Using narrative text and coded data to develop hazard scenarios for occupational injury interventions

A E Lincoln, G S Sorock, T K Courtney, H M Wellman, G S Smith, P J Amoroso

Objective: To determine whether narrative text in safety reports contains sufficient information regarding contributing factors and precipitating mechanisms to prioritize occupational back injury prevention strategies.

Design, setting, subjects, and main outcome measures: Nine essential data elements were identified in narratives and coded sections of safety reports for each of 94 cases of back injuries to United States Army truck drivers reported to the United States Army Safety Center between 1987 and 1997. The essential elements of each case were used to reconstruct standardized event sequences. A taxonomy of the event sequences was then developed to identify common hazard scenarios and opportunities for primary interventions.

Results: Coded data typically only identified five data elements (broad activity, task, event/exposure, nature of injury, and outcomes) while narratives provided additional elements (contributing factor, precipitating mechanism, primary source) essential for developing our taxonomy. Three hazard scenarios were associated with back injuries among Army truck drivers accounting for 83% of cases: struck by/against events during motor vehicle crashes; falls resulting from slips/trips or loss of balance; and overexertion from lifting activities.

Conclusions: Coded data from safety investigations lacked sufficient information to thoroughly characterize the injury event. However, the combination of existing narrative text (similar to that collected by many injury surveillance systems) and coded data enabled us to develop a more complete taxonomy of injury event characteristics and identify common hazard scenarios. This study demonstrates that narrative text can provide the additional information on contributing factors and precipitating mechanisms needed to target prevention strategies.

MODELS OF OCCUPATIONAL INJURY CAUSATION

During 2002 there were over 4.4 million non-fatal injuries in private industry workplaces in the United States. To have an additional impact on workplace safety, new approaches to prevention need to be developed. As Meehan suggests, “Safety professionals... must strive to address the myriad factors that lead to occupational injuries and illnesses. The injured body part is simply the final manifestation of whatever went wrong before the injury-producing event... The focus should be determining what elements combined to produce the [incident] fall.”

Feyer and Williamson proposed a model of causality in occupational accidents that addresses an array of causal factors, their relative sequential relationships, and the relative importance of factors in accident causation. Davies et al advanced the model by recognizing that “…in reality identical accidents are rare and common factors can only be identified by detailed structuring of the data”. Their structured data collection technique captured the components of injury events in a consistent format. Our study extends the use of structured data collection to facilitate analysis.

The statistical method of analysis classifies data about a group of incidents into various categories and bases corrective actions on the most frequent patterns of occurrence. Taxonomies represent a graphical form of this classification and have been used previously to investigate occupational fatalities associated with cranes. “The taxonomic process involves ... classification of data into hierarchical groups according to common patterns and individual differences... The aim is to paint a broad picture of what exists and to indicate the relative importance of different phenomena according to how frequently they occur”.

Another very useful tool for investigating safety report data is the development of hazard scenarios. Drury and Brill recognized that accidents involving a certain product could be classified into major groups, called “hazard patterns”. They recommend grouping events by injury, contingency, or human behavior to identify prototypical accidents describing the victims, products involved, environment, and task. They suggest that deriving hazard scenarios is useful if it results in (A) no more than six scenarios that account for at least 90% of the events, and (B) each scenario results in the identification of at least one feasible and effective intervention strategy.

SAFETY REPORTS: THE BASIS FOR UNDERSTANDING CAUSATION

Safety reports are detailed investigations of workplace injuries that report the basic facts about an incident with an eye toward prevention of future similar injuries. Investigations typically document information regarding the injured employee, the job activity or exposure at the time of injury, and the mechanism, nature, and severity of injury.

Safety reports often include a narrative text field that allows an injured worker, supervisor, or investigator to briefly describe the circumstances of the injury event. Researchers
have strongly encouraged combining narrative text available in safety reports with coded variables to better direct the development of interventions for injury prevention.6–10

OBJECTIVES
The aim of this paper is to determine whether the existing narrative text in a sample of Army safety reports includes sufficient information regarding contributing factors and precipitating mechanisms to complement coded data and enable us to identify hazard scenarios associated with occupational back injury among truck drivers.

METHODS
Our approach involved four distinct components: (1) identify the essential data elements within coded and narrative data to adequately characterize a series of back injuries; (2) reconstruct the injury events using a template to describe the sequence of events in a standardized way; (3) develop taxonomies of the sequences according to essential data elements; and (4) identify hazard scenarios that represent common injury mechanisms and priorities for developing interventions.

Study population
The study population was derived from all mishap investigations reported to the United States Army Safety Center ( Ft Rucker, AL) during the period 1987 to 1997. In order to focus our analysis on a specific occupational group with broad generalizability to civilians, we selected cases with a military occupational specialty of motor transport operator (occupational code 88M) (n = 1585), roughly equivalent to a commercial trucker. Army motor transport operators are responsible for a wide variety of tasks such as routine vehicle maintenance, loading and unloading of cargo, and driving extended distances. In response to several recent studies addressing back pain among professional drivers,11 12 we further limited our population to those where the back was the primary body part injured (n = 130, 8.2%) and the injury occurred while on duty, which resulted in 94 cases, 5.9%.

Data source
The United States Army Safety Center collects data relating to non-battle related accidents/mishaps (that is, unintentional injuries and events) via Form DA 285, which excludes intentional/violent injuries, those resulting from battle/hostile actions, homicides, and suicides, as well as non-occupational diseases. Reports are required in the case of an injury that results in lost time from work, hospitalization, or significant economic losses or property damage of at least $10 000. Either a representative from the injured soldier’s unit or a Safety Center investigator typically completes the report.

These injury reports and pertinent cost information are stored electronically in the Army Safety Management Information System (ASMIS). The cumulative data from these safety databases are used to track and compare frequencies and rates of ground and aviation accidents from year to year.13 From 1980 to 1998, there were over 133 000 ground and aviation accident reports contained in the ASMIS with detailed information that includes personal protective equipment use, drug and alcohol involvement, environmental conditions, and up to 500 words of free text describing the event.14 Typically, however, narratives of 50 to 75 words are found in a given report.

ASMIS data were obtained through the Total Army Injury and Health Outcomes Database (TAIHOD) maintained at the United States Army Research Institute of Environmental Medicine in Natick, MA. The TAIHOD is a collection of personnel, administrative, and health data sets for epidemiological research.14 Data were sorted and all identifying information was removed from both the coded and narrative data fields prior to analysis.

Reconstruction template
We developed a reconstruction template to establish the sequence of events, contributing factors, occupational activity, objects involved, and nature of injury. Specific elements were based on both the well established United States Bureau of Labor Statistics’ Occupational Injury and Illness Classification System (OIICS)15 and recently developed International Classification for External Causes of Injury (ICECI) approved by the World Health Organization.16

The OIICS is used to code: (1) disabling injuries reported in the Bureau of Labor Statistics’ annual Survey of Occupational Injuries and Illnesses, and (2) fatalities reported to the Bureau of Labor Statistics’ Census of Fatal Occupational Injuries programs.15 The OIICS manual contains the rules of selection, code descriptions, code titles, and indices, for the following code structures: nature of injury or illness, part of body affected, source of injury or illness, event or exposure, and secondary source of injury or illness. The ICECI is a multiaxial, modular, hierarchical system designed to aid researchers and prevention practitioners throughout the world in describing, measuring, and monitoring injury occurrence.17 ICECI consists of a core set of seven data elements: intent, mechanism of injury, object/substance producing injury, place of occurrence, activity when injured, alcohol use, and psychoactive drug or substance use.

The elements selected for this study (table 1) were considered the most valuable data elements to develop the capacity to reconstruct the injury event using a standardized template. We included the primary or underlying mechanism (“precipitating mechanism”) and object (“secondary source”) that initiated the injury producing event as well as the direct mechanism (“injury event/exposure”) or object (“primary source”) that resulted in injury. The elements were selected by three investigators (AL, GSik, TC). Coding of the nine elements for all cases was performed by a single coder (AL).

Table 1: Key data elements used to reconstruct injury event

<table>
<thead>
<tr>
<th>Data element</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>The type of broad activity the injured person was engaged in when the injury occurred (for example, maintenance)</td>
</tr>
<tr>
<td>Task</td>
<td>The specific activity engaged in when injury occurred providing additional detail (for example, inspecting engine)</td>
</tr>
<tr>
<td>Contributing factor</td>
<td>The key element that increased the risk such that what is normally completed without incident resulted in injury</td>
</tr>
<tr>
<td>Precipitating mechanism</td>
<td>The cause that initiated the chain of events leading to the injury; those mechanisms involved at the start of the injury event</td>
</tr>
<tr>
<td>Primary source</td>
<td>The object, substance, bodily motion, or exposure that directly produced or inflicted the previously identified injury or illness</td>
</tr>
<tr>
<td>Secondary source</td>
<td>The object, substance, or person that generated the source of injury or illness or that contributed to the event or exposure</td>
</tr>
<tr>
<td>Injury event/exposure</td>
<td>The manner in which the injury or illness was produced or inflicted by the source of injury or illness</td>
</tr>
<tr>
<td>Nature of injury</td>
<td>The principal physical characteristic(s) of the injury or illness</td>
</tr>
<tr>
<td>Outcome</td>
<td>The medical, functional, and/or financial results of the injury</td>
</tr>
</tbody>
</table>
Our reconstruction template was composed of nine essential elements (in italic type below) as an extension of the structured data collection technique used by Davies et al. The format of the template with the embedded elements in italic type is as follows:

During (1. activity) activities when (2. task), a/an (3. contributing factor) contributed to a/an (4. precipitating mechanism) event involving (5. primary source) and (6. secondary source) that caused a/an (7. injury/exposure) event resulting in a (8. nature of injury) and (9. outcome).

An application of the methodology involving the template using key data elements from coded data and narrative text is illustrated in box 1 and table 2. This is an example of an occupational back injury to one Army truck driver as represented by key data elements of coded data and narrative text from mishap investigation applied to the reconstruction template (United States Army Safety Center).

The result of applying the key data elements to the template are presented in the following case reconstruction.

Case reconstruction

During (1) maintenance/repair/servicing activities when (2) inspecting the truck’s engine, (3) greasy hands contributed to a (4) slip that caused a (7) fall from elevation to (5) hard top from (6) truck’s bumper resulting in a (8) contusion and (9) two lost work days.

Taxonomy

Once data from coded and narrative sources had been organized, they needed to be presented in a scientific and useful way. Various taxonomies of the data elements were developed to identify common hazard scenarios associated with occupational low back injury among motor transport operators. In the final model, data were sorted based on the data elements “injury event/exposure”, “precipitating mechanism”, and “task” using Microsoft Excel 2000 software. Additionally, the element “contributing factor” was sorted manually to provide a fourth level of categorization. The first level of breakdown was selected to be “injury event/exposure” because this best captured the various forms of damaging energy involved and, as such, represents the focus of prevention efforts. The second (“precipitating mechanism”), third (“task”), and fourth levels (“contributing factor”) were determined based on their usefulness for isolating patterns of injuries and developing relevant prevention strategies.

RESULTS

Coded data and narrative text

The combination of coded and narrative elements of the safety report contributed information to enable a thorough reconstruction of the injury events using the reconstruction template. For example, 89% of the safety reports contained coded information on the injury event/exposure while 98% of reports included such information directly in the narrative text (table 3). Although there was some variation throughout the 94 cases, coded data typically supplied five data elements to the template: broad activity, task, event/exposure, nature of injury, and outcomes. Narratives consistently contributed six elements: broad activity, task, contributing factor, precipitating mechanism, primary source, and event/exposure. There was redundancy among three elements (activity, task, injury event/exposure) while the remaining factors offered unique information. All elements were well populated with the exception of secondary source (19%). On many occasions no secondary source was involved, which is often the case for back injuries, especially those resulting from bodily reactions or overexertions.

The event reconstruction and sorting of various elements led to the development of a taxonomy of injury event scenarios. The taxonomy helped to organize a large amount

<table>
<thead>
<tr>
<th>Data element</th>
<th>Coded data</th>
<th>Narrative text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>Maintenance/repair/servicing</td>
<td>Maintenance</td>
</tr>
<tr>
<td>Task</td>
<td>Not elsewhere classified</td>
<td>Inspect the engine</td>
</tr>
<tr>
<td>Contributing factor</td>
<td>Not coded</td>
<td>Greasy hands</td>
</tr>
<tr>
<td>Precipitating mechanism</td>
<td>Not used</td>
<td>Lost hold (slip)</td>
</tr>
<tr>
<td>Primary source</td>
<td>Not used</td>
<td>Hard-top</td>
</tr>
<tr>
<td>Secondary source</td>
<td>Not used</td>
<td>Bumper of the vehicle</td>
</tr>
<tr>
<td>Injury event/exposure</td>
<td>Fall from elevation</td>
<td>Fall off of the vehicle</td>
</tr>
<tr>
<td>Nature of injury</td>
<td>Contusion</td>
<td>Not stated in text</td>
</tr>
<tr>
<td>Outcome</td>
<td>Includes $750 medical cost</td>
<td>Not stated in text</td>
</tr>
<tr>
<td></td>
<td>and 2 lost work days</td>
<td></td>
</tr>
</tbody>
</table>
of information into pathways whereby the frequencies suggested how the most common events occurred.

**Taxonomic analysis**

The taxonomy represented in fig 1 suggests three primary hazard scenarios associated with acute back injuries among Army truck drivers that account for 83% of the cases:

1. Struck by/against events during motor vehicle crashes (37%);
2. Falls resulting from slips/trips or loss of balance (32%); and
3. Overexertion from lifting activities (14%).

Of the 35 motor vehicle crashes, 60% occurred during convoys or normal driving duties while another 17% were associated with towing activities. Among the factors contributing to the crash, speeding/driver judgment (31%) was most often cited. Other contributing factors include icy/slicker road conditions (14%), obstructed visibility (14%), and fell asleep/intoxicated (14%).

The 30 injuries resulting from falls were nearly equally divided between loss of balance (43%) and slips/trips (40%). The tasks most often involved in falls were entering or exiting the vehicle (23%) and loading or unloading objects (17%). Standing on vehicle (20%) was the most common contributing factor.

A distant third injury event/exposure was overexertions, most of which were associated with lifting activities (69%). In 56% of the cases involving lifting, unsafe technique was cited as a contributing factor.

**DISCUSSION**

Based on our analysis of coded and narrative data from United States Army safety reports and identification of essential data elements, we developed a taxonomy of occupational back injuries occurring to motor transport operators. This taxonomy helped to identify three primary hazard scenarios accounting for 83% of the cases, which approaches the “usefulness” threshold as proposed by Drury and Brill using just three of the allotted six categories. These include struck by/against events during motor vehicle crashes, falls resulting from slips/trips or loss of balance, and overexertion from lifting activities.

The key to developing the hazard scenarios was using the narrative text to identify the contributing factors, precipitating mechanisms, and primary sources to complement the

<table>
<thead>
<tr>
<th>Data element</th>
<th>Coded data (%)</th>
<th>Narrative text (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Task</td>
<td>100</td>
<td>97.9</td>
</tr>
<tr>
<td>Contributing factor</td>
<td>19.1</td>
<td>95.7</td>
</tr>
<tr>
<td>Precipitating mechanism</td>
<td>0</td>
<td>98.9</td>
</tr>
<tr>
<td>Primary source</td>
<td>0</td>
<td>96.8</td>
</tr>
<tr>
<td>Secondary source</td>
<td>0</td>
<td>19.1</td>
</tr>
<tr>
<td>Injury event/exposure</td>
<td>89.4</td>
<td>97.9</td>
</tr>
<tr>
<td>Nature of injury</td>
<td>100</td>
<td>6.4</td>
</tr>
<tr>
<td>Outcome</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

| Table 3 Percentage of data elements identified within coded variables and narrative text for back injuries among motor transport operators, n = 94, United States Army Safety Center, 1987–97 |

![Figure 1](https://www.injuryprevention.com)
One of the limitations of this study is the recognized undercount of back injuries associated with the ASMIS. The less severe back injuries are more likely to be undercounted than the cases reported in ASMIS based on entry criteria alone. The proportion (sensitivity) of occupational injuries maintained by the ASMIS is unknown, so a more complete surveillance might alter the patterns that were identified or add some new ones. Another limitation concerns the analyst’s dependence on the accuracy of the narrative description and any inherent biases associated with the reporting and determination of causality and contributing factors. The variation in the length and information included in each narrative could potentially be addressed by explicitly requesting the inclusion of essential data elements in a checklist format as a cue for more complete and standardized narratives for the “accident description” field of Accident Report DA Form 285.

One of the major shortcomings of hazard scenario analysis is the exclusive reliance on frequencies without consideration of occupational exposure. While it is useful to recognize how the most common injuries occur, this approach does not indicate the true risk associated with a specific occupational task. Nonetheless, from a public health perspective the data are still useful in targeting more common events in this population.

**Future research**

The taxonomy presented here (fig 1) included only four of the nine elements comprising the reconstruction template (box 1 and tables 2 and 3). However, many other taxa were created using various combinations of elements. Our experience indicated that neither the selection of elements nor their sequence in the taxonomy produced changes to the identification of hazard scenarios. However, the consistency of identified patterns may vary with other data elements, other types of injuries, or other sources of data. The nine elements created for the reconstruction template or the four elements composing the taxonomy are not necessarily the most appropriate for all analyses. Instead, we offer them for consideration and critique to the occupational safety and injury research communities in an effort to advance the use of narrative text and coded data from safety reports or other sources to refine and develop additional taxa.

**CONCLUSIONS**

Coded data from safety investigations lacked sufficient information to complete our reconstruction template. The
addition of existing narrative text, similar to that collected by many injury surveillance systems, and a standard combination of elements enabled us to develop a more complete taxonomy of injury event characteristics and identify common hazard scenarios. This study demonstrates that narrative text can provide the additional information on contributing factors and precipitating mechanisms needed to target prevention strategies.

ACKNOWLEDGEMENTS AND DISCLAIMER
This research was supported by a research fellowship from the American Society of Safety Engineers (ASSE) and Liberty Mutual Research Institute for Safety and funding from NIOSH RO1 OH03703-01A1. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Army, Department of Defense, Department of Veterans Affairs, or the United States Government. The analyses conducted for this paper adhere to the policies for protection of human subjects as prescribed in Army Regulation 70-25 and with the provisions of 45 CFR 46.

Authors’ affiliations
A E Lincoln, War-Related Illness and Injury Study Center, Washington DC Veterans Administration Medical Center, Department of Veterans Affairs, Washington, DC, and Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
G S Sorock, Johns Hopkins Bloomberg School of Public Health, Baltimore, Liberty Mutual Research Institute for Safety, Hopkinton, Massachusetts, USA
T K Courtney, Liberty Mutual Research Institute for Safety, Hopkinton, Massachusetts, USA
H M Wellman, Liberty Mutual Research Institute for Safety, Hopkinton, Massachusetts, USA
G S Smith, Liberty Mutual Research Institute for Safety, Hopkinton, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
P J Amoroso, US Army Research Institute of Environmental Medicine, Natick, Massachusetts, USA

REFERENCES
2 Meehan TE. A commentary on occupational injury. Prof Safety 1999;April:40–2.