Final Report

Principal Investigator: Arthur Oleinick, MD,JD,MPH
Institution: University of Michigan
Department: Environmental Health Sciences
Ann Arbor, MI 48109
Phone: 734-665-7693
e-mail: aoleinic@umich.edu
Project Title: Trucking firm characteristics, driver injury and outcome.
Date: September 27, 2005
Co-Investigators: Julie Horrocks, PhD, biostatistician, 07/01/01-07/31/02
Daniel F Blower, PhD, co-investigator, 07/01/01-1/30/05
Kenneth E Guire, MS, biostatistician, 07/01/01-06/30/04
Sponsor: National Institute of Occupational Safety and Health
Grant Number: R01 OH03804-01A2
Grant Period: July 1, 2001-June 30, 2005

Table of Contents

Abstract:

Problem: In the period 1993-2001(later years are not directly comparable) truck drivers had the highest number of injuries and illnesses producing days away from work of any occupation in the Bureau of Labor Statistics Annual Survey of occupational injury and illness (BLS Survey). Counts were 50-100% higher than those in the next two highest groups (laborers, nonconstruction, and nursing aides and orderlies). In addition to the specialized occupational risks attributable to driving large vehicles and road accidents, truck drivers are also exposed to risks of traumatic and repetitive motion injuries from materials handling as they load, rearrange and unload shipments. However, aside from the greater risk of back injuries among professional drivers and the obvious risk of vehicular accidents, very little is known about whether truck drivers are at higher risk for injuries at other sites, or the days away from work produced by the injuries. There are two reports that truckers' erratic work schedules may affect receipt of health services.

Approach: Ohio is one of only five jurisdictions that designate a state agency as the exclusive provider of workers' compensation insurance under state law. In Ohio, employers may qualify for self-insurance based on size of the company ($\geq$500 employees) and financial criteria. Data indicate that the Ohio Bureau of Workers' Compensation (OBWC) insured employers employ at least two-thirds of the civilian employed population in Ohio. In the 2000 census, the civilian employed population totaled 5.4 million individuals or about 4.2% of the U.S. total. In FY2002, OBWC reported that there were 213,227 allowed claims for medical and/or wage loss associated with work injuries or illnesses.

After review and approval of the proposal by the Ohio Bureau of Workers' Compensation (OBWC) legal department and the University's Institutional Review Board, the OBWC provided a data extract in nine files for 24,131 claims (35,688 diagnoses accepted for payment) from the for-hire carrier industry. Most cases were from for-hire trucking firms, although bus, taxi and emergency medical companies were included. The complexity of the extraction process for the required data was reflected in the fact that it took nine months from the grant start date before we received the final study files. After removing what appeared to be duplicate claims and claims from the same individual with the same date of injury but with differing diagnoses accepted for payment, the initial study population consisted of 23,965 cases with 35,501 separate diagnoses accepted for payment.

As we worked with the data extract, it became clear that the data could be used to achieve two of the three original aims:

1. To identify factors associated with the incidence of injuries by occupation, the use of medical care for such injuries and the duration of lost worktime produced by these injuries; and,

2. To evaluate the predictive value of various types of models (an exercise now grouped under the rubric of "concordance statistics").

3. The data identifying medical care providers by type of organization and specialty were simply too sparse for use in achieving objective three, i.e., detecting associations between regularity of job schedules as reflected by occupation and source of services.

The final study population was further reduced to 23,491 cases with 34,165 diagnoses by removing superficial injury diagnoses (910.*-919.*). These diagnoses were

excluded to avoid diluting the effort to classify more serious diagnoses by functional area affected.

Substantial technical editing was required to convert the nine administrative files into analytic files suitable for modern modeling techniques. This editing consumed almost all the grant time from receipt of the data extract through the one-year no-cost extension, a period of three years and three months. The completed work included seven software modules plus a partial module (all ©University of Michigan) that link multiple claims to a single claimant (20,802 individuals/23,491 claims over three years), track claim number changes over time (approximately 800 changes), remove duplicate or inconsistent claims (166 claims/187 diagnoses), classify diagnoses on several dimensions (functional area affected, severity and whether index injuries had comorbidities) and link diagnostic groups to health services use (the last is the partial module). In addition, manual coding of occupation and industry from text narratives, assignment of race using geocoding and 2000 census data and more complete ascertainment of occupation for truck drivers by using additional Ohio and federal files was completed.

The ninth module, under construction, will yield summary tables of health services utilization by diagnosis and permit analysis of the course of diagnosis and treatment. A tenth module, not begun, would provide estimates of another outcome variable, the number of days away from work derived from wage compensation paid. This last problem is not trivial technically because of the number and legal definition of payment codes used. All software requires only that specified data elements be available somewhere in the data file.

Key Findings: The work that has been completed already provides a more accurate assessment of the medical burden imposed by work injuries on employed populations. We have compared our data, using all medical diagnoses accepted for payment, with BLS Survey data that use mostly lay narratives to characterize nature of injury and part of body affected. In addition, BLS requires either that a case be assigned to the most serious injury or that it be placed in a multiple injury group, either confined to an extremity or simply designated as multiple injuries, with consequent loss of detail.

Our data show that major joint areas such as the back, shoulder and knee, all important functional areas for the occupations involved, are more frequently involved than shown in the BLS Survey. Further, classifying the diagnoses by severity shows that the diagnoses are more severe than indicated by the BLS nature of injury code. Third, comorbidities for severe index diagnoses are themselves often severe. Fourth, exploratory data analyses indicate that the distribution of functional area affected varies by occupation, but modeling is simply too early to be reported. Fifth, preliminary analyses of health services utilization for back disc injuries and sprains yields data on medical costs by condition that should prove useful in cost-benefit analysis of competing primary and tertiary prevention intervention strategies, when systematically available.

Using the Results: Study findings are not yet ready for workplace application. The remaining software modules must be completed and statistical modeling to identify risk factors done. However, our work documents that a statewide administrative database of work injuries can be converted to a rich source of analytic files. These analytic files are relatively inexpensive and results can be easily referenced to the medical literature because they rely on standard medical terminology. Study results can serve as input to ergonomic evaluations of high risk occupation/injury combinations and aid in the design

of work environments, in particular truck cab design. Information on the medical care costs and, ultimately, days away from work (DAFW) costs for specific diagnoses or related diagnoses will prove critical in designing and choosing primary and tertiary prevention activities.

Highlights/Significant Findings:

Using the software modules noted, we have mapped the work injury diagnoses to the functional area affected, grouped diagnoses within five of nine areas by relative severity and then indexed each diagnosis to every other diagnosis in the case by anatomic proximity to identify comorbidities. 894 diagnoses in the study group were categorized to one of twelve functional areas, further subdivided within the functional area into either 23 (back) or 13 (other specific areas such as shoulder, knee, etc.) severity groups and then further categorized with respect to proximity to every other diagnosis (proximal= adjacent functional area v. distal= distant functional area). These 894 diagnoses accounted for 93.90% of the 34,165 diagnoses in the study cohort and accounted for all diagnoses in slightly over 90% of cases. The classification algorithms appear to capture a high percentage of medical information in the files.

These three variables- functional area, severity and injury comorbidities, all based on diagnosis, relate to the two grant objectives noted above. In particular, the variables can be used to identify the strength of associations between occupation and injury site or injury severity or the likelihood of comorbidities from the same event, adjusting for a variety of demographic covariates. We have undertaken preliminary analyses of the relationship between occupation and injury area and also of the occurrence of comorbidities by area and occupation, but the work is to early to be reported. Concordance statistics to evaluate model utility are generated by standard SAS programs.

However, an important finding that was not anticipated at the start of the study is that the medical burden of work injuries is currently underestimated using the BLS nature of injury coding scheme in the Annual Survey. The difference is explained by the fact that the BLS generally codes from a lay narrative of the incident while we used diagnoses accepted for payment. The differences between the two approaches produce some fairly dramatic differences. For example, approximately 2% of all cases in all occupations in our study cohort (approximately 40% truck drivers) have a displaced/herniated intervertebral disc compared to about 0.5% of BLS cases with at least one DAFW (not counting the day of injury) among all occupations, US or Great Lakes States, or among truck drivers at the national level. Since at least two sources indicate that some two-thirds of work injury cases do not experience time away from work beyond the day of injury, the difference between the two approaches is actually more dramatic.

In bivariate distributions injuries of the back, shoulder and knee are more common among truck drivers than among laborers or clerical occupations. We have begun to use functional area and comorbidity variables created by the software as outcomes in logistic regression and Mantel-Haensel chi-square analyses, but results are too preliminary to be reported. Unfortunately, denial of a second no-cost extension with sufficient funds remaining to complete the work has slowed the pace of analysis considerably.

As noted, some work is needed to complete the software modules linking diagnosis to health services utilization. These modules can be expanded in a straightforward fashion to investigate injuries in other areas that have analogous

continuums, e.g., shoulder enthesopathies v. sprains. This would permit achieving the aims noted above. In addition, having the actual costs of the medical procedures allows us to identify the costs per case with a particular diagnosis and we have preliminary data for disc injuries. This information is clearly of use in considering primary and tertiary prevention strategies. Unfortunately, denial of a second annual no-cost extension with adequate funds remaining has slowed work in this area.

The preliminary analyses of disc injuries and back sprains indicate that the average medical cost of care for a case with a disc injury, with or without additional diagnoses, was approximately $23,500 and that this injury group of 477 cases was responsible for approximately 21% of all health care costs for the study cohort. A tertiary prevention effort that was even modestly effective would be cost-effective on a statewide basis.

Translation of Findings:

The findings to date in the present study can not be directly applied to the workplace at this time. However, the development of the three classification taxonomies-functional area affected, severity based on both underlying pathology and clinical course and classification of associated diagnoses by proximity to index injuries, make it possible to organize the massive amounts of data in administrative databases for analytic purposes. Each of these variables can be used to model the data. The diagnosis classification algorithms underlying the software are independent of occupation or industry. Occupation, race and trucking firm characteristic variables have also been created by technical editing of the various study files. More precise determination of occupation for other than truck drivers will require an increase in the detail provided for occupation on the text description. Ascertainment of truck drivers was better because we had three additional files to check against, as noted above.

Outcomes/Relevance/Impact:

Even though the work proved to require more time than we budgeted, the work we've completed has yielded some important outcomes. First, and most important in the Principal Investigator's view, the creation of the three classification modules and completion of the other tasks described have already produced some important findings. When medical diagnoses are used to classify the medical burden of work injuries instead of the quasi-medical taxonomy employed by the BLS Annual Survey to code lay narratives, the burden on employed populations is greater than previously appreciated. Taking into account all medical diagnoses resulting from an injury event instead of confining classification to the "most serious" injury as the highest coding priority, serious comorbidities are identified as imposing substantial additional burdens on injuries that are already significant by themselves. The size and particular distribution of the medical burden of work injuries should be of use in discussions of research priorities and the design of research protocols.

In early analyses, the distribution of injuries by functional area affected appears to differ by occupation. If the concordance statistics for the models warrant it, these differences should be exploited to identify differences in work environment that may be associated with the risks of injuries at different sites. The ability to identify such differences could lead to the development of primary and tertiary prevention activities to reduce the risk. Moreover, the ability to link medical diagnosis to health care utilization

costs (partly developed) and wage compensation costs will provide a basis for choosing the most cost-effective interventions.

The current study establishes that it is possible to use medical diagnoses to describe the work injury experience of a major industrialized state's civilian employed population that represents one in 40 American workers, when the self-insured are omitted from the count. The use of standard medical diagnostic terminology is of some importance because all results of epidemiologic studies on this workforce can immediately be integrated into the substantial occupational medical literature through the common terminology. The powerful positive effect on the pace and output of research of having available a large population-based database employing medical terminology can be seen in the experience of the National Cancer Institute's Surveillance, End Results and Epidemiology (SEER) registries of cancer cases around the country. A recent Ovid© Medline© search yielded an estimated 1200+ articles and numerous intramural government reports using data from this source since their establishment in 1973 compared to 77 reports that cite the BLS data since the establishment of the Supplementary Data System injury characteristic reporting in 1976.

Scientific Report:

Background: Over the period 1993-2001, truck drivers had the unenviable distinction of being the occupational group with the highest number of injuries and illnesses involving time away from work (DAFW)(counts since 2001 are not comparable).[1] While the counts for the two occupations with the next highest DAFW (nursing aides, orderlies; laborers, nonconstruction) decreased monotonically over this time period to a level 60-70% of their 1993 value, the count in truck drivers decreased erratically and only some 15%. Moreover, among selected occupations for whom BLS reported the percent distribution and median for days away from work, truck drivers tied for the second highest median number of days away from work and had the third highest fraction (29.9%) with DAFW exceeding 30 days.

While the risk from transportation accidents is one obvious risk for this occupation, transportation accidents account for a little less than 10% of the events producing the injuries and illnesses in the transportation and public utilities industry, an industry that includes trucking and warehousing (SIC code 42).[2] The truck driver is also exposed to the risks involved in loading, arranging and unloading their truck and in securing the load for driving. Such activities as setting chains to secure the load present the risk of falls from heights, for example, as the truck driver performs the job duties.

An OVID© Medline© search combining the key words "truck driver" and the exploded terms "Workers' Compensation" or "Wounds and Injuries' or "Accidents, Occupational" or "Occupational Medicine" or "Occupational Diseases" yielded approximately 40 original articles over the past ten years. About 25% of these dealt with back and other musculoskeletal injuries. All of these studies focused on a single injury area such as low back pain or work-related musculoskeletal disorders. The modal group, 14/40, dealt with working conditions or road accidents. Most of the remaining studies described the risk of myocardial infarction or cancer.

There have been two reports from the same group of the effect of preexisting comorbidities on the severity of road accidents,[3,4] and only one (discussed below) that describe the impact of comorbidities arising from the same injury event on recovery from

particular index injuries.

There have now been two reports using questionnaires that document that truck drivers have difficulty accessing the health care system.[5,6] Neither study had as an aim linking specific injuries and health care utilization data to identify the manner in which diminished access affected the duration of disability.

At the time we undertook the present study, there had been no work providing a comprehensive view of the distribution of all injuries and illnesses among truck drivers that could be used to identify risk factors for these conditions or to set primary and tertiary intervention strategies. No work was reported during the grant period that changed that. We are recently submitted a paper that provides a comprehensive picture of the range of morbidities that occur in employees of for-hire carriers.

Methods:

1. The data extract.

Ohio is one of five state jurisdictions that require that workers' compensation insurance be obtained exclusively through a state agency unless the employer elects to self-insure and satisfies the eligibility (>500 employees) and financial stability requirement. In FY2002, 0.43% of Ohio employers were self-insured and they accounted for 20-22% of all benefit payments.[7] It is estimated that this reflects coverage for more than two-thirds of the 4.2 million participants in the Ohio civilian employed workforce in the 2000 census.[8] Taking account of the extent of coverage, Ohio can provide information on work injuries and illnesses for approximately one in 40 American workers. Ohio is representative of heavily industrialized states in this country, accounting for almost 6.0% of the workforce in manufacturing, 3.7% in construction and 4.0% in transportation and warehousing.

Because the Ohio Bureau of Workers' Compensation is responsible for both administering the law and supervising the provision of medical services for the Ohio workers that it ensures, it maintains unusually detailed records on cases. This is in contrast to states where a state agency administers the law and commercial insurers provide the insurance. In this latter situation, the agency generally maintains detailed computer records for all cases only for the administrative facts such as type and duration of payments of wage compensation and gets medical information only for contested cases. The commercial insurers (over 400 in Michigan in 2005) maintain detailed medical and payment records on those they insure but differences in the data format between companies and the difficulty of arranging collaboration to produce what would be a representative population appear almost insurmountable. Ohio provides what is a rare opportunity to study a statewide population.

OBWC initially provided test runs of the data extract that were checked for content by the Principal Investigator. Changes in the request were made, where possible, and the final data extract was obtained in March of 2002, almost nine months after the grant began.

The data extract was prepared from OBWC's comprehensive relational database. All accepted claims for the study period 1997-1999 for work injuries from employees of for-hire carriers with establishments in Ohio were included. The National Council on Compensation Insurance (NCCI) rating manual codes[9] were used to identify the study population. The codes employed were: 7219 (Truck NOC all); 7222 (Truck oilfield); 7228-7232 (Truck local, long haul, parcel, mail/parcel, mail/package delivery); 7370

(Taxi); 7380 (Driver NOC); 7382 (Bus).

The data extract consisted of nine files in text format, all linked through the claim number of the index cases in the study period. These nine files consisted of the following: demographic data; a First Report of Injury file containing text narrative describing the injured worker's occupation and a narrative of the injury occurrence; a list of International Classification of Diseases-9th Edition-Clinical Modification[10] (ICD-9-CM) diagnoses accepted for payment; a list of ICD-9-CM diagnoses accepted for payment for claims by study cohort members over as long as 20 years preceding the index injury date; two health services utilization files for cohort claims reflecting two administrative systems for payment; two health services utilization files for prior claims by study cohort members; and, a file with wage compensation paid by interval of payment. The data extract at delivery consisted of 24,131 approved claims containing 35,688 approved diagnoses.

Substantial technical editing of each of these files was necessary to convert the administrative data into analytic data. The technical editing was completed by creating software, coding text files, linking to federal and state files containing firm and truck driver information and by subcontract to the UM Institute for Survey Research to geocode addresses to block and to add selected variable from the 2000 Census.

All technical edit work was directed toward the creation of covariates and outcome measures for identifying risk factors through modeling. The technical editing was necessary to accomplish the study aims of identifying risk factors for the occurrence of injuries, to determine whether health care for the same conditions differed by occupation and to determine the demographic, job, diagnoses and health care factors that showed strong associations with the duration of work disability measured by time away from work paid for by wage compensation.

2. The conversion software

Developing the algorithms reflected in the software modules described below (©University of Michigan), creating the modules (©University of Michigan) and then confirming that the output conformed to the intended algorithms occupied the major portion of grant time. Seven of ten modules are complete, one is partially complete, one is under construction, and a tenth is yet to be started. The modules are:

1. One module assigned claims to individual workers on the basis of an algorithm based on last name, date of birth, first name, middle initial, and zip code.
2. One module identified duplicate claims. Claims were considered duplicate if they involved the same worker, had the same date of injury and involved the same diagnoses or, in the case of the last criterion, had irreconcilable medical diagnoses. 166 claims involving 187 diagnoses were removed at this stage.
3. One module tracked changes in claim number over the life of the claim so that health services utilization could be correctly allocated to individual claims. Approximately 800 claim number changes were identified for later use in allocating health services utilization by claim.
4. One module had the same goal as #2 but required additional programming before it could be applied to prior medical diagnoses.

All outputs of the first four modules have been checked against the intended algorithms by the Principal Investigator and are believed to be correct.

The next three modules converted administrative information into analytic data by

creating a classification scheme that could be represented by design variables in statistical modeling. These three modules were:

1. A functional area classification module. Two simplifying medical criteria were used to create this module. First, because all physical work involves major joint movement and all mental work involves the head and five senses, nine functional areas were created. These are the head, back, shoulder, elbow, carpal tunnel syndrome (this category was considered separately because of the current intense interest in this condition), hand/wrist, hip, knee and ankle/foot. In addition, three general functional areas were created: burns, toxic exposures and herniae. All diagnoses were used to assign functional area so that a case could, and did, appear in more than group.

   The second simplifying assumption was to remove all ICD-9-CM diagnoses in the range 910.*-919.*. This range applies to superficial injuries, i.e., injuries that are abrasions, blisters, insect bites, and splinters. These diagnoses were excluded to facilitate classification by area of functional effect. This resulted in the exclusion of 1,336 diagnoses, of which there were 474 claims with diagnoses only in this group. Since 2.3% of this group did receive wage compensation for more than seven days away from work, compared to 22.2% overall, some potentially useful information was lost. In addition, we included in an "other" category, but did not further categorize, mental health diagnoses or diagnoses involving the thorax and abdomen, except for the back. Once again, these diagnoses were treated differently so as to facilitate area assignment for traumatic and cumulative trauma injuries involving the designated functional areas. These diagnoses are retained and could be assigned to individual claims in the future if studies of vehicle accidents were undertaken, for example.

   A small group of diagnoses, generally regarded as complications of an original injury were also removed at the start of the classification, but were returned to the original claim if an acute diagnosis was consistent with the complication. For example, a malunion of a facture (ICD-9-CM 733.81) was reattached to the claim data if an acute fracture was recorded.

   Rather than assign only those diagnoses included in the study cohort to functional area, the Principal Investigator decided to map all potential work-related diagnoses in the ICD-9-CM to the module classification scheme. Although some additional time was required by this strategy, the structure of the ICD-9-CM taxonomy actually facilitated the mapping because individual diagnoses or groups of diagnoses incorporate anatomic areas as part of the code and the diagnoses are grouped by etiologic or anatomic area.

   OBWC also provided a data element called "Body Object" that contained information on anatomic localization for "general" diagnoses. This last group consists of diagnostic terms that specify a diagnosis but do not specify a location, e.g., 729.2- neuralgia, neuritis, and radiculitis, unspecified.

   Altogether, 3,419 diagnoses were available to classify study case diagnoses using both the medical diagnosis and "Body Object" data element. A module subroutine allowed the investigator to examine on-line the FROI for assignment purposes, but the diagnostic yield was so low that the subroutine

was commented out. It could be restored to the module if the FROI usefulness were improved.

The classification module was applied to both the current and prior claim diagnostic file. The second module was a variant of the first modified to work with the data in the prior claim file.

2. A second module was developed to classify diagnoses within a functional area by severity. Severity classification of diagnoses was done by the Principal Investigator, an experienced epidemiologist with substantial experience and interest in classification issues, and a senior PM&R clinician in consultation with special expertise in work-related disorders of the back and upper extremity (Robert A Werner, MD, MS, Professor of Physical Medicine and Rehabilitation and Chair, PM&R, VA Hospital, Ann Arbor).

   Severity groupings were based on the nature of the underlying pathology and the usual clinical course for the diagnosis. Permanent tissue changes were rated more severe than acute tissue changes and diagnoses requiring chronic medical management more severe than conditions that require a limit number of medical contacts. A more detailed first stage severity classification (23 potential groups) was developed for back injuries than for the extremity areas (13 potential groups). A severity grouping for head injuries remains to be done.

   Severity groupings are available for four functional areas: back, shoulder, elbow and knee. Together these four functional areas account for 48.53% of the diagnoses in the study groups. The shoulder, knee and elbow severity classifications are all based on the same disease groupings and could be extended to the remaining joint areas with a modest amount of work. In reaching agreement on the severity ratings Drs. Oleinick and Werner reconfirmed the accuracy of the initial functional area assignment and came to consensus on the severity rank. Second and third level groupings require additional work.

3. A third module also localized each diagnosis in relation to every other diagnosis in the claim. The localization was dichotomous. Injuries in functional areas adjacent to the index case area were labeled as proximal, all other functional areas were designated as distal. Burns, toxic exposures and hernia were not included in this effort. The resulting classification, together with severity designation, provides one approach to the consideration of the influence of comorbidities on outcome. This topic is discussed in the Results section.

An eighth module, partially completed, and a ninth, yet to be done, are intended to link diagnoses to health services utilization generally regarded as appropriate for the diagnosis or treatment of the conditions. All back injuries, intervertebral disc injury and back sprains have been preliminarily linked to the diagnostic radiologic procedures for these conditions. The prototype linkage can be extended to allow comparison of analogous conditions in other body regions, e.g., all shoulder injuries, shoulder dislocations and/or shoulder enthesopathies compared to shoulder sprains. The eighth module is intended to allow analysis of the diagnostic evaluation process. The ninth module yet to be developed is intended to provide data on the information underlying the

diagnosis at the time the agency accepted the diagnosis for payment.

A tenth, and final, module also remains undeveloped. This module would gather information on the duration of work disability as measured by days away from work that were covered by wage indemnification. The problem is far from trivial because OBWC employs a large number of payment codes to identify wage payments and the administrative meaning of these codes will require a fair amount of work from an individual with legal and scientific training to lay out an algorithm that correctly reflects the payment history.

All algorithms and software modules developed under the grant are copyrighted by the University of Michigan.

3. Coding text narrative relating to occupation and injury characteristics

To code occupation and injury characteristics according to the BLS coding scheme, we used the "train the trainer" approach often used in occupational health training. Ms. Clarissa Simon attended the BLS training program in Occupational Injuries and Illnesses Case Coding.[11] In turn, she trained a panel of largely graduate students in the methodology. Each coder was provided with copies of the Occupational Injury and Illness Classification Manual,[12] the Occupational Illness and Coding System Guidelines (in the Training Manual), the Occupational Injury and Illness Coding Interpretations Manual,[13] and the Occupational Injuries and Illnesses Occupational Coding Manual.[14] Personal identifiers were blinded to the coders by use of a key. An important feature of the BLS coding protocol is that it requires, where more than one injury is noted, that the coder code to the "most serious" injury so that a case is allocated only once to a category (see, for example, Nature of Injury, Rule of Selection 1.3). Subsequent coding of other categories is based upon this initial choice. When a "most serious" injury cannot be chosen by the lay coder, the rules provide for allocation to an extremity or multiple group.

There was a one week training period (these were all part time personnel). Once accuracy was near 100%, coders began coding cases at a rate of 12/hr. Quality checks on 15 cases for each coder were done each week. Only coders who maintained near 100% accuracy were retained to complete the work.

In contrast to BLS, who are able to follow-up to sample participants for clarification, we did not have follow-back authorization. This lack clearly impacted the quality of the data we had to work with. Using the occupation text description, 38.7% of occupations were nonclassifiable (code 9999)(27.1% after additional data sources used to identify truck drivers). In the case of nature of injury, the category intended to describe the morbidity, only 2.33% were nonclassifiable but 47.92% could not be more accurately described than "traumatic injuries and disorders, unspecified" (code 00). The latter provides little useful analytic information. The percent nonclassifiable in the remaining injury characteristic categories were: part of body affected (22.67%), source of injury (10.99%) and injury event (8.50%). The last two categories do provide potentially useful information.

Using only the occupation descriptive text and using alternative terminology or spelling variants to identify truck drivers, we identified 3,981 truck drivers. Using the injury text narrative to supplement ascertainment (e.g., "flipped truck"), an additional 523 truck drivers were identified. Other large occupation groups identified using only the occupation text included 5316 laborers (codes 864-889) and 1827 clerical (various

codes).

4. Obtaining occupation and truck firm characteristics through linkage to federal and state files.

To ascertain additional truck drivers in the study cohort, we used files from three additional classification systems. The first system was the NCCI system used originally to identify cases for the data extract. National Council of Compensation Insurers (NCCI) codes 7228-7232 were considered trucking firms for purposes of classification. The second file contained Ohio residents with Commercial Driver's Licenses (CDL) that were current through March 27, 2002. Federal law since 1986 limits commercial drivers to one driver's license issued in their state of legal residence.[15] The license is required for driving vehicles that have a combined weight of 26,000 pounds, carry in excess of a specified number of passengers for hire or carry hazardous materials.[16] The third file was the Motor Carrier Management Information System (MCMIS) maintained by the Office of Motor Carriers in the federal Department of Transportation[17] which contains firm information for for-hire truck firms that engage in interstate transportation of goods. The CDL and MCMIS files were linked to study case files by means of alphameric matching algorithms developed for the project. Data comparable to that provided by MCMIS are not available for intrastate carriers.

44.99% of cases were employed in firms in NCCI classes 7228-7232. An additional 46.88% of the study group were employed in firms with an NCCI code of 7219 (Truck NOC all) or 7232 (Driver NOC). 35.66% of study cases had Ohio CDLs. 57.30% of study cases were matched to the MCMIS file. By requiring that a study case satisfy three criteria- CDL + driver (from FROI)+ employed by a truck firm identified in either NCCI or MCMIS- or two criteria- driver + truck firm- we identified an additional 1,565 truck drivers. We confirmed the validity of this identification by calculating the fraction of drivers already identified as truck drivers with these same attributes in the firms in which this group of 1,565 were employed. The percent ranged from 79% (two criteria) to 97.03% (three criteria). In addition, approximately two-thirds of study cases that had a CDL and worked for a truck firm were estimated to be truck drivers on the basis of other study cases in the same firms, but we were unable to identify them individually.

Based on the foregoing, we concluded that use of the occupation text together with the injury narrative should be used to classify individuals as definite truck drivers but that additional truck driver categories could be created whose members appeared to have a lower positive predictive value for actually being truck drivers.

From the MCMIS file we linked the following aggregate data elements to individual claims: number of drivers per firm, the number of trucks and tractor units per firm, the number of miles driven by all types of firm trucks, the total number of recordable accidents, the total number of preventable recordable accidents as determined by periodic federal inspection and the time from the most recent inspection of the firm to the date of injury of firm employees (this last to evaluate the currency of firm data).

5. Geocoding worker address to assign race by Census block characteristics

Geocoding was performed under subcontract by the University's Institute for Survey Research (ISR), an organization recognized for its expertise in this area. Personal identifiers other than address were removed from the file before it was sent to ISR. Addresses were linked to census sampling units, either block (the smallest) or block group (the next smallest grouping). After linkage was complete, ISR appended the racial

distribution from Census 2000 to the file and returned the linked data.

The results of the geocoding were quite good. 91.8% of addresses were matched to the Census block by exact address, 1.3% were matched to centroid of the block face (zip+2). Thus, in 93.1% of study cases we obtained racial composition by block, the smallest unit for which it is available. In 6.9% of cases, we matched to the centroid of the 5-digit zipcode.

In 77.47% of cases matched by 5-digit zipcode, the cause was use of a Post Office box to receive mail. In 527 cases, the Census indicated no population in the Census unit. In 375 cases, the match was to exact street address, in 18 cases to the block face centroid, and in 134 cases the match was to 5 digit zipcode. In this last group of 134 cases, 106 gave a PO box number as the mail receipt address. This represented 8.24% of the cases matched on zipcode. Since Post Offices providing PO Box service may be in areas without any resident population, this may explain the 106 cases. Seventeen cases gave a RR address. These areas may be sparsely populated and the study cases may have left the area by the time of the Census. However, it is not clear how the remaining 404 cases could have had addresses in Census units that had no recorded population in the 2000 Census.

The 2000 Census, for the first time, gave respondents the option of indicating all racial groups to which they belonged. In the present study, the percent non-white refers to the fraction of respondents who indicated “black” in any of the possible combinations or singly.

Results:

The title of this section is actually a misnomer. The work we undertook had, as a precondition, the completion of a substantial amount of infrastructure development. The massive structure of the OBWC relational database and the nature of the data elements provided required a parallel massive effort at construction of a suitable infrastructure. However, the results of infrastructure development really represent methodologic achievements and are most appropriately considered with a discussion of the methodologic issues. That has been done.

By the close of the grant, we had either been given or had created a number of covariates suitable for generalized linear modeling using logistic, Cox proportional hazards analysis or Poisson regression. The variables that we were given that we could use unedited were really quite few- age, gender, marital status, and an injury date that could be transformed directly to create annual or seasonal cohorts. We created a classification system for individual medical diagnoses that classified the diagnosis on three dimensions simultaneously- functional area affected, severity and, if there was more than one diagnosis in the claim, on their anatomic proximity to each other as a comorbidity measure. We also created an occupation variable and cross-checked our assignment against three other databases to increase sensitivity of ascertainment for truck drivers, added a proxy for race by identifying aggregate Census data at the block level and, at least for the prototype comparison of intervertebral disc injury to back sprains, preliminarily linked diagnostic group to the use of diagnostic radiologic procedures. There were still a few tasks to complete, but we could begin some intense substantive analysis. Denial of a second no-cost extension, with adequate funds remaining to complete the analysis, put a brake on the work. We still hope to undertake some of the formal modeling but financial constraints will dramatically slow the pace. The denial

was unfathomable from this Principal Investigator's perspective.

With the work completed to date, some important and useful conclusions have been reached. First, the classification scheme by functional area captures a very high percent of the medical information in the claim files. The following table is adapted from a paper submitted for publication that compares our results to the BLS reports:

Percent Distribution of OBWC Study Claims by Number of Diagnoses, Employees of Ohio For-hire Carriers, 1997-1999

Number of diagnoses per claim	Percent of diagnoses	Percent of claims
One	48.29	70.23
Two	22.02	16.63
More than two	23.31	9.42
All diagnoses in claim unclassified	5.92	3.61[1]
Diagnoses indicating a complication of earlier injury	0.48	0.1
Total[2]	100.01	99.99

1. In an additional 3.48% of cases, at least one diagnosis is unclassified.
2. Total due to rounding.

In a little over 90% of study cases all medical information, representing almost 85% of all diagnoses, is assigned to one, or more, functional areas.

These results have important implications for our understanding of work injuries. In contrast to the current BLS practice in coding Annual Survey data of having lay coders preferentially assign each case to the "most serious" injury, the data indicate that work injuries are actually more complex when viewed from a medical perspective that used diagnoses accepted for payment. Second, the fact that the medical information in more than 90% of cases can be assigned to one, or more, discrete functional areas means that modeling of injury outcome, health services utilization and days away from work can be done with the assurance that all medical consequences of the injury event have been identified. Severity classification and identification of comorbidities allow these crucial variables to be added to the model.

Since we did not attempt to categorize soft tissue injuries of the thorax or abdomen, accidents causing these types of injuries would be incompletely characterized. While it might appear at first thought that this group would contain a large number of vehicle accidents as the injury event, that hypothesis turns out not to be true. While vehicular accidents as the injury event account for 12.42% of all study cases, only a little over 10% of injury events in this category (315/2981) produce "other injuries" diagnoses. Cases with "other" diagnoses are scattered among approximately 110 injury event

categories, with no single event category having more than 116 cases (overexertion in lifting).

At least one diagnosis in the back, shoulder or knee area occurs in 44.22% of study cases. Cases are counted in each area indicated by a claim diagnosis.

A detailed analysis of the severity of injuries and concurrent comorbidities is currently underway and, to date, reinforces the conclusion that work injuries are more complex than indicated in the BLS Annual Survey reports. As an example, the next table shows the cross-classification of the first and second most severe diagnoses recorded for each claim for back injuries, using the level 3 groupings (most highly aggregated). 72.15% of back injury cases have a single diagnosis. 8.44% of all back injuries result in some type of diagnosed neurologic injury. The vast majority of the nerve injuries are produced by displaced intervertebral discs ("slipped discs") which account for almost 86% of cases in this category.

It is important to note that 11.49% of injuries in Level 3 severity group one have serious comorbidities in group 2 as well, conditions known to complicate recovery. Third, many individuals in group 3, overwhelmingly composed of back sprains, have multiple sprain areas (20.53%). Analysis of comorbidities in other functional areas reveals that about one in eight back injuries also have an injury in a functional area proximal to the back- the head, shoulder and hip and one in eleven have injuries in more distal functional areas. The comorbid conditions themselves may be severe. Approximately 10% of Level 3 severity group one back injuries have a diagnosed arthropathy, dislocation or enthesopathy of the shoulder.

Table. Cross-classification of all back diagnoses per claim, first and second most severe diagnosis, Ohio For-hire Carrier Employees, 1997-1999

	Level 3 Severtity Grouping 2nd Severity Group						
Level 3 Severity Grouping 1st Severity Group	Single diagnosis	1	2	3	4	5	Total
0. Catastrophic neurologic injury		1					1
1. Disc displacement, neuralgia/neuritis/radiculitis, vertebral fracture with myelopathy, myelopathy without spinal bone injury	70	115	64	296	9	3	557
2. Spinal stenosis, spondylolisthesis, spondylitis, spinal curvatures, spondylosis, chondropathy, disc degeneration	10		14	55			79
3. Sprains, vertebral fractures without spinal injury	4318			1148	113	12	5591
4.Other back injuries- backache, wounds, contusions, crush injuries, etc.	357				4		361
5.Findings of uncertain significance	4					3	7
Total	4759	116	78	1499	126	18	6596

We are currently working on a paper reporting the results of severity grouping and comorbidity identification for back, shoulder and knee injuries. The precise contribution of occupation to areas in particular functional areas is underway but results are too early to report.

The algorithm to link medical diagnosis to health care utilization has been partially implemented for a prototype comparison of intervertebral disc injury and back sprain. When completed, the algorithm is intended to permit evaluation of the sequence of events leading from the accident through definitive diagnosis. Results indicate that, of the 477 study cases with a diagnosis of disc injury, 434 underwent at least one diagnostic radiologic study at some point after the injury. This same group had an average of 2.7 such diagnostic radiologic procedures through follow-up to March, 2002 and that less than one in three, on average, had a diagnostic radiologic procedure to rule out involvement of a second back region. While the disc injury cases represented 2.03% of all cases, they accounted for 21.51% of all monies paid for health services utilization at an average cost of $23,541 per case. In contrast, 11.44% of cases with a diagnosis of back sprain had a similar diagnostic radiologic procedure that, presumably, ruled out a disc injury or other cause. Together, these two diagnostic groups accounted for just over 25% of claims but accounted for some 43.91% of total medical costs.

Conclusions:

Even though we did not complete the proposed work during the grant period, a number of important conclusions are possible based on the work that is done.

First, while it is unclear in retrospect whether the final IRG comment that the grant would lead to a reduction in work injuries "if successfully completed" was expressing skepticism or was remarkably prescient on the workload being undertaken, it is now clear at the end of the grant that the goal is achievable. It is clearly possible to convert a statewide administrative workers' compensation database into analytic files capable of yielding important information on the association between occupation and specific injuries, the effect, if any, of occupation on health services utilization and the risk factors associated with prolonged disability. It just couldn't be done in three years and three months of work. It's not really a question of more money or personnel. We had enough money to support the second one-year no-cost extension that was denied. Nor is it a question of having thrown too few personnel at the task. It turns out that rather few people can do this kind of work. What was needed was simply more time and we ran out.

The ability to compare health service use by occupation is especially topical because, since the grant was funded, two surveys have reported that long haul truck drivers have significant difficulty arranging health care.[5,6] In a survey with 521 completed questionnaires, Solomon reported that nearly half of long haul truck drivers did not have a regular physician or provider and that in almost 40% of this group the reason was difficulty making an appointment/an irregular schedule or that care was not convenient. 26% of the surveyed group reported a back problems diagnosed by a professional though there was no cross-classification by whether or not a regular health professional was used. Reed reported that almost 25% of 284 female truck drivers had no usual place of medical care. While lack of insurance and lack of paid sick leave confound the question of schedule regularity, some 40% were unsatisfied with care on the road. The ability to link medical diagnosis to health care use by occupation should

provide some help to clarify just how occupation affects medical use.

The 3-dimensional injury classification scheme created for this project has already yielded some important results. In a paper submitted for publication we document that the current BLS Annual Survey coding practice underestimates the involvement of several large joint areas and the severity of injuries as measured by medical diagnoses. A second paper is near completion that illustrates that comorbidities (either concurrent injuries or conditions made worse by the occupational injury event) are common in work injuries. So far as we can determine, Nordin's study[18] is the only study of the effect of concurrent comorbidities with work injuries. Nordin shows that documented comorbidities (differs from our definition in that the diagnoses were already in the medical record at the time of injury) and any additional injuries at the time of injury predict a delayed return to work. Both Laberge-Nadeau[3] and Dionne[4] report that preexisting morbidities predict more serious injuries. Clearly, this is a potentially useful covariate for modeling, one that can not be obtained from BLS.

The preliminary data available for health services utilization for disc injuries suggest that even modestly successful primary and tertiary prevention programs would be extraordinarily cost-effective. One doesn't have to prevent very many disc injuries at the average medical cost observed (quite aside from the cost of wage compensation) to more than pay for the prevention effort. Moreover, the OBWC database will permit identification of "hot spot" occupations or industries where the prevention efforts are most needed and, if effective, most likely to produce important cost reductions.

Publications:

1. Oleinick A, Gandra CR, Simon C. A Comparison of the BLS Injury Characteristic and ICD-9-CM Clinical Diagnostic Taxonomies for Estimating the Severity of Work Injuries in Employed Populations, Ohio For-hire Carrier Employees, 1997-1999. Submitted for publication.

Inclusion of gender and minority study subjects: The study was based on secondary data analysis on data provided by the OBWC. Therefore, there was no way to plan for gender and minority study subject inclusion. The study group was 23.89% female (1.9% unspecified), aggregated 2000 Census data was used as a proxy for race, and 1.3% were minors below the age of 18.

References:

[1] Bureau of Labor Statistics (BLS). 2004,2005. Lost-Worktime Injuries and Illnesses: Characteristics and Resulting Days Away from Work, 2003,2003. Washington, DC: U.S. Department of Labor. URL: http:/www.bls.gov/iif/oshcdnew.htm/osnr.0019,0021.

[2] Bureau of Labor Statistics (BLS). 2003. Lost-worktime Injuries and Illnesses: Characteristics and Resulting Days Away from Work, 2001. Washington, DC: U.S. Department of Labor. URL: http://www.bls.gov/iif/oshcdnew.htm/osnr0017.pdf

[3] Laberge-Nadeau C, Dionne G, Maag U, Desjardins D, Vanasse C, Ekoe JM. 1996. Medical Conditions and the Severity of Commercial Motor Vehicle Drivers' Road Accidents. Accident Analysis & Prevention. 28:43-51.

[4] Dionne G, Desjardins D, Laberge-Nadeau C, Maag U. 1995. Medical Conditions, Risk Exposure, and Truck Drivers' Accidents: an Analysis with Count Data Regression Models. Accident Analysis & Prevention. 27:295-305.

[5] Solomon AJ, Doucette JT, Garland E, McGinn T. 2004. Healthcare and the Long Haul: Long Distance Truck Drivers- A Medically Underserved Population. Am J Ind Med 46:463-471.

[6] Reed DB, Cronin JS. 2003. Health on the Road: Issues Faced by Female Truck Drivers. AAOHN Journal. 51:120-125.

[7] Ohio Bureau of Workers' Compensation (OBWC).BWC Library. 2005. Facts and figures: 2002-2004 Fact Sheet. URL: http://www.ohiobwc.com/home/current/FactFigLnks.asp

[8] U.S. Census Bureau. 2005. All Data Sets. Census 2000 Summary File 3 (SF-3)- Sample Data. Table P49.URL: http://factfinder.census.gov/servlet/DatasetMainPageServlet?_lang=en

[9] National Council on Compensation Insurance. 2005. Classification Codes and Statistical Codes for Workers' Compensation and Employers Liability Insurance Manual. Boca Raton, FL: National Council on Compensation Insurance.

[10] National Center for Health Statistics. 2004. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). Hyattsville, MD: National Center for Health Statistics. URL: http://www.cdc.gov/nchs/about/otheract/icd9/abticd9.htm

[11] Bureau of Labor Statistics (BLS). 2002. Introduction to Occupational Injuries and Illnesses Case Coding Training, San Francisco, California. Washington, DC: U.S. Department of Labor.

[12] Bureau of Labor Statistics (BLS). 1992. Occupational Injury and Illness Classification Manual. Washington DC: U.S. Department of Labor

[13] Bureau of Labor Statistics (BLS). 2000. Occupational Injury and Illness Coding Interpretations Manual. Washington, DC: U.S. Department of Labor

[14] Bureau of Labor Statistics (BLS). 2000. Occupational Injury and Illness Occupational Coding Manual. Washington, DC: U.S. Department of Labor

[15] United States Code. 2000.49 U.S.C. 31302. Washington, DC: Gov. Printing Office. URL: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=browse_usc&docid=Cite:+49USC31302.

[16] United States Code. 2000.49 U.S.C. 31301. Washington, DC: Gov. Printing Office. URL: http://frwebgate.access.gpo.gov/cgi-

bin/getdoc.cgi?dbname=browse_usc&docid=Cite:+49USC31301.

[17] Federal Motor Carrier Safety Administration. 1996. Data Dissemination Program Catalog. FHWA-MC-96-032. Washington, DC: Federal Highway Administration URL: http://www.fmcsa.dot.gov/facts-research/facts-figures/analysis-statistics/mcmis/mcmiscatalog.htm

[18] Nordin M, Hiebert R, Pietrek M, Alexander M, Crane, M, Lewis, S. 2002. Association of Comorbidity and Outcome in Episodes of Nonspecific Low Back Pain in Occupational Populations. J Occ Env Med. 44:677-684.