www.cpwr.com • www.elcosh.org

CPWR

THE CENTER FOR CONSTRUCTION
RESEARCH AND TRAINING

# Data Linkage of State Registries for Assessment of Construction Injuries

Lee Friedman, PhD
Peter Ruestow, MSc
Linda Forst, MD, MPH

January 2013

8484 Georgia Avenue
Suite 1000
Silver Spring, MD 20910

PHONE: 301.578.8500
FAX: 301.578.8572

Research for this report was funded by CPWR – The Center for Construction Research and Training, using grant U60 OH009762 from the National Institute of Occupational Safety and Health (NIOSH).

The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIOSH.

CPWR is a 501(c)(3) nonprofit research and training institution created by the Building and Construction Trades Department, AFL-CIO, and serves as the research arm of the BCTD.

CPWR provides safety and health research and information for the construction trades and industry. For more information, visit www.cpwr.com.

**Final Report to CPWR: The Center for Construction Research and Training**

**Small Study #10-1-PS**

**Dates: July 1, 2010 – June 30, 2011**

# Data linkage of state registries for assessment of construction injuries

Lee Friedman, PhD

Peter Ruestow, MSc

Linda Forst, MD, MPH

UIC SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF ILLINOIS AT CHICAGO

**Executive Summary**

The goal of this research project was to investigate a system for occupational injury surveillance linking public health and employer-based databases, and we developed a method for linking three existing databases collected by legal mandate in Illinois. These databases, which have not previously been used for occupational surveillance, include two medical record databases from the Illinois Department of Public Health -- the Illinois Trauma Registry (ITR), and the Hospital Discharge database (HD) -- and one database from the Illinois Workers' Compensation Commission, Workers' Compensation Claims (IWCC -C). We linked the two Illinois Department of Public Health databases to the workers' compensation file.

In our previous work, we identified 19,734 construction workers filing claims through the IWCC. 16,794 of the cases contained sufficient information on the key variables to be used for the final data linkage. Workers' compensation data includes a wide range of injuries and illnesses, most of which would not result in hospitalization or treatment in a trauma unit (e.g. chronic back pain, cumulative trauma disorders). We linked 1,664 workers' compensation records to the two hospital registries – a linkage rate of 9.91%. We expected a linkage rate of approximately 10-15% with the HD and ITR, based on our pilot study and analysis of the type of injuries contained in the workers' compensation database.

A key finding of this project was that white, non-Hispanic construction workers were consistently awarded higher monetary settlements than black and white Hispanic construction workers within specific injuries, despite a mean percent permanent partial disability that was equivalent or lower for the non-Hispanic whites. In the final multivariable models, the difference in the level of total monetary compensation persisted. In the final multivariable robust regression model, compensation was $5,824 higher (p=0.030; CI95%: 551, 11097) for white non-Hispanic workers than for minority workers when controlling for other covariates. Further investigation of these ethnic disparities is necessary. Moreover, the higher compensation was not explained by differences in injury severity, affected body region, type of injury, or common factors associated with higher compensation – number of dependents, average weekly wage, injuries resulting in death or the use of an attorney.

A second important finding of this research project was that temporary total disability was positively associated with two important in-hospital measures of injury severity – length of stay

and discharge status. Temporary total disability indicates how long an injured worker is unable to return to work or placed on restricted work duty. However, temporary total disability was not significantly associated with the injury severity score – a common measure of the severity of an injury used in trauma research. The final multivariable model indicated that each day a patient remained in the hospital was associated with 1.2 weeks' increase of temporary total disability awarded by the workers' compensation commission ($p<0.001$). Furthermore, patients requiring continued care at another facility after discharge from the hospital received 10.75 more weeks of temporary total disability than those discharged directly home ($p<0.001$). Our research indicated that some in-hospital measures are good predictors of future missed work and total temporary disability.

This study provided insights into the strengths and limitations of data linkage methods that can be applied to occupational injury surveillance. Data linkage between workers' compensation records and hospital records provides details that workers' compensation records alone often do not provide. This information allows researchers to better understand the relationship between the acute injury and long-term impairment and disability. This work is part of a larger program at the University of Illinois School of Public Health to establish the use of previously unavailable datasets for research on contributing causes and conditions, injury severity, treatment effectiveness, and short and medium-term health and economic outcomes of occupational injuries. The methodology for obtaining, cleaning, linking, analyzing and reporting as reported in this study are applicable to most other states in the U.S. where similar datasets are available.

**Table of Contents**

**Background**

Construction is one of the most hazardous industries in the United States. Each year several hundred thousand construction workers become ill or are injured as a result of on-the-job hazards. Estimated rates for injuries, illnesses and fatalities among construction workers are consistently among the highest of any economic sector (NIOSH, 2004).

Although the federal government collects data on occupational injuries, there is growing evidence that the Bureau of Labor Statistics data substantially underreports injuries and illnesses, primarily due to inconsistencies in employer reporting and changes to OSHA's recordkeeping standard (Rosenman, 2006; Friedman, 2007; CPWR, 2002). This limits the value of BLS data in describing injuries in the construction sector. Also, there is a dearth of data on the workplaces, working conditions, and mechanisms that lead to injuries in construction as well as predicting disability following injury. There is a need for alternative data sources to help provide a better picture of the pre-event and event factors, as well as the magnitude and trend of injuries.

State-based data repositories can fill in the gaps left by federal surveillance programs. Currently, BLS data are the primary source of occupational surveillance data for Illinois. However, there are several alternative resources that can be used to help better describe construction injuries in the State. These include the Illinois Workers' Compensation Commission (IWCC) Claims data, the Illinois Trauma Registry (ITR), and the Illinois Hospital Discharge database. These databases have been underutilized for occupational surveillance.

The overall goal of this research project is a system of occupational injury surveillance that links medical records for construction workers injured on the job with workers' compensation databases. We developed a method for an ongoing linkage of three existing databases collected by legal mandate in Illinois. These databases, which have not been previously used for occupational surveillance, include two medical record databases from the Illinois Department of Public Health -- the Illinois Trauma Registry (**ITR**), and the Hospital Discharge database (**HD**) -- and one database from the Illinois Workers' compensation Commission, Workers' compensation Claims (**IWCC -C)**. Corresponding databases are maintained in most other states.

We developed and tested the data linkage methodology by focusing on construction injuries. This provided important and previously unavailable information by combining data on treatment (ITR

and HD), long-term disability (IWCC) and level of compensation (IWCC) for individuals. In our previous work with CPWR, we identified 19,734 workers' compensation claims made by injured construction workers between 2000 and 2005.

The specific aims of this project were as follows:

1. Test different methods for linking public health and workers' compensation databases to assemble a mega-dataset with a wide array of variables related to exposure hazards, injury diagnoses, injury severity, short- and long-term outcomes, and cost. We will test these methods by linking records across the listed databases for construction workers. This will allow us to design a uniform platform whereby data repositories with historical data can be efficiently linked. We will also be able to recommend an ongoing system for linking data at the source in the future that will assure compatibility with data collected in the past, and will provide timely information on injury incidence and trends.

2. Conduct a comprehensive descriptive and multivariable analysis of construction injuries in the final merged dataset using simple descriptive statistics and relating risk factors (e.g., job title, description of the event, blood alcohol) and risk modifiers (e.g., demographics, time of injury, wage) to health outcomes (e.g., type of hand injury, injury severity, functional impairment status on discharge, short and long-term disability, post-discharge death) and cost (hospitalization, lost wages, workers' compensation settlements) in more complex statistical models. Within this framework we proposed to investigate two specific hypotheses:

> **Hypothesis #1** – Hispanic construction workers have a higher incidence of work-related injuries than non-Hispanics, and receive lower levels of compensation through the workers' compensation claim process than their non-Hispanic counterparts.
>
> **Hypothesis #2** – Measures of in-hospital severity of injury (injury severity scores, disability at discharge, medical complications) do not correspond with percent disability ratings determined during the Workers' Compensation claims process.

We completed this work on June 30, 2011. Our methods and findings follow..

## Methods

### Data Sources

We have a data sharing agreement with both the Illinois Department of Public Health and the Illinois Workers' Compensation Commission. The IRB at UIC has approved this work, as have the Research Committees at the Illinois Department of Public Health and IWCC.

*Illinois Trauma Registry*

The Illinois Trauma Registry (ITR) was mandated by the state legislature and is managed by the Illinois Department of Public Health. All of the state's level 1 and 2 trauma centers (N=67) are required to report all patients (1) sustaining traumatic injuries (ICD-9-CM external injury codes E800-995) and admitted to a trauma center for > 12 hours, (2) transferred to a level I or II center or (3) are dead-on-arrival (DOA) or die in the emergency department. Level III and IV centers are not included in the ITR. Level III and IV trauma centers primarily provide advanced life support prior to transfer to a level I or II center, or serve communities with no access to level I and II centers. We provide an assessment of data quality in the ITR in a previous paper of ours (Friedman, 2007b). The ITR contains data on demographics (age, gender, race/ethnicity), exposure (mechanism of injury), and health outcomes (diagnoses, measures of injury severity, hospital procedures, disability status on discharge), and economics (payer source).

*Hospital Discharge Database*

The University of Illinois at Chicago Hospital is a member of the Illinois Hospital Association, the organization that compiles and manages the hospital discharge database. As a member of the Illinois Hospital Association, the PI's research team has ongoing access to the hospital discharge database. The hospital discharge database is based on billing records. It includes all patients treated for more than 23 hours in any Illinois hospital (i.e. inpatients only) for any medical reason. The hospital discharge database includes variables on patient demographics (age, gender), exposure (mechanism of injury), health outcomes (diagnoses, hospital procedures, discharge status), and economics (hospital charges, payer source).

*Illinois Workers' Compensation Commission Claims Database*

In existence since 1913, the IWCC operates the administrative court system for workers' compensation cases in Illinois. There are approximately 70,000 claims filed for financial compensation each year with IWCC . Unlike in single carrier states with a well organized and

centralized reporting system (e.g. Washington), in Illinois the IWCC only handles disputed claims in which the employee and employer are unable to resolve injury compensation issues. The IWCC claims dataset only includes information about disputed elements of the compensation claim. Any element of a claim settled prior to initiating a claim through the IWCC and not disputed by either party will not be litigated through IWCC or reported in the dataset. An arbitrator initially hears a workers' compensation claim. The arbitrator's decision can subsequently be appealed before a panel of three commissioners. At any point, the injured worker and employer can settle the claim independently of the IWCC.

We obtained a dataset of all claims in the Illinois Workers' Compensation Commission, which included information on both active and closed claims. The dataset contains an array of information including employer information, employee demographics (age, gender, marital status, number of dependents), cause and type of injury or illness, level of temporary and permanent disability, and details on the compensation costs associated with the injury. For this study, we used a dataset created in our previous work with CPWR. The dataset contains records of 19,734 workers' compensation claims by injured construction workers between 2000-2005. Compensation costs were not adjusted for inflation. The minimum age in this study group was 16 years. (Illinois law prohibits persons under the age of 16 from working in construction.) A full description of the workers' compensation dataset and assessment of data quality has been reported on in our previous work (Friedman, 2009).

The table 1, p.10, shows the data elements in each of the three databases that were used for this investigation.

**Data Cleaning**

In previous work with CPWR, we identified 19,734 workers' compensation claims made between 2000-2005 by injured construction workers and cleaned the IWCC dataset (Friedman, 2009). In the trauma registry and hospital discharge datasets, patients are transferred between hospitals would potentially be reported more than once, so we identified and removed duplicates by matching cases on at least seven variables (e.g., name, date of birth, hospital, gender, ethnicity, injury date, hospital admission date, external cause of injury). A full description of the deduplication process has been previously reported (Friedman, 2007b).

**Data Linkage**

The workers' compensation dataset of 19,734 claims (Friedman, 2009) was the primary dataset. After exclusion of claims without valid dates of birth or dates of injury, the final dataset contained 16,794 claims. We used probabilistic linkage methodology to link the three datasets. We used a date of birth, date of injury, gender and residential ZIP code to link the data. Figure 1 illustrates the general linkage methodology used.

Figure 1: Diagram of the data linkage methodology

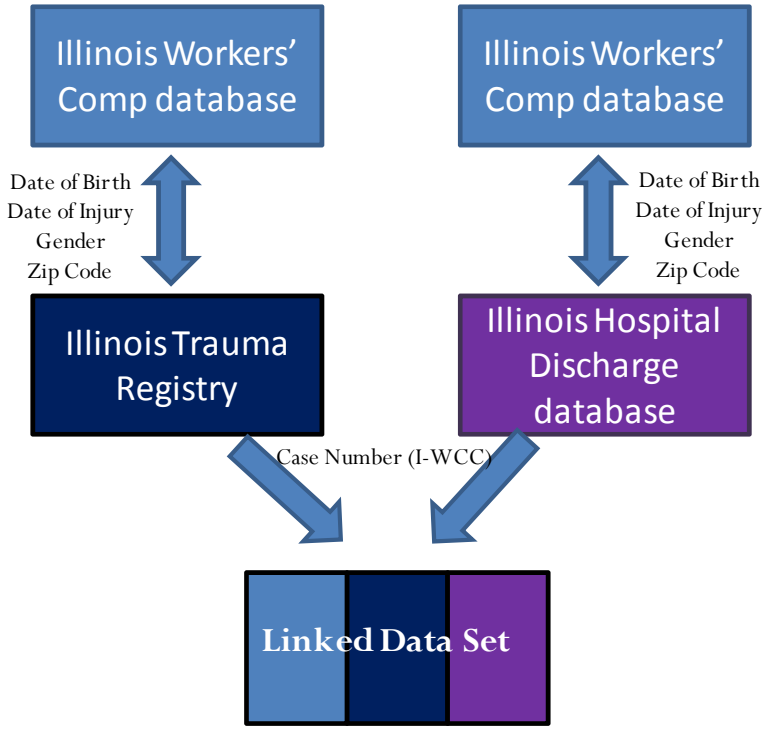


**Table 1: Description of Three State Datasets Used for Data Linkage Project: Illinois Trauma Registry, Hospital Discharge Database and the Illinois Workers' Compensation Commission Claims Database**

| Database | Data Elements | | | | |
|---|---|---|---|---|---|
| | **Inclusion Criteria** | **Demographics** | **Exposure Data** | **Health Data** | **Economic Variables** |
| **Trauma Registry (ITR)** | Persons treated in level 1 or 2 trauma unit for $\geq$ 12 hours (~45,000/yr) | Name<br>Gender<br>Age-DOB<br>Race/ethnicity | ICD9 E-codes E849, showing location where injury occurred<br>Time, day, date of injury | ICD9 N & E-codes<br>Body site<br>Severity<br>Hospital procedures<br>Treatment<br>Disability status on discharge<br>Blood alcohol | Cost of hospitalization<br>Hospital procedures<br>Hospital days |
| **Hospital Discharge (HD)** | All individuals hospitalized in Illinois | Gender<br>Age-DOB<br>Race/ethnicity | ICD-9 N and E codes | ICD 9 codes<br>Hospital procedures<br>Hospital cost<br>Discharge status | Cost of hospitalization<br>Hospital days<br>Payer source |
| **Workers' Compensation Claims (IWCC - C)** | Persons filing workers' compensation claims for arbitration through IWCC (~70,000/yr) | Gender<br>Age-DOB | Employer Name<br>Nature of injury<br>Narrative of injury circumstances | ICD 9 codes<br>Hospital procedures<br>Level of disability | Total medical costs<br>Lost wages<br>Cost of compensation<br>Payer source |
| *ICD 9 External Injury Code E 849 has decimal points that give the place injury occurred, e.g., E849.1 is Farm | | | | | |

Using the four linking variables, cases were matched from the IWCC to both the ITR and the IHD separately to ensure that only construction worker claims that matched to at least one other data set were included in the final data set. A combination of SQL and SAS code was written to carry out all steps in the linkage, using SAS Version 9.1 (See Appendix 1). All linked cases were validated by manually reviewing cause of injury and type of injury codes. In the trauma registry and hospital discharge datasets, cause of injury and type of injury are coded using ICD-9 coding. The workers' compensation dataset uses a different coding system for both cause and type of injury.

Table 2, p.12, outlines the linkage procedure. We ran multiple passes for the linkage between the workers' compensation and the two separate medical record datasets. After each pass, the matched records were removed from the datasets before continuing with the next pass if the two records matched by type of injury. The initial pass identified matches between the datasets that matched exactly on date of birth, date of injury, gender and residential ZIP code. In the second pass, we omitted ZIP code. (The average interval between the date of injury and the date a claim was filed was approximately nine months (276.6 days; sd=296.6 days), and it took more than one and a half years on average from the time of filing a claim to a decision, so if a worker moves he may have a different ZIP code at settlement than at the time of the injury) . In subsequent passes, we allowed for variation for the date of injury and date of birth. The concept of using multiple passes is to begin with the highest level of precision and then modify the precision with each subsequent step. Thereby, ensuring that higher probability matches are identified in the initial step and removed from the datasets prior to subsequent steps so that they don't match again to another case when the matching criteria is less rigorous. In the linkage process, 64.6% of cases were matched during the 1$^{st}$ pass, 28.4% in the second pass, 2.5% in the third, 3.8% in the fourth and 0.7% in the fifth.

We expected a linkage rate of approximately 10-15%, based on our pilot study and analysis of the type of injuries contained in the workers' compensation database. Most claims for compensation involve injuries and illnesses that do not result in hospitalization or acute trauma. We linked 1664 workers' compensation records to the two hospital registries – a linkage rate of 9.9%. One-thousand thirty-nine cases of 16,975 claims in the IWCC data set matched to the ITR (6.1%), while 1378 matched to the IHD (8.1%). There were 753 cases identified that matched across all three data sets.

**Table 2: Summary of the Data Linkage Steps Used for this Project**

| | DOB | Date of Injury | Gender | ZIP Code | Type of Injury |
|---|---|---|---|---|---|
| **Trauma registry to WC** | | | | | |
| 1st pass | Exact | Exact | Exact | Exact | Match |
| 2nd pass | Exact | Exact | Exact | Omit | Match |
| 3rd pass | Exact | plus/minus 1 | Exact | Exact | Match |
| 4th pass | Year | Exact | Exact | Exact | Match |
| **HD to WC** | | | | | |
| 1st pass | Exact | Exact | Exact | Exact | Match |
| 2nd pass | Exact | Exact | Exact | Omit | Match |
| 3rd pass | Exact | plus/minus 1 | Exact | Exact | Match |
| 4th pass | Year | Exact | Exact | Exact | Match |
| 5th pass | Exact | plus/minus 2 | Exact | Exact | Match |

**Comparison of Matched and Unmatched Cases**

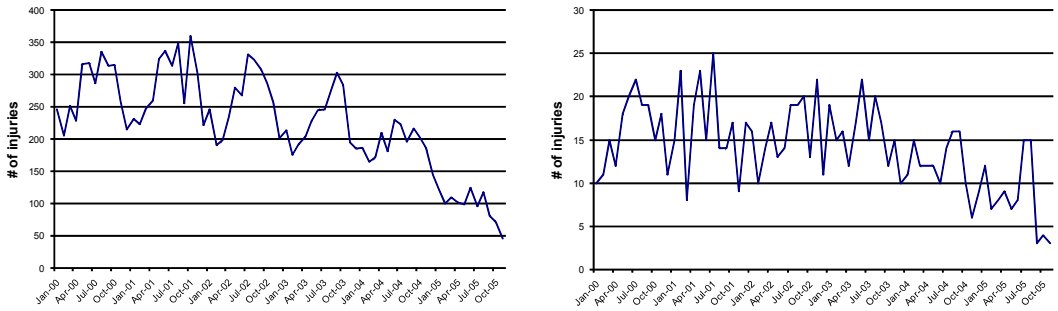
We compared cases in the linked data set with those that did not match in the IWCC to establish whether the linked set represented an acceptable sample of the overall data. If the linked data mirrored the overall data demographically, and if the number of cases followed a similar distribution over time, we felt that a meaningful analysis could follow. These demographic and time-trend comparisons are shown on the following pages (Table 3; Figure 2 and 3). The matched and unmatched cases were very similar by gender, marital status, number of dependents, age, and population density of employment area. Some of these demographic variables – average weekly wage, marital status and number of dependents – play a major role in determining the level of financial compensation to a worker. Approximately the same proportion of matched and unmatched cases had a final decision on their workers' compensation claim, as opposed to dismissed cases or new cases still in progress. In addition, when we looked at the distribution of injuries across time (January 2000 through December 2005), the distributions were very similar (Figure 2 and 3).

**Table 3: Comparison of Matched and Unmatched Cases, Demographic Characteristics Illinois Workers' compensation Claims Data, 2000-2005**

| | Unmatched (N=18070) | Matched (N=1664) |
|---|---|---|
| **Gender** | | |
| Male | 17224 (95.3%) | 1624 (97.6%) |
| Female | 828 (4.6%) | 40 (2.4%) |
| Unspecified | 18 (0.1%) | 0 (0%) |
| **Marital Status** | | |
| Single | 6768 (37.5%) | 651 (39.1%) |
| Married | 11034 (61.1%) | 995 (59.8%) |
| Widowed/Divorced | 40 (0.2%) | 1 (0.1%) |
| Unspecified | 228 (1.3%) | 17 (1.0%) |
| **Number of Dependents** | | |
| 0 | 9454 (52.3%) | 864 (51.9%) |
| 1 | 2941 (16.3%) | 255 (15.3%) |
| 2 | 3135 (17.3%) | 307 (18.4%) |
| 3 | 1639 (9.1%) | 151 (9.1%) |
| 4 | 587 (3.2%) | 54 (3.2%) |
| 5 or more | 310 (1.7%) | 33 (2.0%) |
| Unspecified | 4 (0.0%) | 0.00% |
| **Mean Age (SD)** | 39.6 (sd=10.4) | 38.6 (sd=10.8) |
| 16 to 24 years | 1483 (8.2%) | 191 (11.5%) |
| 25 to 34 years | 4713 (26.1%) | 474 (28.5%) |
| 35 to 44 years | 6284 (34.8%) | 528 (31.7%) |
| 45 to 54 years | 3828 (21.2%) | 333 (20.0%) |
| 55 to 64 years | 1372 (7.6%) | 120 (7.2%) |
| 65 and older | 101 (0.6%) | 11 (0.7%) |
| Unspecified | 289 (1.6%) | 7 (0.4%) |
| **Population Density: Place of Accident (persons/sq.mi)** | | |
| Rural (0-499) | 556 (3.1%) | 39 (2.3%) |
| Mid range (500-999) | 758 (4.2%) | 80 (4.8%) |
| Urban ($\geq$1000) | 16466 (91.1%) | 1529 (91.9%) |
| Out of State or Unspecified | 290 (1.6%) | 16 (1.0%) |
| **Attorney Representation Used** | 13922 (77.0%) | 1394 (83.8%) |
| **Claim w/Final Decision** | 14580 (80.7%) | 1318 (79.2%) |
| **Average weekly wage** | 930.3 (sd=367.4) | 882.4 (sd=378.1) |

**Figures 2 and 3: Distribution of Injuries by Date of Accident: Unmatched and Matched Cases in Data Linkage**

**Unmatched Cases** **Matched Cases**




Matched cases were slightly more likely to use attorney representation, but this is likely related to the fact that matched cases suffered more severe injuries. We expected a greater proportion of the matched cases to report permanent partial disability and different injury profiles because the matched cases represent injuries requiring hospitalization (e.g. multiple injuries, severe fractures, traumatic brain injuries, burns, etc.), whereas the unmatched cases involve injuries which can be disabling but generally do not require hospitalization (e.g. chronic back pain, cumulative trauma disorders, torn ligaments, hearing loss, hernias, etc.) (Table 4). Unmatched cases disproportionately involved back, spine and upper extremity injuries, while matched cases were more likely involve injuries to multiple body parts and deaths (Table 5). Matched cases also had a higher mean percent permanent partial disability (13.9% vs. 10.4%), in particular matched cases were more likely to result in permanent partial disability greater than a 50% loss of function.

The differences between matched and unmatched cases were in line with expectations regarding injury severity, but key demographic variables between the matched and unmatched cases were nearly identical. The similarities between the linked data set and the overall IWCC data set indicated that the linked set was suitable for meaningful analysis.

**Table 4: Comparison of Matched and Unmatched Cases, Injury Type**

**Illinois Workers' compensation Claims Data, 2000-2005**

| | Unmatched (N=18070) | Matched (N=1664) | Ratio of Percent Difference |
|---|---|---|---|
| **Injuries more likely to required hospitalization** | | | |
| Heart Attack with Complications | 6 (0.03%) | 10 (0.6%) | 20 |
| Deaths | 75 (0.42%) | 28 (1.68%) | 4 |
| Concussion | 23 (0.13%) | 7 (0.42%) | 3.23 |
| Amputation | 93 (0.51%) | 23 (1.38%) | 2.71 |
| Fracture | 1672 (9.25%) | 310 (18.63%) | 2.01 |
| Multiple Injuries | 1679 (9.29%) | 290 (17.43%) | 1.88 |
| Burn | 119 (0.66%) | 17 (1.02%) | 1.55 |
| **Injuries less likely to require hospitalization** | | | |
| Hearing Loss | 22 (0.12%) | 1 (0.06%) | 0.5 |
| Chronic Illness / Syndrome | 22 (0.12%) | 1 (0.06%) | 0.5 |
| Internal Derangement | 287 (1.59%) | 11 (0.66%) | 0.42 |
| Spinal Injury | 842 (4.66%) | 26 (1.56%) | 0.33 |
| Sprains and Strains | 943 (5.22%) | 29 (1.74%) | 0.33 |
| Hernia | 199 (1.1%) | 5 (0.3%) | 0.27 |
| Tear/Avulsion | 938 (5.19%) | 22 (1.32%) | 0.25 |
| Carpal Tunnel Syndrome | 446 (2.47%) | 6 (0.36%) | 0.15 |
| Other Cumulative Trauma Disorders | 51 (0.28%) | 0 (0%) | 0 |
| Dermatitis | 5 (0.03%) | 0 (0%) | 0 |
| Vision Loss | 8 (0.04%) | 0 (0%) | 0 |

**Table 5: Comparison of Matched and Unmatched Cases, Injury and Compensation Illinois Workers' compensation Claims Data, 2000-2005**

| | Unmatched (N=18070 | Matched (N=1664) |
|---|---|---|
| **Body Part** | | |
| Head and Neck | 1156 (6.4%) | 121 (7.3%) |
| Back and Spine | 3798 (21.0%) | 183 (11.0%) |
| Upper Extremities | 6141 (34.0%) | 364 (21.9%) |
| Torso | 421 (2.3%) | 22 (1.3%) |
| Lower Extremities | 4478 (24.8%) | 414 (24.9%) |
| Internal | 71 (0.4%) | 14 (0.8%) |
| Multiple Extremities Unspecified | 3257 (18.0%) | 641 (38.5%) |
| **Number of Body Parts Affected** | | |
| 0 | 79 (0.4%) | 6 (0.4%) |
| 1 | 16699 (92.4%) | 1560 (93.8%) |
| 2 | 1138 (6.3%) | 83 (5.0%) |
| 3 or more | 154 (0.9%) | 15 (0.9%) |
| **Percent Permanent Partial Disability** | | |
| No Disability | 9738 (53.9%) | 956 (57.5%) |
| 1 to 25 Percent | 5765 (31.9%) | 316 (19.0%) |
| 26 to 50 Percent | 2170 (12.0%) | 271 (16.3%) |
| 51 to 100 Percent | 397 (2.2%) | 121 (7.3%) |
| **Fatalities** | 75 (0.4%) | 28 (1.7%) |
| **Median Total workers' compensation ($USD)** | 10936.0 | 19936.5 |
| **Mean percent permanent partial disability** | 10.4 (sd=15.7) | 13.9 (sd=21.8) |
| **Median days from filing to decision** | 921 | 1070 |

**General Analytic Approach**

The final linked data set was analyzed to address our specific aim and hypotheses (see specific aims). We used SAS software for all statistical analyses (v.9.1; Cary, NC). Distribution of injuries by age, gender and ethnicity were determined as well as frequency of injuries overall. We also described type of injuries and disparities between groups in terms of injury severity, external cause of injury, work location, duration of hospitalization, level of disability, in-hospital mortality, percent disability and level of monetary compensation through workers' compensation.

Appropriate parametric (Pearson's chi-square) and non-parametric tests (Wilcoxon Rank Sum) were used to evaluate bivariate relationships. Student's t-test was used to compare mean differences in continuous metrics such as ISS scores, percent disability awarded and level of

monetary compensation. The Levenne test was used to test for equal variance between samples to determine whether to used equivariant or non-equivariant statistical measures of significance.

ICD-9 N codes were used to assess body region and type of injury based on the Barell classification matrix (Barell, 2002; Baker, 1974). We calculated injury severity scores (ISS) using Abbreviated Injury Scale (AIS) scores for six major body regions. The AIS scores are calculated using a standardized computer algorithm based on discharge records rather than admission status in order to reduce miscoding (Jurkovich, 1999). We report the number of hospitalization days (LOS; Length of Stay), days of treatment in intensive care units (ICU), days on a ventilator, and disability at discharge.

We calculated the cumulative percent permanent partial disability for this study. Permanent partial disability involves partial loss of body function at the point of maximum medical improvement. We used the statutory formula to calculate cumulative percent disability when more than one body part was injured and limited in function. An example of the statutory formula for computing cumulative disability is A + (1-A)*B, where A is the percent disability for a specific injury involving a specific body part and B is the percent disability for a second specific injury involving a specific body part.

***Hypothesis #1****: Hispanic construction workers have a higher incidence of work-related injuries than non-Hispanics, and receive lower levels of compensation through the workers' compensation claim process than their non-Hispanic counterparts.*

Only the trauma registry contained information regarding race and ethnicity. Therefore, this data analysis was restricted to the 1,039 claims that linked to the trauma registry. This analysis only includes cases of acute traumatic injuries treated initially within one of the 67 level 1 and 2 trauma units in the State of Illinois. The first aspect of the hypothesis relating to differences in work-related injury rates was previously reported on (Friedman, 2007b). We found that the incidence rates of traumatic injuries among Hispanics treated in Level I and II trauma centers were more than two times higher than among non-Hispanic Caucasians (Friedman, 2007b). For the regression analysis, the dependent variable (total monetary compensation) was not normally distributed (kurtosis=34.4; skewness = 4.0). In scenarios with extreme or many outliers causing the data to be skewed, ordinary least squares (OLS) regression will produce biased parameter estimates. This is because in OLS the parameter estimates will be weighted towards the outliers,

which also inflates the variance. However, we did not transform the dependent variable because back transformation of log transformed data leads to biased estimates. (Parkhurst, 1998; Huybrechts, 2002) While the log transformation makes the data less skewed, it changes the relationship between the dependent and independent variables. (Parkhurst, 1998; Huybrechts, 2002)

Therefore, for the multivariable regression analysis, we used robust M-estimation as implemented in SAS Version 9 (PROC ROBUSTREG; SAS Institute, Inc., Cary, NC). The parameter estimates derived from robust regression are less influenced by outliers. This is generally achieved by weighting observations whose residuals are large and does not remove them.

The multivariable model included demographic variables, wage, injury outcome, and attorney representation. The outcome variable was total financial compensation of decided claims, excluding claims in progress and dismissed claims. Total financial compensation included medical costs, dependent benefits, survivor benefits, settlement payments, attorney fees, and other miscellaneous costs. We used a manual stepwise selection method to identify the best model fit for the predictors. Akaike (AIC) criterion and Schwarz information criteria (BIC) were also used for model selection and to identify the best weighting function. In the final model, gender, age at the time of accident, number of dependents, and interval from day of accident to day of filing were highly insignificant, and therefore were excluded. Because of the small sample size of the final linked dataset we did not have the statistical power to look at individual ethnic groups. Therefore, we compared white non-Hispanic construction workers with all the other ethnic groups combined. The final model included the following covariates (data format, data source): employee's weekly wage (continuous; wc data), injury severity score (continuous; tr), days of initial hospitalization (continuous; tr), discharged from the hospital to an acute/intermediate care facility or rehabilitation center (dichotomous; tr), attorney representation (dichotomous; wc), total weeks of total temporary disability (continuous; wc), and percent permanent partial disability (continuous; wc).

To validate our primary multivariable robust regression model, we also ran a linear regression model on a truncated dataset in which outliers in monetary compensation were removed. Based on an analysis of the distribution of total financial compensation overall and within specific ethnic subgroups, we removed all cases with total financial compensation exceeding $225,000.

The removal of the outliers made the dependent variable near normally distributed. We ran the same multivariable model used in the primary analysis.

***Hypothesis #2*** *– Measures of in-hospital severity of injury (injury severity scores, disability at discharge, medical complications) do not correspond with percent disability ratings determined during the Workers' Compensation claims process.*

The linked data set was also used to assess whether injury severity score (ISS) is a good predictor of the disability rating assigned by the Illinois Workers' Compensation Commission. As such, the major variables of interest included permanent partial disability and temporary total disability from the IWCC and the ISS from the ITR. Permanent disability is defined in the IWCC as the partial or complete loss of body function at the point of maximum medical improvement, and is measured as a percentage. Temporary total disability also reported in the IWCC, indicates for how long an injured worker is unable to return to work or should be placed on restricted work duty. The ITR's ISS measure uses six body regions (head, face, chest, abdomen, extremities, and external) to classify an injury. Though the ISS was not originally intended as a measure of injury outcome, it is strongly correlated with 24-hour survival, 30-day survival, duration of hospital stay, duration of rehabilitation, and long-term disability.

To assess the relationship between the two disability ratings and ISS, univariate and bivariate distributions of the variables were assessed. All three variables were not normally distributed (heavily skewed to the right), and thus, subsequent statistical testing utilized non-parametric methods and test statistics. Because injury cases resulting in death were not assigned injury severity scores at the trauma center, they were excluded from the analysis (n = 20). In addition, because we analyzed the role total compensation played in every phase of the analysis, only claims for which a decision was reached were included in the analysis.

Spearman rho coefficients of correlation were calculated to assess the relationship between the disability ratings and potential in-hospital measures of injury severity (ISS, hospital length of stay). Correlation of disability ratings and measures of injury severity with total compensation received from the claims process was also evaluated.

Disability ratings and ISS were also summarized by nature of injury (as defined in the I-WCC), part of body affected, and diagnosis category (defined by ICD-9 codes, assigned upon discharge from a trauma center).

We conducted an analysis of variance (Kruskal-Wallis test) to assess differences in disability rating by ISS category (0-9, 10-16, 17-25, > 25). Subsequently, we characterized the relationship between disability ratings and trauma center severity scores using a regression model to adjust for confounding and identify interactions. The multivariable model included demographic variables, wage, injury outcome, and attorney representation. The outcome variable was temporary total disability in terms of weeks of missed work for decided claims, excluding claims in progress and dismissed claims. Temporary total disability also reported in the IWCC, indicates for how long an injured worker is unable to return to work or should be placed on restricted work duty. Because temporary total disability was not normally distributed we used a robust regression model for the analysis. We used a manual stepwise selection method to identify the best model fit for the predictors. Akaike (AIC) criterion and Schwarz information criteria (BIC) were also used for model selection and to identify the best weighting function. In the final model, gender, age at the time of accident, number of dependents, average weekly wage, number of body parts affected, and interval from day of accident to day of filing were highly insignificant, and therefore were excluded. The final model included the following covariates (data format, data source): injury severity score (continuous; tr), days of initial hospitalization (continuous; tr), discharged from the hospital to an acute/intermediate care facility or rehabilitation center (dichotomous; tr), attorney representation (dichotomous; wc), and injuries to the lower extremities (dichotomous; tr).

**Results - Ethnic Disparities in Workers' Compensation**

In the final linked dataset, the ethnic distribution among traumatically injured construction workers was as follows: black, N=68; white Hispanic, N=168; white non-Hispanic, N=724; and other ethnicities, N=79. Nearly all the injured workers in the dataset were male (98.5%). Injured Hispanic workers were disproportionately more likely to be married (73.8%) than either black (41.2%) or white non-Hispanic workers (58.4%). Hispanics were also significantly younger than white non-Hispanics ($p<0.05$; Tukey's comparison of means), but not blacks. White non-Hispanic workers were more likely to have no dependents. The mean number of dependents per worker was 1.2 among injured black workers, 1.6 among Hispanic workers and 0.9 among white non-Hispanics. The mean number of dependents among injured Hispanic workers was significantly greater than white non-Hispanics ($p<0.05$; Tukey's comparison of means). No significant differences in mean number of dependents were observed between the other ethnic groups.

The average weekly wage reported to the IWCC was highest for white non-Hispanic construction workers. The average weekly wage for white non-Hispanics was significantly higher than the weekly wage reported for white Hispanics ($p<0.05$; Tukey's comparison of means), but not for the other groups. The proportion of construction workers who used attorney representation for their claim process instead of advocating on one's own behalf was as follows: black, 63 (92.6%); white Hispanic, 146 (86.9%); white non-Hispanic, 604 (83.4%); and other ethnicities, 73 (92.4%).

**Cause of Injury**

Black construction workers were disproportionately more likely to suffer injuries as a result of assaults and motor vehicle crashes compared to other ethnic groups (Table X). Whereas, the cause of injury between white Hispanic and white non-Hispanic workers were very similar, with the majority of injuries caused by falls (Table 6).

**Table 6: Cause of Injury Among Construction Workers Filing Workers' Compensation Claims by Ethnicity**

| | Black (N=68) | White, Hispanic (N=168) | Other (N=79) | White Non-Hispanic (N=724) |
|---|---|---|---|---|
| **Cause of Injury** | | | | |
| Assault | 7 (10.3%) | 3 (1.8%) | 2 (2.5%) | 6 (0.8%) |
| Motor vehicle crash | 11 (16.2%) | 17 (10.1%) | 6 (7.6%) | 77 (10.6%) |
| Cutting or Piercing Instrument | 3 (4.4%) | 11 (6.5%) | 5 (6.3%) | 44 (6.1%) |
| Electrocution | 0 (0.0%) | 1 (0.6%) | 2 (2.5%) | 15 (2.1%) |
| Falls | 27 (39.7%) | 89 (53.0%) | 45 (57.0%) | 394 (54.4%) |
| Machinery | 4 (5.9%) | 15 (8.9%) | 6 (7.6%) | 64 (8.8%) |
| Struck by or against an object | 9 (13.2%) | 20 (11.9%) | 5 (6.3%) | 68 (9.4%) |
| Caught between objects | 1 (1.5%) | 6 (3.6%) | 2 (2.5%) | 20 (2.8%) |
| Other | 6 (8.8%) | 6 (3.6%) | 6 (7.6%) | 36 (5.0%) |

**Type of Injury**

The most common injury types across all workers were fractures, internal injuries and open wounds (Table 7). Black workers disproportionately suffered internal injuries and open wounds compared to other groups. White Hispanics were more likely to suffer amputations and the vast majority of white non-Hispanics suffered fractures (Table 7). Black workers disproportionately suffered type 1 traumatic brain injuries, the most severe type of brain injury, of which most occurred from falls (N=9) and assaults (N=5). The cause of traumatic brain injuries was primarily attributed to falls in all groups. Overall the distribution of injuries by body part was nearly identical between white non-Hispanic and white Hispanic construction workers.

**Table 7: Injury Type and Body Part Affected Among Construction Workers Filing Workers' Compensation Claims by Ethnicity**

| | Black (N=68) | White, Hispanic (N=168) | Other (N=79) | White Non-Hispanic (N=724) |
|---|---|---|---|---|
| **Injury Type** | | | | |
| Fracture | 34 (50.0%) | 107 (63.7%) | 47 (59.5%) | 532 (73.5%) |
| Dislocation | 6 (8.8%) | 7 (4.2%) | 4 (5.1%) | 51 (7.0%) |
| Sprain or Strain | 4 (5.9%) | 13 (7.7%) | 4 (5.1%) | 46 (6.4%) |
| Internal Injury | 26 (38.2%) | 34 (20.2%) | 21 (26.6%) | 160 (22.1%) |
| Open Wound | 30 (44.1%) | 50 (29.8%) | 31 (39.2%) | 211 (29.1%) |
| Amputation | 1 (1.5%) | 11 (6.5%) | 0 (0.0%) | 25 (3.5%) |
| Blood Vessel | 3 (4.4%) | 3 (1.8%) | 4 (5.1%) | 26 (3.6%) |
| Crush | 1 (1.5%) | 4 (2.4%) | 2 (2.5%) | 16 (2.2%) |
| Burns | 4 (5.9%) | 6 (3.6%) | 6 (7.6%) | 28 (3.9%) |
| Nerve Injury | 5 (7.4%) | 10 (6.0%) | 3 (3.8%) | 34 (4.7%) |
| **Body Part Injured** | | | | |
| Type 1 Traumatic Brain Injury | 20 (29.4%) | 19 (11.3%) | 10 (12.7%) | 77 (10.6%) |
| Type 2 Traumatic Brain Injury | 3 (4.4%) | 16 (9.5%) | 10 (12.7%) | 52 (7.2%) |
| Face and Neck | 26 (38.2%) | 42 (25.0%) | 26 (32.9%) | 194 (26.8%) |
| Spinal Column Injury | 0 (0.0%) | 6 (3.6%) | 3 (3.8%) | 16 (2.2%) |
| Vertebral Column Injury | 8 (11.8%) | 26 (15.5%) | 7 (8.9%) | 139 (19.2%) |
| Torso | 20 (29.4%) | 46 (27.4%) | 21 (26.6%) | 209 (28.9%) |
| Upper Extremity | 31 (45.6%) | 75 (44.6%) | 42 (53.2%) | 284 (39.2%) |
| Lower Extremity | 25 (36.8%) | 63 (37.5%) | 34 (43.0%) | 274 (37.8%) |
| System Wide or Unspecified | 12 (17.6%) | 23 (13.7%) | 10 (12.7%) | 95 (13.1%) |

***Totals in each column may exceed column totals because any individual workers may have suffered more than one injury**

### Measures from Hospital Stay

Injured black construction workers stayed in the hospital slightly longer than other workers, and they were more likely to be sent to an intensive care unit and be put on a ventilator, but none of these differences were statistically significant (Table 8). Mean injury severity scores did not significantly differ between the ethnic groups. In addition, both black and white non-Hispanic workers were slightly more likely to be discharged to an intermediate care or rehabilitation facility than were white Hispanic workers.

**Table 8: Measures of Injury Severity and Discharge Status Among Construction Workers Filing Workers' compensation Claims by Ethnicity**

| | Black (N=68) | White, Hispanic (N=168) | Other (N=79) | White Non-Hispanic (N=724) |
|---|---|---|---|---|
| **Hospital Treatment** | | | | |
| Mean Days in Hospital | 4.7 (sd=6.8) | 4.0 (sd=7.5) | 7.1 (sd=25.5) | 4.2 (sd=6.7) |
| Sent to Intensive Care Unit (ICU) | 19 (27.9%) | 27 (16.1%) | 15 (19.0%) | 132 (18.2%) |
| Mean Days in ICU | 3.4 (sd=5.9) | 4.3 (sd=5.6) | 4.3 (sd=9.1) | 3.3 (sd=6.6) |
| Put on Ventilator | 6 (8.8%) | 9 (5.4%) | 3 (3.8%) | 35 (4.8%) |
| Mean Days on Ventilator | 2.4 (sd=4.6) | 6.3(sd=6.4) | 4.1 (sd=4.6) | 2.3 (sd=6.1) |
| Underwent Operation | 27 (39.7%) | 90 (53.6%) | 38 (48.1%) | 389 (53.7%) |
| **Severity of Injury** | | | | |
| Penetrating Injuries (N=) | 8 (11.8%) | 16 (9.5%) | 8 (10.1%) | 65 (9.0%) |
| Mean Injury Severity Score | 8.0 (sd=6.4) | 8.1 (sd=7.4) | 7.8 (sd=6.8) | 8.3 (sd=8.4) |
| ISS 16 and higher | 9 (13.2%) | 23 (13.7%) | 11 (13.9%) | 96 (13.3%) |
| **Discharge Status** | | | | |
| Discharged Home | 57 (83.8%) | 144 (85.7%) | 62 (78.5%) | 587 (81.1%) |
| Acute Care / Inpatient Rehab / Skilled Nursing | 9 (13.2%) | 16 (9.5%) | 16 (20.3%) | 113 (15.6%) |
| Left Against Medical Advice | 1 (1.5%) | 4 (2.4%) | 0 (0%) | 4 (0.6%) |
| Morgue | 1 (1.5%) | 4 (2.4%) | 1 (1.3%) | 20 (2.8%) |

**Workers' compensation Claims**

The proportion of workers whose workers' compensation claims had been decided by the court system was nearly identical between the various ethnic groups: black, N=52 (76.5%); white Hispanic, N=132 (78.6%); white non-Hispanic, N=578 (79.8%); and other, N=58 (73.4%). The median total monetary compensation awarded to the workers differed significantly by ethnicity (p=0.01) (Table 9). Despite a difference in median monetary compensation, total percent permanent partial disability and weeks of total temporary disability did not significantly differ between the ethnic groups (Table 9).

**Table 9: Measures of Injury Severity and Discharge Status Among Construction Workers Filing Workers' compensation Claims by Ethnicity**

| | Black (N=68) | White, Hispanic (N=168) | Other (N=79) | White Non-Hispanic (N=724) |
|---|---|---|---|---|
| **Total Workers' compensation (USD)** | | | | |
| Median | $ 26,385.29 | $ 27,121.19 | $ 23,768.67 | $ 33,056.93 |
| Mean | $ 47,935.00 | $ 48,519.70 | $ 43,048.43 | $ 60,431.34 |
| **Mean Temporary Permanent Disability (Weeks)** | | | | |
| Median | 12.4 | 12.1 | 8.9 | 12.8 |
| Mean | 21.8 | 26.7 | 28.3 | 29.5 |
| **Mean Percent Permanent Partial Disability** | | | | |
| Median | 15.0% | 18.9% | 11.3% | 20.0% |
| Mean | 22.7% | 24.8% | 21.8% | 25.7% |

The difference in monetary compensation persisted when the data was stratified by injury type. Table 10 shows the mean total monetary compensation and mean percent permanent partial disability awarded for these specific injuries. White non-Hispanic construction workers consistently were awarded higher monetary settlements, despite in most cases within specific injuries the mean percent permanent partial disability was equivalent or lower than black and white Hispanic construction workers.

**Multivariable Models**

In the final multivariable robust regression model, compensation was $5,824 higher (p=0.030; CI95%: 551, 11097) for white non-Hispanic workers when controlling for other covariates compared to minority workers ($R^2$=0.25). In the second validation model in which the extreme outliers were removed, compensation was $6,844 higher (p=0.027; CI95%: 781, 12907) for white non-Hispanic workers when controlling for other covariates compared to minority workers ($R^2$=0.32).

**Table 10: Median Monetary Compensation and Average Permanent Partial Disability Among Construction Workers Filing Workers' Compensation Claims by Ethnicity**

| | Black (N=68) | White, Hispanic (N=168) | White Non-Hispanic (N=724) |
|---|---|---|---|
| **Fracture of Back and Spine** | | | |
| Number of cases | 6 (8.8%) | 25 (14.9%) | 131 (18.1%) |
| Median monetary compensation (USD) | $ 16,843.00 | $ 24,872.00 | $ 34,852.00 |
| Average Permanent Partial Disability (%) | 24.4% | 17.1% | 21.6% |
| **Open Wound Upper Extremities** | | | |
| Number of cases | 5 (7.4%) | 21 (12.5%) | 71 (9.8%) |
| Median monetary compensation (USD) | $ 30,856.87 | $ 6,407.19 | $ 17,235.51 |
| Average Permanent Partial Disability (%) | 25.5% | 37.7% | 29.0% |
| **Amputation Upper Extremities** | | | |
| Number of cases | 1 (1.5%) | 11 (6.5%) | 24 (3.3%) |
| Median monetary compensation (USD) | $ 10,000.00 | $ 9,571.00 | $ 14,517.00 |
| Average Permanent Partial Disability (%) | ~ | 49.3% | 42.0% |
| **Internal Injury Torso** | | | |
| Number of cases | 5 (7.4%) | 13 (7.7%) | 80 (11.0%) |
| Median monetary compensation (USD) | $ 14,498.00 | $ 20,386.63 | $ 27,034.42 |
| Average Permanent Partial Disability (%) | 28.8% | 26.6% | 21.7% |
| **Type 1 and 2 Traumatic Brain Injuries** | | | |
| Number of cases | 23 (33.8%) | 35 (20.8%) | 129 (17.8%) |
| Median monetary compensation (USD) | $ 31,324.02 | $ 28,687.59 | $ 45,600.88 |
| Average Permanent Partial Disability (%) | 16.0% | 16.5% | 18.8% |

**Results - Injury Severity Score (ISS) and Disability Rating**

Demographics for cases used in this analysis are shown in Table 3 above. The majority of cases were 25-54 year old males. About 60% were married at the time of injury, with slightly less than 50% having at least one dependent. Average weekly earnings for the injured construction workers in the data set were $882.00. Spearman rho correlation coefficients were calculated to provide a crude understanding of the relationship between the disability ratings and in-hospital severity. These coefficients, along with associated p-values, are shown in Table 11. Total compensation had a positive, statistically significant association with temporary disability, ISS, and hospital length of stay (LOS). Temporary disability and ISS were also positively correlated. Permanent disability was found to be negatively correlated with temporary disability and ISS.

**Table 11: Correlation Between Disability Ratings and in-Hospital Indicators of Severity Among Construction Workers Filing Workers' Compensation Claims**

| Factor (source) | Total Compensation (I-WCC) | Temporary Disability (I-WCC) | Permanent Disability (I-WCC) | Injury Severity Score (ITR) | Hospital Length of Stay (HD) |
|---|---|---|---|---|---|
| Total Compensation (I-WCC) | 1.000 | **0.183** | 0.047 | **0.250** | **0.226** |
| | - | **(< .0001)** | (0.198) | **(< .0001)** | **(< .0001)** |
| Temporary Disability (I-WCC) | | 1.000 | **-0.405** | **0.269** | 0.082* |
| | | - | **(< .0001)** | **(< .0001)** | (0.047) |
| Permanent Disability (I-WCC) | | | 1.000 | **-0.111** | -0.089* |
| | | | - | **(.013)** | (.030) |
| Injury Severity Score (ITR) | | | | 1.000 | **0.305** |
| | | | | - | **(< .0001)** |
| Hospital Length of Stay (HD) | | | | | 1.000 |
| | | | | | - |

Spearman rho coefficient, associated p-value shown in table
Significant positive correlations bolded in green, significant negative correlations bolded in red
*Though statistically significant at the .05 level, the low correlation coefficient has little practical significance
Note: Mean compensation for only cases with a decision. New filings and dismissed cases are excluded
Injuries resulting in death excluded (n = 20)

Analysis of variance using the non-parametric Kruskal-Wallis test demonstrated a statistically significant difference in temporary total disability at different categories of ISS ($p<0.001$) and hospital length of stay ($p<0.001$). Temporary total disability increased with an increase in ISS and length of stay. This pattern was consistent with the correlation analysis results. Temporary disability rating was highest for back and spine injuries, and for head and neck injuries (median =

7.50 and 7 weeks respectively). Head and neck injuries had the highest ISS assigned in the trauma center, with multiple extremities unspecified following (median = 9.00 and 8.00 respectively). Total compensation, temporary disability rating, and ISS all increase with increasing number of body parts affected. All three measures were relatively high for intracranial injuries, internal torso injuries, and blood vessel injuries. The large majority of cases had fracture injuries, leaving smaller n-sizes within other diagnosis categories.

In the crude robust regression models, ISS, hospital length of stay and whether a patient was discharged to a rehabilitation or intermediate care facility were all significantly associated with total temporary disability (Table 12). However, in the multivariable model after controlling for multiple covariates, ISS was not significantly associated with total temporary disability, but hospital length of stay and whether a patient was discharged to a rehabilitation or intermediate care facility continued to be significantly associated with the dependent variable (Table 12). The final multivariable model indicates that for each day a patient remains in the hospital, it is associated with an increase in 1.2 weeks of temporary total disability as awarded by the workers' compensation commission. Patients requiring continued care after discharge from the hospital at another facility received 10.75 more weeks of temporary total disability than those discharged home. In the model looking at permanent partial disability, ISS, hospital length of stay and whether a patient was discharged to a rehabilitation or intermediate care facility were not statistically associated with the dependent variable.

**Table 12: Association Between Temporary Total Disability (Weeks) and in-Hospital Indicators of Severity Among Construction Workers Filing Workers' Compensation Claims**

| | Crude Models (SE) | P-value | Final Multivariable Model (SE) | P-value |
|---|---|---|---|---|
| **Injury Severity Score** | 0.34 (0.15) | 0.024 | -0.05 (0.16) | 0.738 |
| **Hospital Length of Stay** | 0.88 (0.12) | <0.001 | 1.20 (0.19) | <0.001 |
| **Discharged to a Rehabilitation or Intermediate Care Facility** | 11.69 (1.87) | <0.001 | 10.75 (2.36) | <0.001 |

## Discussion

The value of a data linkage strategy rests on the supposition that more can be learned from multiple information sources together than from a single source. By focusing on construction workers, we feel that we demonstrated a few of the ways health information may relate to or differ from the information collected during the legal process in a high-risk group. However, to achieve the statistical power necessary to draw confident conclusions using this information, the public databases must offer ways to more efficiently link data across them.

Though the overall linkage rate was slightly lower than expected (9.9%), the demographic distributions among linked cases were very similar to those IWCC cases that did not link to either of the other databases. We feel that the similarities suggest that our linked data set was an acceptable sample of the overall population, and that the analysis results are meaningful. The fact that proportions by type of injury differed between the linked and non-linked cases is not surprising. Not every worker who filed a claim necessarily sought care at a trauma center or inpatient setting, and those who did may have had more serious injuries.

### *Ethnic Disparities in total monetary compensation*

White non-Hispanic construction workers consistently were awarded higher monetary settlements than black and white Hispanic construction workers despite (in most cases) equivalent or lower mean percent permanent partial disability within specific injuries. In the final multivariable models, the difference in the level of total monetary compensation persisted. In the final multivariable robust regression model, compensation was $5,824 higher (p=0.030; CI95%: 551, 11097) for white non-Hispanic workers when controlling for other covariates compared to minority workers ($R^2$=0.25). In the second validation model, in which the extreme outliers were removed, compensation was $6,844 higher (p=0.027; CI95%: 781, 12907) for white non-Hispanic workers when controlling for other covariates compared to minority workers ($R^2$=0.32). Further investigation of these ethnic disparities is necessary. Because the data were restricted to construction workers who were treated in level one and two trauma units, which generally are composed of the most severe acute traumatic injuries, it is possible that the ethnic disparities in compensation are specific to traumatic injuries suffered by these construction workers alone. In a broader dataset involving workers of all industries suffering both acute injuries, chronic injuries and illnesses, the disparities observed in this analysis may not persist. This study does not

explain why white non-Hispanic construction workers would receive higher compensation. All that can be stated is that the higher compensation is not explained by differences in injury severity, affected body region, type of injury, or common factors associated with higher compensation – number of dependents, average weekly wage, injuries resulting in death or in the use of an attorney.

***Injury Severity and Percent Partial Disability Ratings***

The correlation analysis and analysis of variance (Kruskal-Wallis) tests provided a crude understanding of the relationship between temporary total disability and hospital measures of severity. Temporary total disability was found to be positively associated with two in-hospital measures of severity – length of stay and discharge status. The final multivariable model indicates that each day a patient remains in the hospital is associated with 1.2 weeks' additional temporary total disability as awarded by the workers' compensation commission. The median length of stay was two days and the 95$^{th}$ percentile was 13 days. The difference between median stay and the 95$^{th}$ percentile group was approximately 13 weeks of total temporary disability. Furthermore, patients requiring continued care at another facility after discharge from the hospital received 10.75 more weeks of temporary total disability than those discharged home.

Head, back and spine injuries among construction workers stood out with the highest temporary total disability. They also had the highest mean severity scores. Future studies aimed at further characterizing these subsets of construction injuries would be useful in designing more effective safety interventions. These interventions could have positive effects on both the compensation process and the health care system in terms of costs avoided.

Two overarching themes seemed to emerge during this analysis. One was that the injury severity did not do a good job of predicting temporary total disability in the claims process. This suggests that information on lost work days (temporary total disability) should not be used as a proxy for the medical condition of the worker during a trauma visit or hospital stay. Thus, the data linkage provides new information about the injured worker not contained within either the IWCC or ITR alone. In addition, in the workers' compensation dataset, the nature of injury was poorly specified. By linking the datasets, we were able to gather much more precise information about the nature of injury and associated medical complications and procedures.

The major limitation of the present study was low sample size. This may have been a result of the probabilistic linkage strategy that had to be employed in the absence of a shared case identifier among the data sets. Along the same lines, there was a level of uncertainty that could not be avoided because of this methodology. However, we feel that the linkage variables used resulted in highly accurate matches among the cases that did link across at least two of the databases. This was checked manually by comparing ITR and HD diagnosis codes with IWCC nature of injury and body-part-affected fields.

Data linkage methodologies show promise in the field of injury epidemiology. By linking various existing data sources, interested parties can greatly increase the amount of information available for injury cases. For instance, by linking the three Illinois databases mentioned at the beginning of this review (ITR, HD, I-WCC), one would be able to follow an injured worker through both the health care system and the workers' compensation claims structure. The different information contained within each database might help us to better characterize risk factors and design more effective interventions.

## References


Baker, S.P., O'Neill, B., Haddon, W., Jr., and Long, W.B. The Injury Severity Score: A Method for Describing Patients with Multiple Injuries and Evaluating Emergency Care. J Trauma 14:187-196, 1974.

Barell V, Aharonson-Daniel L, Fingerhut LA, Mackenzie EJ, Ziv A, Boyko V,Abargel A, Avitzour M and Heruti R. An introduction to the Barell body region by nature of injury diagnosis matrix. Inj Prev. 2002: 8; 91-96.

Bureau of Labor Statistics. 2007. Census of Fatal Occupational Injuries. 2007. http://www.bls.gov/news.release/pdf/cfoi.pdf, accessed 9/25/07

Center to Protect Workers Rights. The Construction Chart Book. 2002; http://www.cdc.gov/eLCOSH/docs/d0100/d000038/pdfs/page%2015.pdf, accessed 11/2/07

Friedman LS, Forst L. The Impact of OSHA Recordkeeping Regulation: Changes in Occupational Injury and Illness Trends in the US: A Time Series Analysis. Occup Environ Med 2007(a); 64(7):454-60

Friedman LS and Forst L. Occupational Injury Surveillance of Traumatic Injuries in Illinois, Illinois Trauma Registry: 1995-2003. J Occ Environ Med. 2007(b) Apr;49(4):401-10.

Friedman LS and Forst L. Workers' Compensation Costs Among Construction Workers: A Robust Regression Analysis. J Occ Environ Med. 2009; 51(11): 1306-1313.

Huybrechts T, Thas O, Dewulf J, and Van Langenhove H. How to estimate moments and quantiles of environmental data sets with non-detected observations? A case study on volatile organic compounds in marine water samples. 2002; 975 (1): 123-133.

Jurkovich, Gregory J. MD; Mock, Charles MD, PhD. Systematic Review of Trauma System Effectiveness Based on Registry Comparisons. J Trauma. 1999 Sep;47(3 Suppl):S46-55.

Manufacturers' News, Inc. Evanston, Illinois. State Manufacturers Database of the Construction Industry: Illinois. 2007. http://www.manufacturersnews.com/, accessed 5/1/09.

NIOSH. 2004. High Risk Injuries and Occupations. Chapter 4. http://www.cdc.gov/niosh/docs/2004-146/ch4/ch4.asp.htm , accessed 5/1/09.

Parkhurst DF. Arithmetic versus geometric means for environmental concentration data. Environmental Science and Technology. 1998; 32: 92A-98A.

Rosenman K, Kalush A, Reilly MJ, Gardiner JC, Reeves M, Zhewui L. How Much Work-Related Injury and Illness is Missed by the Current National Surveillance System? J Occup Environ Med 2006;48(4):357-65

US Bureau of Labor Statistics. US Department of Labor. Current Employment Survey. State and Area Employment, Hours and Earnings. Available at: http://data.bls.gov/PDQ/outside.jsp?survey=sm, accessed, March 12, 2009

# APPENDIX 1

## Data Linkage Code

**Step 1:** Clean the variables that will be linked and create a linkage key unique to each dataset before the linkage begins. This unique linkage key will be used to link the data variables from that dataset after linkage process is complete.

In this case we used date of birth, date of injury, gender and residential ZIP code. We converted all four to the same format in all three datasets. We also cleaned the injury variables in each dataset. In the hospital discharge dataset there are 9 variables describing the injury/diagnosis and are based on ICD-9 codes. In the trauma registry dataset there are 25 variables describing the injury/diagnosis and are based on ICD-9 codes. In the workers' compensation dataset there are two variables describing the injury/diagnosis and are based on a coding system specific to IWCC.

**Step 2: Link the data on the four linkage variables**

Below is an example of the SAS code for the first pass between the workers' compensation and hospital discharge datasets. All matches involve exact matches.

```
LIBNAME t "C:\…LOCATION OF YOUR FILE COPY AND PASTE HERE";

PROC SQL;
CREATE TABLE WCC_HD_Pass1 as
SELECT wcc.CASENO, wcc.DOBDate, wcc.DOADate, wcc.Pet_ZIP, wcc.PET_SEX,
wcc.natureinjury, wcc.bodypart, wcc.accdtype,
hd.HDID,hd.date_birth,hd.date_admit,hd.date_discharge,hd.Patzip,hd.patsex,hd.dx1,hd.dx2,hd.d
x3,hd.dx4,hd.dx5,hd.dx6,hd.dx7,hd.dx8,hd.dx9
FROM t.wccdata as wcc INNER JOIN t.hddata as hd
ON wcc.DOBDate = hd.date_birth
AND wcc.PET_SEX = hd.Patsex
AND wcc.Pet_ZIP_Num = hd.Patzip
AND wcc.DOAdate = hd.date_admit;
QUIT;

***Where
WCC_HD_Pass1 is the name of the new dataset created which contains only the cases that linked
during the first pass
```

The variables included in the linkage from the workers' compensation dataset are wcc.CASENO, wcc.DOBDate, wcc.DOADate, wcc.Pet_ZIP, wcc.PET_SEX, wcc.natureinjury, wcc.bodypart, wcc.accdtype

The variables included in the linkage from the hospital discharge dataset are hd.HDID,hd.date_birth,hd.date_admit,hd.date_discharge,hd.Patzip,hd.patsex,hd.dx1,hd.dx2,hd.dx3,hd.dx4,hd.dx5,hd.dx6,hd.dx7,hd.dx8,hd.dx9

FROM t.wccdata as wcc -- this is saying bring in the original wccdata file in library t and call it "wcc". This is the abbreviation you will use before each variable in this dataset.

INNER JOIN -- this command means that only matches to the workers' compensation will be included in the new dataset "WCC_HD_Pass1". Cases that don't match will be dropped.

t.hddata as hd -- this is saying bring in the original hddata file in library t and call it "hd". This is the abbreviation you will use before each variable in this dataset.

ON wcc.DOBDate = hd.date_birth -- This is the linkage command. It is saying to look for exact matches on date of birth in both datsets.

;

Below is an example of the SAS code for the third pass between the workers' compensation and hospital discharge datasets. All matches involve exact matches, except date of admission/accident which allows for plus/minus one day.

```
LIBNAME t "C:\…LOCATION OF YOUR FILE COPY AND PASTE HERE";

PROC SQL;
CREATE TABLE WCC_HD_Pass3 as
SELECT wcc.CASENO, wcc.DOBDate, wcc.DOADate, wcc.Pet_ZIP, wcc.PET_SEX,
wcc.natureinjury, wcc.bodypart, wcc.accdtype,
hd.HDID,hd.date_birth,hd.date_admit,hd.date_discharge,hd.Patzip,hd.patsex,hd.dx1,hd.dx2,hd.d
x3,hd.dx4,hd.dx5,hd.dx6,hd.dx7,hd.dx8,hd.dx9
FROM t.wccdata as wcc INNER JOIN t.hddata as hd
ON wcc.DOBDate = hd.date_birth
AND wcc.PET_SEX = hd.Patsex
AND wcc.Pet_ZIP_Num = hd.Patzip
AND wcc.DOAdate BETWEEN (hd.date_admit  + 1) AND (hd.date_admit  - 1);
QUIT;

***Where
WCC_HD_Pass3 is the name of the new dataset created which contains only the cases that linked during the first pass
The variables included in the linkage from the workers' compensation dataset are wcc.CASENO, wcc.DOBDate, wcc.DOADate, wcc.Pet_ZIP, wcc.PET_SEX, wcc.natureinjury, wcc.bodypart, wcc.accdtype
The variables included in the linkage from the hospital discharge dataset are
hd.HDID,hd.date_birth,hd.date_admit,hd.date_discharge,hd.Patzip,hd.patsex,hd.dx1,hd.dx2,hd.d
x3,hd.dx4,hd.dx5,hd.dx6,hd.dx7,hd.dx8,hd.dx9
FROM t.wccdata as wcc -- this is saying bring in the original wccdata file in library t and call it "wcc".  This is the abbreviation you will use before each variable in this dataset.
INNER JOIN -- this command means that only matches to the workers' compensation will be included in the new dataset "WCC_HD_Pass1".  Cases that don't match will be dropped.
t.hddata as hd -- this is saying bring in the original hddata file in library t and call it "hd".  This is the abbreviation you will use before each variable in this dataset.
```

ON wcc.DOBDate = hd.date_birth  -- This is the linkage command.  It is saying to look for exact matches on date of birth in both datsets.
AND wcc.DOAdate BETWEEN (hd.date_admit  + 1) AND (hd.date_admit  - 1) - this is the linkage command that allows for fuzzy matching on date of admission and date of accident to be off by plus/minus one day.

**Step 3.** Merge all the subsets with matched cases, then verify that they are matches by looking at type of injury.  This was done manually because the workers' compensation system uses a different coding system for injury and cause of injury.  In data systems that use more common coding systems like ICD-9, this process can be automated.

CPWR THE CENTER FOR CONSTRUCTION RESEARCH AND TRAINING

www.cpwr.com • www.elcosh.org