Group Medical Claims as a Source of Information on Worker Health and Potentially Work-Related Diseases

P. Timothy Bushnell, PhD, MPA, Jia Li, MS, and Deborah Landen, MD, PhD

Objective: To help address underrecognition of occupational illnesses and support planning of workplace health initiatives. **Methods:** Data from Highmark Inc., a health care insurer headquartered in Pittsburgh and Camp Hill, Pennsylvania, was used to calculate age and gender-adjusted rates of 15 diseases by industry and broad industry sector. **Results:** Significant industry differences in disease rates were observed, some corresponding to known differences in workplace risk factors. **Conclusion:** Group medical claims offer timely, relatively low cost, longitudinal data on rates of physician-diagnosed cases and costs of many diseases for large populations. Limitations of medical claims data include inaccuracies in industry coding, lack of occupation codes, and lack of key variables that affect health. Yet, some elevated industry rates suggest potential new targets for wellness programs and evaluation of possible workplace health risks.

This article describes a project that uses health insurance claims data to identify patterns of disease in working populations. The aims are to help target disease prevention programs and research that may encompass both occupational and nonoccupational risk factors. The National Institute for Occupational Safety and Health (NIOSH) and Highmark Inc. of Pittsburgh have worked together to assemble the data, and have developed written agreements that govern data protection and reporting of results. We report highlights of findings on the industry rates of 15 conditions, including respiratory, musculoskeletal, and circulatory conditions, depression, and several other types of conditions.

The primary goal of this study is to help address the well-known underrecognition and underreporting of occupational illness cases, which we review in brief later. We do this by identifying industries with increased rates of disease, which may be due, at least in part, to workplace exposures. Another broader goal is to understand and address the health challenges of employees in specific industries, whether or not they are related to workplace factors. This goal is aligned with NIOSH's Total Worker Health program,[1] which is premised on the idea that nonoccupational and occupational health risks of employees are best addressed together. A number of conditions are significantly influenced by both occupational and nonoccupational risk factors. Examples include musculoskeletal disorders that may arise in part from obesity and in part from cumulative trauma from repetitive work tasks, and respiratory conditions that may be exacerbated by smoking as well as workplace respiratory exposures.

Another goal of the project is to enhance the information that health care insurers may provide to employers on the health of their employees. This goal is not the focus of this article. Nevertheless, the concept is that employers of sufficient size may benefit by knowing rates of specific diseases among their employees, and having the opportunity to interpret these rates in the context of regional rates and industry-level variation in rates.

The Problem of Underrecognition and Underreporting of Work-Related Disease

The underrecognition and reporting of occupational disease is partly a result of the same obstacles that lead to very substantial underreporting of occupational injuries, for which there is considerable evidence.[2–8] In comparison with injuries, however, there is much more limited ascertainment and tracking of occupational illness cases in the United States, especially for the many important diseases that also commonly have nonoccupational causes and long latency periods.[9,10] A fundamental reason is that work relatedness is often difficult to determine in individual cases, both for employees and for physicians.[3,11] This basic problem is compounded by the fact that physicians usually receive little or no training in occupational health,[11] and have little time to determine work relatedness. We cite just three important examples.

First, musculoskeletal disorders, such as back pain and carpal tunnel syndrome (CTS), are often classified as illnesses, and make up a large portion of the occupational injuries and illnesses reported by the Bureau of Labor Statistics (BLS) (28.8% of days away from work cases in 2007).[12] Nevertheless, a broad spectrum of research and testimony assembled by the Occupational Safety and Health Administration led to the conclusion that a substantial percentage of cases are not recorded by the annual BLS survey of occupational injuries and illnesses which is based on employer logs of work-related injuries and illnesses required under the Occupational Safety and Health Administration record-keeping standard.[13,14] The often chronic and recurrent episodic nature of occupational back pain not only makes it difficult to trace symptoms to a single event at work but also leads to a significant share of the medical care of recognized and unrecognized occupational cases being paid by medical rather than workers' compensation insurance.[15] The possible extent of underreporting in the workers' compensation system was suggested in a Connecticut population survey in which only 11% of self-reported cases of work-related cumulative trauma musculoskeletal disorders of the hand, arm, and neck had filed a workers' compensation claim.[16]

A second example of underrecognition and reporting is asthma. During 2003 to 2008, BLS reported an annual average of 400 days-away-from-work cases of extrinsic asthma (allergic asthma) in private industry.[17]† Yet, the estimated total number of prevalent cases of occupational asthma in the United States was 2.8 million in 2007, based on an estimate of the total current adult asthma cases in 2007 (18.6 million),[18] and the proportion that may be due to occupational causes (15%).[19] A third example is circulatory system diseases for which the BLS survey of occupational injuries and illnesses reported an average of 702 cases per year of nonfatal, circulatory system

From the National Institute for Occupational Safety and Health (Dr Bushnell and Ms Li), Division of Surveillance, Hazard Evaluations and Field Studies, Cincinnati, OH; and National Institute for Occupational Safety and Health (Dr Landen), Office of Mine Safety and Health Research, Pittsburgh, PA.

The study was funded by the National Institute for Occupational Safety and Health.
The findings and conclusions in this report are those of the authors and do not necessarily represent the views of NIOSH.

Address correspondence to: P. Timothy Bushnell, PhD, MPA, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, 4676 Columbia Parkway, MS R17, Cincinnati, OH 45226 (PLB4@cdc.gov).

DOI: 10.1097/JOM.0b013e3182363bbe

†There are also special programs for reporting occupational asthma in four states (CA, MA, MI, NJ). An average of 413 cases per year were reported in these states from 1993 to 2002. (NIOSH. U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention. Work-related Lung Disease Report, electronic version (eWoRLD) Table 9-3. http://www2a.cdc.gov/drds/WorldReportData/pdf/2007T09-03.pdf. Accessed December 2009)

diseases of all kinds during 2003 to 2008. This contrasts with an estimated 4500 to 12,900 coronary heart disease deaths attributable to occupation in 1997.[20]

Advantages and Disadvantages of Medical Claims Data

Disease rate information on the basis of medical claims may offer some helpful clues to the presence and possible magnitude of some occupational illnesses. Prospective advantages and disadvantages of group medical claims data for this purpose include the following.

Advantages

1. Accurate denominators for calculating disease rates, based on insurance enrollment.
2. Objective information on physician diagnoses and treatments.
3. Information on drug prescriptions, physician specialty, and type of facility can also be used for case definitions and severity measures. Commonly, however, drug claims do not include diagnosis codes, so prescriptions must be linked to medical diagnoses through medical knowledge and the timing of prescriptions with respect to other treatments.
4. Costs paid by both insurer and insured for treatments and prescriptions are available.
5. Databases of sufficient size to provide samples for relatively specific industries, and that contain enough employers to protect the confidentiality of individual employer data.
6. Potentially less expensive to collect than employee survey data, and without response rate concerns (although claims data from more than one plan may be needed).
7. Data are continuously updated and potentially available with a short time lag.
8. Data are longitudinal for those workers who stay insured with the same insurer.
9. Employer codes allow insurers to produce reports for individual employers, and improve the analysis of industry disease rates.

Disadvantages

1. Work-relatedness of occupational conditions usually not indicated in claims records.
2. No information on occupation.
3. Diagnosis codes may not always be accurate because of diagnosis uncertainty, time pressures, transcription errors, lack of full knowledge of available codes by health care providers and coders, and choice of codes to increase payment for treatments or to respond to patients' wishes.[21] Nevertheless, many of these inaccuracies would tend to vary by individual patient and provider and make less difference to the comparison of disease rates of large, industry groups.
4. Industry rate differences may also reflect differences in the tendency to seek and receive medical care, perhaps due in turn, to differences in insurance plans, or to cultural and economic differences.
5. Data include only employed persons who obtain insurance through their own employers. Those who obtain insurance through their spouses may work for employers who offer less generous health benefits and may also tend to have different working conditions and health characteristics.

This last disadvantage points to a basic limitation to the potential for medical claims data to characterize working populations. On the basis of Current Population Survey data, in 2008, 53.3% of US workers aged 18 to 64 had health insurance in their own name and through their own employer.[22] Medical claims data also provide more information on some groups of workers than others. For example, 66% of full-time, full-year workers had coverage through their employer, versus 22% of part-time, full-year workers. The percentage also increases with employer size (eg, 37% in employers with 10 to 24 employees versus 64% in employers with 1000 or more employees) and varies by industry (eg, 39% of workers in agriculture, forestry, fishing, mining, and construction, versus 67% in manufacturing and 73% in the public sector).

Use of Medical Claims Data for Occupational Health Surveillance: Other Examples

To our knowledge, this is the first effort in the United States to use a large health care insurer database to examine variation in rates of specific diseases by industry. Some large employers use employee medical claims to look for patterns that could suggest a role for workplace factors that can be controlled. Nevertheless, we have identified few published reports. One example is a study using claims data for a large US auto manufacturer, along with detailed information on work histories, to identify specific plant departments with statistically significant elevations of rate of five types of musculoskeletal disorder. They were able to estimate the percentage of cases attributable to occupational exposures, which varied between 19% and 47%.[23,24] Another example is the Duke University Health System, which has used medical claims to identify occupational groups at higher risk for preterm delivery, musculoskeletal disorders, and health consequences of work-related stress.[25–27]

An additional example of use of medical claims data to address occupational health is the analysis of health and welfare fund medical claims of union carpenters in Washington State. The researchers observed that individuals who filed musculoskeletal workers' compensation claims also tended to have higher group medical claim costs for musculoskeletal disorders after their injury, and that an increasing share of occupational cases resulted in medical rather than workers' compensation claims.[28,29]

In British Columbia in Canada, more extensive use has been made of large medical claims databases for occupational health analysis.[30] Health care and workers' compensation records have been linked for the entire working age population, supporting studies of the effects of specific work exposures in subpopulations of interest.

DATA AND ANALYSIS

The Multi-Year Highmark Database

The National Institute for Occupational Safety and Health worked with Highmark to assemble a multi-year database for the purpose of calculating industry-specific, disease prevalence rates. A written agreement governs the protection of data and reporting of results. All data received were "limited data sets" as defined under the Health Insurance Portability and Accountability Act.[31,32] No personal identifiers were included in the data provided to NIOSH, although unique "dummy identifiers" were used to link claim and enrollment records of individuals. Unique "dummy" employer numbers were also included in the database so that employer rates could be computed and shared by the insurers with interested employers. A file of dummy identifiers was kept by the insurer to link to individuals and employers but was not shared with NIOSH. Therefore, no employees or employers were identifiable by NIOSH through the data sets provided. Members of the NIOSH Human Subjects Review Board determined that because of the nature of the data used in the project and its classification as public health practice, the project was not subject to the board's review.

The database contains enrollment information and medical claims data for individuals who received or purchased health insurance through their employer during 2002 to 2005. Enrollment information included birth year, gender, months of insurance coverage, and employer industry. Claims information consisted of number and

cost of claims for each disease in each year, as well as total annual health care cost and share of cost paid by the employee.

The following diseases and health conditions were included: hypertension, low back pain, depression, cardiovascular disease, diabetes, asthma, carpal tunnel syndrome, dermatitis, osteoarthritis of load-bearing joints, chronic obstructive pulmonary disease, hearing loss, bladder cancer, Parkinsonism, asbestosis, and coal workers' pneumoconiosis. Each was chosen in consultation with NIOSH researchers and Highmark, with consideration given to disease prevalence and research and prevention priorities. Some diseases, such as coal workers' pneumoconiosis, were chosen because of their known links to certain types of workplaces. Others, such as back pain and asthma, were chosen because they have been linked to certain work exposures, but more knowledge is needed about where such exposures may be increasing prevalence rates. Still others, such as Parkinsonism, have been less frequently linked to the workplace, but industry variation in rates could yield new research leads.

Individuals with at least 44 of 48 possible months of enrollment during 2002 to 2005 were included in the database. With few exceptions, each employer is assigned a single, four-digit Standard Industrial Classification (SIC) code. Thus, some employees of employers with multiple locations and industries are not accurately classified. Highmark has contracted with Dunn and Bradstreet to supply and perform quality control for the SIC code field. Employers may request a change in code (which can affect premiums), but Highmark reports that less than 2% of codes are changed due to the resulting investigations of code accuracy.

Standard industrial classification was determined for each year on the basis of retrieval of employer SIC for a single month in the year. Only individuals classified in the same two-digit SIC in each of the 4 years were included in the database, thus increasing the share of long-term employees in the data. This was desirable, because several of the diseases included in the project develop only with long-term exposure to risk factors or often with a significant lag from time of exposure.

To maintain confidentiality and prevent attribution of industry disease rates to any single employer, all industry samples were required to contain at least three times as many individuals as the largest employer in the industry. Therefore, a random sample of employees of the largest employers was deleted, where necessary, so that all industry samples met the requirement. Industries with fewer than four employers were retained in the data, but rates for them are not reported.

Disease Case Definitions

Case definitions for diseases and health conditions were based on the *International Classification of Diseases, Ninth Revision,* diagnosis codes listed in Table 1. We generally chose to focus on broad categories of conditions and to err on the side of including all codes that might be used for a condition. For low back pain, which includes a variety of disease processes, but which is a symptom-based category, codes were included which indicate that (1) the patient probably had low back pain, and (2) symptoms may have resulted from cumulative trauma or strain events (eg, not fractures), and (3) do not indicate a specific source of pain without potential relationship to physical and psychosocial stressors in the workplace (eg, a tumor).

The standard case definition is two or more claims during 2002 to 2005 that contain an *International Classification of Diseases, Ninth Revision,* diagnosis code for the disease. Individuals with only one claim for a disease are more likely to have received only a rule-out diagnosis. The first-listed diagnosis code on a claim is called a primary diagnosis code and the other codes are secondary codes. There is often a presumption that the primary code represents the most important diagnosis for an encounter or treatment, but there is no assurance that this is the case. Primary and secondary diagnosis codes were considered equally valid.

Employee Share of Total Medical Claim Costs

Benefit plan details for each individual were not available. Nevertheless, for individuals who received health care, we did have information on total yearly medical claim cost and share of cost paid by the employee (not including drug prescription costs). This information was used to create a share-of-cost variable on the employer level. Although insurance plans can vary among employees of the same employer, it was assumed that their plans tend to be similar. The share of cost is affected by co-pays, coinsurance, and deductibles, which are expected to have their greatest effect on share of cost in the low to moderate range of total cost. Each employee has up to four annual measurements of percentage of cost. For each employer, we fit a generalized linear model of the percentage of cost on the logarithm transformed total cost with a logit link function.[33] Multiple measures of percentage of cost on the same member were treated as repeated measures. We then estimated the expected share of cost for each employer at the median total cost of $859 (2009 dollars, adjusted with GDP national accounts price index).[34]

Calculation of Disease Rates

Industry disease prevalence rates were estimated with logistic regression, adjusting for differences between the age and gender distribution of each industry's employees and that of the overall database population, as well as differences between industries in employee share of medical claims costs. The following model was fit for each disease.

$$\begin{aligned}\mathrm{LOGIT} &= \mathrm{LOG}\left(\frac{P}{1-P}\right)\\ &= \beta_0 + \beta_1\mathrm{SIC} + \beta_2\mathrm{GENDER} + \beta_3\mathrm{AGE} + \beta_4\mathrm{AGE}^2\\ &\quad +\beta_5\mathrm{GENDER} * \mathrm{AGE} + \beta_6\mathrm{GENDER} * \mathrm{AGE}^2\\ &\quad +\beta_7\mathrm{COSTSHARE} + \beta_8\mathrm{COSTSHARE}^2\end{aligned}$$

Interaction and quadratic terms increased model flexibility to better reflect true relationships among variables. To properly account for the hierarchical nature of our data, we used multilevel modeling, treating employer as a cluster variable. This allowed us to assume correlation among individuals from the same employer, which can arise from having similar workplace exposure, insurance plan, socioeconomic status, and so on. The generalized estimating equation approach was used for parameter and standard error estimation. Industries with samples less than 200 were too small to produce reliable estimates for low prevalence diseases and are less likely to be representative of heterogeneous industries. These were pooled in a separate, multi-industry group. When the disease outcome was depression, an additional confounder, "behavioral carve-out," was included in the model. Employees with a behavioral care "carve-out" have their behavioral claims processed by a separate organization, and thus, many of their claims with depression codes do not appear in the claims data available from Highmark.

The adjusted rates were estimated using the method described by Lee,[35] and Lane and Nelder.[36] For each SIC $= i$,$\mathrm{LOGIT}_{\mathrm{SIC}=i}$ was estimated for each of the N individuals in the data set using the regression coefficient β_1 of SIC $= i$, and $P_{\mathrm{SIC}=i} = \sum \frac{\mathrm{EXP}(\mathrm{LOGIT}_{\mathrm{SIC}=i})}{1+\mathrm{EXP}(\mathrm{LOGIT}_{\mathrm{SIC}=i})}/N$. Calculated in this way, the adjusted rate for a certain industry represents the average predicted disease rate if everyone in the database population had been in this industry. A one-sample z test was used for testing each adjusted rate against the population rate.

Because the chance of false positive findings is high, when many tests of statistical significance are performed simultaneously,

TABLE 1. Case Definition Diagnosis Codes*

Disease/Condition	*ICD-9* Codes
Hypertension	401 and all subcodes†, 402 and all subcodes, 403 and all subcodes, 404 and all subcodes, 437.2, 642.0 and all subcodes, 642.2 and all subcodes, 642.7 and all subcodes
Low back pain	721, 721.3, 721.4, 721.42, 721.8, 721.9, 721.90, 721.91, 722, 722.1, 722.10, 722.2, 722.5, 722.51, 722.52, 722.6, 722.7, 722.70, 722.73, 722.8, 722.80, 722.83, 722.9, 722.90, 722.93, 724, 724.0, 724.00, 724.02, 724.09, 724.2, 724.3, 724.4, 724.5, 724.6, 724.8, 724.9, 738.4, 739.3, 739.4, 839.2, 839.20, 839.3, 839.30, 846, 846.0, 846.1, 846.2, 846.3, 846.8, 846.9, 847, 847.2, 847.3, 847.9
Depression	296, 296.2 and all subcodes, 296.3 and all subcodes, 296.5 and all subcodes, 296.6 and all subcodes, 296.7 and all subcodes, 296.8 and all subcodes, 296.9 and all subcodes, 300.4, 301.1 and all subcodes, 309.0, 309.1, 309.28, 311 and all subcodes
Cardiovascular disease	410 and all subcodes, 411 and all subcodes, 412 and all subcodes, 413 and all subcodes, 414 and all subcodes, 428 and all subcodes, 429.2, 430 and all subcodes, 431 and all subcodes, 432 and all subcodes, 433 and all subcodes, 434 and all subcodes, 435 and all subcodes, 437.0, 437.1, 438 and all subcodes, 440 and all subcodes, 443.0, 443.9, V434
Diabetes	250 and all subcodes
Asthma	493 and all subcodes
Carpal tunnel syndrome	354 and all subcodes
Contact dermatitis	692 and all subcodes
Osteoarthritis of load-bearing joints	715.15, 715.16, 715.17, 715.95, 719.96, 719.97
Chronic obstructive pulmonary disease	491 and all subcodes, 492 and all subcodes, 496 and all subcodes
Hearing loss (noise-induced)	388.1, 388.10, 388.12, 388.3, 388.4, 389, 389.1, 389.10, 389.11, 389.9
Bladder cancer	188 and all subcodes, 233.7, 236.7, 239.4
Parkinsonism	332, 332.0, 333.0
Coal workers' pneumoconiosis	500 and all subcodes

*Number of claims with one or more of the listed codes was counted during 2002 to 2005. Claims with a case definition code appearing as a primary diagnosis code (first listed code) or as a secondary code were counted equally. Standard case definition used in analysis is two or more claims during 2002 to 2005.
†Subcodes are all codes that begin with the same digits.
ICD-9, International Classification of Diseases, Ninth Revision.

we calculated adaptive false discovery rate *P* values.[37,38] The total number of tests performed in this study, over all conditions, was 711, excluding SICs with zero cases for some conditions. The false discovery rate *P* value was used instead of a conventional Bonferroni-type adjustment, because it provides a lower upper bound on the adjusted *P* value.

Calculation of Annual Costs Per Case

The insurer-paid costs of all claims in which a case definition *International Classification of Diseases, Ninth Revision,* code appears as the primary diagnosis were summed. We did not include the costs of claims on which a case definition code appears only as a secondary code, because the costs of these claims are better attributed to the primary diagnosis. Drug costs were not included because we did not have complete drug claim data.

RESULTS

Table 2 gives the characteristics of the population in the multi-year Highmark database, and compares them to the US working population. There were 55 two-digit SICs with samples of 200 or more of a total of 82 two-digit industries in the SIC system. Of the 55 SICs, 44 had samples of 500 or more. Samples of 200 or more were available for 65 of 1004 four-digit SICs, but results for these samples are not reported here. To limit the share of individual employers to less than one third of the sample for their industry, records for 48,266 of 262,679 individuals were deleted from the database to compute rates for two-digit SIC industries. To maintain confidentiality, the impact of this on the individual industry level cannot be reported, but the deleted population as a whole was similar to the database population in the subset of industries from which employees were deleted. Deleted employees were slightly older (mean age 44.9 vs 44.5 years) and a greater proportion were women (32.7% vs 30.7%). Disease prevalence rates in the deleted population were lower by 8% or less among conditions with prevalence more than 3% and were within 16% of the database population rate for most of the other conditions.

Table 3 reports the number of cases for each disease, on the basis of two different case definition criteria: the standard definition of two or more claims and, for comparison, an alternative definition of one or more claims. Average number of claims per case is also reported. Table 4 reports prevalence rates for each condition by broad industry sector (cases defined as two or more claims). Table 5 reports the highest and lowest-rate industries on the two-digit SIC level for each disease, using the two or more claim case definition. Industries listed have rates that are at least 20% above or below the database population rate and a false discovery rate $P < 0.25$ for test of difference from the population rate. For conditions with prevalence more than 1%, Table 6 compares the cost per case for high-rate industries to that of the database population. Costs of each disease are also compared.

As described earlier, an employee share-of-cost variable on the employer level was used to adjust rates for industry differences in insurance plans. It is negatively associated with all of the higher prevalence diseases (prevalence > 1%). Cost share is most strongly associated with depression and dermatitis, where the 25th percentile share of cost is associated with a 23% higher rate than the 75th percentile share of cost. Among diseases with prevalence more than 1%, cost share is least associated with diabetes, where the 25th percentile share of cost is associated with a statistically nonsignificant 3% higher rate than the 75th percentile share of cost.

DISCUSSION

Interpretation of Disease Rates

There are some general points to be kept in mind when interpreting industry disease rate differences. First, the meaning of a

TABLE 2. Population Characteristics: Highmark Database and US Workers

	Highmark Database*		US Working Population†
Sex	%		%
Male	61.1		53.6
Female	38.9		46.4
Age, yrs			
Median	45		41
Mean	45		41
	%		**%**
18–29	8.7		23.3
30–39	22.2		23.5
40–49	34.7		26.0
50–59	29.2		19.1
60–69	4.9		6.5
70+	0.4		1.6
NORA Industry Sector‡	***N***	**%**	**%**
Agriculture, forestry, and fishing	454	0.2	1.6
Construction	7,746	3.0	7.7
Health care and social assistance	20,563	7.9	12.1
Manufacturing	62,855	24.2	12.0
Mining	3,752	1.4	0.4
Services	127,555	49.1	46.2
Transportation, warehousing, and utilities	14,545	5.6	5.2
Wholesale and retail trade	22,337	8.6	14.9
Total	259,807	100.0	100.0

*Mean values for 2002 to 2005. Database randomly excludes 2872 individuals from large employers so that all sector samples are over three times as large as largest employer in sector.

†Mean values for 2003 to 2005, (2002 excluded due to change in industry coding in 2003). (US Department of Labor, Bureau of Labor Statistics, Current Population Survey).

‡Industry sectors as defined by the NORA of the National Institute for Occupational Safety and Health (www.cdc.gov/niosh/nora/sector.html). NORA definitions of sectors are based on NAICS industry codes. Because NAICS codes were unavailable, SIC codes were used to define sectors as follows: Mining: 1011–1499, Construction: 1521–1799, 6552, Manufacturing: 2011–3999 (except 2411, 2711, 2721, 2731, 2741), 8072, Wholesale and retail trade: 5012–5999 (except 5812, 5813), Transportation, warehousing, and utilities: 4011–4971 (except 4493, 4724, 4725, 4812-4841, 4953), Services except health care and social assistance: 741, 742, 781–783, 2711, 2721, 2741, 4493, 4724, 4725, 4812–4841, 4953, 5812, 5813, 6011–7999 (except 6552), 8099–8322, 8399-9721, and Health care and social assistance: 8011–8093 (except 8072), 8331–8361.

SIC, Standard Industrial Classification; NORA, National Occupational Research Agenda.

high industry rate depends upon how different the rate is from the comparison population rate in both absolute and relative terms. For example, population rates of hypertension and low back pain are near 27%. Therefore, a modest elevation of an industry rate to 32% may not suggests that the overall level of risk factors in that industry are greatly elevated. Nevertheless, one additional employee out of 20 has the condition. By contrast, an industry rate that is double the population rate for a lower-prevalence condition such as chronic obstructive pulmonary disease (2.3%) would much more strongly suggest relatively high-risk factors in that industry but represents a lower estimate of one additional case for every 43 employees.

Second, it may be more appropriate to compare all industries to the industries with the lowest rates, because these industries may better represent a low-risk environment. The drawback to that approach is that some industries may have very low rates because of personal rather than workplace risk factors.

Third, for most diseases, costs per case appear to have little relationship to disease rate, suggesting that relative disease rates are fairly good indicators of relative disease costs. Yet, there were some diseases for which costs per case were somewhat higher for high-rate industries. This might result from risk factors that increase both the rate and average severity of cases, or a greater tendency to seek medical care that increases both diagnosis rates and intensity of care.

Examples of Specific Results of Interest

Several of the results support well-known associations between disease and industry (eg, coal mining and coal worker's pneumoconiosis, chronic obstructive pulmonary disease, and hearing loss; construction and musculoskeletal disorders), but others may help focus attention on underrecognized workplace health issues.

Industry differences in disease rates may be attributed to differences in behavioral risk factors (independent of the work environment or possibly influenced by it), to differences in work-related risk factors, to other environmental factors, or to some combination of these. An example is cardiovascular disease, for which the potential work-related risk factors include exposure to certain chemicals (eg, carbon disulfide), particulates, workplace stress, shiftwork, and physical inactivity. Few previous studies have compared cardiovascular disease rates among industries, and so several of the findings for this condition contribute to guidance for further examination of the possible need for targeted workplace wellness efforts and the potential role of work-related risks. For example, we could find few or no precedents for the high rates found in retailing (general

TABLE 3. Number of Disease Cases in Database Population*

Case Definition†	No. of Cases		Average Claims Per Case
	1 or More Claims	2 or More Claims	2 or More Claims
Hypertension	88,284	71,558	8.5
Low back pain	90,845	70,466	15.2
Depression	36,846	26,592	14.0
Cardiovascular disease	32,988	20,752	11.9
Diabetes	27,368	20,325	16.3
Asthma	19,790	12,107	7.3
Carpal tunnel syndrome	18,811	11,523	5.3
Dermatitis	31,263	10,689	3.0
Osteoarthritis of load-bearing joints	12,221	6,638	7.3
Chronic obstructive pulmonary disease	11,686	5,875	7.0
Hearing loss	7,051	2,104	2.8
Bladder cancer	760	467	17.9
Parkinsonism	311	203	12.5
Asbestosis	166	70	3.9
Coal workers' pneumoconiosis	73	20	3.1

*Database randomly excludes some individuals from large employers so that all NORA industry sector samples are over three times as large as largest employer in sector.

†Cases are defined as one or more claims in 2002 to 2005, or two or more claims in 2002 to 2005.

merchandise stores and miscellaneous retail), personal services,[39] oil and gas extraction,[40] general building contractors[41,42] and auto repair, services and parking.[43] The highest rate industry, local and interurban passenger transit, contains bus drivers which have frequently been observed to have elevated rates of heart disease, hypertension, or stroke, often attributed in part to work stress.[44,45] Elevated rates in the textile mill products industry have also been reported several times previously.[42,46–48]

Industry disease rate results may help point to industries that often are not well recognized as including work activities that may have health risks. For example, the high rate of low back pain observed for real estate contrasts with rates below or near the private industry average for work-related back injuries and musculoskeletal disorders that have been reported for real estate by the Bureau of Labor Statistics (based on employer reports) and observed in workers' compensation data.[49–51] Part of the real estate industry group, however, is devoted to building operations, including cleaning and maintenance occupations,[52] a possible source of stressors to the back. Some subclassifications of the real estate sector have sometimes been reported to have elevated rates of back injuries or musculoskeletal disorders, though not consistently.[49,50] It is notable that a related industry classification, services to buildings and dwellings, which is outside of real estate, had a nearly 50% greater rate of back injuries than private industry as a whole in national BLS data,[49] about twice the average rate of back pain among female workers in a national survey,[53] and over twice the average rate of workers' compensation claims for back musculoskeletal disorders in Washington State.[51]

Another somewhat novel finding is the elevated rate of CTS observed in mining, and in coal mining in particular, which contrasts with BLS data showing relatively low rates in mining on the basis of employer logs,[54] and which is not reflected in other studies of CTS by industry.[55–57] Nevertheless, one study did find an elevated rate of CTS in a study of miners with compensation claims for hand-arm vibration syndrome,[58] and there is additional evidence that hand-arm vibration is associated with CTS and is found in mining.[58–61] Suggestively, Dillon et al[61] also report that mining had the highest industry odds ratio for handwrist arthritis, which was also observed to be associated with repetitive handbending and twisting, another recognized risk factor for CTS,[59] which has also been observed in mining.[63]

An elevated rate of bladder cancer was observed among workers in the fabricated metal products industry. With the exception of aluminum, metalworking is not one of the industrial activities known to be most strongly associated with bladder cancer, but there are several studies that have found increased risks among fabricators, machinists, and metalworkers.[64]

Several industries have among the highest rates of several conditions. Most saliently, local and interurban passenger transit had the highest rate of hypertension, depression, cardiovascular disease, and diabetes. Some of these results are in line with previous research,[44] but they suggest that this industry may need special focus on wellness and health improvement efforts. Although coal mining is well known for several health hazards, social services, personal services, and auto repair, services and parking also had high rates of multiple conditions.

Limitations

One limitation of the data is classification of industry on the basis of a single business address for each employer. This means that some workers in diversified employers are not classified correctly. The employees of only a few large employers were classified to more than one SIC based on specific location or organizational unit. The four-digit SIC codes of a sample of 65 employers were manually checked by Highmark for the year 2008. The five largest employers in seven industries with high rates of more than one disease were included, as well as an additional random sample of 30 employers (selection probability proportional to number of employees in the data). There was a clear match between known employer activities and four-digit code in most cases. There were two or three exceptions in the public sector, where activities are often quite varied, and it appeared that the assigned code might not be the most appropriate. In addition, one group of secondary education employees was clearly misclassified as employed by a business association that sponsored their insurance plan.

Classification of cases by diagnosis code was also imperfect. Treatments and drug prescriptions were not used to make more

TABLE 4. Adjusted Prevalence of Diseases by NORA Industry Sector, 2002–2005 (%)

Disease	Mining	Construction	Manufacturing	Wholesale and Retail Trade	Transportation, Warehousing and Utilities	Services*	Health care and Social Assistance
Hypertension	31.6 (29.8, 33.4)	26.5 (25.1, 28.0)	26.5 (25.1, 27.9)	29.3 (28.2, 30.4)	30.4 (27.4, 33.5)	27.1 (25.9, 28.3)	29.0 (27.7, 30.3)
Low back pain	25.3 (21.7, 29.0)	30.8 (28.1, 33.5)	27.2 (25.6, 28.7)	28.3 (27.0, 29.5)	26.0 (24.5, 27.5)	27.1 (25.1, 29.1)	25.6 (23.7, 27.5)
Depression	8.4 (6.8, 10.1)	10.6 (9.2, 11.9)	8.9 (7.9, 9.9)	10.2 (9.5, 10.9)	10.2 (8.8, 11.6)	10.6 (9.5, 11.6)	11.5 (10.7, 12.3)
Cardiovascular disease	8.5 (7.7, 9.2)	8.4 (7.6, 9.2)	8.0 (7.6, 8.4)	9.0 (8.4, 9.6)	8.8 (7.7, 10.0)	7.6 (7.3, 8.0)	8.3 (7.8, 8.8)
Diabetes	7.8 (7.0, 8.5)	6.6 (5.8, 7.5)	7.6 (7.2, 8.0)	8.3 (7.7, 8.9)	9.0 (7.0, 11.0)	7.7 (7.1, 8.2)	8.7 (8.2, 9.2)
Asthma	4.6 (3.7, 5.4)	4.4 (3.9, 5.0)	4.1 (3.8, 4.4)	4.5 (4.1, 4.9)	4.6 (3.6, 5.5)	4.8 (4.6, 5.1)	5.1 (4.8, 5.5)
Carpal tunnel†	4.6 (3.2, 6.0)	3.3 (2.8, 3.7)	3.3 (2.9, 3.6)	3.4 (3.2, 3.7)	2.8 (2.3, 3.3)	2.8 (2.6, 2.9)	3.1 (2.8, 3.4)
Dermatitis	4.1 (3.0, 5.1)	4.6 (4.1, 5.2)	4.0 (3.7, 4.3)	4.1 (3.7, 4.4)	4.3 (3.9, 4.6)	4.2 (3.9, 4.4)	3.7 (3.3, 4.2)
Osteoarthritis of load-bearing joints	2.5 (2.1, 2.8)	2.9 (2.3, 3.4)	2.3 (2.1, 2.5)	2.5 (2.1, 2.8)	2.4 (2.0, 2.7)	2.6 (2.4, 2.8)	2.9 (2.6, 3.2)
Chronic obstructive pulmonary disease	4.2 (2.8, 5.6)	2.7 (2.3, 3.1)	2.6 (2.3, 2.8)	2.8 (2.5, 3.2)	3.1 (2.5, 3.6)	1.9 (1.7, 2.1)	2.2 (1.9, 2.5)
Hearing loss	1.63 (1.04, 2.22)	0.78 (0.61, 0.95)	0.78 (0.67, 0.89)	0.70 (0.55, 0.86)	0.63 (0.45, 0.81)	0.84 (0.77, 0.92)	0.75 (0.60, 0.90)
Bladder cancer	0.26 (0.07, 0.45)	0.20 (0.10, 0.30)	0.21 (0.17, 0.25)	0.16 (0.11, 0.22)	0.17 (0.11, 0.22)	0.16 (0.14, 0.19)	0.21 (0.15, 0.28)
Parkinsonism		0.13 (0.04, 0.22)	0.07 (0.05, 0.10)	0.07 (0.04, 0.11)	0.06 (0.02, 0.11)	0.08 (0.06, 0.10)	0.08 (0.03, 0.12)
Asbestosis	0.02 (0.00, 0.05)	0.07 (0.00, 0.13)	0.04 (0.02, 0.07)	0.02 (0.00, 0.04)	0.03 (0.00, 0.07)	0.01 (0.01, 0.02)	0.02 (0.00, 0.04)
Coal workers' pneumoconiosis‡	0.21 (0.08, 0.33)	0.02 (0.00, 0.06)	0.01 (0.00, 0.01)		0.01 (0.00, 0.03)	0.01 (0.00, 0.01)	

Note: Adjusted for sex, age, and share of costs paid by insured. Case definition is two or more claims during 2002 to 2005 (95% confidence intervals). Industry sectors as defined by the NIOSH National Occupational Research Agenda (NORA) (See notes to Table 2). The agriculture, forestry, and fisheries sector was not included in the Table, because the sample for this sector was small and likely to be particularly unrepresentative. Database randomly excludes some individuals from large employers so that all sector samples are over three times as large as largest employer in sector.

*Services other than health care and social assistance.

†One 4-digit, service sector industry whose results are not reportable due to confidentiality protections, had the highest rate of carpal tunnel syndrome and a large share of the cases in the data. Data for this industry were excluded in the rate calculations, because its inclusion had undue influence on the difference between the service sector and other sectors.

‡Analysis of coal workers' pneumoconiosis restricted to men, because there were no female cases.

NORA, National Occupational Research Agenda.

TABLE 5. Two-Digit SIC Industry Groups With Adjusted Prevalence More Than 120% or Less Than 80% of Population Average and FDR $P < 0.25$

Disease or Condition (Database Population Rate)	Adjusted Prevalence More Than 120% of Population Rate	Rate (%) (95% CI)	Adjusted Prevalence Less Than 80% of Population Rate (Up to Five SICs listed)	Rate (%) (95% CI)
Hypertension (27.6)	Local and interurban passenger transit	41.5 (38.9, 44.2)*	Amusement and recreation services	20.2 (16.4, 24.0)*
	Social services	33.3 (30.6, 36.1)*		
Low back pain (27.3)	Real estate	34.9 (31.4, 38.5)*	Legal services	21.8 (20.3, 23.3)*
	Holding and other investment offices	34.1 (27.5, 40.8)		
	Special trade contractors	34.1 (31.3, 36.9)*		
	Building materials and garden supplies	32.8 (29.4, 36.1)*		
Depression (10.5)	Local and interurban passenger transit	16.2 (14.0, 18.3)*	Amusement and recreation services	6.9 (3.9, 9.9)
	Real estate	15.7 (13.2, 18.1)*	Stone, clay, and glass products	7.4 (5.9, 8.9)*
	Social Services	14.6 (12.9, 16.3)*	Heavy construction, except building	7.5 (5.4, 9.6)*
	Miscellaneous manufacturing industries	14.2 (12.2, 16.3)*	Coal mining	7.9 (6.3, 9.4)*
	Personal services	14.2 (10.9, 17.6)	Transportation equipment	8.1 (6.0, 10.2)
	Legal services	13.4 (12.2, 14.7)*		
	Membership organizations	13.3 (11.6, 14.9)*		
	Security and commodity brokers	12.6 (10.7, 14.5)		
Cardiovascular disease (8.1)	Local and interurban passenger transit	14.5 (11.8, 17.3)*	Transportation by air	5.2 (2.2, 8.2)
	General merchandise stores	12.5 (10.3, 14.8)*		
	Oil and gas extraction	12.0 (7.8, 16.2)		
	Textile mill products	11.1 (9.3, 12.9)*		
	Auto repair, services, and parking	11.0 (8.5, 13.5)		
	Personal services	10.6 (8.3, 12.9)		
	General building contractors	10.2 (8.7, 11.8)*		
	Miscellaneous retail	10.0 (8.7, 11.2)*		
Diabetes (7.8)	Local and interurban passenger transit	17.3 (12.2, 22.4)*	General merchandise stores	5.2 (3.1, 7.2)*
	Auto repair, services, and parking	11.3 (8.7, 14.0)*	Legal services	5.3 (4.7, 6.0)*
	Eating and drinking places	11.1 (9.3, 12.9)*	Heavy construction, except Building	5.6 (4.0, 7.2)*
	Social services	10.2 (8.5, 11.9)*		
	Personal services	10.0 (7.6, 12.3)		
Asthma (4.7)	Social services	6.9 (5.6, 8.1)*	General merchandise stores	2.5 (1.7, 3.3)*
	Holding and other investment offices	6.8 (4.8, 8.8)	Furniture and home furnishings stores	2.8 (1.7, 3.9)*
			Miscellaneous repair services	2.8 (1.0, 4.7)
			Electronic and other electric equip.	3.4 (2.8, 4.1)*
			Textile mill products	3.6 (2.7, 4.5)*
Carpal tunnel syndrome[a] (3.0)	Coal mining	5.2 (3.0, 7.4)	Transportation by air	1.8 (0.7, 2.9)
	Administration of economic programs	4.6 (3.4, 5.9)*	Engineering and management services	2.2 (1.9, 2.6)*
	Lumber and wood products	4.5 (3.2, 5.7)		
	Wholesale trade-nondurable goods	4.3 (3.1, 5.5)		
Dermatitis (4.2)	Transportation by air	6.4 (4.3, 8.5)	[None]	
	Legal services	5.8 (5.1, 6.5)*		
	Administration of economic programs	5.3 (4.2, 6.5)		
	Eating and drinking places	5.1 (4.2, 6.0)		
	Engineering and management services	5.0 (4.4, 5.7)*		

(*Continued*)

TABLE 5. (*Continued*)

Disease or Condition (Database Population Rate)	Adjusted Prevalence More Than 120% of Population Rate	Rate (%) (95% CI)	Adjusted Prevalence Less Than 80% of Population Rate (Up to Five SICs listed)	Rate (%) (95% CI)
Osteoarthritis of load-bearing joints (2.7)	Oil and gas extraction	6.2 (3.0, 9.4)	Transportation equipment	1.7 (0.9, 2.4)*
	Special trade contractors	3.3 (2.7, 4.0)	Wholesale trade-durable goods	1.7 (1.3, 2.1)*
			Transportation services	1.7 (1.1, 2.3)*
			Trucking and warehousing	1.9 (1.5, 2.2)*
			Rubber and miscellaneous plastics products	1.9 (1.3, 2.5)*
Chronic obstructive pulmonary disease (2.3)	Coal mining	4.7 (3.1, 6.2)*	Educational services	1.5 (1.3, 1.7)*
	Textile mill products	4.4 (2.9, 5.9)*	Security and commodity brokers	1.6 (1.1, 2.0)*
	Auto repair, services, and parking	4.1 (2.5, 5.8)	Depository institutions	1.6 (1.3, 1.8)*
	Heavy construction, except Building	3.9 (2.7, 5.1)*	Legal services	1.7 (1.3, 2.1)*
	Trucking and warehousing	3.5 (3.1, 3.9)*	Engineering and management services	1.8 (1.3, 2.3)
	Automotive dealers and service stations	3.4 (2.6, 4.2)*		
	Eating and drinking places	3.3 (2.3, 4.3)		
	Food stores	3.0 (2.4, 3.7)		
	Fabricated metal products	3.0 (2.6, 3.4)*		
	Industrial machinery and equipment	2.9 (2.4, 3.5)		
	Stone, clay, and glass products	2.8 (2.4, 3.3)		
Hearing loss (0.79)	Coal mining	1.77 (1.24, 2.29)*	Rubber and miscellaneous plastics products	0.32 (0.01, 0.64)*
	Depository institutions	1.06 (0.81, 1.31)	Textile mill products	0.36 (0.06, 0.67)*
	Administration of economic programs	1.05 (0.87, 1.22)*	Printing and publishing	0.50 (0.18, 0.81)
	Executive, legislative, and general	0.96 (0.83, 1.08)*	Automotive dealers and service stations	0.50 (0.23, 0.77)
			Stone, clay, and glass products	0.54 (0.33, 0.75)
Bladder cancer (0.19)	Administration of human resources	0.57 (0.17, 0.96)	Depository institutions	0.13 (0.09, 0.18)*
	Fabricated metal products	0.35 (0.23, 0.46)*		
Parkinsonism (0.08)	Membership organizations	0.16 (0.10, 0.22)*	[None]	
Asbestosis (0.03)	[None]		Educational services	0.02 (0.00, 0.03)
Coal workers' Pneumoconiosis[d] (0.02)	Coal mining	0.29 (0.05, 0.53)	[None]	

Note: Disease rate based on case definition of two or more claims during 2002 to 2005. Rates are adjusted for age, gender, and share of costs paid by insured. 95% confidence interval does not include multiple testing adjustments. Database randomly excludes some individuals from large employers so that all two-digit industry samples are over three times as large as largest employer in SIC.

*FDR $P < 0.1$.

[a]Rate calculations excluded one large four-digit SIC with a high rate, so that other industry rates could be more usefully compared to the remainder of the database population. See notes to Table 4.

[b]Analysis of coal workers' pneumoconiosis restricted to men, because there were no female cases.

SIC, Standard Industrial Classification; FDR, false discovery rate.

exact determinations. Although unintentional or intentional choice of inappropriate codes may be of some concern, these sources of error should not differ greatly among industries.

Several factors independent of work affected industry rates. Medical claims miss cases that are untreated, and so relative industry rates would be affected by differences in the tendency to seek medical care, which could be affected partly by deductibles, co-pays, and coinsurance rates. We did not have data on these factors, but we employed a summary, share-of-cost variable on the employer level that had an association with most of the included diseases. Although industry rates were adjusted for age and gender, they could not be adjusted for other important factors such as smoking, body mass index, education, race and ethnicity, or household income that might be regarded as partly or wholly independent of industry of employment. Also, industries in which people become ill may have had their disease rates reduced if their employees seek jobs in other industries with working conditions that are better for their health. The rates presented here are for long-term industry employees, so workers who stop working or who switch jobs for health or other reasons are underrepresented.

There are several reasons why the results are not fully representative of the working population in the region of the insurer, or in individual industries and employers. A single insurer provides only a portion of the insurance coverage in its region. At the same time,

TABLE 6. Cost of Claims With Primary Diagnosis Code in Case Definition (2009 Dollars)*

	Average Annual Cost for Database Population ($)	Average Cost Per Claim ($)	Total Annual Cost Per Case ($)	
			Database Population Mean (Median)	SICs With Rates >120% of Population† Mean (Median)
Hypertension	5,100,738	94	86 (39)	85 (32)
Low back pain	17,138,914	379	293 (114)	276 (109)
Depression	6,530,337	452	291 (101)	343 (122)
Cardiovascular disease	38,040,938	906	2,195 (385)	2,199 (449)
Diabetes	6,005,084	132	361 (128)	292 (120)
Asthma	1,657,803	201	164 (52)	139 (49)
Carpal tunnel‡	2,164,296	602	342 (174)	464 (196)
Dermatitis	460,316	63	51 (25)	51 (26)
Osteoarthritis of load-bearing joints	2,920,136	615	514 (183)	505 (176)
Chronic obstructive pulmonary disease	1,160,698	304	232 (41)	270 (47)

*Only costs of claims that had a case definition code listed as the primary diagnosis code are included. Based on data employed for calculation of two-digit SIC rates. Costs adjusted to 2009 levels, using GDP price index (US Bureau of Economic Analysis. National Income and Product Accounts. Table 1.4.4. 2010).

†SICs with false discovery rate $P < 0.25$ for test of difference of SIC rate from population rate. See Table 5.

‡Rate calculations excluded one large four-digit SIC with a high rate, so that other industry rates could be more usefully compared to the remainder of the database population. See notes to Table 4.

SIC, Standard Industrial Classification.

it insures employees outside its region if they work for regionally based employers. Samples may also be biased at the industry and employer level, because some employers may provide employees a choice of insurer. As noted earlier, the data do not include workers who do not obtain health insurance through their employer and who may have different characteristics and working conditions. Finally, to protect employer confidentiality, random samples of employees from some large employers were deleted, potentially biasing some industry samples. Limitations to the estimates of medical cost of disease include the omission of drug costs and of costs of relevant claims whose primary diagnosis codes are not for the disease.

Next Steps

Next steps are of two kinds—more in-depth data analysis, and exploration of the use of results to enhance health care insurer reporting to employers. Each disease has its own set of occupational and nonoccupational risk factors that are distributed differently across industries. Focused examination of these distributions in relation to the observed disease rates is needed to develop reasonable hypotheses about the sources of the observed high rates. Disease rate reports for individual employers do not provide immediate guides to action but are a tool that may help lead to more focused attention on key health problems and opportunities for improving health and productivity. Statistical reports need to be supplemented with guidance on how to go about investigating the reasons for a high disease rate and identifying and implementing effective prevention measures. Consultation with occupational disease specialists may be needed.

CONCLUSION

Industry and employer disease rate analysis of medical claims appears to offer broad but useful indications of where more attention is needed to occupational and nonoccupational risk factors for some diseases, especially those with high prevalence. Medical claims analysis may be particularly useful given the great shortcomings of current employer-based systems of identifying and recording individual cases of occupational disease. Its main limitations are the lack of industry coding on the subemployer level, lack of information on occupation and exposure, restriction to treated cases, and lack of key variables that affect health, such as smoking and body mass index. In addition, results may not be generalizable to workers not insured through their employer. Its strengths are that it offers timely, relatively low cost, longitudinal data on physician-diagnosed cases and costs for large populations.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the collaboration and support of Highmark Inc., headquartered in Pittsburgh and Camp Hill, Pennsylvania, whose data are analyzed in this article. We thank Brian Day, Andrea DeVries, and Cara Hirsch at Highmark for their assistance in establishing and guiding the project, and Jerry O'Donnell, also at Highmark, for his assistance with the data. The authors also thank Eva Hnizdo, Lee Petsonk, Ahmed Gomaa, Tom Hales, Toni Alterman, Mark Stephenson, and Tania Carreon-Valencia at NIOSH, Ray Ten Eyck at Wright State University School of Medicine, and Lawson Wulsin at the University of Cincinnati Medical School for their contributions to project planning and formulation of disease case definitions. Kim Jinnett of the Integrated Benefits Institute, Phillip Hunt of Social Sectors Development Strategies, Marie Haring Sweeney, and Aaron Sussell of NIOSH, and an anonymous reviewer for the journal provided useful editorial suggestions.

REFERENCES

1. NIOSH. US Department of Health and Human Services, Centers for Disease Control and Prevention. Total Worker Health web page. http://www.cdc.gov/niosh/TWH/. Accessed October 11, 2011.
2. Leigh JP, Marcin JP, Miller TR. An estimate of the U.S. Government's undercount of nonfatal occupational injuries. *J Occup Environ Med.* 2004;46:10–18.
3. Ruser J. Examining evidence on whether BLS undercounts workplace injuries and illnesses. *Monthly Labor Rev.* 2008;131:20–32.
4. Boden L, Ozonoff A. Capture-recapture estimates of nonfatal workplace injuries and illnesses. *Ann Epidemiol.* 2008;18:500–506.
5. Biddle J, Roberts K. Claiming behavior in workers' compensation. *J Risk Insur.* 2003;70:759–780.
6. Rosenman KD, Kalush A, Reilly MJ, Gardiner JC, Reeves M, Luo Z. How much work-related injury and illness is missed by the current national surveillance system? *J Occup Environ Med.* 2006;48:357–365.

7. Oregon Department of Consumer and Business Services, Information Management Division. Research alert: workplace injuries and workers' compensation claim filing: results from the 2006 Oregon population survey. http://actprod.cbs.state.or.us/iportal/report_catalog.htl. Accessed October 11, 2011.
8. Won JU, Dembe AE. Services provided by family physicians for patients with occupational injuries and illnesses. *Ann Fam Med.* 2006;4:138–147.
9. Pollack ES, Deborah GK, eds. *Counting Injuries and Illnesses in the Workplace: Proposals for a Better System*. Panel on Occupational Safety and Health Statistics, Seymour Geisser, Chair. Committee on National Statistics. Commission on Behavioral and Social Sciences and Education. National Research Council. Washington, DC: National Academy Press; 1987.
10. Barth P. Some reflections on the National Commission and its legacy. *Workers' Compensation Policy Rev*. 2004;4:21–27.
11. Azaroff LS, Levenstein C, Wegman D. Occupational injury and illness surveillance: conceptual filters explain underreporting. *Am J Public Health*. 2002;92:1421–1429.
12. US Department of Labor. Bureau of Labor Statistics. 2007 supplemental Table 10: number, percent, and rate of cases of musculoskeletal disorder. (Reissued March 2009). http://stats.bls.gov/iif/oshcdnew2007.htm#07 Supplemental_Tables. Accessed October 11, 2011.
13. US Department of Labor, Occupational Safety and Health Administration. 29 CFR Part 1910. Ergonomics program final rule. November 14, 2000. [OSHA Archive Document] pp. 68386-68388. http://www.osha.gov/FedReg_osha_pdf/FED20001114.pdf. Accessed October 11, 2011.
14. US Department of Labor, Occupational Safety and Health Administration. Recordkeeping Web page. www.osha.gov/recordkeeping/index.html. Accessed October 11, 2011.
15. Johnson WG. Back pain: acute injury or chronic disease? *Workers' Compensation Policy Rev.* 2004;4(5):9–18.
16. Morse TF, Dillon C, Warren N, Levenstein C, Warren A. The economic and social consequences of work-related musculoskeletal disorders: the Connecticut Upper-Extremity Surveillance Project (CUSP). *Int J Occup Environ Health.* 1998;4:209–216.
17. US Department of Labor. Bureau of Labor Statistics. Survey of Occupational Injuries and Illnesses. Industry division or selected characteristic by detailed nature of condition. Series ID CHU00 × 1443XX6N100. http://data.bls.gov/cgi-bin/dsrv. Accessed January 19, 2011.
18. US Department of Health and Human Services. Centers for Disease Control and Prevention. Behavioral risk factor surveillance system. 2007 adult asthma data: prevalence tables and maps. Table C1. Adult self-reported current asthma prevalence rate (percent) and prevalence (number) by State or Territory. http://www.cdc.gov/asthma/brfss/07/current/tableC1.htm. Accessed October 11, 2011.
19. Balmes J, Becklake M, Blanc P, et al. American Thoracic Society Statement: occupational contribution to the burden of airway disease. *Am J Respir Crit Care Med*. 2003;167:787–797.
20. Steenland K, Burnett C, Lalich N, Ward E, Hurrell J. Dying for work: the magnitude of US mortality from selected causes of death associated with occupation. *Am J Ind Med.* 2003;43:461–482.
21. O'Malley KJ, Cook KF, Price MD, Wildes R, Hurdle JF, Ashton CM. Measuring diagnosis: ICD code accuracy. *Health Serv Res*. 2005;40:1620–1639.
22. Fronstin P. Sources of health insurance and characteristics of the uninsured: analysis of the March 2009 current population survey. Washington, DC: Employee Benefit Research Institute; 2009. Issue Brief No. 334.
23. Park RM, Nelson NA, Silverstein MA, Mirer FE. Use of medical insurance claims for surveillance of occupational disease: an analysis of cumulative trauma in the auto industry. *J Occup Med.* 1992;34:731–737.
24. Park RM, Krebs JM, Mirer FE. Occupational disease surveillance using disability insurance at an automotive stamping and assembly complex. *J Occup Environ Med.* 1996;38:1111–1123.
25. Dement JM, Pompeii AP, Østbye T, et al. An integrated comprehensive occupational surveillance system for health care workers. *Am J Ind Med.* 2004;45:528–538.
26. Schoenfisch AL, Dement JM, Rodríguez-Acosta RL. Demographic, clinical, and occupational characteristics associated with early onset of delivery: findings from the Duke Health and Safety Surveillance System, 2001–2004. *Am J Ind Med.* 2008;51:911–922.
27. Pompeii LA, Lipscomb HJ, Dement JM. Surveillance of musculoskeletal injuries and disorders in a diverse cohort of workers at a tertiary care medical center. *Am J Ind Med.* 2008;51:344–356.
28. Lipscomb HJ, Dement JM, Silverstein B, Kucera KL, Cameron W. Health care utilization for musculoskeletal back disorcers, Washington State Union Carpenters, 1989–2003. *J Occup Environ Med*. 2009;51:604–611.
29. Lipscomb HJ, Dement JM, Silverstein B, Cameron W, Glazner JE. Who is paying the bills? Health care costs for musculoskeletal back disorders, Washington State Union Carpenters, 1989-2003. *J Occup Environ Med*. 2009;51:1185–1192.
30. Smith PM, Stock SR, McLeod CB, Koehoorn M, Marchand A, Mustard CA. Research opportunities using administrative databases and existing surveys for new knowledge in occupational health and safety in Canada, Quebec, Ontario, and British Columbia. *Can J Public Health.* 2010;101: S46–S52.
31. Centers for Disease Control and Prevention (CDC). HIPAA privacy rule and public health. Guidance from CDC and the U.S. Department of Health and Human Services. *Morb Mortal Wkly Rep*. 2003;52: 1–20.
32. US Department of Health and Human Services. Health information privacy web page. http://www.hhs.gov/ocr/privacy/. Accessed July 8, 2010.
33. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *J Appl Econometrics*. 1996;11:619–632.
34. US Bureau of Economic Analysis. National Income and Product Accounts. GDP price index. Table 1.4.4. http://www.bea.gov/national/nipaweb/index.asp. Accessed October 11, 2011.
35. Lee J. Covariance adjustment of rates based on the multiple logistic regression model. *J Chron Dis.* 1981;34:415–426.
36. Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. *Biometrics.* 1982;38:613–621.
37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statist Soc B.* 1995;57:289–300.
38. Benjamini Y, Hochberg Y. 2000. The adaptive control of the false discovery rate in multiple testing with independent statistics. *J Educ Behav Statist* 25:60–83.
39. Tüchsen F, Anderson O, Costa G, Filakti H, Marmot MG. Occupation and ischemic heart disease in the European Community: a comparative study of occupations at potential high risk. *Am J Industrial Med.* 1996:30:407–414.
40. Landen D, Wang RC, Makowski BD, Tuchman RJ. *Injuries, Illnesses, and Hazardous Exposures in the Mining Industry, 1986–1995: A Surveillance Report*. US Department of Health and Human Services, PHS, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. Pittsburgh, PA: DHHS (NIOSH) Publication No. 2000-117, May 2000.
41. Torén K, Bergdahl IA, Nilsson T, Järvholm B. Occupational exposure to particulate air pollution and mortality due to ischaemic heart disease and cerebrovascular disease. *Occup Environ Med*. 2007;64:515–519.
42. Tüchsen F, Bach E, Marmot M. Occupation and hospitalization with ischaemic heart diseases: a new nationwide surveillance system based on hospital admissions. *Int J Epidemiol*. 1992;21:450–459.
43. Hansen ES. Mortality of auto mechanics. A ten-year follow-up. *Scand J Work Environ Health*. 1989;15:43–46.
44. Tse JLM, Flin R, Mearns K. Bus driver well-being review: 50 years of research. *Transport Res.* Part F. 2006;9:89–114.
45. Tüchsen F, Hannerz H, Roepstorff C, Krause N. Stroke among male professional drivers in Denmark, 1994–2003. *Occup Environ Med*. 2006;63:456–460.
46. Schilling R, Goodman N. Cardiovascular disease in cotton workers: part I. *Br J Ind Med*. 1951;8:77–90.
47. Sweetnam PM, Taylor SW, Elwood PC. Exposure to carbon disulphide and ischaemic heart disease in a viscose rayon factory. *Br J Ind Med*. 1987;44:220–227.
48. Zanardi F, Harris EC, Brown T, Rice S, Palmer KT, Coggon D. Mortality from diabetes and ischaemic heart disease in textile workers. *Occup Environ Med*. 2011;68:172–175.
49. US Department of Labor. Bureau of Labor Statistics. Survey of occupational injuries and illnesses. Detailed industry by selected parts of body affected. Resource Table R6. Various years. http://www.bls.gov/iif/oshcdnew.htm. Accessed July 11, 2011.
50. NIOSH CDC. Worker Health eChartbook. Data table: number and rate of nonfatal occupational musculoskeletal disorder cases with days away from work by service sector, US private industry, 2003–2007. http://wwwn.cdc.gov/niosh-survapps/echartbook/Chart.aspx?id=11641&cat=1272. Accessed July 13, 2011.
51. Silverstein B, Adams D. Work-related musculoskeletal disorders of the neck, back, and upper extremity in Washington State, 1997–2005 [Technical Report No. 40-11-2007]. December 2007. http://www.lni.wa.gov/Safety/Research/Files/2007WmsdRpt.pdf. Accessed October 11, 2011.

52. US Department of Labor. Occupational employment statistics. May 2010 National industry-specific occupational employment and wage estimates; NAICS 531000-Real Estate. http://www.bls.gov/oes/current/naics3_531000.htm. Accessed July 14, 2011.
53. Guo H-R, Tanaka S, Halperin WE, Cameron LL. Back pain prevalence in US industry and estimates of lost workdays. *Am J Public Health*. 1999;89:1029–1035.
54. US Department of Labor. Bureau of Labor Statistics. Survey of occupational injuries and illnesses. Detailed industry by selected natures. Resource Table R5. Various years. http://www.bls.gov/iif/oshcdnew.htm. Accessed July 11, 2011.
55. Shuford H, Restrepo T. Carpal tunnel claims rank second among major lost time diagnoses. *NCCI Research Brief*. 2005;3. https://www.ncci.com/nccimain/IndustryInformation/ResearchOutlook/ArchivedArticles/Pages/research-carpal-tunnel.aspx. Accessed October 11, 2011.
56. Tanaka S, Wild DK, Seligman PJ, Halperin WE, Behrens VJ, Putz-Anderson V. Prevalence and work-relatedness of self-reported carpal tunnel syndrome among U.S. workers: analysis of the Occupational Health Supplement Data of the 1988 National Health Interview Survey. *Am J Ind Med.* 1995;27:451–470.
57. Silverstein B, Viikari-Juntura E, Kalat J. Use of a prevention index to identify industries at high risk for work-related musculoskeletal disorders of the neck, back, and upper extremity in Washington State, 1990–1998. *Am J Ind Med.* 2002;41:149–169.
58. Burke FD, Lawson IJ, McGeoch KL, Miles NV, Proud G. Carpal tunnel syndrome in association with hand-arm vibration syndrome: a review of claimants seeking compensation in the Mining Industry. *J Hand Surg Br.* 2005;30:199–203.
59. van Rijn RM, Huisstede BM, Koes BW, Burdorf A. Associations between work-related factors and the carpal tunnel syndrome–a systematic review. *Scand J Work Environ Health*. 2009;35:19–36.
60. Oddo R, Loyau T, Boileau PE, Champoux Y. Design of a suspended handle to attenuate rock drill hand-arm vibration: model development and validation. *J Sound Vib*. 2004;275:623–640.
61. McPhee B. Ergonomics in mining. *Occup Med*. 2004;554:297–303.
62. Dillon C, Petersen M, Tanaka S. Self-reported hand and wrist arthritis and occupation: data from the U.S. National Health Interview Survey-Occupational Health Supplement. *Am J Ind Med*. 2002;42:318–327.
63. Winn FJ, Biersner RJ, Morrissey S. Exposure probabilities to ergonomic hazards among miners. *Int J Ind Ergonom.* 1996;18:417–422.
64. Ruder A, Carreon T, Ward E, Schulte P, Halperin W. Bladder cancer. In: Rosenstock L, Cullen M, Brodkin CA, Redlich C, eds. *Textbook of Clinical Occupational and Environmental Medicine.* Philadelphia, PA: Elsevier Saunders; 2005.

