Incidence of Workers Compensation Indemnity Claims Across Socio-Demographic and Job Characteristics

Juan Du, PhD[1] and J. Paul Leigh, PhD[2]*

Background *We hypothesized that low socioeconomic status, employer-provided health insurance, low wages, and overtime were predictors of reporting workers compensation indemnity claims. We also tested for gender and race disparities.*
Methods *Responses from 17,190 (person-years) Americans participating in the Panel Study of Income Dynamics, 1997–2005, were analyzed with logistic regressions. The dependent variable indicated whether the subject collected benefits from a claim.*
Results *Odds ratios for men and African-Americans were relatively large and strongly significant predictors of claims; significance for Hispanics was moderate and confounded by education. Odds ratios for variables measuring education were the largest for all statistically significant covariates. Neither low wages nor employer-provided health insurance was a consistent predictor. Due to confounding from the "not salaried" variable, overtime was not a consistently significant predictor.*
Conclusion *Few studies use nationally representative longitudinal data to consider which demographic and job characteristics predict reporting workers compensation indemnity cases. This study did and tested some common hypotheses about predictors.*
Am. J. Ind. Med. 54:758–770, 2011. © *2011 Wiley-Liss, Inc.*

KEY WORDS: economics; workers compensation

INTRODUCTION

Roughly 4–5 million new job-related injuries and illnesses are recorded each year by the Bureau of Labor Statistics [Bureau of Labor Statistics, 2009a]. The annual cost for workers compensation in 2007 was $85 billion [Sengupta et al., 2009]. Whereas there are some economic studies [Butler and Worrall, 1983; Guo and Burton, 2010], we are not aware of any epidemiologic studies that use reporting of workers compensation claims as dependent variables. Economists typically use aggregate state-by-state workers compensation data, and do not focus on demographic characteristics, and with exceptions [Lakdawalla et al., 2007; Guo and Burton, 2010] most are conducted on data from prior to 1995. On the other hand, a few epidemiological studies have addressed personal and job characteristics that are correlated with reporting a job-related injury or illness using national data [Oh and Shin, 2003; Dembe et al., 2004, 2005; Strong and Zimmerman, 2005]; but reporting an injury is not the same as reporting a claim.

This study presents socio-demographic and job characteristics data from a nationally representative sample of Americans who reported workers' compensation indemnity claims. Determining who reports claims is important. One issue involves health disparities. [Kilbourne et al., 2006]. Considerable attention is placed on establishing whether and how significant gender or race differences are in the incidence of many different diseases and injuries as

[1]Department of Economics, Old Dominion University, Norfolk, Virginia
[2]Center for Healthcare Policy and Research and Department of Public Health Sciences, University of California, Davis, California
Contract grant sponsor: National Institute for Occupational Safety and Health; Contract grant number: OH008248-01.
Disclosure Statement: The authors report no conflicts of interests.
*Correspondence to: J. Paul Leigh, PhD, Center for Healthcare Policy and Research, UC Davis Medical School, MS1C, Davis, CA 95616-5270. E-mail: pleigh@ucdavis.edu

Accepted 2 June 2011
DOI 10.1002/ajim.20985. Published online 14 July 2011 in Wiley Online Library (wileyonlinelibrary.com).

well as participation in government programs such as unemployment insurance [Blank and Card, 1991] and whether these differences are affected by levels of education. Another issue involves prevention. If groups can be identified, prevention resources can be more efficiently targeted. Hypotheses have also been advanced regarding socioeconomic status and job characteristics as predictors of claim filing and, in related literature, predictors of work-related injury. Several studies have considered whether employees in jobs with employer-provided health insurance are more or less likely to report workers compensation claims [Card and McCall, 1996; Lakdawalla et al., 2007; Lipscomb et al., 2009]. In addition, studies have considered whether low wages or overtime is predictive of experiencing work-related injury or reporting claims [Leigh, 1986; Robinson, 1989; Dembe et al., 2005; Kaufman and Hotchkiss, 2005].

DATA AND METHOD

Data

The Panel Study of Income Dynamics (PSID), which began in 1968, is a longitudinal, representative US sample of men, women, and their children. The PSID contains rich information including, for example, respondents' workers compensation claims, age, race, gender, education, and region. We combined data on household heads (most frequently men) with what the PSID referred to as "wives", if any, for five recent waves: 1997, 1999, 2001, 2003, and 2005.

Our dependent variable was derived from this question: "Did you receive any income (in 1998) from workers compensation?" The PSID, therefore, only measured claims with indemnity benefits, not claims that were "medical only." The year in parentheses varied depending on the wave: 1998 for the 1999 wave; 2000 for the 2001 wave, etcetera. These were the possible PSID responses: 1 = yes, 5 = no, 8 = do not know, 9 = not available, refused, and 0 = wild code. We scored the responses for our workers compensation variable as "1" if the PSID response was "1" and as "0" for PSID responses of "5." Only eight PSID responses were neither "1" or "5" among respondents who met our criteria (below), and these eight observations were dropped.

Since our dependent variable involved workers compensation, we limited the sample to adults who were working prior to their possible claims. This included employees between and including ages 21–65. We did not include the self-employed since they are not eligible for workers compensation [Sengupta et al., 2009]. We required that persons must have worked at least 500 annual hours (roughly 10 hr per week) in the year preceding their possible workers compensation claims. The great majority worked 2,000+ annual hours, or roughly 40+ weekly hours In our preferred regressions, we excluded persons with missing data. For our entire sample, 17,190 person-years had no missing values on any variable and 336 person-years had at least one missing value for at least one variable. For persons who reported work hours for the year prior to the interview year, or in the previous waves if he/she did not work the year prior to the interview, the sample size contained: 3,861 respondents in 1999; 4,235 in 2001; 4,484 in 2003; and 4,610 in 2005. We did not collect workers compensation data from the 1997 wave. The 1997 wave was only used for covariates.

As Table I illustrates, our binary workers compensation variable did not have many "1's." This was expected. In any given year, only a fraction of persons are injured on jobs and then apply for workers compensation indemnity payments. In our sample, the percentage of workers with workers compensation claims ranged from 1.09% to 1.68% depending on the year.

We intentionally restricted the number of covariates. We created four race/ethnic categories: white non-Hispanic, African-American non-Hispanic, Hispanic, and "all others." Age was divided into five categories (Table I). Married was binary and equaled "1" if the respondent was married with spouse present. Two education variables were created. One measured whether respondents never completed high school and another whether they completed high school, but not college. Rural residence indicated population <20,000 and not adjacent to a metropolitan area. Four Census regions were included.

Wages reflected respondents' annual income from work divided by annual work-hours in the year before the interview. This wage variable measured all earnings, including "second jobs," bonuses, overtime pay, tips, commissions, and other miscellaneous labor income. Wages were adjusted for inflation with the Consumer Price Index for all urban consumers [Bureau of Labor Statistics, 2009b]. Regional differences per year in the cost of living were accounted for in the East, Midwest, and West, using South as the base region. We created an indicator wage variable that equaled "1" for persons whose wages were in the bottom quartile of the wage distribution of PSID respondents. We also used wages as a continuous variable in separate analyses.

The overtime variable was an indicator and equaled "1" if the respondent reported working any overtime in the previous year; 20.3% reported overtime. The "not salaried" variable was also an indicator and equaled "1" if the respondent reported that he or she was not paid a weekly or monthly salary, but rather an hourly wage or by commission and tips. Roughly 58.3% reported "not salaried."

Following Dembe et al. [2005], a measure of high-risk occupations was included as a covariate. Unlike

TABLE I. Descriptive Statistics

	Frequency (n = 17,190)	Percent for binary variables and mean for continuous variables (n = 17,190)
Panel A	Report of workers compensation claim, panel A only	Report of workers compensation claim, panel A only
1999 interview year with workers compensation data from 1998	68	1.68
2001 interview year with workers compensation data from 2000	48	1.16
2003 interview year with workers compensation data from 2002	60	1.09
2005 interview year with workers compensation data from 2004	69	1.45
Workers compensation all years combined	245	1.34
Panel B covariates (all years combined)		
Socio-demographic variables		
Male	8,114	49.89
Female	9,076	50.11
African-American, non-Hispanic	5,120	11.78
Hispanic	1,197	6.06
All other races, non-Hispanic	872	4.50
White, non-Hispanic	10,001	77.66
21–24 years[a]	1,482	5.19
25–34 years[a]	4,634	24.51
35–44 years[a]	5,268	29.82
45–54 years[a]	4,102	28.04
55–64 years[a]	1,704	12.44
Married, spouse present	10,209	65.00
All other marital status	6,981	35.00
Rural residence (population $<$ 20,000, not adjacent to metropolitan area)	3,408	20.73
Northeast	2,513	18.79
South	7,032	31.66
Midwest	4,212	27.40
West	3,432	22.14
Education		
$<$High school education	2,971	13.54
High school graduate but not college	9,953	56.64
Job characteristics		
Hourly wage in lower 25th percentile	3,557	25.00
Employer-provided health insurance	13,433	83.08
Total number of work hours	n/a	2042.66
Overtime	3,736	20.30
Not salaried	11,050	58.30
BLS injury rate	n/a	1.07

[a]Ages correspond to the year before workers compensation claims, i.e., for interview year 2005, workers compensation claims were measured in year 2004, and ages were measured in year 2003.

Dembe et al., we constructed a continuous variable: BLS injury rate, which measured number of annual injuries divided by 100 fulltime workers. BLS only started calculating and publishing rates for occupations in 2006 (personal communication with Louis Martinez, September 23, 2010.) The data applied to non-fatal injuries and illnesses that resulted in at least 1 day of work loss. We assumed that the relative rankings of occupations would not change a great deal in these few years and therefore applied the 2006 data to each year. For the PSID occupations from 2003 and 2005, the matching was exact because the PSID used the same categories as BLS. We matched to 24 two-digit occupations on the BLS list: Management Occupations through Transportation and Material Moving

Occupations. We did not have any matches for Military Specific Occupations because the PSID did not interview anyone in the military. The 1999 and 2001 PSID waves used an older 1990 Census scheme for classifying occupations. To make matches, we consulted a cross-walk, selected "large" occupational categories with the greatest number of people, and used judgment based on category names.

The workers compensation variable was from waves 1999 to 2005, and all the covariates, including job characteristics, were from previous waves (1997–2003) except for the region and rural covariates. Had we used covariates from the same year, the estimates would be biased due to reverse causality and interpretations of results would have been more difficult. For example, a worker's injury can result in work-loss and therefore a loss of opportunity to work overtime. In this case, if overtime and workers compensation were positively correlated, it would be problematic to assert that overtime predicted workers compensation claims.

We divided covariates into two categories: (1) sociodemographic and region (gender, race, age, marital status, education, rural residence, and Census region); and (2) job characteristics (wages, employer-provided health insurance, work hours, overtime, salaried, and BLS injury rate).

Statistical Method

The PSID contains data with and without population weights. Because the PSID is longitudinal, persons without weights (non-sample persons) are those who enter the PSID through marriage or residency over time. They are not assigned weights because the time series data for non-sampled persons are left-censored from the year they enter the PSID, and it is unlikely this censoring is random. In our years of 1997–2005, roughly 35% of the sample did not have PSID weights. We slightly modified the PSID weights for our years. We first multiplied the PSID individual weight by the numbers of waves each person appeared. To normalize, we then multiplied by the sample size and divided by the sum of all weights in the sample.

Sample means and percents in Table I were estimated with sampling weights controlling for stratum and primary sampling unit. We used logistic regressions to estimate odds ratios [Stata, 2005]. We combined data across five waves and used random effects estimation procedures with weights to account for the lack of statistical independence for measures on the same individual across the years. The PSID is a geographically structured survey. The Stata technique, which accounts for geographic clusters was not available for these longitudinal data that covered five waves. (As far as we know, this Stata technique has not yet been written).

RESULTS

Table I panel A, provides data on the four waves with workers compensation data. The small percentage of respondents with workers compensation each year demonstrates why we combined waves. Table I, panel B, provides frequencies and means on covariates for all waves combined.

Table II presents results from running logistic regressions with workers compensation as the dependent variable. The first column of numbers was drawn from single-category regressions, i.e., from regressions with only one covariate (e.g., male) or one set of similar covariates (e.g., age categories). For example, in the first row, first column of numbers, the odds ratio for men compared to women without adjusting for any other covariates was 1.78. Many covariates in the first column garnered statistical significance at 5% except these: "other race," age groups 20–24, 25–34, and 45–54, married, rural residence, Northeast, Midwest, low-wage, employer-provided health insurance, and work hours.

But simple single-category regressions likely give biased estimates because there are many well-known sociodemographic and job-related predictors. Table II, second column of numbers, provides odds ratios and 95% confidence intervals (CIs) from a multiple logistic regression that included all covariates. Whereas most covariates maintained statistical significance, odds ratios changed when compared to the single-category regressions in the first column. For example, whereas the odds ratio for Hispanic in simple regression was 2.669, it was 1.756 in the multiple regression. A number of differences between the two columns were noteworthy. The difference between the simple regression odds ratio for education less than high school minus the multiple regression ratio was 2.8. Subtracting the two odds ratios, other noticeable differences occurred: Hispanic (0.91), high school without college (1.25), and "not salaried" (1.06). In addition, a number of covariates were statistically significant at the 0.05 level in the univariate analyses, but not in the multivariable analyses including Hispanic and overtime. Moreover, some variables such as male, African-American and age 55–64 were notable because they maintained strong statistical significance and odds ratios were very similar in both univarite and multivariable analyses.

Employer-provided insurance generated an odds ratio below one with simple regression (0.88) and an odds ratio above one (1.39) in the multiple regression, but was never statistically significant. Odds ratios on the wage variable were always less than one. The low-wage variable was neither significant in the single-category regression nor in the multiple regression. Workers from the West and Northeast were more likely than those from the South to report workers compensation claims in multiple regression.

TABLE II. Single-Category (Simple) and Multiple Regression Results

	Single-category regression: odds ratio (95% CI) (separate regressions for gender only, race only, age only, etc.)	One (long) multiple regression, including all covariates: odds ratio (95% CI)
Covariates (and sample size)		
Socio-demographic variables		
Male	1.781*** (1.294, 2.451)	1.810*** (1.267, 2.583)
Female	Referent	Referent
African-American, non-Hispanic	2.070*** (1.343, 3.184)	2.167*** (1.377, 3.410)
Hispanic	2.669*** (1.571, 4.537)	1.756* (0.965, 3.194)
All other races, non-Hispanic	1.439 (0.703, 2.952)	1.427 (0.691, 2.949)
White, non-Hispanic	Referent	Referent
20–24 years	1.496 (0.817, 2.738)	1.418 (0.767, 2.622)
25–34 years	0.667* (0.431, 1.033)	0.686* (0.443, 1.063)
35–44 years	Referent	Referent
45–54 years	1.322 (0.919, 1.903)	1.575** (1.093, 2.269)
55–64 years	0.454** (0.240, 0.855)	0.525** (0.279, 0.989)
Married, spouse present	0.850 (0.621, 1.164)	0.881 (0.632, 1.228)
All other marital status	Referent	Referent
Rural residence (population $<$ 20,000, not adjacent to metropolitan area)	1.328 (0.923, 1.910)	1.494** (1.010, 2.208)
Northeast	1.440 (0.897, 2.313)	2.051*** (1.248, 3.371)
West	2.249*** (1.472, 4.436)	2.382*** (1.508, 3.762)
South	Referent	Referent
Midwest	1.279 (0.824, 1.986)	1.353 (0.863, 2.121)
Education		
$<$High school education	7.554*** (4.232, 13.484)	4.750*** (2.508, 8.993)
High school graduate, but not college	4.297*** (2.578, 7.162)	3.048*** (1.787, 5.198)
Job characteristics		
Hourly wage in lower 25th percentile	1.019 (0.688, 1.510)	0.730 (0.476, 1.118)
Employer-provided health insurance	0.879 (0.601, 1.287)	1.388 (0.912, 2.114)
Total number of work hours (in 1,000 hr)	0.848 (0.668, 1.076)	0.785* (0.602, 1.024)
Overtime	1.508** (1.088, 2.090)	1.285 (0.925, 1.784)
Not salaried	2.656*** (1.463, 2.031)	1.594** (1.079, 2.355)
BLS injury rate	1.724*** (1.868, 3.774)	1.208* (0.995, 1.466)
Log-of-likelihood and P-value for overall significance of multiple regression	n/a	−1114.794 ($P < 0.001$)

Sample size is 17,190. 95% confidence interval is in the parenthesis.
*Indicates $0.05 \leq P < 0.1$.
**Indicates $0.01 \leq P < 0.05$.
***Indicates $P < 0.01$.

Odds ratios for the overtime variable differed: 1.51 in the univariate analysis and 1.29 in the multivariable analysis. Non-salaried workers were 59% more likely to report workers compensation based on the multiple regression.

We found the BLS injury rate to be highly correlated with workers compensation, although, again, the sizes of odds ratios were different in single-category versus multiple regression.

In additional analyses available from the authors, numerous multiple logistic regressions were run that were designed to address this question: were logistic multiple regression findings on gender and race influenced by the inclusion or exclusion of education, wage, employer-provided insurance, work hours, overtime, salaried, or BLS injury rate ether singly or as a group? We wanted to test how robust our results on gender and race were to the inclusion of education and the six job characteristics. For example, men typically are paid more than women so that whether wages were included in the regressions might influence the odds ratio for men. The results of this analysis were remarkably consistent. Sizes of odds ratios and levels of statistical significance were largely unaffected by the addition of education through the injury rate variables mentioned above for men and African-Americans. Hispanic, on the other hand, was especially sensitive to the inclusion of the education variables. If the education variables were excluded, Hispanic generated statistical significance at the 0.05 level with odds ratios well above 2.0. But when the education variables were included, P-values were in the range 0.05–0.10 and odds ratios fell to well below 2.0.

We next ran 20 multiple regressions that provided a sensitivity analysis for seven ''policy'' variables: education and the six job characteristics. This sensitivity analysis was important because we knew that all of these were correlated (leading to problems of multicollinearity) and each was associated with different government policies. For example, levels of education may be influenced by Head Start programs or government funding of public schools. Wages may be affected by minimum wage laws and individual businesses. Employer-provided insurance can be mandated; amounts of overtime and definitions of who is a salaried employee may be influenced by policies of individual businesses, labor laws, and enforcement of those laws. The BLS injury rate may be influenced by OSHA policies. Each of these regressions also included the socio-demographic variables as well as rural residence and regions. Covariates were entered in combinations (e.g., wages and employer-provided insurance; work hours and overtime) as well as singly (e.g., BLS injury rate). Because the two education variables appeared to be such strong predictors, most of these 20 regressions included both education variables.

Table III presents results on 14 of these 20 regressions. Neither employer-provided health insurance or low wage ever achieved statistical significance nor did the inclusion or exclusion of either of these variables noticeably affect odds ratios or confidence intervals on other covariates. Consequently, we did not include results on either employer-provided health insurance or low wage in Table III. In separate regressions, available from the authors, wage was entered as a continuous variable but results were not substantially different.

In Table III, regressions 1–5 added education, work hours, overtime, not salaried, and the BLS injury rate separately. Regressions 6–14 contained the base set and the two education variables together with paired combinations of work hours, overtime, salaried, and BLS injury rate. Comparisons should be made across rows within the same column. We found some differences in odds ratios. We have 10 regression results for the education variables. Odds ratios for the less than high school variable varied from 4.89 (95% CI: 2.64–9.07) to 6.47 (95% CI: 3.56–11.74). For example, it was 4.89 when overtime and ''not salaried'' were included (Table III, regression #11). The odds ratios for education more than high school but less than college ranged from 3.29 (95%CI: 1.95–5.58) to 4.08 (95%CI: 2.44–6.81); it also became the lowest when overtime and ''not salaried'' were included (Table III regression #11). The odds ratios for overtime ranged from 1.31 (95%CI: 0.95–1.82) to 1.45 (95%CI: 1.05–2.01). The lower bound on the 95% CI was above 1.0 only when education was excluded (regression #3); but a 90% CI was above 1.0 when education was included. The continuous work hours variable generated upper bound CIs below 1.0 at the 95% (regression #2) and 90% level (all other regressions). The odds ratio of 0.782 (95% CI:0.61–1.01) for work hours (regression #12) was interpreted as follows: persons who worked 2,000 hr in the year before the possible workers compensation claim were 1–0.782 or 21.8% less likely than persons who worked 1,000 hr in the same year to report a claim.

The policy ''importance'' of covariates in logistic regression is typically measured with odds ratios far away from 1.0. The statistically significant ''policy'' variables that consistently generated the largest odds ratios were, in order, education $<$ high school (95%CI odds ratios of 2.64–11.74, 2.64 is the lower bound for the point estimate of 4.9 and 11.74 is the upper bound for the point estimate of 6.5), high school graduate (95%CI 1.81–6.81) (3.3–4.1), not salaried (95%CI 0.98–3.55) and BLS injury rate (95%CI 0.98–1.83).

Separate questions involve whether working longer hours or reporting overtime affected the probability of claims differently for salaried and non-salaried workers. We investigated two interaction terms: one between salaried and overtime, and the other between salaried and total hours. These two interaction terms were entered in all regressions in which overtime, total hours, and salaried were included. We did not find the interaction terms to be statistically significant in any specifications. In addition, when the interaction terms were added, salaried and total hours became statistically insignificant. We believe the lack of statistical significance for the interactions as well as for the overtime and salaried variables when the interactions were included as covariates were the result of collinearity.

DISCUSSION

We used nationally representative data from 1997–2005 (17,190 person-years) to describe the worker and job characteristics of persons who reported collecting benefits from workers compensation indemnity claims. Findings pertaining to demographics indicated that men and African-Americans were more likely to report claims in both univariate and multivariate analyses. Confounding from education level suggested Hispanics were only marginally

TABLE III. Logistic Multiple Regression Results on Education, Work Hours, Overtime, Salaried, and BLS injury rate

Regression number and categories of covariates included (n = 17,190)	Odds ratio and (95%CI)					
	Education < high school	High school graduate, but not college	Total work hours	Overtime	Not Salaried	BLS injury rate
1. Demographics, geography, education	6.466*** (3.561, 11.739)	4.078*** (2.443, 6.805)				
2. Demographics, geography, work hours			0.745** (0.580, 0.958)			
3. Demographics, geography, overtime				1.451** (1.048, 2.009)		
4. Demographics, geography, not salaried					2.486*** (1.739, 3.554)	
5. Demographics, geography, BLS injury rate						1.531*** (1.285, 1.825)
6. Demographics, geography, education, work hours	6.192*** (3.412, 11.239)	3.992*** (2.395, 6.652)	0.793* (0.615, 1.021)			
7. Demographics, geography, education, overtime	6.348*** (3.497, 11.523)	3.970*** (2.377, 6.630)		1.362* (0.983, 1.885)		
8. Demographics, geography, education, not salaried	4.918*** (2.652, 9.120)	3.346*** (1.976, 5.666)			1.772*** (1.222, 2.571)	
9. Demographics, geography, education., BLS injury rate	5.156*** (2.773, 9.590)	3.504*** (2.072, 5.928)				1.272** (1.057, 1.530)
10. Demographics, geography, education, work hours, overtime	6.068*** (3.344, 11.010)	3.883*** (2.328, 6.473)	0.784* (0.607, 1.012)	1.379* (0.996, 1.909)		
11. Demographics, geography, education, overtime, salaried	4.891*** (2.639, 9.067)	3.294*** (1.945, 5.579)		1.314 (0.949, 1.820)	1.737*** (1.197, 2.520)	
12. Demographics, geography, education, work hours, BLS injury rate	4.905*** (2.638, 9.119)	3.416*** (2.022, 5.769)	0.782* (0.607, 1.008)			1.280*** (1.063, 1.541)
13. Demographics, geography, education, overtime, BLS injury rate	5.103*** (2.745, 9.485)	3.342*** (2.082, 5.807)		1.341* (0.969, 1.858)		1.264** (1.050, 1.522)
14. Demographics, geography, education, salaried, BLS injury rate	5.156*** (2.314, 8.186)	3.504*** (1.809, 5.270)			1.627** (0.984, 1.334)	1.192* (0.984, 1.445)

Note: All regressions also included these covariates: male, African-American, Hispanic, all other races, age categories, married, rural residence, and regions.
95% confidence interval is in the parenthesis. Sample size is 17,190.
*Indicates $0.05 \leq P < 0.1$.
**Indicates $0.01 \leq P < 0.05$.
***Indicates $P < 0.01$.

more likely to report claims than non-Hispanic whites. Findings pertaining to job characteristics included these: employer-provided health insurance and low wages were not consistently statistically significant predictors; the ''not salaried'' variable was a strong and consistent predictor; and the predictive capacity of the overtime and BLS injury rate variables were muted by confounding from the ''not salaried'' variable. All job variables became less significant after the education variables were included. A ranking by odds ratios revealed that among variables of greatest concern in this study, the following were the covariates that were statistically significant across numerous regression specifications: education less than high school, high school graduate, African-American, male, ''not salaried,'' and BLS injury rate.

Our study can be compared to the incidence of workers compensation cases per 100 employed persons in recent estimates. Bonauto et al. [2010] analyzed over 31,000 responses from employed persons across 10 states who participated in the 2010 Behavioral Risk Factor Surveillance System (BRFSS) survey. Individuals were asked whether they experienced a work-related injury that required medical advice or attention during the past 12 months and whether the medical costs were paid by workers compensation insurance. By multiplying the rate per-100 employed persons with the percentage paid by workers compensation, they derived a workers compensation incidence rate. The highest rate was 3.45 from New York and the lowest was 2.52 from Massachusetts. But the Bonauto et al. [2010] study did not distinguish between ''medical only'' and indemnity claims; these 2.52 to 3.45 figures therefore included ''medical only'' as well as indemnity cases that paid workers wage-replacement benefits. In aggregate, across all states, approximately 77% of workers compensation cases were estimated to be ''medical only'' [Sengupta et al., 2009]. But simply multiplying (1–0.77) with these figures would not be accurate because it is likely that respondents recall bias would be greater for the less severe injuries in the for ''medical only'' category than in the more severe indemnity category. If we assume that one-half of reported cases were ''medical only'' claims and the other half were indemnity claims, and then multiply the 2.52 and 3.45 figures by 0.5, we generate rates of 1.26–1.73. These 1.26–1.73 rates compare favorably to our 1.34 estimate in Table I.

A literature search uncovered three recent studies and two older ones that used large national data sets on individuals to predict the incidence of workers compensation benefits. A recent study by Lakdawalla et al. [2007] analyzed data from the National Longitudinal Survey of Youth (NLSY). The NLSY contains a nationally representative sample of persons aged 14–22 years old in 1979. Lakdawalla et al. [2007] used data from 1988 to 1998 in which the mean age was 32. Lakdawalla et al. [2007] were concerned with estimating correlations between whether employers paid for private health insurance for employees and whether any of those same workers applied for and received benefits from workers compensation. They only provided univariate information on covariates of most concern to us. Their multivariable analysis used the fixed effects technique that sweeps out time-invariant characteristics such as gender and race. An advantage of their study was that the NLSY asked respondents who reported on-the-job injuries whether they or their employer filed for workers compensation claims and whether benefits were received from those claims. Walters et al. [2010], a second recent study, looked at the occupational injuries for people younger than 25 years old in Oregon using workers' compensation claims data during 2000–2007. A third recent study by Luckhaupt and Calvert [2010] used older data—the Occupational Health Supplement 1988 National Health Interview Survey—and reported the prevalence of work related injuries.

The most widely cited older study used data from the March Current Population Surveys for 1983–1985 (n = 19,082 person-years) [Krueger, 1990]. Krueger's design was similar to ours: base year characteristics from 1983 to 1984 were used to predict subsequent participation in workers compensation for 1984 and 1985. Whereas his emphasis was on economic variables reflecting the level of benefits and legal restrictions, he nevertheless included demographic variables as covariates. A second older study, by Leigh [1986], used PSID data (n = 4,962) in a similar design with 1978 treated as the base year and 1979 as the subsequent year.

There are numerous studies involving related research. The BLS produces myriad reports and raw data tables for non-fatal cases in its Survey of Occupational Injuries and Illnesses (SOII), 2009c [Hoskin, 2005; Rogers and Wiatrowski, 2005; Pierce, 2008]. The SOII is independent from workers compensation systems, but there is overlap between the two. Boden and Ozonoff [2008] linked data reported to workers compensation systems with data from private firms reporting to the BLS within six states for 1998–2002. Only BLS-SOII data with workdays lost (at least 3–7 days depending on the state) were matched to workers compensation indemnity cases. Most cases reported to BLS were also reported to workers compensation systems. Percentages ranged from a low of 65% in Minnesota to a high of 95% in Washington. Nestoriak and Pierce [2009] also found significant overlap. In addition, there are studies that use nationally representative data and regression models on respondents' answers to survey questions regarding job-related injury and illness [Oh and Shin, 2003; Dembe et al., 2004, 2005; Strong and Zimmerman, 2005]. But caution should be exercised when comparing our results with BLS-SOII and those on survey respondents. The Boden and Ozonoff [2008] percentages strictly only apply to cases with at least 3–7 days of work loss, not ‘‘medical only’’ workers compensation cases or BLS-SOII cases involving only 1–2 days of work loss. In addition, most BLS reports typically only use univariate correlations or cross-tabulations. Regarding studies with survey data, as the Bonauto et al. [2010] study demonstrates, sizeable percentages of respondents work injuries are not reported to workers compensation. Even though the Lakdawalla et al. [2007] study is limited to relatively young persons, in their Table II, depending on the year, from 49% to 63% of respondents who reported a work-related injury also reported filing a workers compensation claim.

Gender differences are addressed in the literature. Our results that male is a strong predictor is broadly consistent with the literature. In Lakdawalla et al. [2007] and Krueger’s [1990] univariate analyses, male was a strongly statistically significant and positive predictor. In multivariable analysis accounting for other demographic variables and job characteristics, Krueger [1990] did not find male to be a significant predictor whereas Leigh [1986] and Walters et al. [2010] did. Turning to studies on reporting injuries as opposed to workers compensation claims, Pierce [2008], using BLS data and univariate methods, found men roughly 37% more likely than women to report a non-fatal case, whereas Oh and Shin, using the 1988 Occupational Health Supplement to the National Health Interview Survey, and using multiple regressions, found men roughly 28% more likely than women to report an injury. For non-fatal injuries resulting in days away from work, the BLS reported men with 29% greater chances than women [Bureau of Labor Statistics, 2009c]. These 37%, 28%, and 29% were smaller than our estimated odds ratios of 78% with univariate methods and 81% with multiple regressions; the discrepancies may be due to differences in reporting injuries versus reporting workers compensation claims. Finally, it is noteworthy that among persons who qualified for unemployment compensation benefits, Blank and Card [1991] found men were more likely than women to receive those benefits.

We found non-Hispanic African-Americans with 76–167% statistically significant greater chances of reporting a workers compensation claim than non-Hispanic whites. We found that the Hispanic variable was statistically significant at the 5% level only in univariate and some multivariable analyses. For most multivariable analyses, especially those that combined the education variables with others, Hispanic was not statistically significant at the 5%. Findings in the literature are conflicting. In studies for which dependent variables reflected workers compensation claims: Lakdawalla et al. [2007] found blacks to have slightly lower rates than whites; Krueger [1990] did not find any statistical significance for African-American or Hispanic; and Leigh [1986] did not find any significance for the only race/ethnicity variable he included—non-white. Dong et al. [2007] found that Hispanics were more likely to report medical conditions caused by work-related injuries, but they were less likely to report workers compensation payments. Findings also varied from studies in which dependent variables were job injuries. Dembe et al. [2004], Oh and Shin [2003], and Luckhaupt and Calvert [2010] did not find any statistically significant differences for any race or ethnicity. Robinson [1989] found both non-Hispanic blacks and Hispanics had higher risk of occupational injury than non-Hispanic whites in a large sample from California. Friedman and Forst [2008] using state injury registry data in 1997–2003 from Illinois, found higher incidence rates for Hispanics than non-Hispanic whites but a lower rate for non-Hispanic African-Americans than non-Hispanic whites. But all of these data were old. The Dembe et al. data were from 1996 to 1998, and the Oh and Shin [2003] and Luckhaupt and Calvert [2010] data were from 1988; patterns change over time [Berdahl, 2008]. Economic conditions such as rising long-term unemployment, outsourcing dangerous work to small firms that frequently go out-of-business, and declines in employer-provided medical insurance are more prevalent in the 2000s than in the 1980s and 1990s. We therefore conclude our findings that being African-American was predictive but being Hispanic does not reflect current economic and employment conditions. We also conclude that most of the Hispanic and non-Hispanic white disparity was explained by varying levels of education between those two groups.

Our results for age group 55–64 being a negative predictor has at least two explanations. First, assuming some correlation between reported injuries and claims, it is likely that older workers with greater experience are better able to avoid injuries than younger, inexperienced workers [Laflamme and Menckel, 1995]. Second, our results are consistent with the hypothesis that there is cost-shifting from workers compensation systems to Social Security Disability Insurance (SSDI), because this age group is by far the most prevalent among persons receiving SSDI [Guo and Burton, 2010].

Various measures of lower socioeconomic status have been found to be positive and statistically significant predictors. For dependent variables measuring workers compensation cases, both Krueger [1990] and Leigh [1986] found employment in blue-collar jobs was predictive. Lakdawalla et al. [2007] found college graduates with reported work injuries were less likely than high school dropouts to file claims. Krueger [1990] found that lower levels of education were strongly predictive of claim filing. Leigh [1986] found education to be predictive only in analyses that excluded measures of occupation and injury rates. For dependent variables measuring respondents’ reports of work injuries, low education levels and blue-collar employment as operative, laborer, or production worker were positive predictors [Oh and Shin, 2003; Dembe et al., 2004; Strong and Zimmerman, 2005; Luckhaupt and Calvert, 2010]. Our findings for the two education variables were consistent with these previous findings. Moreover, our odds ratios were large, ranging from 3.3 to 6.5 depending on which other covariates were accounted for. Finally, it is worth noting that these education odds ratios were roughly two to three times the size of odds ratios for either male or African-American.

But results on socioeconomic status as measured by low wages were conflicting in the literature. Krueger [1990] only indirectly measured wages as the ratio of indemnity benefits in the state divided by the workers preinjury wage. He found strong and positive correlations reporting workers compensation claims. Holding indemnity benefits constant, this result suggested that increases in wages resulted in decreased reporting of workers compensation claims. Leigh [1986] did not find any correlation between wage levels and reports of workers compensation. Turning to survey data on reports of work injuries, Berdahl and McQuillan [2008], using the 1988 National Longitudinal Survey of Youth (NLSY) and multiple regressions, found that hourly wages were not statistically significant predictors of workplace injury. Strong and Zimmerman [2005] also used the NLSY for 1988–2000 and found that for men, higher wages were statistically significant and were associated with fewer injuries; they found no statistical associations for women.

These conflicting results and findings could be explained by the conflicting theories regarding associations between wages on the one hand and either reports of workers compensation or workplace injury on the other. As a measure of socioeconomic status, we expected a negative association: the lower the wage, the greater the chance of either workers compensation claims or injuries [Robinson, 1989]. But the "compensating wage" hypothesis from economics suggests a positive association with chances of workplace injury [Kaufman and Hotchkiss, 2005] but an ambiguous association with reports of workers compensation [Arnould and Nichols, 1983]. This hypothesis states that "the market" generates a positive association because well-informed workers will only take dangerous jobs if they receive some "extra" compensation—a compensating wage—for doing so. But depending on the size and probability of receiving workers compensation benefits in the event of an injury, the size and even the existence of the compensating wage is in doubt [Arnould and Nichols, 1983]. The conflicting results from the literature and predictions from conflicting theories are consistent with our statistically insignificant findings for the low wage variable.

A separate indicator of socioeconomic status was whether the worker had employer-provided health insurance. Again, as a measure of socioeconomic status, we expected a negative correlation. But there are other hypotheses. Workers with employer-provided health insurance may prefer not to become entangled in a contentious battle with an employer or workers compensation insurance company attempting to prove that the injury occurred at work. Workers who become injured off the job and who do not have employer-provided health insurance may nevertheless seek coverage from workers compensation [Card and McCall, 1996]. For both of the reasons above, we would expect a negative correlation. Lipscomb et al. [2009] analyzed data on claims to private health insurers among union carpenters over 15 years in the state of Washington. They found evidence for cost-shifting from workers compensation claims to private insurance claims. This cost-shifting is consistent with negative correlations between employer-provided health insurance and workers compensation claims [Lipscomb et al., 2009]. Card and McCall [1996] tested whether uninsured workers nevertheless gained workers compensation coverage for injuries sustained off the job. Card and McCall [1996] did not find evidence for this effect, however. Lakdawalla et al. [2007], on the other hand, found a strong positive association and attributed it to employer behavior that helps employees file claims. Neither Krueger [1990] nor Leigh [1986] considered employer-provided health insurance. Our statistically insignificant results were consistent with those of Card and McCall [1996] and generally reflect the conflicting results in the literature.

A different finding in the literature is that working overtime was associated with reporting workers compensation claims and workplace injuries. [Leigh, 1986; Dembe et al., 2005]. These findings are consistent with the hypothesis that fatigue plays a role in increasing the chances of experiencing a workplace injury. But overtime wages are generally not paid to salaried personnel [Dembe et al., 2005]. Our "not salaried" variable proved to be valuable. Our results suggested that overtime was statistically significant at 5% only when education and "not salaried" were excluded; it was significant at the 10% level in all regressions, except for regressions with "not salaried". In virtually all regressions, overtime odds ratios were in the 1.3–1.5 range. We conclude that any future analysis that attempts to estimate the effects of overtime on reporting workers compensation claims needs to account for confounding from whether workers are salaried.

Several hypothesis surround total annual work hours. Workers who work full-time, by definition, have more job-risk exposure than workers who work part-time. There is evidence that longer hours per day or per week are associated with reporting work injuries [Dembe et al., 2005]. But, again, reporting injuries is not the same as reporting workers compensation claims. Part-time workers may view workers compensation differently than full-time workers. The only study with which we are familiar that used work hours to predict filing claims did not find it to be a statistically significant predictor [Leigh, 1986]. Finally, work hours may be correlated with whether or not the person holds a salaried job. We found that annual work hours were frequently statistically significant at the 10% level in multiple regressions but never at the 5% level as long as either the education or ''not salaried'' variables were included as covariates.

Finally, our results on the BLS injury rate variable suggested that persons holding jobs with high injury rates were more likely to report workers compensation claims.

Strengths and Limitations

There are strengths for our data and method. First, the PSID is a valuable data set. It is longitudinal, enabling us to measure job related variables in years preceding possible reports on workers compensation. The PSID sample size is large, nationally representative, and contains detailed demographic and socioeconomic variables. The PSID is highly regarded and widely used, especially by social scientists. From its inception to 2009, roughly 332 papers had been published under the broad category of ''health'' [Panel Study of Income Dynamics, 2009]. Moreover, numerous authors have successfully analyzed the same PSID workers compensation variable we used [Schoeni, 1997; Charles, 2003; Ziliak, 2003]. Second, our data are from 1997 to 2005, significantly more recent than data from 1980 to 1998, which are most common in the literature. Third, as indicated above, our results are consistent with literature demonstrating links between gender and socioeconomic status on the one hand and reporting workers compensation claims on the other. The fact that many of our results are consistent with extant literature augers well for the credibility of our PSID samples and method.

There are also limitations. First, our workers compensation variable is self-reported. Fraud undoubtedly exists. But we are aware of no scientific studies of workers compensation fraud committed by workers. If workers compensation is similar to unemployment compensation, then perhaps only 2% of claims might be fraudulent [US General Accounting Office, 2002]. A different limitation is that our dependent variable measured reports of workers compensation indemnity benefits, not ''medical only'' claims, and the latter comprise roughly 77% of all claims [Sengupta et al., 2009]. We are not aware of any studies, however, that address the demographic or job predictors of solely ''medical only'' claims or both ''medical only'' and indemnity claims. A related limitation is that even though recall bias may be less for indemnity claims than ''medical only'' claims, there are still likely some respondent errors in memory for indemnity claims.

Our overtime variable is limited. It does not measure frequency. Moreover, because of legal definitions, our overtime and workers compensation correlations were confounded by the ''not salaried'' variable.

An additional limitation is that we might be underestimating the effects of wages and low socioeconomic status. Attrition might bias our findings. It is likely that some persons in low wage jobs and with low socioeconomic status died over these years since low income and status are associated with premature death [McDonough et al., 1997; Duleep, 1986]. Had there been no premature death, the number of respondents in low wage jobs and with low status would have increased and the estimated incidence of reporting of workers compensation benefits within those low wage jobs would have been higher. Another problem is that some of our covariates—education, BLS injury rate, and wages—may be correlated with unobserved ''third variables'' such as time preference. But our random effects model is designed to control for possible spurious correlations that might arise from unobserved individual characteristics such as time preference that does not change for the same person over the 8 years of the data, 1997–2005.

Whereas some studies have considered more than six job characteristics, we felt limited, given the relatively few persons reporting workers compensation cases in any given year. We nevertheless picked six job characteristics that have received considerable attention in the literature: wages, employer-provided insurance, work hours, overtime, salaried, and BLS injury rate.

Because we combined data across several years and because we used logistic regression, we were unable to adjust for geographic clusters for the combined years of data. On the other hand, we were able to adjust for geographic clusters with single wave cross-sectional data and four separate logistic regressions for 1999, 2001, 2003, and 2005. Most statistically significant results on demographic categories remained, as did results on most job characteristics for 3 years. We prefer our reported results for all years combined, again, because of the relative infrequency of persons reporting workers compensation claims in any 1 year.

A final limitation is that the types of injury or illness (musculoskeletal conditions, fractures, cut, burns, etc.) and work-related conditions (lifting, carrying, standing, etc.) were not collected by PSID. Our study cannot examine the correlations between demographic groups and job characteristics on the one hand and types of injuries or work-related conditions on the other.

CONCLUSION

Each year, millions of workers file for workers compensation claims. Few studies address socio-demographic characteristics and job traits as predictors of workers compensation claims. This is unfortunate since with greater knowledge we would be better able to assess the incidence across the population, and better able to tailor interventions to targeted groups.

ACKNOWLEDGMENTS

JP Leigh acknowledges partial financial support from the National Institute for Occupational Safety and Health (OH008248-01).

REFERENCES

Arnould RJ, Nichols LM. 1983. Wage-risk premiums and workers' compensation: A refinement of estimates of compensating wage differential. J Polit Econ 91(2):332–340.

Berdahl TA. 2008. Racial/ethnic and gender differences in individual workplace injury risk trajectories: 1988–1998. Am J Public Health 98(12):2258–2263.

Berdahl TA, McQuillan J. 2008. Occupational racial composition and nonfatal work injuries. Soc Probl 55(4):549–572.

Blank RM, Card D. 1991. Recent trends in insured and uninsured unemployment: Is there an explanation? Quart J Econ 106(4):1157–1189.

Boden LI, Ozonoff A. 2008. Capture–recapture estimates of nonfatal workplace injuries and illnesses. Ann Epidemiol 18(6):500–506.

Bonauto DK, Fan JZ, Largo TW, Rosenman KD, Green MK, Walters JK. 2010. Proportion of workers who were work-injured and payment by workers' compensation systems—10 States, 2010. Morb Mort Weekly Rep 59(29):897–900.

Bureau of Labor Statistics. 2009a. http://www.bls.gov/iif/oshwc/osh/os/osh06_appa.pdf (accessed June 30, 2009).

Bureau of Labor Statistics. 2009b. Consumer Price Index for all urban consumers. http://www.bls.gov/bls/inflation.htm (accessed November 20, 2009).

Bureau of Labor Statistics. 2009c. http://www.bls.gov/news.release/archives/osh2_12042009.pdf (accessed October 7, 2010).

Butler RJ, Worrall JD. 1983. Workers' compensation: Benefit and injury claims rates in the seventies. Rev Econ Stat 65(4):580–589.

Card D, McCall BP. 1996. Is workers' compensation covering uninsured medical costs: Evidence from the "Monday Effect". Ind Labor Relat Rev 49(4):690–706.

Charles KK. 2003. The longitudinal structure of earnings losses among work-limited disabled workers. J Hum Resour 38(3):618–646.

Dembe AE, Erickson JB, Delbos R. 2004. Predictors of work-related injuries and illnesses: National Survey Findings. Occup Environ Hyg 1(8):542–550.

Dembe AE, Erickson JB, Delbos RG, Banks SM. 2005. The impact of overtime and long work hours on occupational injuries and illnesses: New evidence from the United States. Occup Environ Med 62(9):588–597.

Dong XW, Ringer K, Men YR, Fujimoto A. 2007. Medical costs and sources of payment for work-related injuries among Hispanic construction workers. J Occup Environ Med 49(12):1367–1375.

Duleep HO. 1986. Measuring the effect of income on adult mortality using longitudinal administrative record data. J Hum Resour 21(2):238–251.

Friedman LS, Forst L. 2008. Ethnic disparities in traumatic occupational injury. J Occup Environ Med 50(3):350–357.

Guo X, Burton JF. 2010. Workers compensation: Recent developments in moral hazard and benefit payments. Ind Labor Relat Rev 63(2):340–355.

Hoskin AB. 2005. Occupational injuries, illnesses, and fatalities among women. Mon Labor Rev 128(10):31–37.

Kaufman BE, Hotchkiss J. 2005. The economics of labor markets. 7th edition. New York: Thompson-Southwest.

Kilbourne AM, Switzer G, Hyman K, Crowley-Matoka M, Fine MJ. 2006. Advancing health disparities research within the health care system: A conceptual framework. Am J Public Health 96(12):2113–2121.

Krueger AB. 1990. Incentive effects of workers' compensation. J Public Econ 41(1):73–99.

Laflamme L, Menckel E. 1995. Aging and occupational accidents – A review of the literature of the last three decades. Safety Sci 21(2):145–161.

Lakdawalla DN, Reville RT, Seabury SA. 2007. How does health insurance affect workers' compensation filing? Econ Inq 45(2):286–303.

Leigh JP. 1986. Individual and job characteristics as predictors of industrial accidents. Accident Anal Prev 18(3):209–216.

Lipscomb HJ, Dement JM, Silverstein B, Cameron W, Glazner JE. 2009. Who is paying the bills? Health care costs for musculoskeletal back disorders, Washington state union carpenters, 1989–2003. J Occup Environ Med 51(10):1185–1192.

Luckhaupt SE, Calvert GM. 2010. Work-relatedness of selected chronic medical conditions and workers' compensation utilization: National health interview survey occupational health supplement data. Am J Ind Med 53(12):1252–1263.

McDonough P, Duncan GJ, Williams D, House J. 1997. Income dynamics and adult mortality in the United States, 1972 through 1989. Am J Public Health 87(9):1476–1483.

Nestoriak N, Pierce B. 2009. Comparing workers compensation claims with establishments' responses to the SOII. Mon Labor Rev 132(5):57–64.

Oh JH, Shin EH. 2003. Inequalities in nonfatal work injury: The significance of race, human capital, and occupations. Soc Sci Med 57(11):2173–2182.

Panel Study of Income Dynamics. PSID bibliography http://psidonline.isr.umich.edu/Publications/Bibliography/BrowseKeywords.aspxcome (accessed November 21, 2009).

Pierce B. Methods underlying new workplace injury and illness rates by demographic group. *Compensation and Working Conditions*, on-line journal, US Bureau of Labor Statistics, published 3/26/ 2008, http://www.bls.gov/opub/cwc/sh20080320ar01p1.htm (accessed October 12, 2009).

Robinson JC. 1989. Exposure to occupational hazards among Hispanics, Blacks, and Non-Hispanic whites in California. Am J Public Health 79:629–630.

Rogers E, Wiatrowski WJ. 2005. Injuries, illnesses and fatalities among older workers. Mon Labor Rev 128(10):24–30.

Schoeni RF. 1997. Private interhousehold transfers of money and time: New empirical evidence. Rev Income Wealth 4:423–448.

Sengupta I, Reno V, Burton JF. 2009. Workers compensation: Benefits, coverage, and costs 2007. Washington, DC: National Academy of Social Insurance. http://www.nasi.org/usr_doc/Workers_Comp_Report_2007.pdf (accessed October 28, 2009).

StataCorp. 2005. Stata User's Guide College Station, TX: STATA Press.

Strong LL, Zimmerman FJ. 2005. Occupational injury and absence from work among African American, Hispanic, and Non-Hispanic white workers in the National Longitudinal Survey of Youth. Am J Public Health 95(7):1226–1232.

US General Accounting Office. 2002. Unemployment insurance: Increased focus on program integrity could reduce billions in overpayments (GA0-02-697). Washington, DC: US Government Printing Office.

Walters JK, Christensen KA, Green MK, Karam LE, Kincl LD. 2010. Occupational injuries to oregano workers 24 years and younger: An analysis of workers' compensation claims, 2000–2007. Am J Ind Med 53(10):984–994.

Ziliak JP. 2003. Income transfers and assets of the poor. Rev Econ Stat 85(1):63–76.