Near-Miss Narratives from the Fire Service: A Bayesian Analysis

FINAL REPORT
Grant: R03OH009984
Project Period: 09/01/2011—08/31/2013

PI: Jennifer A. Taylor, PhD, MPH
(215) 762-2590
Jat65@drexel.edu

Project Manager: Alicia Lacovara, MPH, CPH
(215) 762-8453
Avl24@drexel.edu

Affiliation/Award Institution:
Drexel University School of Public Health
1505 Race St. MS 1034
Philadelphia, PA
19102-1192

Co-Investigators:
Gordon S. Smith, MB, ChB, MPH; University of Maryland, School of Medicine
Mark Lehto, MSIE, PhD; Purdue University, School of Industrial Engineering

Date of Final Report completion: September 30, 2013

Table of Contents

Closeout Document 1: *The Final Progress Report*

List of Terms and Abbreviations

NIOSH: National Institute for Occupational Safety and Health
NFFNMRS: National Fire Fighter Near-Miss Reporting System
IAFC: International Association of Fire Chiefs
ICD-9-CM: International Classification of Diseases, 9th Revision, Clinical Modification
NHIS: National Health Interview Survey

Abstract

Background:

In occupational safety research, narrative text analysis has been combined with coded surveillance data to improve identification and understanding of injuries and their circumstances. Injury data give information about incidence and the direct cause of an injury, while near-miss data enable the identification of various hazards within an organization or industry. Further, near-miss data provide an opportunity for surveillance and risk reduction. The National Firefighter Near-Miss Reporting System (NFFNMRS) is a voluntary reporting system that collects narrative text data on near-miss and injurious events within the fire and emergency services industry. In recent research, autocoding techniques using Bayesian models have been used to categorize/code injury narratives with up to 90% accuracy, thereby reducing the amount of human effort required to manually code large datasets. Autocoding techniques have not yet been applied to near-miss narrative data. Furthermore, the data collected in the Contributing Factors (CF) and Loss Potential (LP) fields of the NFFNMRS have not been analyzed. We sought to examine the utility of these quantitative variables, particularly in relation to injuries and near-misses.

Methods:

We manually assigned mechanism of injury codes to previously un-coded narratives from the NFFNMRS and used this as a training set to develop two Bayesian autocoding models, Fuzzy and Naïve. We calculated sensitivity, specificity, PPV, ROC curves, confusion matrices, and top predictor word lists. We also evaluated the effect of training set size on prediction sensitivity and compared the models' predictive ability as related to injury outcome. We cross-validated a subset of the prediction set for accuracy of the models when coding de novo, and cross-validated the training set to assess variation in the wordlists and probabilities as a function of the training set. To examine CFs in relation to injury outcome, we performed descriptive analyses between the variables and completed a 3-model latent class analysis to identify whether the 21 CFs could be reduced to fewer categories. For LP, we performed descriptive analyses as well.

Results:

Overall, the Fuzzy model performed better than Naïve, with a sensitivity of 0.74 compared to 0.678. Cross-validation of the prediction set showed sensitivity reached 0.602, where Fuzzy and Naïve had the same prediction. As the number of records in the training set increased, the models performed at a higher

sensitivity, suggesting that both the Fuzzy and Naïve models were essentially "learning." Injury records were predicted with greater sensitivity than near-miss records. There was no evidence of a relationship between CFs or LP and injury outcome. Furthermore, the CF categories did not reduce into fewer categories, as the categories showed an overall poor fit to the model.

Conclusion:

The application of Bayesian autocoding methods can successfully code both near-misses and injuries in longer-than-average narratives with non-specific prompts regarding injury. Such coding allowed for the creation of two new quantitative data elements for injury outcome and injury mechanism. With the CFs and LP as currently captured in the data system, we conclude that these variables and answer selections lack definition and are poorly understood by the individuals reporting to the system and add little value to the system as a result.

Section 1 of the Final Progress Report

Significant (Key) Findings.

1. Bayesian autocoding methods were successful in assigning mechanism of injury codes to near-miss and injury narratives from the NFFNMRS, reaching a sensitivity of 74% and 68%, for Fuzzy and Naïve, respectively. Additionally, the autocoding was successful at assigning injury outcome (yes/no) to narratives, able to correctly assign injury outcome to 92% of narratives. This created two new quantitative variables within the data system: mechanism of injury and injury outcome.
2. The Fuzzy Bayesian model performed at a higher level of sensitivity than the Naïve model.
3. The algorithm's performance improved as the size of the training set increased.
4. The Contributing Factors (CFs) variable within the NFFNMRS has no relationship to the injury outcomes determined in Key Finding 1. This is because a standard definition does not exist for each of the CFs, thereby allowing each reporter to interpret it for him or herself. A Latent Class Analysis confirmed this finding, in that there was little relationship between any of the CFs, and they had poor loading coefficients.
5. Within the Loss Potential (LP) variable, none were found to have any relationship with injury outcome; the ratio of injuries to near-misses varied very little between Loss Potential categories.

Translation of Findings. Computer processing of narratives from workers makes their stories more efficiently translated into discrete injury and cause categories. Now that these new data elements have been added to the NFFNMRS reports, the fire service can use them to focus on preventing injuries in its membership. Furthermore, all employers collecting narratives from workers can use this technology to create new variables and conduct hazard analyses in their data systems. The methods in this project provide proof of concept that not only actual injuries, but near-misses can be accurately coded by machine learning computer algorithms. This is significant because near-misses represent upstream opportunities for prevention (before an injury occurs). Employers may find significant cost savings in workers compensation claims if they focus on hazard mitigation in their enterprises instead of treatment for injuries that could have been prevented.

Outcomes/ Impact. As the only data system of its kind, the NFFNMRS has collected valuable information about the occupational hazards within fire and emergency services and has the potential to be useful, but requires modifications to improve the robustness and data quality. Drexel has previously made recommendations to improve the system in a Data Quality Report to the International Association of Fire Chiefs (IAFC). This was done under a contact between IAFC and Drexel. IAFC has expressed interest in making these changes but is currently waiting for funding in order to proceed. The NIOSH RO3 award enabled us to not only conduct the machine learning analysis, but review other variables in the NFFNMRS that were areas of concern. Our current research presented here has provided additional confirmation of the need for changes in how the data is collected. New variables are necessary for collecting information regarding the Contributing Factors and Loss Potential of an event, and better prompts are needed to obtaining information about how an event happened, whether an injury occurred,

and what the injury was. This is presented below in our results sections. This additional information has also been shared with the IAFC and is awaiting their consideration for modification to the system.

The recommended changes will ensure that future data provides specific information about how the event occurred, whether an injury resulted, and what the injury was. Information provided about the various contributing factors (once the field is refined) will give information about other aspects of the fire scene that played a part in the chain of events leading up to the near-miss or injury, thus providing the potential for intervention.

Section 2 of the Final Progress Report

Background

There are approximately 1.1 million firefighters in the U.S., of whom 336,000 are career professionals and 812,000 are volunteers (National Institute for Occupational Safety and Health, 2012). Over the past ten years, an annual average of 100 firefighters died in the line-of-duty, 50% as a result of injuries (Karter & Molis, 2007). The National Fire Protection Association (NFPA) estimates that approximately 80,000 firefighters are injured in the line of duty each year (Karter & Molis, 2007) but this is believed to be an undercount. These data indicate that almost half of firefighter injuries in 2007 occurred during firefighting operations, an estimated 15,435 injuries were sustained at non-fire emergency incidents, and 13,665 during other on-duty activities. Strain/sprain or muscular pain was the most common injury for both firefighting (45.1%) and non-firefighting activities (57.8%). An estimated 20.4% of all injuries resulted in lost time in 2007 (Karter & Molis, 2007).

Knowledge is advancing regarding firefighter line of duty deaths, but understanding of near misses and nonfatal injuries lags behind. Near miss data are especially valuable because studying only injuries focuses attention downstream (outcome bias) whereas near miss data indicate problems in the process upstream of injury (Rivard, Rosen, & Carroll, 2006)and may also identify the strategies that prevented an incident from becoming more serious. Further, one need not capture every single near miss in order to improve safety; it has been shown in patient safety, for example, that even a few voluntary reports can be sufficient to detect a new hazard (Leape, 2002). Other industries have instituted near miss systems resulting in significant improvements in prevention.

The Aviation Safety Reporting System (ASRS) at NASA identifies hazards related to aviation operations and safety culture, such as a cockpit interaction incident that caused deviation from the 'sterile cockpit' protocol required during takeoff and landing. Study of this deviation is then used for recommendations, training, and policy modification by the Federal Aviation Administration (Sumwalt, 1993).

Barach and Small detail the advantages of analyzing near misses over studying only actual injury or fatality outcomes (Barach & Small, 2000):

1. Near misses occur up to 300 times more often than adverse events, enabling quantitative analysis;
2. Fewer barriers to data collection exist, allowing analysis of interrelations of small failures;
3. Recovery strategies can be studied to enhance proactive interventions and to de-emphasize the culture of blame; and
4. Hindsight bias is more effectively reduced.

Since near miss events happen more frequently than injuries, "near misses allow a more quantitative and predictive representation of the system's potential failure points…[and] provide data to study human recovery within the system," providing a powerful opportunity for hazard surveillance and risk reduction (Kaplan, 2003). The common cause hypothesis states that near misses and accidents have the same relative causal patterns, and research has supported this hypothesis in fields including childhood injury, railway safety, and patient safety(Alamgir, Yu, Gorman, Ngan, & Guzman, 2009; Kaplan, 2003; Wright

& van der Schaaf, 2004). Near miss events reveal hazards that could have resulted in injury, but did not, calling attention to opportunities for safety interventions at an earlier point in the process than can be identified by data describing injuries and their consequences. Many near misses are potentially fatal events, and the narratives contained in these reports are opportunities to examine vital causal information that would not have been available had the reporter died.

In order to improve understanding of the circumstances leading to firefighter injuries, the International Association of Fire Chiefs (IAFC) (with funding from the Assistance to Firefighters Grant Program of the U.S. Department of Homeland Security (DHS)) launched the National Fire Fighter Near-Miss Reporting System (NFFNMRS) in 2005. NFFNMRS was modeled after NASA's ASRS and aspires to the respected utility ASRS has achieved. While it took ten years for ASRS to reach a steady state level of reporting (Connell, 2008), NFFNMRS can learn from ASRS to craft its own analytic plan while reporting levels are still growing. This proposed study will create a method for systematic and sustainable analysis of NFFNMRS data that will be applicable from the nascent stage (3,695 reports since 2006) through maturity (ASRS currently receives ~40,000 reports annually).

Reporting to NFFNMRS is voluntary and non-punitive. A near miss is defined as "an unintentional, unsafe occurrence that could have resulted in an injury, fatality, or property damage" (www.firefighternearmiss.com). Despite this definition, the NFFNMRS captures a number of actual injuries, including fractures, back injuries, hypothermia, burns, and cyanide poisoning, as well as melted equipment and destroyed engines. In 2009, 1,116 near miss reports were submitted to NFFNMRS, 726 of which were identified as having the potential for life threatening injury. The reports were evenly divided between incidents during firefighting and non-firefighting, with 48% of reports from suburban fire departments, 33% urban, and 19% rural. Standard Operating Procedures (SOPs) were followed in 63% of the reports, in 20% SOPs were not followed, and in 17% there were no SOPs in place that addressed the situation. Twelve percent of reports involved improper use of equipment, and 6% involved mechanical failure of equipment (National Fire Fighter Near-Miss Reporting System, 2010). Given the current volume of reporting, it is clear that most near miss events are not reported, and that the existing reports are therefore affected by volunteer bias. As with the development of ASRS, the system is expected to experience significant reporting growth over the next 5 years, providing a more representative data source and a stronger basis for statistical analysis.

In occupational safety research outside the fire service, narrative text analysis has been combined with coded surveillance data to improve identification and understanding of injuries and their circumstances. The large scale study of narrative text has only recently been made possible by advances in computerized information retrieval techniques. Narrative text analysis identifies more target events than can be found using injury codes alone, thus reducing the problem of undercounting – a critical concern in injury surveillance. Further, narrative text analysis provides a means to check coding accuracy, and provides important information on circumstances surrounding injuries and unknown risk factors (Bondy, Lipscomb, Guarini, & Glazner, 2005; Bunn, Slavova, & Hall, 2008; Lipscomb, Glazner, Bondy, Lezotte, & Guarini, 2004; Smith et al., 2006). Information about the context and circumstances surrounding a near miss offers more useful clues for injury prevention than are found in the structured areas of a reporting form. Event factors may be organized according to their relationship to error/failure detection,

error localization, and correction of the problem (Kaplan & Fastman, 2003). New risk factors identified through narrative text analysis are an important source of variables to be added to administrative coding systems (Bunn et al., 2008). Narrative data analysis can also be a basis for comparing data among systems and countries that use different coding schemes, or to study historical data that include narrative text (Stout, 1998).

NASA staff has developed proprietary text-mining software to analyze ASRS narrative reports through keyword search, phrase search, phrase generation, and phrase discovery (McGreevy, 2001). The coded quantitative data, coupled with analysis of narrative text fields through which the reporter describes the near miss in his or her own words, "provide an exceptionally rich source of information for policy development and human factors research" (http://asrs.arc.nasa.gov/overview/database.html). NASA research shows that 70% of aviation
accidents involve human error (Helmreich, 2000), highlighting the importance of understanding the human factors, such as communication, teamwork, decision-making and fatigue, central to the nature and circumstances of these errors. Helmreich points out that aviation has improved safety through error management in which corrective actions are tailored to the key human factors involved, for example errors in communications are corrected by team training, while violations are addressed through cultural change coupled with procedural review and improvement (Helmreich, 2000). In the reports submitted to NFFNMRS, the most frequently cited contributing factors were decision-making, human error, and individual action, followed by situational awareness, communication, and equipment. All but the final item on this list are human factors, which are best understood by analyzing the rich detail contained in narrative text reports.

Keyword analysis can efficiently and comprehensively identify cases of interest from large databases. Smith et al. studied a year's data from Workers Compensation claims from a large insurance company, using narrative text analysis to identify ladder fall cases, confirm fracture diagnoses, and code injury circumstances from among 535,605 claims. 705 cases with fractures were identified among the potential ladder cases, of which 120 (17%) were identified only by the narrative text search. This study successfully linked injury event and outcome with information on mechanisms, cost, and disability (Smith et al., 2006).

Near miss incidents may be reconstructed by creating a template to standardize the description of the chain of events. This standardized template must be able to encompass all data elements needed to characterize the target incidents. The elements used in the template can then be used as a basis for creating taxonomies of near misses. An incident taxonomy allows the identification of different hazard scenarios, utilizing narrative text to identify contributing factors, precipitating mechanisms, and primary sources as a complement to administratively coded data. The prevalence of different hazard scenarios reveals common mechanisms and priorities for prevention (Lincoln et al., 2004). However, one cannot create a taxonomy of hazard scenarios without having already identified the mechanisms of injury involved. Review of the literature, and the work of our consultant Mark Lehto, suggest that fuzzy Bayesian analysis is the most efficient approach to classifying the mechanisms within free-text narratives due to its capacity to consider multiple word combinations, estimate the likely accuracy of its

predictions, and its low demand on computational resources. An in-depth review of the Bayesian methods literature most relevant for our research is provided in Table 1 (p.9-10).

Table 1 Comparison of Previous Auto-coding Studies

	NFFNMRS Narratives (current study)	Motor Vehicle Accident Data (Lehto & Sorock 1996)	NHIS (Wellman et al. 2004)	Worker's Comp (Lehto et al. 2009)	Worker's Comp (Wellman et al. 2011)	Worker's Comp (Ohio) (Bertke et al. 2012)
Study Purpose/Objective	Manually assign injury codes to near-miss FF narratives; apply Bayesian methods to autocode narratives.	Proof of concept—could a Bayesian model learn from keywords to search & accurately classify narratives.	Investigate accuracy of computerized coding to classify injury narratives into E –codes.	To compare Fuzzy and Naïve methods for classifying injury narratives.	To assess performance of combined Fuzzy/Naïve approach for classifying injury narratives.	To evaluate computerized coding of injury narratives.
Narrative type	Fire fighter-occupation specific narratives, with near-misses & injury	Insurance company automotive accident narratives	General population injury narratives	Worker's compensation injury narrative	Worker's compensation injury narrative	Ohio Bureau of Worker's Compensation Claims
Narrative Characteristics	Long narratives (avg. 216 words)	Short narratives (2-3 sentences long)	Short narratives (avg. 11 words)	Short narratives (avg. 20 words)	Short narratives (avg. 20 words)	
# Cause Categories	14 coding categories	9 coding categories (2 main groups)	13 coding categories	21 coding categories used (out of 40 OIICS codes)	21 coding categories used (out of 40 OIICS codes)	3 broad coding categories; 8 specific coding categories
Coding Scheme	ICD9-CM (3-digit)	2 categories: Pre-crash (5 codes) and Crash (4 codes)	ICD9-CM (2-digit)	OIICS classification (2-digit)	OIICS classification (2-digit)	OIICS classification
Size of Dataset	2,280 (uncoded)	3,686	5,677 (all pre-coded)	17,000 (uncoded)	17,000 (uncoded)	10,132 (uncoded)
Training Set Size; % of dataset	Total of 1,000 manually coded with 764 used to train the algorithm; 43.4%	3,686 narratives; training set was a set of keywords, not coded narratives.	5,677; 100%	11,000 (manually coded)	11,000 (manually coded); Training set 367% larger than prediction set.	2,240 (2,400, minus 160 due to coder disagreement or NOC); 22.1%
Time required to code data	25 hours for 1,000 records (40 records/hour)	Unknown	Unknown	Unknown	Unknown	10 hours for 2,400 records (240 records/hour)
Coder agreement	Final coder agreement greater than 79% (κ > 0.75).	Only 1 coder.	n/a—records pre-coded	Overall 1-Digit agreement of 87%; 2-digit agreement of 75%	Overall 1-Digit agreement of 87%; 2-digit agreement of 75%	Overall agreement of 93.8%.
Prediction Set	2,285 (includes training set)	419	5,677 (same as training set)	3,000 (pre-coded)	3,000	7,732

Table continued on next page.

Table 1 Comparison of Previous Auto-coding Studies (*continued*)

	NFFNMRS Narratives (current study)	Motor Vehicle Accident Data (Lehto & Sorock 1996)	NHIS (Wellman et al. 2004)	Worker's Comp (Lehto et al. 2009)	Worker's Comp (Wellman et al. 2011)	Worker's Comp (Ohio) (Bertke et al. 2012)
Training set modifications	Drop words occurring fewer than 3 times. No synonyms, or stop words.	Keyword list of 2,619 was morphed, endings removed (ing, ed), articles removed, misspellings corrected.	Creation of Keyword list-- words occurring more than 3 times in dataset; drop word lists; synonym words.	Drop word list; drop words occurring fewer than 3 times; remove punctuation and non-alphanumeric characters.	List of Keywords and Drop words was generated; transformation of synonyms; correction of misspelling.	None described.
Analyses	Fuzzy Bayesian and Naïve Bayesian models using Single Word predictor; comparison of predictive ability as it relates to injury or near-miss	Leave-one-out/Naïve Bayesian and Fuzzy Bayesian	Single word Fuzzy; Multiple word Fuzzy (single words, up to 4-word combos).	Naïve and Fuzzy Bayes (comparison)	Naïve and Fuzzy Bayes (combined); 1st Strategy: assign cases for manual review if Fuzzy and Naïve models disagree; 2nd Strategy: selection of additional cases for manual review from Agree dataset using prediction strength to reach level of 50% computer and 50% manual coding.	Assessed number of categories and size of training set on prediction set sensitivity. Assessed use of training set from one sector upon another sector.
Distribution of Codes	Weighted mostly to Fire-Fall (25.7%), Fire-Struck-By (24.1%) and Fire-Burn (22.1%).	Not provided in results.	Heavily weighted to Falls (35%), followed by Struck-by (16%), and Overexertion (12%)	Weighted towards Overexertion (17.8%), Falls (17.4%) and Struck-by (9.8%).	Weighted towards Overexertion (17.8%), Falls (17.4%) and Struck-by (9.8%).	Weighted towards Contact with Object or Equipment (49.3%), Slips, Trips and Falls (23.8%), and Musculoskeletal Disorders (18.0%)
Results	**See results section of AIM 1 in document.**	Keyword based classification results consistently good. Fuzzy Bayes can augment results in cases where keyword classification failed and in categories where keyword classification performed poorly.	A computer program based on fuzzy Bayes logic is capable of accurately categorizing cause-of-injury codes from injury narratives. The ability to set threshold levels significantly reduced the amount of manual coding required, without sacrificing accuracy.	Single-digit codes predicted better than double-digit; Naïve slightly more accurate than Fuzzy; Naïve had sensitivity of 80% and 70% (for one and two digit codes respectively), Fuzzy Bayes had a sensitivity of 78% and 64%. Specificity and PPV was higher in Naïve than Fuzzy.	1st Strategy: Agreement alone as filtering strategy left 36% for manual review (computer coded 64%, n=1928). Overall combined sensitivity was 0.90 and PPV>0.90 for 11 of 18 2-digit categories.	Naïve Bayesian auto-coding of narrative text and injury diagnosis showed up to 90% accuracy, improvement in performance with increased training size, and training sets with broader coding performed as well or better to predict more specific sector claims.

Specific Aims

AIM 1 - To use recently developed autocoding methods to characterize firefighter near miss narratives and classify these narratives into mechanisms of risk/injury. This analysis applied the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), using Bayesian machine learning techniques to identify the various mechanisms captured in the near miss narratives and their relative prevalence.

AIM 2 - To identify the correlation between each mechanism of risk/injury and each of the "Contributing Factors" listed on the NFFNMRS reporting form. Using Latent Class Analysis, we analyzed the "Contributing Factors" variable with regards to the manually assigned code of injury/no-injury, to see if any relationships existed.

AIM 3 - To use manual coding to identify actual injury incidents contained within a random sample of 1,000 near miss narratives and correlate these injuries with the "Loss Potential" categories on the NFFNMRS reporting form. We analyzed the "Loss Potential" variable with regards to the manually assigned code of injury/no-injury, to see if any relationships existed.

For clarity and simplicity, we have presented each AIM as its own section, which contains the methods, results and discussion specific to that aim.

AIM 1

*The major findings of what is presented below with regards to AIM 1 was submitted in June 2013 to the journal *Accident Analysis and Prevention.* On September 17, 2013, this paper was accepted for publication (AAP3289) and is currently in press. It will be published in open access format in order to ensure the widest dissemination and to remove any access barriers the fire service could experience.

AIM 1 METHODS:

Data Source

In order to improve understanding of the circumstances leading to firefighter injuries, the International Association of Fire Chiefs (IAFC) (with funding from the Assistance to Firefighters Grant Program of the U.S. Department of Homeland Security) launched the NFFNMRS in 2005. Reporting to the system is voluntary and non-punitive. The NFFNMRS defines a near-miss as "an unintentional, unsafe occurrence that could have resulted in an injury, fatality, or property damage" (www.firefighternearmiss.com). Despite this definition, the NFFNMRS captures a number of actual injuries, including fractures, back injuries, hypothermia, burns, and cyanide poisoning, as well as melted equipment and destroyed engines.

The reporting form consists of 22 fields. Two of these fields are narrative sections, asking the reporter to "Describe the event," and to share "Lessons Learned." Within these fields, reporters can submit as much text as they wish.

Selection of narratives for manual coding

The quantitative component of the near-miss forms contains a field called "Event Type" in which the reporter selects whether the incident occurred during a fire emergency event, a vehicle event, a training activity, etc. (the form can be viewed at

http://www.firefighternearmiss.com/Resources/NMRS-Mail.pdf). In order to reduce cognitive shifts required for coding of different event types (hazards described in vehicle event narratives are different than those in fire event narratives), we limited our analysis to only include those indicated as fire emergency events, as identified by the reporter. This data set contained 2,285 narratives. Of these "Fire Emergency Events," we manually coded 1,000 narratives, which resulted in 764 fire-related events considered suitable as training narratives for the algorithm. The 236 narratives discarded from the training set were not "Fire" related cases (e.g., neither the precipitating nor proximal cause was a fire event), or they were fire-related but lacked specific information for sub-categorization (e.g., Fire-Burn, Fire-Struck-by/Against), or they fell into a category that ended up having fewer than 5 narratives (e.g., Motor Vehicle-Rollover, Hot substance or object, caustic or corrosive material, and steam). Figure 1(p.13) shows the case inclusion criteria for our analysis.

Figure 1 Case Inclusion Flow Chart

Total narratives in
Near-Miss Dataset
n= 4,814

All other Event Types
excluded from study[‡]
n= 2,529

Event Type: Fire
Emergency Events
n= 2,285

"Training" Set: Records assigned
a Mechanism of Injury Code for
Precipitating and Proximal cause,
and Injury Status Code (yes/no)
n= 1,000

"Prediction" Set: Records
not assigned Mechanism
of Injury Code
n= 1,285

Records excluded
from "Training" Set*
n= 214

Records included in
"Training" Set**
n= 786

Training of Algorithm

"Prediction" Set
Coded by
algorithm
n= 1,285

Cross Validation of
"Prediction" Set
n= 300

‡ Other event types in the reporting system
include: Non-fire emergency events, On-duty
activities, Training activities, and Vehicle events.

* Records did not receive cause of "Fire" for either
Precipitating or Proximal code.

**Records were assigned a cause of "Fire" in either
Precipitating or Proximal code.

16

Manual coding rubric

The initial rubric was a set of mechanism of injury codes from the International Classification of Disease 9 Clinical Modification Manual (ICD-9-CM), selected by the Principal Investigator (JAT) as codes that were possible within the fire-fighting/EMS occupational field. The rubric was modified over time in an iterative, consensus-driven process. Whenever a change was made the Project Manager (AVL) went back over the previously coded narratives and amended the code in accordance with the revised rule when necessary. A precipitating mechanism (what set the injury chain of events in motion) and a proximal mechanism (what caused the injury or near-miss) were assigned to each narrative.

In creating our coding rubric, it became evident that the ICD-9-CM is not granular enough for firefighting. For example, since fire fighters encounter fire frequently, coding the majority of cases to Conflagration (i.e., E890-E899) would mask hazards that occur during fires such as electrocutions, falls, smoke inhalation, struck-by motor vehicles, etc. Therefore, we created subcategories within Conflagration (Fire) to further capture specific firefighting hazards (Table 2). The resultant coding scheme extracted more detail from each narrative while honoring the ICD-9-CM hierarchy by retaining the overall cause category as Conflagration (Fire). Because this process was iterative, we re-coded previous cases as necessary updating them to the newer rubric.

Table 2 Firefighter-specific List of Mechanism of Injury Categories used to classify narratives

Original List of Cause Codes (pre-coding)	Final List of Cause Codes
Accidents caused by machinery	Accidents caused by explosive material (Gas leak, Dynamite, etc)
Air and Space Transport Accidents	Accidents caused by machinery
Caught accidentally in or between objects	Air and space transport accidents
Cutting and piercing instruments or objects	Caught accidentally in or between objects
Drowning/Submersion	Cutting and piercing instruments or objects
Electric current	Drowning/Submersion
Exposure to radiation	Electric current
Explosive material	Exposure to radiation
Explosion of pressure vessel	Fall
Fall	Fire
Fire	Fire - Burn
Firearm	Fire - Caught-in/Between
Hot substance or object, caustic or corrosive material, and steam	Fire - CO, Smoke, Fumes from PVC, etc
Motor Vehicle Non-Traffic Accident	Fire - Collapse
Motor Vehicle Traffic (MVT)	Fire - Electric Current
Natural/Environmental	Fire - Equipment/Machinery
Other	Fire - Explosion caused by Fire
Other Road Vehicle Accidents	Fire - Fall (through floor, from ladder, jump)
Overexertion	Fire - Medical Condition (MI, Asthma, etc)
Poisoning	Fire - Struck-by
Railway accidents	Fire - Vehicle
Struck by, against	Fire - Wildland, etc
Suffocation	Firearm/Ammunition
Water transport accidents	Hot substance or object, caustic or corrosive material, steam

	Motor Vehicle Non-Traffic Accident
	Motor Vehicle Traffic (MVT)
	MV - Collision
	MV - FF Struck by vehicle
	MV - Other
	MV - Rollover
	N/A
	Natural/Environmental
	Other
	Other Road Vehicle Accidents
	Overexertion
	Poisoning
	Railway Accidents
	Struck-by, against
	Suffocation
	Water Transport Accidents

Manual coding of narratives

In the field of autocoding, there has not been an established minimum size of the training set with regard to the total dataset. Therefore, we decided to code a minimum of 20% of our dataset to act as the training set for the algorithm, similar to Bertke et al. (2012). Based on these recommendations, we calculated that we needed to manually code a minimum of 456 narratives for our training set and aimed to complete more than this.

Three of the authors (JAT, AVL, GS) coded each narrative for a 1) whether an injury occurred (yes/no), 2) the cause of the injury/near-miss (proximal cause), and 3) what lead to the injury/near-miss (precipitating cause). By asking the above three questions in this order, we were able to consistently evaluate each narrative for Injury Outcome, Proximal Cause, and Precipitating Cause. The order was important because in near-miss narratives, the proximal cause is often difficult to discern since no actual injury occurred. It took each coder approximately 25 hours to assign mechanism of injury codes to 1,000 narratives. The narratives were coded in seven batches. After each batch, the three coders reconciled their individual scores for each narrative, assigning a final Mechanism of Injury code. Reconciliation of the seven batches took approximately 25 hours. The entire coding, reconciliation, and rubric revision process occurred over a one year interval. Overall coder agreement statistics were calculated and kappa values were obtained.

The final set used as the training set consisted of 764 narratives. A total of 236 narratives were not included in the training set because they were not assigned a code of "fire" for either the precipitating or proximal code (n=214), or they were assigned a mechanism of injury code that existed in fewer than 5 total narratives (n=22). For example, many of the narratives were categorized by the reporter as "Fire emergency events," but the narrative actually describes a motor vehicle accident on the way to a structure fire. Other narratives lacked enough detail or information to either classify them as a fire event, or assign a mechanism of injury code. The categories with fewer than 5 narratives were not included in the analysis because after dropping rare words, which is standard practice to reduce model noise, these small categories would no longer have strong predictor words.

Model Development

Two different Bayesian models, referred to as Naïve Bayes and Fuzzy Bayes, were developed and evaluated using the TextMiner program (developed by author ML). The models and software have been described elsewhere (Lehto, Marucci-Wellman, & Corns, 2009). Both models used the statistical relationship between the words present in the injury narratives of the training set (n=764) and the manually assigned Mechanism of Injury code to predict a particular code for a new narrative. This prediction is essentially the probability of a particular code given the words within the new narrative. The two models differ in that the Naïve Bayes prediction is a weighted function of all the words present, while the Fuzzy Bayes prediction is based on the single strongest predictive word for each category. Specifically, the Naïve Bayes model calculates the probability of an injury code using the following expression:

$$P\left(E_i|n\right) = P(E_i)\Pi_j \frac{P(n_i|E_i)}{P(n_j)} \qquad (1)$$

where P (E_i|n) is the probability of event code category E_i given the set of n words in the narrative. P (n_i| E_i) is the probability of word n_j given category E_i. P(E_i) is the probability of category E_i, and P(n_j) is the probability of word n_j in the entire list of keywords.

The Fuzzy model is similar, except that it estimates P (E_i|n) using the 'index term' most strongly predictive of the category, instead of multiplying the conditional probabilities as in the Naïve model:

$$P\left(E_i|n\right) = MAX_j \frac{P(n_i|E_i)\,P(E_i)}{P(n_j)} \qquad (2)$$

The two models were both tested using the TextMiner Software which runs on a Microsoft Access platform. After all the Fire-Events narratives were manually coded, the database was prepared for analysis in TextMiner. Narratives that were non-fire related (as coded by the researchers, see Figure 1) were removed from the dataset. For the remaining narratives, all non-alphanumeric symbols were removed (e.g., Fire-Eqpt/Mach became FireEqptMach). A training flag was used to denote all manually coded narratives that were part of the training set.

Once the training set (n=764) and prediction set (n=1, 285) were divided, the words from the narratives within the training set were used to generate a wordlist. The wordlist was contained in a table listing every word in the entire dataset, starting with the first word in the first narrative and ending with the final word of the last narrative. The dataset was cleaned by removal of words occurring fewer than three times. Each narrative was edited for spelling mistakes during the initial report submission process. No additional modifications were made such as assigning synonyms to words or removing common stop words such as "A,THE, …" The purpose of this was to see how well the algorithm could perform on a raw dataset with little to no human input.

Model Evaluation

Training set

The two Bayesian models generated predictions for every narrative in our Fire Emergency Events dataset, including the training set. The results of the predictions were compared to the manually assigned "gold standard" codes. Manual review also helped accomplish our fourth objective: to calculate sensitivity, specificity, and positive predictive value (PPV) of our Bayesian analysis. Confusion matrices were calculated for each model and ROC (Receiver Operating Characteristics) Curves and lists of top predictor words were generated for each code category using the TextMiner software.

We then evaluated the change in sensitivity of the models as the size of the training set was increased (by increments of 100). The latter analysis was done to both help determine possible over-fitting effects and assess the expected improvement due to increasing sample size. In addition, we compared model performance using different training sets (each equaling 100 narratives) to measure the variation in model performance as a function of the training set.

Five-Fold Cross Validation of Training Set

To estimate prediction reliability and robustness, we conducted a 5-fold cross validation in which we took a random selection of 619 narratives from our manually coded dataset (80%) to generate predictions for the remaining 167 narratives in the manually coded dataset (20%). A random number generator was used to select the 619 narratives for each of the five iterations, from the master dataset containing the total 786 manually coded narratives. Each of the five datasets was then prepared for analysis by TextMiner (addition of a training flag to denote the 619 training narratives). Each of the five samples was then run through TextMiner, generating predictions for the remaining 167 narratives. Sensitivity and PPV (Positive Predictive Value) were calculated for each trial.

Cross Validation of the Prediction Set

The Fuzzy and Naïve Bayes models were also both run on a Prediction Set of 1,285 previously unclassified narratives (Figure 1). In order to test the accuracy of the algorithm's predictions for these new (not originally manually classified cases), we performed a cross validation study in which 300 narratives from the prediction set were manually coded by the reviewers. The cases in the cross validation set were equally divided into 3 categories: 1) strongly predicted, 2) moderately predicted, and 3) poorly predicted cases. The cases were assigned based on prediction strength and whether the Fuzzy and Naïve predictions agreed.

The strongly predicted cases corresponded to narratives in which the Fuzzy and Naïve predictions agreed (n=475). For this category, the distribution of narratives to be included in the cross validation set matched that of the distribution of in the original sample of 475.

The poorly and moderately predicted categories corresponded to cases where the Fuzzy and Naïve predictions disagreed. The latter cases were further subdivided based on prediction strength. Prediction strength was simply the probability assigned by the respective model to its prediction (see

equations (1) and (2)). The poorly predicted cases were those where the Fuzzy and Naïve models disagreed on the prediction, and both had strength predictors in the top 50% of their respective distributions. For example, Fuzzy might predict "Fire-Fall" with a prediction strength of 0.99, while Naïve predicted "Fire-Burn" with a strength of 0.97. They disagree, and both predictions are strong. The moderately predicted cases were those cases where Fuzzy and Naïve disagreed, and one had a strength predictor in the top half of their distribution, and the other had a strength predictor in the bottom half of their distribution. We considered the percentile ranks of these strength predictions to build our poor and moderate samples.

One-hundred narratives were randomly selected from each of the 3 categories. Each narrative was assigned a mechanism of injury code by each coder. The 300 narratives were then reconciled so that each narrative received a single code. These codes were then compared to the codes predicted by the Fuzzy and Naïve algorithms.

Proximal Cause Prediction by Injury Outcome

Finally, we wanted to determine how well each model was able to correctly predict a mechanism of injury code, according to injury outcome. After both models had been run, the training set was separated by injury outcome (injury vs. near-miss), and sensitivity was obtained for each. The effect of increased training set size (in iterations of 100 narratives) was also evaluated by calculating sensitivity separately for each sized training set (for injury vs. near-miss).

AIM 1 RESULTS:

Characteristics of narratives

Within the Fire Emergency Event narratives (n=2,285), the mean word count was 216, with a median count of 156 words and a range from 2 words to 2,420 words.

Intra- and Inter-rater Reliability

Agreement between coders improved substantially with an overall agreement above 79%. Agreement between coders 1 and 2 improved 12% (κ=0.785), coders 1 and 3 improved 8% (κ=0.75), and coders 2 and 3 improved 13% (κ=0.774). Each of the three coders had substantial agreement with their original scores when coding the same narratives a second time (0.68<κ<0.80).

Modification of the coding rubric

Creation of the coding rubric was an iterative process. With each narrative read, common themes occurred and thus informed the creation of specific sub-categories. For example, when there was a roof collapse, we assumed the mechanism to be "struck by/against" unless the reporter specified otherwise. We reached saturation of repetitive events after batch 3 and the rubric did not change for coding of the remaining batches.

Performance of automated coding

Training Set

Table 3 Top 3 Predictor Words for Fuzzy Bayes for Largest 5 Cause Categories

|			
Fire – *Fall*	Pit (0.86)	Stories (0.83)	Spongy, Waist (0.78)
Fire – *Struck By/Against*	Strut (0.90)	Cracking (0.86)	Effect (0.83)
Fire – *Burn*	Burns (0.91)	Flashed (0.84)	Intense (0.81)
Fire – *Electric Current*	Energized (0.93)	Arcing (0.9)	Volt, Arced (0.89)
Fire – *CO, Smoke, Fumes, etc.*	Inhalation (0.88)	Inhaled (0.86)	Speak (0.83)

Overall, Fuzzy Bayes performed better than Naïve Bayes. Table 3 shows the top predictor words when applying the Fuzzy model. Fuzzy outperformed Naïve Bayes with a sensitivity of 0.74 compared to 0.678 (Table 4). The Fire-Burn category was well predicted by both Naïve and Fuzzy, though the specificity and PPV was higher with Fuzzy. For the categories of Fire-Fall and Fire-Struck-By, Fuzzy had better sensitivity while Naïve had better PPV. In general, Fuzzy performed with higher sensitivity, specificity and PPV, particularly in the larger categories. Naïve performed a bit better with the smaller categories.

Table 4 Fuzzy and Naïve Bayesian Analyses: Sensitivity, Specificity, and PPV

Mechanism of Injury Category	N	Fuzzy Model			Naïve Model		
		Sensitivity	*Specificity*	*PPV*	*Sensitivity*	*Specificity*	*PPV*
OVERALL	764	0.740	—	—	0.678	—	—
Fire – *Fall*	196	0.745	0.887	0.695	0.561	0.995	0.973
Fire – *Struck By/Against*	184	0.728	0.933	0.775	0.342	1	1
Fire – *Burn*	169	0.941	0.877	0.685	1	0.652	0.449
Fire – *Electric Current*	68	1	0.974	0.791	0.853	1	1
Fire – *CO, Smoke, Fumes, etc.*	48	0.521	0.997	0.926	0.917	1	1
Fire – *Explosion Caused by Fire*	25	0.12	0.999	0.75	0.84	0.996	0.875
Fire – *Equipment/Machinery*	17	0	1	—	0.412	1	1
Fire – *Medical Condition*	9	0.889	1	1	1	0.999	0.9
Fire – *Caught In/Between*	6	0	1	—	1	0.992	0.5
MV – *FF Struck-by Vehicle*	26	0.577	0.999	0.938	0.692	1	1
Firearm/Ammunition	9	0.778	0.999	0.875	1	0.966	0.257
Cutting/Piercing Instruments/Objects	7	0	1	—	0.571	1	1

Increasing the size of the training set improved the performance of the algorithm (Figure 2). For example, using a training set of 100 narratives to predict the entire dataset of 764, the Fuzzy model had a sensitivity of 43%. From the initial training set of 100 to the final training set of 764, the algorithm improved by 31% for Fuzzy and 35% for Naïve. The algorithm appeared to be learning with each additional batch of narratives added to the training set. It is possible we were approaching a threshold with the Fuzzy sensitivity, judging by the incremental gains as the training set progressed past 700 narratives. Naïve appeared to still be improving by the final training set.

Figure 2 Model Sensitivity with Increasing Size of Training Set

Model Sensitivity with Increasing Training Size
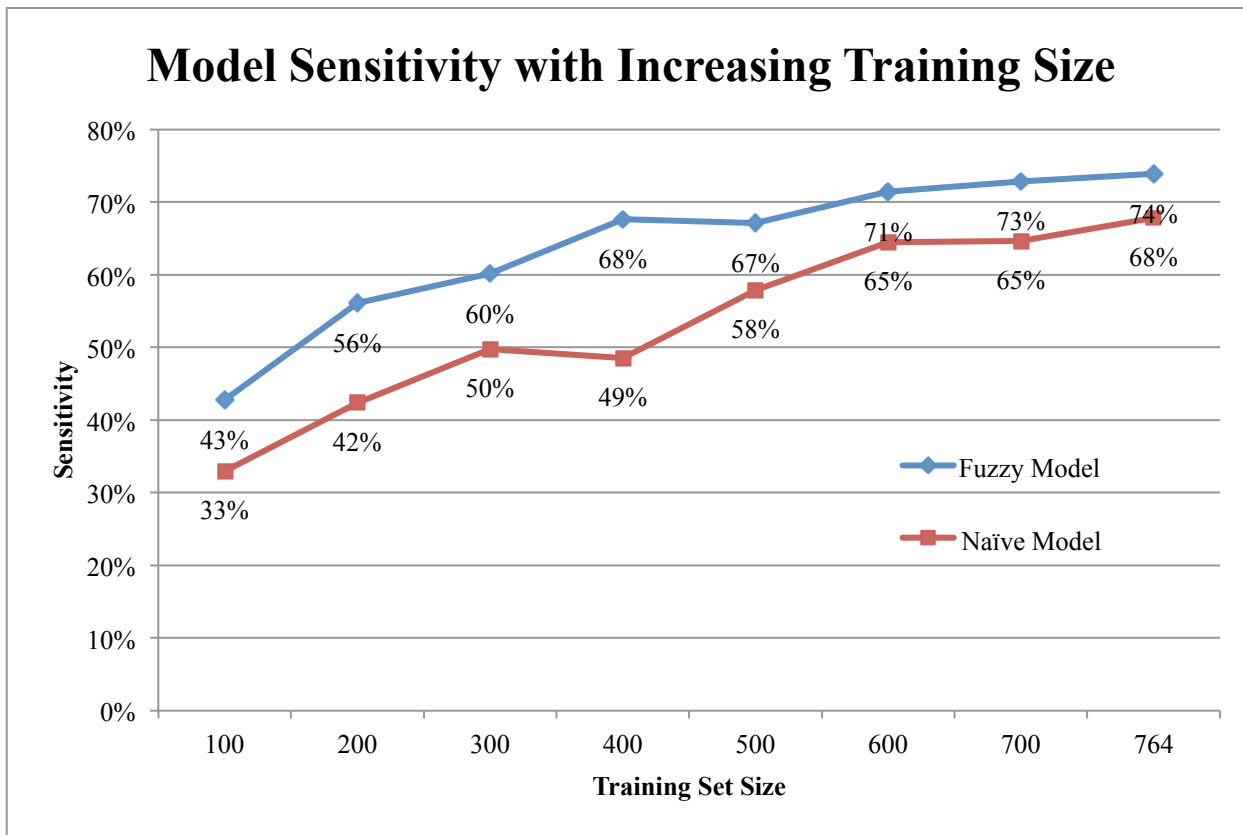
The confusion matrices were generated for each model, confirming the findings of Table 4. The Fuzzy matrix showed significant over-prediction in two categories: Fire-Burn and Fire-Electric Current. Fuzzy predicted the same number of categories for Fire-Fall (n=196), but only correctly predicted 151 of those, incorrectly assigning 45 other narratives to the Fire-Fall category. The Fire-Struck-by/Against category was overall under-predicted by Fuzzy (only predicted 172 times, as compared to the actual incidence of 184), with mistaken assignment of this category to other narratives (30 incorrectly coded). Interestingly, the Fire-Struck-by/Against Category was mistakenly assigned 20 times to narratives that were actually part of the Fire-Fall category. This is not entirely surprising, as the circumstances that led up to each of these outcomes (Fall or Struck-by) often had similar precipitating events, such as a structural collapse. With the Naive matrix, we found the categories Firearm/Ammunition and Fire-Burn to be highly over-predicted (37 to 9 for Firearm/Ammunition and 372 to 169 for Fire-Burn), which explains the low PPV. Although Naïve often under-predicted other categories, it did not have much misclassification. All other categories, including the smaller ones were relatively well predicted by Naïve as compared to Fuzzy, having higher PPV.

Receiver Operating Characteristic (ROC) curves showed better performance by the Fuzzy model in the category Firearm/Ammunition and only slightly better in the Fire-Medical Condition and Fire-CO, Smoke, Fumes. Naïve performed better in the Fire-Cutting/Piercing Instruments/Objects category, and only slightly outperformed Fuzzy in the Fire-Equipment/Machinery category. Both Fuzzy and Naïve performed equally well in the remaining categories.

Figure 3 ROC Curve for Fuzzy Bayesian Model
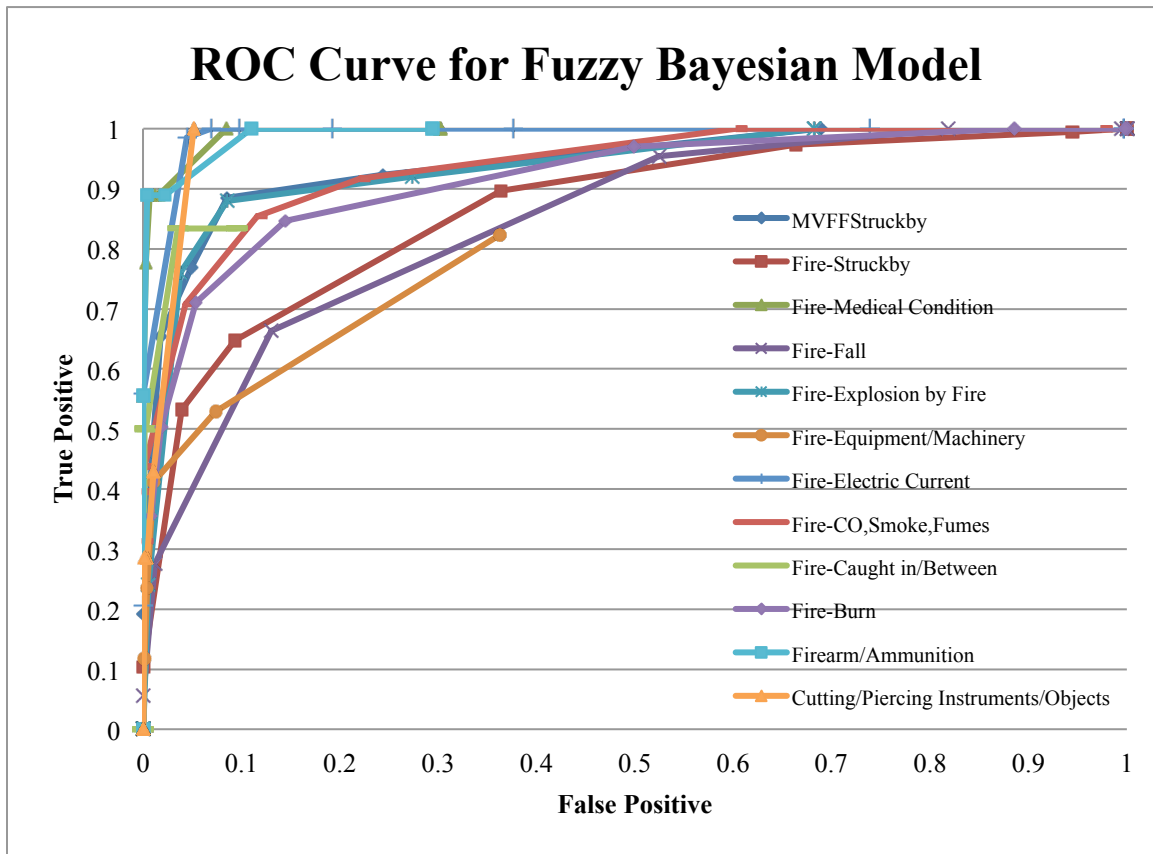
Figure 4 ROC Curve for Naïve Bayesian Model
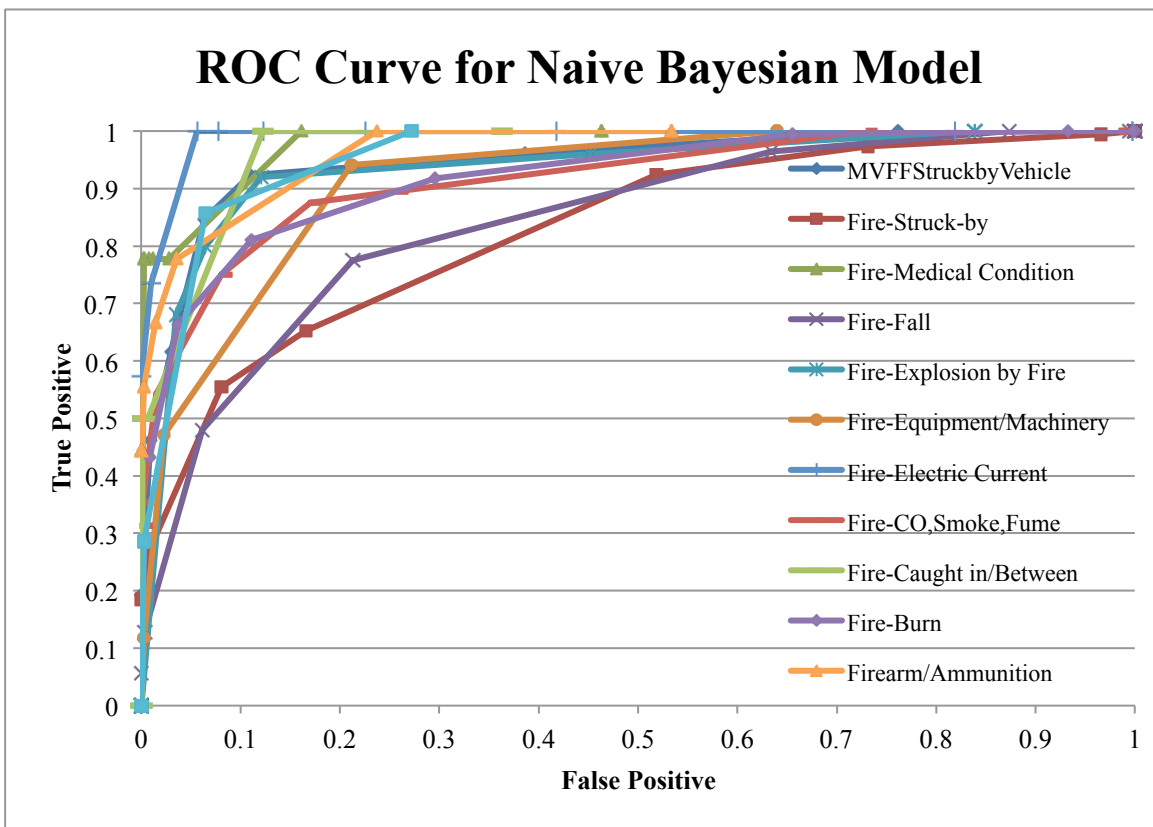
Five-fold cross validation

The five-fold cross validation of the manually coded narratives showed consistent performance of the algorithm across the varied training set. Of the 167 narratives in the prediction set, trial 3 had the highest hit rate (56.3 %) whereas the lowest hit rate was obtained in Trial 1 (50.3 %) (Table 5). Overall, the mean sensitivity was 52.8%.

Table 5 Sensitivity of 5 Cross Validation Trials

	# Correct	Sensitivity
Trial 1	84	50.3%
Trial 2	88	52.7%
Trial 3	94	56.3%
Trial 4	90	53.9%
Trial 5	85	50.9%
Mean	88.2	52.8%

Cross Validation of Prediction Set

Out of the 300 narratives within the cross-validation set, the manual coders identified 7 narratives that were not sufficiently detailed, or were not fire-related, and thus not included in the final analysis. Overall, for the 293 cases examined Fuzzy had a sensitivity of 51.9%, while the sensitivity for Naïve was about half, at 24.9% (Table 6). For those narratives within the strong category, of which Fuzzy and Naïve had the same prediction, the sensitivity was 60.2%. In the moderate and poor categories, Fuzzy performed much better, giving a sensitivity of 40.4% and 55.2% respectively.

Table 6 Cross Validation of the Prediction Set

| Prediction Strength (Proximal cause)	n	Fuzzy Correct Predictions (n)	Fuzzy Sensitivity	Naïve Correct Predictions (n)	Naïve Sensitivity
STRONG - Where fuzzy and naïve predicted the same category	98	59	60.2%	59	60.2%
MODERATE - Where fuzzy and naïve disagreed on the prediction, one had a good strength indicator, the other did not	99	40	40.4%	7	7.1%
POOR - Where fuzzy and naïve disagreed on the prediction, and both had good strength indicators associated with their predictions	96	53	55.2%	7	7.3%
OVERALL	**293**	**152**	**51.9%**	**73**	**24.9%**

Model performance by Injury Outcome

Manually coding the narratives for injury outcome yielded 215 injuries (28%) and 549 (72%) near-misses. Thus, we were able to create a new quantitative variable "Injury (yes/no)". Furthermore, applying the Bayesian models to the training set (n=764) to predict injury outcome, Fuzzy sensitivity reached 92% (data not shown).

Using this new variable, the Fuzzy model predicts the mechanism of injury with a higher sensitivity for Injury narratives (0.823) than Near-Miss narratives (0.707) (Figure 5). In general, the mechanism of injury is correctly predicted more frequently for Injury narratives than for Near-Miss narratives. Regardless, the overall sensitivity of the algorithm improved for both models, regardless of injury outcome.

Figure 5 Algorithm Performance by Injury vs. Near-Miss
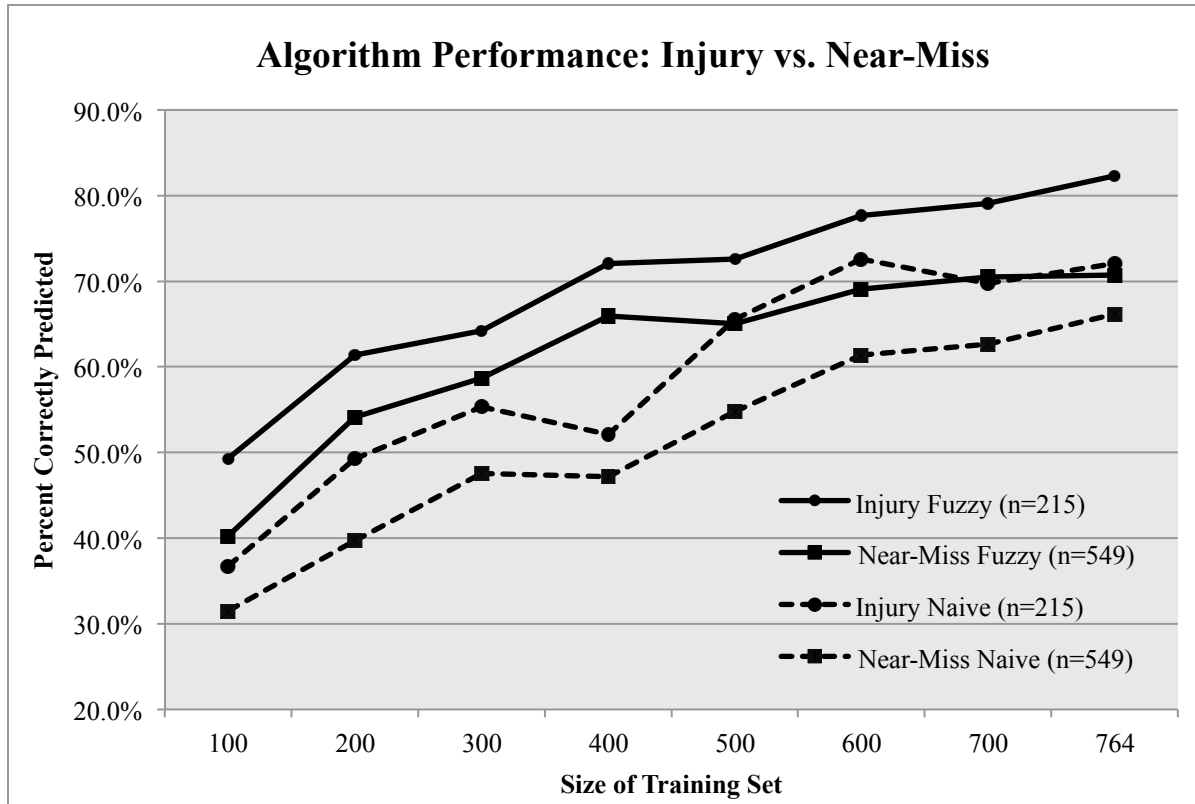
Injury Narratives

				Size of Training Set				
	100	**200**	**300**	**400**	**500**	**600**	**700**	**764**
Fuzzy Sensitivity	0.493	0.614	0.642	0.721	0.726	0.777	0.791	0.823
Naïve Sensitivity	0.367	0.493	0.553	0.521	0.656	0.726	0.698	0.721

*Denominator is total number of injury records, n=215

Near-Miss Narratives

				Size of Training Set				
	100	**200**	**300**	**400**	**500**	**600**	**700**	**764**
Fuzzy Sensitivity	0.403	0.541	0.587	0.659	0.650	0.690	0.705	0.707
Naïve Sensitivity	0.315	0.397	0.475	0.472	0.548	0.614	0.627	0.661

*Denominator is total number of near-miss records, n=549

AIM 1 DISCUSSION:

We found that TextMiner was able to correctly predict a mechanism of injury code for 74% of the narratives using the Fuzzy model and 68% of the narratives using the Naïve model. Injuries were correctly predicted at a higher rate (Fuzzy 0.82, Naïve 0.72) than near-misses (Fuzzy 0.71, Naïve 0.66). Overall, our sensitivity is comparable to the results of Lehto et al. (2009), which saw sensitivity between 70-80% for Naïve and between 64-78% sensitivity for Fuzzy when analyzing injury narratives. To our knowledge, this study is the first of its kind to successfully use machine learning algorithms to assign mechanism of injury codes to near-miss narratives. Previous research has only looked at injury narratives.

Manual coding of near-miss narratives

Coding near-miss narratives is not as straight-forward as coding actual injury narratives. To do so, we must look for the most likely outcome that could have occurred, recognizing that one decision must be made when multiple outcomes are possible. Such decision-making is time-consuming and therefore expensive in terms of human resources. In the Methods, we discussed the importance of and adherence to the coding order of operations: Injury Outcome first, then Proximal Cause, then Precipitating Cause. The challenge of coding near-miss events is that it is often difficult to determine a finite point from which to work backward because there is no injury. For this reason, starting with "Did an injury happen (yes/no)?" was invaluable in helping us determine the mechanism of that injury (or near-miss) and then assess what started the chain of events in motion (precipitating). However, there were times when the coders were often forced to speculate on the outcome and select a code. This was where the majority of disagreement between codes occurred. If too little information was provided or the report did not provide a clear understanding of the potential outcome, we coded it as NOC (not otherwise classifiable) and omitted it from the analysis. Table 7 (p. 24) presents two contrasting narratives that illustrate the challenges of coding.

Table 7 Example Narratives: Near-miss vs. Injury

INJURY OUTCOME: Near-miss	INJURY OUTCOME: Injury
*"On arrival, there was fire showing on the second floor "B" side of a two and one half story wood-frame residential structure. We had been operating a two and one half inch attack line for approximately ten minutes. As Division Two Commander, I felt at that time that we were beginning to lose progress. On orders of the Incident Commander, orders were given to immediately evacuate the second floor. As Division Two Commander, all crews were evacuated, excluding myself and two other crew members in a final effort, despite the Incident Commanders orders. Upon evacuating, after the Incident Commander's second order to evacuate, we observed the second floor **flashover** and **collapse**."*	*"This was an extremely cold night, 0400, -4 degrees. First due companies found a well involved dwelling. The Incident Commander call for a 2nd alarm. I was the lieutenant to the 4th due Engine Company. We were told to search the second floor for victims. We had cars in the driveway and kids' toys in the driveway as well. No one was coming forward to say everyone was out of the building. The main body of fire was knocked down when we made access to the second floor bathroom. One of my crew asked me if I had checked the bath tub because kids hide in tubs. **I reach into the tub just as its weight caused the floor in the bathroom to collapse.***
Coders' determination:	*I don't remember falling but I do remember hearing the Mayday from my crew and wondering to myself "Oh My god who's hurt". I had landed on the first floor on my back with my SCBA on. The crew on the first floor got me up and out of the building. I had no obvious injury and went back to work, until three hours later. I was back at the station making out reports, when a low back spasm caused me excruciating pain. I was taken to the hospital with torn muscles in my lumbar area. I was out of work for 6 months. The people that lived in the house were outside in a squad car. The patrolman never told us everyone was out of the building."*
PRECIPITATING CAUSE: Fire-Collapse
PROXIMAL CAUSE: Fire
In this narrative, the potential resultant injuries included possible struck-by or fall from the collapse, and/or burn injuries from the flashover. The collapse was the most precipitating of causes, while the generic code of "Fire" as the proximal cause indicated that too many possible outcomes were described, preventing a single cause to be identified.	**Coders' determination:**
	PRECIPITATING CAUSE: Fire-Collapse
PROXIMAL CAUSE: Fire-Fall
In this narrative, a clear sequence of events is presented. Working backwards from the injury (torn back muscles), he suffered a fall (proximal), from collapse of second floor bathroom (precipitating).

Given the challenges of coding near-misses compared to injuries we were pleased with our level of substantial agreement. We obtained 79% agreement, which is comparable to research by Lehto et al. (2009) showing 75% agreement. Percentage agreement and kappa statistics were in the lower range of previous studies (Bondy *et al.* 2005, Lehto *et al.* 2009, Lombardi *et al.* 2009, Marucci-Wellman *et al.* 2011). Therefore, we conclude that this method—which has been rigorously applied to injuries—is substantiated for scenarios with less definitive outcomes like near-misses.

Structure of the data system
In NFFNMRS, reporters are asked to "Describe the event," allowing them to say anything. Narratives often begin with information about arrival and staging which are events that precede the beginning of the chain of events leading up to an injury or near-miss. It is important to note that the "Describe

the event" field does not ask specific questions about any injuries that did happen or could have happened. This is different than other data systems like the National Health Interview Survey (NHIS) which asks "How did your injury on [date] happen? Please describe fully the circumstances or events leading to the injury and any objects, substances, or other people involved." In addition to asking how the injury occurred, the NHIS also asks a series of specific prompts to seek for more detailed information for certain causes such as whether the injured individual was in a motor vehicle, on a bike, scooter, skateboard, skates, skis, horse, etc., a pedestrian who was struck by a vehicle such as a car or bicycle, in a boat, train, or plane, suffered a fall, or burned or scalded by substances such as hot objects or liquids, fire, or chemicals (National Center for Health Statistics, 2009). The average length of narratives within the NFFNMRS dataset is quite long (mean word count 216), as compared to other datasets. In contrast, narratives from the NHIS contain 11 words on average (Wellman *et al.* 2004). Furthermore, the time required to manually code our initial 1,000 narratives was approximately 25 hours per coder, with an additional 25 hours required for reconciliation of these 1,000 narratives. Using worker's compensation narratives of approximately 20 words, Bertke and colleagues (2012) stated that it took them 10 hours to code 2,400 worker's compensation claims—which is 2.4 times the number of narratives we coded, in less than half of the time. We observed that coding of lengthy narratives—especially those without known injury outcome—is very time consuming and requires extensive human resources. Therefore the algorithm's high performance is especially welcome for narratives that emanate from generic prompts such as "Describe the event."

Performance of autocoding
Considering that coding of near-miss narratives via automated methods has not been previously described in the literature we were pleased with the performance level of the algorithm on near-miss narratives, reaching above 70% specificity.

The higher performance of the Fuzzy model as compared to the Naïve model was not too surprising, given the longer narratives in our dataset. Previous research has used much shorter narratives and seen exceptional performance by the Naïve model—particularly because with less words in each narrative, there are fewer opportunities for a strong predictive word to outweigh the other words within the narrative. Categories with fewer narratives tended to be better predicted by the Naïve model, likely because it took into account all words, rather than picking words with the single strongest predictor (as in Fuzzy). Previous research done with the TextMiner software has been applied to shorter narratives, predominately using the Naïve model.

To further elucidate why the Fuzzy model was performing better, we checked to see if there was any evidence of overfitting of the data. We analyzed the difference in correct predictions between the training set and prediction set, for both the Fuzzy and Naïve models and found no indication that the Fuzzy model was overfitting the data, suggesting that the Fuzzy model truly does perform better with this particular dataset.

The results of the Cross-validation showed that when the Fuzzy and Naïve models both predicted the same code, the agreement of the autocoding to the manual codes reached 60.2%. In using a similar technique to filter cases for manual review, Wellman and colleagues (2011) reported a

Fuzzy and Naïve agreement of 64%. Applying both the Fuzzy and Naïve models to a dataset could be another way of optimizing the performance and accuracy of autocoding.

The process of adding narratives to the training set in increments of 100 showed marked improvement, suggesting that the algorithm was learning with each addition. It did not appear that the algorithm had yet reached a threshold, suggesting that addition of more cases (beyond 764) will result in improved prediction rate by the models. The work of Bertke et al. (2012) showed increasing improvement in the sensitivity as the training set increased up to 1,000 (with remaining cases as prediction set n=1,240), with marginal returns beyond that.

Using additional modifications such as paired words, word sequences, morphs, and drop words lists would likely improve the hit rate. In fact, in a preliminary analysis using paired words and 3-word sequences, we saw an increase in prediction success by the Fuzzy model (82% and 85% respectively, data not shown). However, this indicates that with minor modifications, the predictive capability of the algorithm can improve to a significantly higher level of sensitivity, thereby reducing the amount of narratives that would need manual review.

Creation of additional quantitative data elements
Applying the Bayesian models enabled us to create two new quantitative data elements: injury outcome and mechanism of injury. This enriches the analysis of existing quantitative data in the NFFNMRS because we can look at differences between near-misses and injuries, and construct hazard scenarios.

AIM 2

AIM 2 METHODS:
For **AIM 2**, we examined the relationship between the manually coded injury outcome field (injury yes/no) with the quantitative field on the NFFNMRS reporting form, "Contributing Factors." The "Contributing Factors" section of the report lists a variety of items related to the firefighters' functioning during the incident (e.g. Command, Communication, Protocol).This analysis was to accomplish our **third objective**: *to reveal and interpret patterns in the distribution of contributing factors within each mechanism of near miss/injury*. We hypothesized that the 20 categories of contributing factors were ill-defined, and therefore poorly understood by reporters. For example, it was not clear how a firefighter would choose between "Protocol" and "Standard Operating Procedure" from the Contributing Factors list. The NFFNMRS itself offers no definition.

Secondly, we conducted a Latent Class Analysis to discern whether the 21 Contributing Factors could be reduced to aggregated latent scores, with the goal to examine the latent scores and their relationship with injury outcome. The Latent Class Analysis was completed using SAS Software (Cary, North Carolina). We used the Proc LCA procedure (downloaded from http://methodology.psu.edu/downloads/proclcalta) with the setting of RHO Prior = 1, as recommended within the Proc LCA User's Guide (Lanza, Dziak, Huang, Xu, & Collins, 2011). The model was run looking for fit with a 3-class model. Each narrative was assigned a score, which was essentially the probability of that narrative being in each of the three latent classes. We ran two

separate logistic regressions: 1) modeling injury with the 21 Contributing Factors, and 2) modeling injury with the three latent class scores.

AIM 2 RESULTS:

A cross-tabulation of each Contributing Factor (CF) with the manually coded Injury Outcome showed little variation of near-miss narratives to injury-narratives. The proportion of near-misses to injuries was similar for nearly all of the CF categories (Table 8).

Table 8 Crosstab analysis of Contributing Factors and Injury Outcome

Contributing Factor	Near Miss	Injury	Total	Near-Miss Row Percent	Injury Row Percent
Situational Awareness	387	121	508	76.2%	23.8%
Decision Making	316	93	409	77.3%	22.7%
Human Error	242	83	325	74.5%	25.5%
Individual Action	195	70	265	73.6%	26.4%
Communication	154	58	212	72.6%	27.4%
Command	120	34	154	77.9%	22.1%
Training Issue	96	25	121	79.3%	20.7%
Equipment	69	30	99	69.7%	30.3%
Procedure	72	23	95	75.8%	24.2%
Accountability	60	31	91	65.9%	34.1%
Teamwork	58	18	76	76.3%	23.7%
SOP/SOG	54	20	74	73.0%	27.0%
Other	49	19	68	72.1%	27.9%
Staffing	44	19	63	69.8%	30.2%
Weather	40	12	52	76.9%	23.1%
Fatigue	29	13	42	69.0%	31.0%
Task Allocation	28	12	40	70.0%	30.0%
Protocol	17	4	21	81.0%	19.0%
Unknown	12	6	18	66.7%	33.3%
Horseplay	2	0	2	100.0%	0.0%
Control & Accountability	0	1	1	0.0%	100.0%
Total	**2,044**	**692**	**2,736**	**74.7%**	**25.3%**

With the Latent Class Analysis, each of the three latent class scores are probabilities, all weights for each class add up to 1. One was dropped due to co-linearity. Neither model was good (Table 9). The coefficients were not significant, and overall model fit was poor.

Table 9 Latent Class Analysis of Contributing Factors

Latent Class A 40.2%		Latent Class B 31.0%		Latent Class C 28.8%	
0.66	*Human Error*	0.60	*Situational Awareness*	0.53	*Situational Awareness*
0.52	*Situational Awareness*	0.60	*Decision Making*	0.20	Other
0.51	*Decision Making*	0.46	*Communication*	0.19	Decision Making
0.51	*Individual Action*	0.45	*Command*	0.14	Equipment
0.20	Communication	0.26	Accountability	0.14	Human Error
0.15	Training Issue	0.20	Individual Action	0.10	Training Issue
0.14	Equipment	0.17	Human Error	0.09	Weather
0.11	Procedure	0.16	SOP/SOG	0.09	Individual Action
0.09	Teamwork	0.14	Training Issue	0.08	Procedure
0.07	SOP/SOG	0.13	Procedure	0.08	Staffing
0.07	Command	0.12	Teamwork	0.07	Unknown
0.06	Fatigue	0.10	Staffing	0.05	Communication
0.04	Staffing	0.06	Task Allocation	0.03	Task Allocation
0.04	Weather	0.05	Weather	0.03	Fatigue
0.04	Task Allocation	0.04	Fatigue	0.02	Teamwork
0.03	Accountability	0.03	Equipment	0.01	SOP/SOG
0.03	Protocol	0.03	Protocol	0.01	Accountability
0.03	Other	0.02	Other	0.01	Protocol
0.00	Unknown	0.00	Control & Accountability	0.01	Horseplay
0.00	Horseplay	0.00	Unknown	0.00	Command
0.00	Control & Accountability	0.00	Horseplay	0.00	Control & Accountability

AIM 2 DISCUSSION:

Since no recognizable pattern emerged in the distribution of near miss/injury mechanisms with
contributing factors, indicating that there was no discernment in their selection by reporters, we
recommended to the IAFC that the NFFNMRS reporting form be modified to improve data quality
and utility. We suggested prompting for contributing factors by changing the list to a free text field
(and then using the TextMiner algorithm to code emerging factors), or adding drop down definitions
for each factor once they were clearly and discretely defined. The later would require cognitive
testing to see if firefighters could clearly discriminate among categories. These recommendations are
currently under consideration by IAFC.

AIM 3

AIM 3 METHODS:

Within the 1,000 manually coded narratives, we analyzed the distribution of injuries with respect to
each of the categories listed under "Loss Potential" on the NFFNMRS reporting form. The "Loss
Potential" section allows reporters to indicate whether a near miss could have resulted in life
threatening injury, lost time injury, minor injury, etc. Given that the reporting form has no field for

identifying injuries, we hypothesized that in cases of actual injury the "Loss Potential" section is being used to describe the injury rather than unrealized potential for harm. This analysis enabled us to achieve our **fifth objective**: *to disentangle true near misses from injury events captured by NFFNMRS, clarifying how the "Loss Potential" section of the report is being used in cases of actual injury*. Within the Loss Potential variable, reporters to the system could enter multiple loss potential answers, such that one narrative can be associated with multiple Loss Potential responses. We transformed the dataset such that for each selection of Loss Potential for a particular narrative, that narrative is repeated multiple times with a different Loss Potential category represented. Thus, the same narrative could be entered as multiple data points could in this dataset.

AIM 3 RESULTS:

The analysis of Loss Potential showed little variation between the Loss Potential categories and the injury outcome. While injuries represented 23.5% of the narratives, the Loss Potential category "Life Threatening Injury" was only selected in 18.4% of injury narratives, as compared to 81.6% of near-miss narratives. In fact, near-miss narratives more frequently had "Lost Time Injury" and "Minor Injury" listed as potential outcomes in 30.5% of narratives and 32.1% of narratives, respectively (Table 10).

Table 10 Crosstab Analysis of Loss Potential with Injury Outcome

Loss Potential Code	Near-Miss	Injury	Total	Near-Miss Row Percent	Injury Row Percent
Life Threatening Injury	526	119	645	81.6%	18.4%
Lost Time Injury	333	146	479	69.5%	30.5%
Minor Injury	199	94	293	67.9%	32.1%
Property Damage	175	37	212	82.5%	17.5%
Environmental	15	2	17	88.2%	11.8%
Unknown	18	1	19	94.7%	5.3%
Other	41	2	43	95.3%	4.7%
Total	1,307	401	1,708	76.5%	23.5%

We also compared our team's manual coding of narratives to that of the report reviewers. In analyzing the Report Type variable (as identified by the IAFC Reviewer) and Injury Outcome, as identified through our manually coded set of Fire-Emergency Event narratives (n=764), we found an additional 52 incidents of injury (Table 11). In the category "Near-Miss Event, no injury or damage," we found 38 cases of injury, which were not identified by the Reviewer. Additionally, in the "Potential Hazard/Unsafe Act" category, 13 injuries were present—which suggests that the way in which this data system is structured (single-select categories that are not mutually exclusive) prevents accurate capture of injury events. Interestingly, we found fewer actual injuries in the two Report Type categories that collect injury, "Near-Miss Event, with Injury Only," and "Near-Miss Event, with Injury and Property Damage." We identified these two categories to have 22 fewer injuries, as compared to the Reviewer's assessment of the narrative.

Table 11 Comparison of Report Reviewer Outcome Assessment with Research Team

Code	Report Type (as identified by Reviewer)	Report Type Code by NFFNMRS Report Reviewers	Coded as Injury by Research Team (JAT, AVL, GSS)	Coded as Near-Miss by Research Team (JAT, AVL, GSS)
0	Near-Miss Event, no injury or damage	431	38	393
1	Potential Hazard/Unsafe Act	68	13	55
2	Suggestion/Observation/Other	9	0	9
3	Near-Miss Event, with injury only	158	139	19
4	Near-Miss Event, with property damage only	43	1	42
5	Near-Miss Event, with injury and property damage	14	11	3
	Total Number of Reports	**764**		

AIM 3 DISCUSSION:

We concluded that Loss Potential is not a good proxy for whether or not an injury occurred. We compared what the reporter said under Loss Potential with what the report reviewer concluded about injury occurrence based on a follow-up conversation with the reporter. Report reviewers concluded injuries happened in 22.5% of Fire-event reports. This is a stark contrast to firefighters reporting injury 81% of the time when using the Loss Potential categories (add first three categories of Table 8). We recommended to IAFC that the categories currently used by report reviewers replace the Loss Potential categories and that all categories be mutually exclusive (only one choice allowed).

We recommended that Loss Potential be changed to a new variable called "What happened?" and the answer categories be changed to those we modified from the Reviewer "Report Type" variable:

> ___*Suggestion/Observation/Other*
> ___*Potential Hazard/Unsafe Act*
> ___*Near-Miss Event, no injury or property damage*
> ___*Property damage only*
> ___*Injury only*
> ___*Both Injury and Property Damage*

This recommendation is currently under consideration by IAFC.

Conclusion

In this study of narratives from the fire service we were able to successfully apply the Fuzzy and Naïve models to injury and near-miss narratives, which were much longer than those that have previously been investigated. While both models had relatively high sensitivity, Fuzzy proved to be the more agile model for very long narratives.

We trained the algorithm to assign a mechanism of injury and an injury outcome for each narrative. This process resulted in the creation of two new quantitative data elements that will empower more in-depth analyses of the National Fire Fighter Near Miss Reporting System.

Previous studies have the benefit of their short narratives emanating from specific questions about how the injury occurred. That the near miss narratives had fairly vague instructions to "describe the

event," and that the machine learning methods were able to assign a specific mechanism of injury code is a testament to the power of Bayesian models. An important point is that no effort was made in the current study to optimize the predictive models used. Additional steps could be taken that would be likely to improve the performance of both models, such as increasing the sample size. Other steps are also likely to lead to significant improvements, such as trimming the word set by dropping common noise words to improve performance of the Naïve Bayes model, or using word combinations and sequences to increase the sensitivity of the Fuzzy model.

This study lays the groundwork for a future grant application to fully characterize firefighter near miss hazard scenarios – a scientific classification that has not been accomplished. In order to establish hazard scenarios, we first needed to populate the NFFNMRS with the injury mechanisms developed through the current study. The future goal is to develop a comprehensive taxonomy for all hazard scenarios contained in the near miss database and to use this taxonomy to further identify high priority areas for safety policy and practice. In addition to publishing our findings in the peer-reviewed literature, such future funding will allow us to develop practical tools such as a "Top 10 List" of causes of firefighter injury which can be disseminated through the major fire service publications, the NFFNMRS website, the NIOSH website, and other relevant forums. The Centers for Disease Control and Prevention's "Ten Leading Causes of Death and Injury" is a useful model for this type of tool (http://www.cdc.gov/injury/wisqars/LeadingCauses.html). A preliminary summary of the mechanisms of injury for near-misses and injuries in descending order is presented in Table 12.

Table 12 Distribution of Mechanism of Injury for Proximal Cause, Fire Events only

Training Set (n=764)					
Near-Miss Narratives	**n**	**%**	**Injury Narratives**	**n**	**%**
Fire - Fall	141	26	Fire - Burn	76	34
Fire - Struck-by, Against	139	26	Fire - Fall	55	25
Fire - Burn	93	17	Fire - Struck-by/Against	45	20
Fire - Electric Current	59	11	Fire - CO, Smoke, Fumes	17	8
Fire - CO, Smoke, Fumes	31	6	Fire - Electric Current	9	4
MV - FF Struck-by	26	5	Fire - Explosion	5	2
Fire - Explosion by Fire	20	4	Cutting/Piercing Object	5	2
Fire - Equipment/Machinery	17	3	Fire - Caught-in, Between	2	1
Fire - Medical Condition	9	2	Firearm/Ammunition	1	0
Firearm/Ammunition	8	1	**Total**	**215**	**100**
Fire - Caught-in, Between	4	1			
Cutting/Piercing Object	2	0			
Total	**549**	**100**			

Comparison of the near-misses and injury causes presented in the table should be interpreted with caution because the end point in the chain of events is different for near-misses and injuries. For example, burns represent 17% of the near-misses, and 36% of the injuries. This may indicate when a burn hazard is present, it is more likely to lead to an injury than other hazards (like fall risk, which

has equal percentages in both near-misses and injuries). Similarly, electrocutions were 11% of all near-misses, yet there were no electrocutions on the injury side. This may be because electrocution hazards are identified earlier in the chain of events and are intercepted before they can cause injury. It is possible that since electrical hazards are usually outside of a building, whereas burn hazards are usually within, the built environment causes the thermal energy of fire to transmit to the fire fighter with less opportunity for intervention or escape. Also, electricity is usually a very contained hazard, while fire is not. The differences in near-misses and injuries in terms of the frequency of causation merit deeper investigation.

"Fire – Medical Condition" included events such as asthma attack or myocardial infarction (MI). While these are not traditionally recognized as injuries, they were included because of the deep concern among firefighters of their impact. In the United States, MIs cause half of all firefighter line of duty deaths. Events included in this analysis were acute events exacerbated by presence at a working fire.

References

Alamgir, H, Yu, S, Gorman, E, Ngan, K, & Guzman, J: [2009]. Near miss and minor occupational injury: Does it share a common causal pathway with major injury? American journal of industrial medicine, 52:(1), 69-75.

Barach, P, & Small, SD: [2000]. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. BMJ, 320:(7237), 759-763.

Bondy, J, Lipscomb, H, Guarini, K, & Glazner, JE: [2005]. Methods for using narrative text from injury reports to identify factors contributing to construction injury. American Journal Industrial Medicine, 48:(5), 373-380.

Bunn, TL, Slavova, S, & Hall, L: [2008]. Narrative text analysis of Kentucky tractor fatality reports. Accid Anal Prev, 40:(2), 419-425.

Connell, L: [2008]. Cross-Industry Applications of a Confidential Reporting Model. Aviation Safety Reporting System, National Aeronautics and Space Administration (62). Retrieved on August 2, 2013, from
http://asrs.arc.nasa.gov/docs/rs/62_Cross_Industry_Applications_of_Reporting_Model.pdf

Helmreich, RL: [2000]. On error management: lessons from aviation. BMJ, 320:(7237), 781-785.

Kaplan, HS: [2003]. Benefiting from the "gift of failure": essentials for an event reporting system. J Leg Med, 24:(1), 29-35.

Kaplan, HS, & Fastman, BR: [2003]. Organization of event reporting data for sense making and system improvement. Qual Saf Health Care, 12, Ii68-Ii72.

Karter, MJ, & Molis, JL: [2007]. US Firefighter Injuries 2006. Quincy, MA: National Fire Protection Association.

Lanza, ST, Dziak, JJ, Huang, L, Xu, S, & Collins, LM: [2011]. PROC LCA & PROC LTA Users' Guide. Retrieved January 3, 2013, from http://methodology.psu.edu/downloads/proclcalta

Leape, LL: [2002]. Reporting of adverse events. New England Journal of Medicine, 347:(20), 1633-1638.

Lehto, M, Marucci-Wellman, H, & Corns, H: [2009]. Bayesian methods: a useful tool for classifying injury narratives into cause groups. Inj Prev, 15:(4), 259-265.

Lincoln, AE, Sorock, GS, Courtney, TK, Wellman, HM, Smith, GS, & Amoroso, PJ: [2004]. Using narrative text and coded data to develop hazard scenarios for occupational injury interventions. Inj Prev, 10:(4), 249-254.

Lipscomb, HJ, Glazner, J, Bondy, J, Lezotte, D, & Guarini, K: [2004]. Analysis of text from injury reports improves understanding of construction falls. J Occup Environ Med, 46:(11), 1166-1173.

McGreevy, MW: [2001]. Searching the ASRS Database Using QUORUM: Keyword Search, Phrase Search, Phrase Generation, and Phrase Discovery.

National Fire Fighter Near-Miss Reporting System: [2010]. Near Miss Calendar and 2009 Annual Report. Fire Rescue Magazine.

National Institute for Occupational Safety and Health: [2012]. The NIOSH Fire Fighter Fatality Investigation and Prevention Program. Retrieved July 25, 2013, from http://www.cdc.gov/niosh/docs/2012-169/pdfs/2012-169.pdf

Rivard, PE, Rosen, AK, & Carroll, JS: [2006]. Enhancing patient safety through organizational learning: Are patient safety indicators a step in the right direction? Health Serv Res, 41:(4p2), 1633-1653.

Smith, GS, Timmons, RA, Lombardi, DA, Mamidi, DK, Matz, S, Courtney, TK, & Perry, MJ: [2006]. Work-related ladder fall fractures: identification and diagnosis validation using narrative text. Accid Anal Prev, 38:(5), 973-980.

Stout, N: [1998]. Analysis of narrative text fields in occupational injury data Occupational Injury: Risk, Prevention and Intervention: Taylor & Francis pp. 15-20.

Sumwalt, R: [1993]. The Sterile Cockpit 24. ASRS Directline(4).

Wright, L, & van der Schaaf, T: [2004]. Accident versus near miss causation: a critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors. J Hazard Mater, 111:(1-3), 105-110.

Publications.

A manuscript describing the results of AIM 1 was submitted in June 2013 to the journal *Accident Analysis and Prevention*. On September 17, 2013, this paper was accepted for publication (AAP3289) and is currently in press. It will be published in open access format in order to ensure the widest dissemination and to remove any access barriers the fire service could experience.

Inclusion of gender and minority study subjects. Gender and minority are not captured in this dataset, thus we have no information about the demographic characteristics of the data.

Inclusion of Children. This research did not involve children.

Materials available for other investigators. The reports may be obtained from the IAFC NFFNMRS website at www.firefighternearmiss.com.

Closeout Document 2: The Final Financial Status Report (FSR)

All financial documents will be submitted by Drexel University's Office of Research Administration.
Contact: Ms. Patricia Groshon, CRA 215-895-2946

CDC Procurement & Grants Office - Branch V Equipment Inventory Listing

Report Date: September 30, 2013   Grant Number: 5R03OH009984-02

Project Title: Near Miss Narratives from the Fire Service: A Bayesian Analysis   Project Period: 09/1/11-08/31/2013

Grantee Name: Drexel University   Project Officer: Mary Pat Shanahan

Grants Management Officer: Larry Guess   Grants Specialist:

Description of Item: i.e. pH Meter	Mfr.[1] i.e. Fischer	Serial Number	Quantity	Condition[2]	Location[3]	Purchase Cost	Date Received [mm/dd/yyyy]
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[1]Mfr. (Manufacturer)

[2]Condition: (Excellent) (Good) (Fair) (Poor) (Inoperable)

[3]Location: complete physical address

For Government Use Only, not to be completed by the Grantee

Property Administrator & PO Disposition Recommendation and Instructions:

Description of Item	Disposition[1]	Address[2]
[Copy from above]	_____Transfer Title _____Retain and Compensate Awarding Agency _____ Return to Program Office _____Other (explain)	Attn: [Project Officer] CDC / NIOSH 1600 Clifton Road, NE MS E-74 Atlanta, GA 30329-4018
[Copy from above]	_____Transfer Title _____Retain and Compensate Awarding Agency _____ Return to Program Office _____Other (explain)	

[1]Check the appropriate disposition

[2]CDC Warehouse is the central receiving point for delivery of all non-hazardous and non-perishable supplies and equipment, CDC –AM–2004-03, update 2010