared error (SSE)	N	Mean squared error (MSE)	Root mean squared error (RMSE)
Stepwise Regression with two-way interactions	21049062	10802	1948.6	44.14
Decision tree model	23047462	10802	2133.6	46.19

To ensure the findings of this study were robust, the model building and evaluation process was repeated by removing all claims with no DAFW. Additionally, a second regression model was built with no interaction terms to evaluate the merit of including the two-way interactions terms in the model. The model comparison results for this extended analysis are shown in Table 4. Results showed that the stepwise regression model with the two-way interactions had the lowest RMSE compared to the other models. It is noteworthy that the regression model even without interaction terms had a lower error value compared to the decision tree model. This finding provides further evidence that for the data in this study

the regression models are an appropriate method to model the DAFW as compared to the decision tree model.

Table 4: RMSE values of regression and tree models for claims where DAFW >0

Model type	Sum of squared error (SSE)	N	Mean squared error (MSE)	Root mean squared error (RMSE)
Stepwise Regression with two-way interactions	19303881	8160	2365.7	48.6
Stepwise Regression with no interactions	21468529	8160	2630.9	51.3
Decision tree model	21641214	8160	2652.1	51.5

Conclusions

The objective of this study was to build and compare predictive models for determining the DAFW using the information recorded during the workers' compensation claims process. The two techniques used for building the predictive models were multiple regression using stepwise variable selection and the decision tree model. Regression models are the most widely used method for predictive analysis. However, for regression models to predict the dependent variable with high degree of accuracy, the underlying data must satisfy certain assumptions. Violating these assumptions can result in erroneous outcome. Data mining techniques such as decision trees are increasingly used for predictive analysis because they are non-parametric methods that do not require underlying assumptions to be satisfied. This study found that the linear regression and the decision tree model could adequately model the DAFW.

While previous studies that investigated the DAFW focused on a single type of industry, this study used data from a wide range of process-oriented businesses to model the DAFW. Injury and incident analysis in businesses such as biofuels, grain handling and food

processing has been sparsely documented. This study augments the existing body of research literature on occupational injuries in hazardous industries such as biofuels and grain handling. Obtaining injury data for analysis in these industries continues to be a major challenge, but the use of workers' compensation claims data provide a good source of injury surveillance and analysis.

Several limitations are noted for this study. Firstly, a major limitation of this study was that the scope of analysis was narrowed by the information available in the dataset. Secondly, while workers' compensation data are extremely useful in injury prevention studies, the recording of information during the workers' compensation claims process could contain errors. Finally, it must be noted that the investigators did not have any control on the data collection process. Future work can augment these models by including other information such as occupation of the injured employee, their level of education, pay rate, or any other variable recorded in the injury reporting process. Also, future studies can evaluate the performance of other prediction modeling techniques such as random forests, and artificial neural network to model the DAFW in addition to the techniques used in this study.

References

- Abdolhamidzadeh, B., Abbasi, T., Rashtchian, D., & Abbasi, S. A. (2011). Domino effect in process-industry accidents - An inventory of past events and identification of some patterns. *Journal of Loss Prevention in the Process Industries*, 24(5), 575-593.
doi:10.1016/j.jlp.2010.06.013
- Agresti, A., & Finlay, B. (2008). *Statistical Methods for the Social Sciences* (4th ed.). NJ: Prentice Hall.

- Anand, S., Keren, N., Tretter, M. J., Wang, Y. J., O'Connor, T. M., & Mannan, M. S. (2006). Harnessing data mining to explore incident databases. *Journal of Hazardous Materials*, 130(1-2), 33-41. doi:10.1016/j.jhazmat.2005.07.046
- Baril, R., Berthelette, D., & Massicotte, P. (2003). Early return to work of injured workers: multidimensional patterns of individual and organizational factors. *Safety Science*, 41(4), 277-300. doi:10.1016/s0925-7535(02)00039-5
- BLS. (2015). Nonfatal Occupational Injuries and Illnesses: Cases with days away from work [Press release]. Retrieved from <http://www.bls.gov/iif/oshwc/osh/case/osch0055.pdf>
- BLS. (2016a). *Case and Demographic Characteristics for Work-related Injuries and Illnesses Involving Days Away From Work*. Retrieved from <http://www.bls.gov/iif/oshcdnew2015.htm#15g>.
- BLS. (2016b). Employer-Reported Workplace Injury and Illness Summary [Press release]. Retrieved from <http://www.bls.gov/news.release/osh.nr0.htm>
- Boden, L. I., O'Leary, P. K., Applebaum, K. M., & Tripodis, Y. (2016). The impact of non-fatal workplace injuries and illnesses on mortality. *American journal of industrial medicine*, 59(12), 1061-1069.
- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199-215. doi:10.1214/ss/1009213726
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.
- Calvo Olivares, R. D., Rivera, S. S., & Nunez Mc Leod, J. E. (2014). Database for accidents and incidents in the biodiesel industry. *Journal of Loss Prevention in the Process Industries*, 29, 245-261. doi:10.1016/j.jlp.2014.03.010

- Calvo Olivares, R. D., Rivera, S. S., & Nunez McLeod, J. E. (2015). Database for accidents and incidents in the fuel ethanol industry. *Journal of Loss Prevention in the Process Industries*, 38, 276-297. doi:10.1016/j.jlp.2015.10.008
- Chang, C. W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65(2), 480-490.
- Chang, L.-Y., & Wang, H.-W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis and Prevention*, 38(5), 1019-1027. doi:10.1016/j.aap.2006.04.009
- Chang, L. Y., & Chen, W. C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 36(4), 365-375.
doi:10.1016/j.jsr.2005.06.013
- Cheng, C. W., Leu, S. S., Cheng, Y. M., Wu, T. C., & Lin, C. C. (2012). Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. *Accident Analysis and Prevention*, 48, 214-222.
doi:10.1016/j.aap.2011.04.014
- De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81(11), 3178-3192.
doi:10.1890/0012-9658(2000)081[3178:cartap]2.0.co;2
- Dement, J. M., Pompeii, L. A., Ostbye, T., Epling, C., Lipscomb, H. J., James, T., . . . Thomann, W. (2004). An integrated comprehensive occupational surveillance system for health care workers. *American Journal of Industrial Medicine*, 45(6), 528-538.
doi:10.1002/ajim.20017

- Dong, X., Largay, J. A., Wang, X., & Windau, J. A. (2014). Fatalities in the construction industry: findings from a revision of the BLS Occupational Injury and Illness Classification System. *Monthly Lab. Rev.*, *137*, 1.
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802-813. doi:10.1111/j.1365-2656.2008.01390.x
- Fayyad, U., PiatetskyShapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *Ai Magazine*, *17*(3), 37-54.
- Fordyce, T. A., Leonhard, M. J., Watson, H. N., Mezei, G., Vergara, X. P., & Krishen, L. (2016). An analysis of fatal and non-fatal injuries and injury severity factors among electric power industry workers. *American Journal of Industrial Medicine*, *59*(11), 948-958. doi:10.1002/ajim.22621
- Fragiadakis, N. G., Tsoukalas, V. D., & Papazoglou, V. J. (2014). An adaptive neuro-fuzzy inference system (anfis) model for assessing occupational risk in the shipbuilding industry. *Safety Science*, *63*, 226-235. doi:10.1016/j.ssci.2013.11.013
- Gnanadesikan, R. (2011). *Methods for statistical data analysis of multivariate observations* (Vol. 321): John Wiley & Sons.
- Honn, K. A., Satterfield, B. C., McCauley, P., Caldwell, J. L., & Dongen, H. P. A. (2016). Fatiguing effect of multiple take-offs and landings in regional airline operations. *Accident Analysis and Prevention*, *86*, 199-208. doi:10.1016/j.aap.2015.10.005
- Janicak, C. (2010). *Safety Metrics: Tools and Techniques for Measuring Safety Performance* Government Institutes. Lanham, MD.

- Khanzode, V. V., Maiti, J., & Ray, P. K. (2012). Occupational injury and accident research: A comprehensive review. *Safety Science*, *50*(5), 1355-1367.
doi:10.1016/j.ssci.2011.12.015
- Kim, H., Dropkin, J., Spaeth, K., Smith, F., & Moline, J. (2012). Patient handling and musculoskeletal disorders among hospital workers: Analysis of 7 years of institutional workers' compensation claims data. *American Journal of Industrial Medicine*, *55*(8), 683-690. doi:10.1002/ajim.22006
- Krause, N., Frank, J. W., Dasinger, L. K., Sullivan, T. J., & Sinclair, S. J. (2001). Determinants of duration of disability and return-to-work after work-related injury and illness: Challenges for future research. *American Journal of Industrial Medicine*, *40*(4), 464-484. doi:10.1002/ajim.1116
- Lander, F., Nielsen, K. J., & Lauritsen, J. (2016). Work injury trends during the last three decades in the construction industry. *Safety Science*, *85*, 60-66.
doi:10.1016/j.ssci.2015.10.013
- Lee, J., Chung, K., & Kang, S. (2016). Evaluating and addressing the effects of regression to the mean phenomenon in estimating collision frequencies on urban high collision concentration locations. *Accident Analysis & Prevention*, *97*, 49-56.
- Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*, *39*(12), 11303-11311. doi:10.1016/j.eswa.2012.02.063
- Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, *1*(1), 14-23. doi:10.1002/widm.8

- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part a-Policy and Practice*, 44(5), 291-305. doi:10.1016/j.tra.2010.02.001
- Mabila, S. L., Gracia, G., Cohen, R., Almberg, K., & Friedman, L. S. (2015). Injury and illness data for Illinois mining industry employees, 1990 to 2012. *Journal of Occupational and Environmental Medicine*, 57(12), 1305-1310.
- Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, 1-22.
- Margolis, K. A. (2010). Underground coal mining injury: A look at how age and experience relate to days lost from work following an injury. *Safety Science*, 48(4), 417-421. doi:10.1016/j.ssci.2009.12.015
- Mistikoglu, G., Gerek, I. H., Erdis, E., Usmen, P. E. M., Cakan, H., & Kazan, E. E. (2015). Decision tree analysis of construction fall accidents involving roofers. *Expert Systems with Applications*, 42(4), 2256-2263. doi:10.1016/j.eswa.2014.10.009
- Nenonen, N. (2013). Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database. *Applied Ergonomics*, 44(2), 215-224. doi:10.1016/j.apergo.2012.07.001
- Nestoriak, N., & Pierce, B. (2009). Comparing Workers' Compensation claims with establishments' responses to the SOII. *Monthly Labor Review*, 132(5), 57-64.
- Oleinick, A., & Zaidman, B. (2004). Methodologic issues in the use of workers' compensation databases for the study of work injuries with days away from work. I.

- Sensitivity of case ascertainment. *American Journal of Industrial Medicine*, 45(3), 260-274. doi:10.1002/ajim.10333
- Onder, S. (2013). Evaluation of occupational injuries with lost days among opencast coal mine workers through logistic regression models. *Safety Science*, 59, 86-92. doi:10.1016/j.ssci.2013.05.002
- Pasman, H. J. (2009). Learning from the past and knowledge management: Are we making progress? *Journal of Loss Prevention in the Process Industries*, 22(6), 672-679. doi:10.1016/j.jlp.2008.07.010
- Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, 29(1), 65-74. doi:10.1016/j.eswa.2005.01.006
- Reiner, A. M., Gerberich, S. G., Ryan, A. D., & Mandel, J. (2016). Large Machinery-Related Agricultural Injuries Across a Five-State Region in the Midwest. *Journal of Occupational and Environmental Medicine*, 58(2), 154-161. doi:10.1097/jom.0000000000000584
- Reville, R. T., Bhattacharya, J., & Weinstein, L. R. S. (2001). New methods and data sources for measuring economic consequences of workplace injuries. *American Journal of Industrial Medicine*, 40(4), 452-463. doi:10.1002/ajim.1115
- Riedel, S., & Field, W. (2011). *Summary of over 800 grain storage and handling-related entrapments and suffocations documented in the US between 1970 and 2010*. Paper presented at the Presentation at XXXIV CIOST-CIGR V Conference, Vienna, Austria.

- Singh, G., Sachdeva, S. N., & Pal, M. (2016). M5 model tree based predictive modeling of road accidents on non-urban sections of highways in India. *Accident Analysis & Prevention, 96*, 108-117.
- Springmeyer, R. R., Blattner, M. M., & Max, N. L. (1992). A CHARACTERIZATION OF THE SCIENTIFIC-DATA ANALYSIS PROCESS. *Visualization 92 : Proceedings*, 235-242.
- Strobl, C., Malley, J., & Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods, 14*(4), 323-348.
doi:10.1037/a0016973
- Tsioras, P. A., Rottensteiner, C., & Stampfer, K. (2014). Wood harvesting accidents in the Austrian State Forest Enterprise 2000-2009. *Safety Science, 62*, 400-408.
doi:10.1016/j.ssci.2013.09.016
- Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32*(9), 1761-1768.
- Utterback, D., Meyers, A., & Wurzelbacher, S. (2014). Workers' Compensation Insurance: A Primer for Public Health. *Department of Health and Human Services: National Institute for Occupational Safety and Health*.
- Utterback, D. F., Schnorr, T. M., Silverstein, B. A., Spieler, E. A., Leamon, T. B., & Amick, B. C., III. (2012). Occupational Health and Safety Surveillance and Research Using Workers' Compensation Data. *Journal of Occupational and Environmental Medicine, 54*(2), 171-176. doi:10.1097/JOM.0b013e31823c14cb

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., . . .

. Rowe, C. M. (1985). STATISTICS FOR THE EVALUATION AND COMPARISON OF MODELS. *Journal of Geophysical Research-Oceans*, 90(NC5), 8995-9005. doi:10.1029/JC090iC05p08995

Wurzelbacher, S. J., Al-Tarawneh, I. S., Meyers, A. R., Bushnell, P. T., Lampl, M. P.,

Robins, D. C., . . . Raudabaugh, J. A. (2016). Development of methods for using workers' compensation data for surveillance and prevention of occupational injuries among State-insured private employers in Ohio. *American journal of industrial medicine*, 59(12), 1087-1104.

CHAPTER 5. GENERAL CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

Conclusions

The overall objective of this study was to characterize occupational injuries in the agribusiness industry by extracting useful information from a large workers' compensation claims dataset using statistical and decision tree modeling techniques. The overall objective was operationalized into three separate studies documented in chapters 2, 3 and 4.

The study in Chapter 2 focused on occupational injuries in biofuel production facilities such as ethanol and bio-diesel plants. Very little has been published on occupational injury patterns in these facilities due to the lack of a centralized injury data source. Workers' compensation claims obtained from these facilities serves as an alternate source to help address some of this knowledge gap. Analyzing workers' compensation claims data provides insights on the characteristics of occupational injuries that occur in ethanol and bio-diesel producing facilities. The primary objective of the study in chapter 2 was to characterize the direct cost of occupational injury using the information on the workers' age and tenure, body part injured, and cause and type of injury obtained from the workers' compensation claims.

Contingency table analysis was used to classify the variable pairs, while chi-square test and chi-square residuals were employed to evaluate the relation between the variable pairs and identify the at-risk groups. Results showed age of injured employee, tenure of employee in the organization, type of claim, body part injured, nature of injury, and the cause of injury all have a significant influence in determining the claim amount. Furthermore, employees aged 46 to 50, less experienced employees, injuries related to strain, fractures, slips, trips, or falls, and injuries to lower extremities were identified as categories with a high likelihood of severe injuries.

The findings of the first study enhance the understanding of the risks of injury in biofuel facilities by providing empirical evidence through analysis of past safety injury data. The analysis of claims, recorded over an extended period of time, characterizes the strength and rigorousness of this study.

The objective of the study in chapter 3 was to characterize workplace injuries in the commercial grain elevator industry. Like the biofuel industry, very few published studies have examined work-related injuries, primarily due to the lack of availability of past injury data. Therefore, workers' compensation claims are an important source of knowledge on occupational injuries in this industry.

The specific objective of the study in chapter 3 was to characterize the direct cost of occupational injury using the information obtained from the workers' compensation claims. Variables examined included age, tenure of employee, body part injured, and cause of injury. Contingency table analysis, chi-square test and chi-square residuals were employed to evaluate the relationship between the variable pairs and identify at-risk groups.

Results showed that age and tenure of employee, cause of injury, and body part injured all have a significant influence on the cost paid for the claim. Furthermore, a strong relationship between age and claim amount was observed across all age groups except for employees in the 36 to 40 years' age group category. With respect to the tenure of the employee, the general trend observed was that as length of tenure increases, the number of claims decreases, indicating that employees with less experience are the most at-risk of occupational injury. Also, the chi-square tests showed that the claim amount is not the same for all types of injuries. Additionally, cause of the injury was a significant factor in determining the claim amount. Similarly, the claim amount varied significantly based on the

body part injured thereby suggesting that injuries to certain body parts have higher claim amounts. In this case, injuries to upper and lower extremities tend to be more expensive as compared to injuries to “head and trunk”.

Despite the existence of numerous workplace hazards in the grain handling industry, most studies investigating injuries and fatalities in grain facilities have focused only on a few safety hazards. The study in chapter 3 investigated occupational injuries across several hazard categories of the grain handling industry.

The findings of the study in chapter 3 will enhance the understanding of recommended areas of preventative intervention in the grain handling environment. Also, the findings can be used to build multivariate models of the claim amounts, which would help commercial grain elevators, as well as the worker’s compensation insurance providers, better analyze the risks contributing to occupational injuries.

The objective of the study in Chapter 4 was to build and compare two multivariate predictive models for determining the number of days away from work of workers who suffered an occupational injury. The number of days away from work (DAFW) is an important metric to measure the severity of a work-related injury. Predicting severity is important because severe injuries have higher economic and social costs and determining the severity of occupational injuries allows supervisors and safety managers to evaluate injury risks, identify problem areas, and make decisions to mitigate the impacts of severe work-related injuries.

The study in chapter 4, built and compared two predictive models for the DAFW, linear regression model and decision tree model. These models were built in the SAS Enterprise Miner application using the data recorded in a large workers’ compensation claims

dataset. Comparing the root mean square error (RMSE) of the models showed that the linear regression model with two-way interactions had a lower RMSE than the decision tree model in predicting the DAFW. Furthermore, since regression analysis is one of the most widely used and understood statistical method; this study suggested that the linear regression is a viable and adequate option for modeling the DAFW.

The study in chapter 4 augments the existing body of research literature on occupational injuries in agribusiness industries such as biofuels and grain elevators. Very little published research has modeled the DAFW in these industries. Also, previous studies that investigated the DAFW focused on a single type of industry, while this study used data from a wide range of process-oriented businesses to model the DAFW. Analysis of a large number of claims obtained from various agribusiness industries and recorded over an extended period of time, characterizes the strength and rigorousness of the study in chapter 4.

Limitations

Several limitations should be considered in interpreting the findings of this study. The data for the study were obtained from a private insurance company. The investigator did not collect the data and no controls were exercised by the investigator during the data collection process. The insurance company collected the data during the workers' compensation claims process and employees, employers, and other involved parties provided the information. While inputs from various entities enriches the data, making the WC data an excellent source of information related to work-related injuries, the recording of information from multiple sources during the WC claims process are prone to human errors. Also, it must be noted that the primary purpose of information collection during the WC claims process is to provide

benefits to the injured employee and not for research purposes. Therefore, the scope of analysis for this research project was narrowed by the information available in the dataset.

Workers' compensation data, like other injury data sources, are also prone to underreporting of injury claims. For example, workers' compensation claims data of temporary and seasonal workers who are not in the payroll on a permanent basis may not be part of this dataset. Hiring temporary and seasonal workers is a common practice in agribusiness industry due to the seasonal nature of their business cycle. Therefore, the actual number of injuries could be higher than is reflected in the dataset.

Finally, this research project is an observational study of the occupational injury data for non-farm agricultural industries. One important limitation of observational studies is that, unlike experimental or intervention studies, evaluating the relationship between two variables by controlling the effects of other variables is not possible. This limited some options for analysis of the dataset.

Recommendations for future research

Based on the findings of this research the following are recommendations for future research:

- In addition to commercial grain elevators and biofuel production facilities, there are other agribusinesses where workers' compensation claims can be utilized to study occupational injuries.
- In this study, the claim amount was used as dependent variable to characterize occupational injuries. Similar analysis can also be conducted using cause of injury, nature of injury or body part injured as dependent variables. Such analysis would

provide further information to agribusiness industries for developing targeted interventions and mitigating occupational injuries.

- Extend the two-way contingency table analysis to three-way or n-ways to investigate if the relationships between the two variables changes based on the values of a third variable.
- The primary objective of this research was to gain a better understanding of occupational injuries in commercial grain elevators and biofuel production. This understanding can be useful for developing more advanced statistical models for predicting the number of claims based on the employee and injury characteristics.
- While this study used the total claim amount as dependent variable, future studies can investigate if there is a relationship between the employee and injury characteristics and a particular type of costs such as medical or indemnity costs. For example, do older workers have higher indemnity costs than younger workers assuming all other factors are kept constant?
- In this study, about 50% of the 34 columns of information available in the workers' compensation claims dataset were utilized. Future work can include other data fields such as accident state, occupation, and claims description to investigate injuries in agribusiness facilities. For example, the description associated with each claim can be mined using text-mining programs to identify patterns associated with these claims.
- In this study, linear regression and decision trees were used to build predictive models for the days away from work. Future studies can use other statistical and data mining techniques such as logistic regression, artificial neural network, and random forests, for building and evaluating models for the number of days away from work.

- Generally, managers and supervisors of the injured employee are actively involved in the injury reporting and workers' compensation claims process. However, very little is known about the cost of time and effort spent by various parties involved in the claim process. Future research can investigate some of these costs. For example, is the time and effort spent by managers in the claim process of severe injuries significantly different than those on less severe injuries?

Most studies that used workers' compensation claims to study occupational injuries investigated only claims related to production agriculture or on-farm work injuries. This study is the first to investigate occupational injuries in non-farm agricultural workplaces such as grain elevators and biofuel producers using workers' compensation claims. Continuous improvement in workplace safety outcomes requires learning from past incidents. The empirical evidence documented in this study can help safety professionals and supervisors in agribusinesses implement targeted injury prevention strategies to mitigate future occurrences of work-related injuries.

APPENDIX

DECISION TREE MODEL































