Human and Ecological Risk Assessment

ISSN: 1080-7039 (Print) 1549-7860 (Online) Journal homepage: https://www.tandfonline.com/loi/bher20

# Relative Risk of Involuntary Injuries Among Currently Employed Adults in the U.S.

Jai Won Choi , Ki Moon Bang & Robert A. Wright

Human and Ecological Risk Assessment: Vol. 4, No. 6, pp. 1353-1370 (1998)

# Relative Risk of Involuntary Injuries Among Currently Employed Adults in the U.S.

**Jai Won Choi,[1,*] Ki Moon Bang,[2,**] and Robert A. Wright[1,*]**
[1]National Center for Health Statistics, Hyattsville, MD; [2]National Institute for Occupational Safety and Health, Morgantown, WV

**ABSTRACT**

It can be very informative to compare the risk of injury for different occupations, places where accidents happen, demographic variables, and other characteristics. However, most epidemiologic literature shows that multiple logistic regression or odds ratio were commonly used to compare the relationship between the variables on two or more levels. A multiple logistic regression could be used for multiple comparisons assuming that the model correctly reflects the situation at hand. But it might not be a true assumption. Assuming no model, we defined the relative risk to compare risks of injury for different occupations, and applied this method to the data collected by the National Center for Health Statistics (NCHS) National Health Interview Survey (NHIS). This application showed that the relative risk of injury varied by occupation and by place of accident, sex, race, and age.

**Key Words:** occupational injuries, model independent analysis, sample estimation, test of hypothesis

## INTRODUCTION

Unintentional injury occurs in a wide variety of places and occupations. Injury is a serious health and economic problem in industry. Injury accounted for 19% of all Worker's Compensation claims in the United States in 1979 (Klein, Jensen, and Senderson, 1984; Rowe, 1983), and Worker's Compensation claims have been increasing continually in recent years (Frank, Pulcins, Kerr, Shannon, and Stansfeld, 1995).

---

* National Center for Health Statistics, 6525 Belcrest Road, Hyattsville, MD 20782
** National Institute for Occupational Safety and Health, Morgantown, WV 26505

A total of 6.6 million injuries and illnesses were reported in workplaces in private industry during 1995 (U.S. BLS, 1997). The estimated cost resulting from injury was approximately $145 billion in the United States (Leigh and Landrigan, 1997).

The National Council on Compensation Insurance estimated that expenditures for Worker's Compensation were increasing every year (Rowe, 1983; Leigh and Landrigan, 1997). Injuries are one of the leading causes of lost work days in the United States (Loser and Volinn, 1991; NCHS, 1995). Leavitt (1992) found that heavy physical work was the most frequent cause of injuries among currently employed.

The plans for the prevention of occupational injuries have to be based on data that illuminate the nature and magnitude of these injuries. Injury epidemiology is an important part of public health concerns and became one of the focus areas in the healthy people 2000 objectives (USDHHS, 1996). In order to prevent these costly injuries, one step is to obtain as much information on occupational injuries as possible. The more we know about them, the better we can prevent them. To this end, we developed a simple method for comparing the relative risk of injury for different occupations, places of work, and other characteristics.

To present the concept and applications of relative risk, we introduced a sampling and classification of occupations in the 2nd section. In the 3rd section, we define the relative risk of involuntary injury, and present an estimator. Its variance and testing of significance are presented in the Appendix. In the 4th section, this method was applied to the NCHS data, and the risks of injury are shown in the ensuing tables. Finally, we have some comments in Section 5.

## SAMPLE

The National Health Interview Survey (NHIS) is a continuous, cross-sectional survey of the U.S. civilian non-institutionalized population (NCHS, 1995). It is conducted annually by the National Center for Health Statistics (NCHS), part of the Centers for Disease Control and Prevention. In 1994, the basic health and demographic questionnaire included questions about the unintentional injuries and other medical conditions of the persons living in the sample households. This survey provided the data for this study. Basically the segments were selected through probability proportional to the population size, and persons in each segment were selected with equal probability. The details of sampling procedures, including the design and the questionnaire, are described in detail elsewhere (NCHS, 1995; U.S. Bureau of the Census, 1978).

The 1994 National Health Interview Survey sample included about 116,000 persons, of which about 53,300 persons were 18 years of age or older and currently employed. Only 3,200 of these 53,000 persons reported an unintentional injury. Each person was counted only once if the person had more than one injury in the same accident. Each sample person was weighted to represent the population from which this person was selected. An estimated 7.41 million workers among 122.9 million currently employed (6.03%) were injured in 1994.

All the variables in the study are based on the family information obtained from one adult sampled from each household. Occupation was coded according to the 1980 Classified Industries and Occupations of the U.S. Bureau of the Census (U.S. Bureau of the Census, 1980). For this study, seven occupation groups were used:

1. ***Managerial*** and professional specialty occupations combining, "Executive, administrative and managerial occupations" and "Professional specialty occupations"

2. ***Technical***, sales, and administrative support occupations combining "Technicians and related support occupations," "Sales occupations," "Administrative support occupations, including clerical."

3. ***Service*** occupations combining "Private household occupations," "Protective service occupations," and " Service occupations, except protective and household."

4. ***Farming***, forestry, and fishing occupations

5. ***Precision*** production, craft and repair occupations

6. ***Operators***, fabricators and laborers combining, "Machine operators, assemblers and inspectors," "Transportation and material moving occupations," and "Handlers, equipment cleaners, helpers and laborers."

7. ***Unknown*** (includes never worked, refused, classified, etc.)

The persons injured were first divided into two groups, those injured while working and those injured while not working. Nine places of injury are included in the NHIS, regardless working or not working:

1. at ***home inside*** a house,

2. at ***home adjacent*** premises,

3. ***street and highway*** including roadway and public sidewalks,

4. ***farm***,

5. ***industrial*** place including premises,

6. ***school*** including premises,

7. ***place of recreation*** and sports, except school,

8. ***other***,

9. ***unknown***.

For completeness, a tenth "place," ***not injured***, was included.

## METHOD

For this study, the relative risk is defined as the ratio of the percent of persons injured in one occupation group to that of other occupation groups. We use the parameter $R_{o/r}$ for the relative risk in the population, and $R_{o/r}$ compares the percent of the population in occupation "o"who are injured to that of the reference occupation "r" when 0 = 1, . . . ,0:

$$R_{o/r} = \frac{\text{percent of persons in the population injured in occupation o}}{\text{percent of persons in the population injured in reference occupation r}}$$

The parameter $R_{o/r}$ is estimated by a sample estimator $r_{o/r}$:

$$r_{o/r} = \frac{\text{percent of persons in the sample injured in occupation o}}{\text{percent of persons in the sample injured in reference occupation r}}$$

We chose one of the occupations as the reference group "r", depending on the needs of the analysis. The percent of persons injured in the reference occupation group were compared to the percent of persons injured in another occupation group "o". There are the same number of relative risks as the number of occupations to be compared. The relative risk of reference occupation is 1, when the percent of reference occupation was divided by itself. The relative risks of the other occupations would fluctuate around 1. If the $r_{o/r}$ of one specific occupation is larger than 1, the persons in that occupation would be more likely to be injured than those in the reference occupation. On the other hand, if the $r_{o/r}$ is smaller than 1, they would be less likely to be injured.

For example, assume 40% of persons from the *manufacturing* occupation group were reported injured, while 10% of persons from the *sales* occupation group were reported injured. If both occurred at an *industrial* place, and the *sales* occupation group was selected as the reference occupation, the relative risk for the two occupation groups was 4 at an *industrial* place. There were 4 injuries per 100 persons in the *manufacturing* occupation group for every 1 injury per 100 persons in the *sales* occupation group, at an *industrial* place. The relative injury risk of injury for *manufacturing* workers was four times that of *sales* workers.

## VARIANCE

The NHIS sample design is complex, the data are collected in houshold interviews, and the estimates are subject to a variety of sampling and non-sampling errors. The actual sampling design and estimation processes must be considered to obtain the variance of the relative risk based on such data. The Appendix shows the calculation of the variance of the estimator $r_{o/r}$ derived under certain assumptions.

There are different approaches which can be used to derive the variance, including resampling methods and model-based methods. The variance derived in the Appendix is based on a design method along with correlation models and the Taylor expansion.

## TESTS OF HYPOTHESIS

We can standardize the relative risk under a null hypothesis with the variance obtained by SUDAAN (Shah, Barnwell, Bieler, 1996) or by the method suggested in the Appendix.

The standardization allows us to assume a normal distribution of the test statistic to test the null hypothesis. Alternatively, an interval estimation could be used to see if the value in the null hypothesis belongs to such an interval. Interval estimation is especially useful when the sample size is too big or too small.

When the relative risks were widely scattered or skewed to one side, or when variables are too large or too small, we can stabilize them by taking the log. The log numbers are often easier to handle than the original numbers, as discussed in Appendix.

We set up a null hypothesis that the log of the relative risk equals the log of the relative risk under the null hypothesis. The variance of the log relative risk is the same as that of original relative risk except R is replaced by log R. The result of both approaches, original relative risk and log of relative risk, is the same. If one test suggests accepting or rejecting the null hypothesis, the other test would suggest the same.

## RELATIVE RISK (RR) VS. ODDS RATIO (OR)

An OR differs from a RR, but they often are very close. The OR compares the odds of two events in a $2 \times 2$ table. Suppose that A is the attribute for two rows, A and not A and B is the attribute for two columns, B and not B. The not A and not B are the complements of A and B, respectively. OR is the ratio of two odds, $p(B|A)/p(\text{not } B|A)$ over $p(B|\text{not } A)/p(\text{not } B|\text{not } A)$.

The RR compares any two probabilities or percents in a K by L matrix ($K, L \geq 2$). It is not restricted to a $2 \times 2$ table. First we calculate the percent in each column, and then compare two percents from different rows as seen in Tables 1 and 1a. The $RR = p(B|A)/p(B|\text{not } A)$ for the $2 \times 2$ table. We use the $2 \times 2$ table to find similarities and differences between OR and RR.

**Table 1. Number of currently employed persons in thousand row-wise, percent injured by occupational groups working or not (standard errors), United States, 1994.**

| Occupation group | No. of employed and injured persons (in thousands) | Percent injured while working | Percent injured while not working |
|---|---|---|---|
| Managerial | 34,940 (446) | 1.735 (0.11) | 3.96 (0.17) |
| Technical | 36,312 (416) | 1.743 (0.11) | 3.64 (0.15) |
| Service | 15,992 (272) | 2.80 (0.22) | 3.69 (0.25) |
| Farming | 3,052 (117) | 2.46 (0.39) | 2.60 (0.43) |
| Precision | 12,549 (223) | 4.09 (0.28) | 3.45 (0.28) |
| Operators | 16,794 (296) | 4.29 (0.25) | 3.06 (0.20) |
| Unknown | 3,274 (157) | 1.03 (0.26) | 2.15 (0.38) |
| Total | 122,912 (989) | 2.46 (0.07) | 3.57 (0.09) |

**Table 1a. Column-wise relative risk among workers by occupational group groups working or not, and (standard errors), United States, 1994.**

| Occupation group | Relative risk while working | Relative risk while not working |
|---|---|---|
| Managerial[a] | 1.00 (0.11) | 1.00 (0.17) |
| Technical | 1.01 (0.16) | 0.92 (0.23) |
| Service | 1.61 (0.25) | 0.93 (0.30) |
| Farming | 1.42 (0.41) | 0.66 (0.46) |
| Precision | 2.36 (0.30) | 0.87 (0.33) |
| Operators | 2.47 (0.27) | 0.77 (0.26) |
| Unknown | 0.59 (0.28) | 0.54 (0.42) |

[a] Reference.

Let the $2 \times 2$ table include four cells, $n_{11}$ and $n_{12}$ for A in the first row, and $n_{21}$ and $n_{22}$ for not A in the second row. Then:

$$OR = (n_{11}n_{22})/(n_{12}n_{21}) \quad \text{and} \quad RR = (n_{11}n_{21} + n_{11}n_{22})/(n_{11}n_{21} + n_{12}n_{21})$$

The OR becomes the RR when we add the same number $n_{11}\ n_{21}$ to the numerator and denominator of the OR. If the occurrence of B is unlikely, both the OR and the RR would not differ much, as seen in the Tables 3b, 4b, and 5b.

**Table 2. Percent distribution of persons injured over places of injury for each occupation, and (standard error), United States, 1994.**

| Place of injury | Managerial | Technical | Service | Farming | Precision | Operators |
|---|---|---|---|---|---|---|
| Home inside | 0.56 (0.06) | 0.55 (0.06) | 0.56 (0.09) | 0.30 (0.15) | 0.19 (0.06) | 0.30 (0.07) |
| Home around | 0.57 (0.06) | 0.45 (0.05) | 0.43 (0.07) | 0.13 (0.09) | 0.62 (0.11) | 0.59 (0.08) |
| *Street and highway* | 1.76 (0.11) | 1.83 (0.11) | 1.92 (0.19) | 1.52 (0.35) | 1.80 (0.21) | 1.67 (0.16) |
| *Farm* | 0.04 (0.01) | 0.03 (0.01) | 0.07 (0.04) | 1.42 (0.30) | 0.13 (0.05) | 0.08 (0.03) |
| *Industrial* | 0.89 (0.08) | 1.10 (0.09) | 1.70 (0.16) | 0.75 (0.22) | 3.25 (0.25) | 3.39 (0.22) |
| *School* | 0.35 (0.05) | 0.14 (0.03) | 0.31 (0.09) | 0.06 (0.07) | 0.06 (0.04) | 0.04 (0.03) |
| *Place of recreation* | 0.68 (0.07) | 0.62 (0.07) | 0.47 (0.10) | 0.30 (0.15) | 0.51 (0.09) | 0.51 (0.09) |
| ***Other*** | 0.81 (0.08) | 0.65 (0.08) | 0.99 (1.13) | 0.45 (0.18) | 0.81 (0.12) | 0.77 (0.12) |
| ***Unknown*** | 0.04 (0.02) | 0.01 (0.01) | 0.05 (0.03) | 0.13 (0.09) | 0.06 (0.03) | 0.01 (0.01) |
| ***No injury*** | 94.30 (0.20) | 94.61 (0.19) | 93.51 (0.32) | 94.95 (0.60) | 92.57 (0.37) | 92.65 (0.34) |
| | 100% | 100% | 100% | 100% | 100% | 100% |

**Table 2a. Relative risks of injury for occupations at each place of injury, United States, 1994. managerial occupation used as reference.**

| Place of injury | Managerial[a] | Technical | Service | Farm | Precision | Operators |
|---|---|---|---|---|---|---|
| Home inside | 1.00 | 0.98 | 1.00 | 0.54 | 0.34 | 0.53 |
| Home around | 1.00 | 0.79 | 0.75 | 0.23 | 1.09 | 1.04 |
| *Street and highway* | 1.00 | 1.04 | 1.09 | 0.87 | 1.02 | 0.94 |
| *Farm* | 1.00 | 0.75 | 1.75 | 35.50 | 3.25 | 2.00 |
| *Industrial* | 1.00 | 1.24 | 1.91 | 0.84 | 3.65 | 3.81 |
| *School* | 1.00 | 0.40 | 0.89 | 0.17 | 0.17 | 0.11 |
| *Recreation* place | 1.00 | 0.91 | 0.69 | 0.44 | 0.75 | 0.75 |
| ***Other*** | 1.00 | 0.80 | 1.22 | 0.56 | 1.00 | 0.95 |
| ***Unknown*** | 1.00 | 0.50 | 1.25 | 3.25 | 1.50 | 0.25 |
| ***Not injured*** | 1.00 | 1.00 | 0.99 | 1.01 | 0.98 | 0.98 |

[a] Reference.

**Table 3. Number of currently employed persons in thousands, percent injured groups working or not, by race, and (standard errors), United States, 1994.**

| Race | No. of currently employed persons | Percent injured while working | Percent injured while not working |
|---|---|---|---|
| Black | 13,249 (425) | 1.18 (0.21) | 3.34 (0.26) |
| Non-black | 109,662 (1,004) | 2.54 (0.08) | 3.59 (0.09) |
| All | 122,912 (989) | 2.46 (0.07) | 3.57 (0.09) |

**Table 3a. Relative risk of currently employed persons working and by race.**

| Race | Injury while working | Injury while not working |
|---|---|---|
| Black[a] | 1.00 | 1.00 |
| Non-black | 2.15 | 1.08 |

[a] Reference.

**Table 3b. Relative risk (RR) vs. odds ratio.**

| Race | All injured Percent | All injured RR | Not injured percent | |
|---|---|---|---|---|
| Black[a] | 4.52 | 1.00 | 95.48 | Odds ratio of injury of non-black = $(6.13 \times 95.48) / (4.52 \times 93.87) = 1.38$ |
| Non-black | 6.13 | **1.36** | 93.87 | Relat. risk of non-black injury = 1.36 |

[a] Reference.

The OR = 1 implies that the odds of the occurrence of B is about the same regardless of the presence of A. The RR = 1 implies that the risk of the occurrence of B is about the same regardless of the pressure of A. The RR can be used to compare any two cells provided both percents are based on the respective total.

**Table 4. Number of currently employed persons in thousands, percent injured groups working or not, by age, and (standard errors), United States, 1994.**

| Age | No. of currently employed persons | Percent injured while working | Percent injured while not working |
|---|---|---|---|
| 18–44 | 83,925 (833) | 2.33 (0.09) | 3.46 (0.10) |
| 45+ | 38,987 (413) | 2.75 (0.12) | 3.79 (0.16) |
| All | 122,912 (989) | 2.46 (0.07) | 3.57 (0.09) |

**Table 4a. Relative risk of injured groups working or not and by age.**

| Age | Injured while working | Injured while not working |
|---|---|---|
| 18–44[a] | 1.00 | 1.00 |
| 45+ | 1.18 | 1.10 |

[a] Reference.

**Table 4b. Relative risk vs. odds ratio of injury.**

| | All injuries | | Not injured | |
|---|---|---|---|---|
| Age | % | RR | % | |
| 18-44 | 5.79 | 1.00 | 94.21 | Odds ratio of injury of old = $(6.54 \times 94.21) / (5.79 \times 93.46) = 1.14$ |
| 45+ | 6.54 | 1.13 | 93.46 | Rel. risk of old injury = 1.13 |

## EXAMPLES

We applied the relative risk (RR) of injury to the NHIS data. Table 1 shows the row-wise percent of currently employed persons injured while working or not working by occupation group. Using Table 1, the column wise relative risks in Table 1a were computed. The *managerial* group was the reference group "r" to which the other occupations were compared.

While working, among the six occupation groups, the highest RR was reported by *operators*, and the lowest RR reported by the *managerial* group. For RR = 1 among *managerial* workers, there would be 1.01 risk of injury for *technical* workers, 1.6 risk for *service* workers, 1.4 risk for *farm* workers, 2.36 risk for *precision* workers, and 2.47 risk for *operators*.

**Table 5. Number of currently employed persons in thousand, percent injured groups working or not, by sex and (standard errors), United States, 1994.**

| Sex | No. of currently employed persons | Percent injured while working | Percent injured while not working |
|---|---|---|---|
| Male | 66,941 (566) | 3.14 (0.10) | 3.37 (0.11) |
| Female | 55,971 (516) | 1.66 (0.08) | 3.80 (0.13) |
| All | 122,912 (989) | 2.46 (0.07) | 3.57 (0.09) |

**Table 5a. Relative risk of injury by whether while working or not and by sex.**

| Sex | Injured while working | Injured while not working |
|---|---|---|
| Male[a] | 1.00 | 1.00 |
| Female | 0.53 | 1.13 |

a Reference.

**Table 5b. Relative risk vs. odds ratio of injury.**

| Sex | All injured Percent | All injured RR | Percent not injured | |
|---|---|---|---|---|
| Male[a] | 6.51 | 1.00 | 93.49 | Odds ratio of injury of female = $(5.46 \times 93.49) / (6.51 \times 94.54) = 0.83$ |
| Female | 5.46 | **0.84** | 94.54 | Rel. risk of female injury = 0.84 |

a Reference.

While not working, compared to the RR of 1 for the reference occupation, the RR of the other occupations fell below 1. The highest RR was reported by managerial workers.

Table 2 includes the column-wise percent of persons injured at different places according to whether working or not working. Each column shows the percent for an occupation group, and the column sum is 100%. For example, the *Operators* are divided among the six places of injury, unknown, and not injured. The percent distribution of injuries among operators is 3.39% in industrial places, 1.67% in streets and highways, 0.89% at home (inside and around the household), 0.77% in other places, 0.51% in recreational places, 0.08% at farming, 0.04% at school, 0.01% for unknown, and 92.65% of them were not injured, and similarly for the other occupation groups.

From Table 2, using the *managerial* occupation group as the reference, we calculate the RRs for each row, dividing the percent injured in a row by the percent injured in the reference occupation. The RR of the reference occupation is 1, and others fluctuate about 1. Each row in Table 2a shows the RR of injury for 6 occupations at a particular place of injury.

For injuries on the *street and highway* in Table 2a, when compared to RR = 1 for the *managerial* occupation group, for the *technical* occupation it was 1.04, for the *service* occupation group it was 1.09, for the *farm* occupation group it was 0.87, for the *precision* occupation group it was 1.02, and for the *operator* occupation group it was 0.95. The RRs in *street and highway* was about the same except the *farm workers* with the lower RR score of 0.87. It appears that most of us were exposed to injuries in the *street and highway* with about the same RR except *farm workers*.

The first row included the RRs *inside home* among six occupations. Compared to the *managerial* occupation, the RRs of injury *inside home* were 0.98 for *technical* occupation, 1.00 for *service* workers, 0.54 for *farm workers*, 0.34 for *precision* workers, and 0.53 for *operators*. Those of *precision* occupation group reported the lowest RR at *home*.

The fifth row showed the RRs at *industrial place*. All the RRs except *farm workers* were higher than one: 1.24 for *technical*, 1.91 for *service*, 0.84 for *farm workers*, 3.65 for *precision*, and 3.81 for *operators*. The persons in the of *precision* and *operator* occupation groups reported a higher RR of injury at *industrial places* with 3.65 and 3.81, respectively, almost four times of the *managerial* workers.

In general, at *industrial places*, the blue color workers - *service, precision, and operator* - had a higher RR of injury and the white color workers of *managerial* occupation and *farm workers* had lower RR . This trend was reversed for the RR *at home*.

*Farm workers'* RR *at a farm* recorded as 35.5 times that of the *managerial* occupation, the highest RR among those of other occupations. It implied that *farm workers* were injured most *at a farm*.

The RR of injury *at school* was very low for most occupations, 0.17 for *farm workers*, 0.17 for *precision* workers, 0.11 for *operators*. Only *service* workers reported 0.89, closer to the RR of the *managerial* occupation.

The RR in the *recreational area* is 0.91 for *technical* workers, 0.69 for *service* workers, 0.44 for *farm workers*, 0.75 for both *precision* workers and *operators*. *Managerial* workers reported the highest rate of injury at *recreational places* among these six occupations.

Table 3 shows the number of employed persons and percent of injuries by race and by working and not-working. Table 3a included the RR obtained from Table 3. Black was the reference point with RR = 1 to which non-black RR was compared. While working, non-black workers reported an RR of injury 2.15

times higher than black workers. But, while not working, the RR of non-black was 9.93 slightly lower than blacks.

Table 4 shows the numbers and percent of persons injured by age, 18-44 years and 45+.

Table 4a was constructed from the percents in Table 4. The younger group was used as reference to which the older group compared. The RR of injury for the older group was 1.18 while working and 1.10 while not working; both RRs higher than those of the younger group.

Table 5 included the numbers and percents of persons injured by sex, men and women. The corresponding RRs were shown in Table 5a. Compared to men, women reported a lower RR of 0.53 while working, but women reported a slightly higher RR while not working with the score of 1.13. This may imply that men are more likely injured while working, and women are more likely injured while not working.

The Table 3b, Table 4b, and Table 5b are the $2 \times 2$ tables by injured or not injured and by race, age, or sex, respectively. We calculated OR and RR for each table for comparison. The respective scores of RR and OR were 1.36 and 1.37 in Table 3b for the race, 1.13 and 1.14 in Table 4b for the age, and 0.84 and 0.83 in Table 5b for the sex. These scores are fluctuating little around one. They remained very close to one for both OR and RR because the percent of injury is very low, compared to that of not injured.

One RR score or the difference between two RRs may be tested by interval or a null hypothesis. Log of RR might stabilize the RR scores when they are skewed to one side or scattered around widely as suggested in the Appendix.

## DISCUSSION

We may choose the overall percent of persons injured as the reference point instead of the percent of persons injured in one specific occupation group. In this case, we are comparing the percent of persons injured of one occupation group to the percent, of persons injured of all other occupation groups. For example, in Table 1a we choose the managerial occupation as the reference point, but we could have chosen the total of all occupation groups as the reference point. This reference would be the percent of injuries among all currently employed persons. We may also consider a weighted average of the percent of all occupation groups who were injured as a reference point.

The SUDAAN is used to calculate the variances, assuming that the correlation in the sampling cluster is negligible. This may not be true for certain variables. For instance, there would be a strong correlation among members of the family or cluster when we consider the family income or race. In this case, we have to use the variance formula shown in the Appendix.

The relative risk can be applied to other areas in business, mortality, or diseases. For instance, before picking a stock to buy, one may compare the percent of relative profit of this stock to those of other candidates, and then purchase the stock with better relative profit. Another example is the relative risk of getting lung cancer among miners. Considering the percent of miners getting lung cancer as the reference, we may compare it to the percent of miners getting other diseases.

As the NHIS collects yearly data on injury, one might evaluate the relative risk of injury of the working population over time, and develop information systems for surveillance. This would exert pressure for new preventive efforts to target high-risk work sites, and develop specific strategies for injury prevention. This method may help to provide helpful information to develop injury reduction strategies at a workplace.

## ACKNOWLEDGMENT

Authors appreciated the help of Dr. Christina Park for her helpful comments, and Drs. Van Parson and Wayne Johnson for help with SUDAAN language, and Ms. Viona Brown and MS. Arlene Siller in obtaining data and SAS programing.

## REFERENCES

1. Choi, J.W. and Landis, J.R. 1997. Correlation Estimator for Unbalanced Data. Proceedings of the Section on Survey Research Methods. pp 858–863.
2. Bell, C.A., Stout, N.A., Bender, T.R. *et al.* 1990. Fatal occupational injuries in the United States, 1980 through 1985. *JAMA* **263,** 3047–3050.
3. Frank, J.W., Pulcins, I.R., Kerr, M.S., Shannon, H.S., and Stansfeld, S.A. 1995. Occupational back pain — an unhelpful polemic. *Scand. J. Work Environ. Health.* **21,** 3–14
4. Klein, B.P., Jensen, R.C., and Sanderson, L.M. 1984. Assessment of workers' compensation claims for back strains/sprains. *J. Occup. Med.* **26,** 443–
5. Leavitt, F. 1992. Physical exertion factor in compensable work injuries. *Spine* **17,** 307–10.
6. Leigh, J.P. and Landrigan, P.J. 1997. Occupational Injury and Illness in the United States — estimates of costs, morbidity, and mortality. *Arch. Intern. Med.* Vol. 157, July 28, 1997.
7. Loser, J.D. and Volinn, E. 1991. Epidemiology of low back pain. *Neurosurg. Clin. North Am.* **2,** 713–8.
8. McCurdy, S.A., Schenker, M.B., and Lassiter, D.V. 1989. Occupational injury and illness in the semiconductor manufacturing industry. *Am. J. Ind. Med.* **15,** 499–510.
9. National Center for Health Statistics. 1995. Vital and Health Statistics Series 10, No. 193. Current Estimates from the National Health Interview Survey, U. S. Department of Health and Human Services, Hyattsville, Md. December 1995.
10. Park, C.H., Wagener, D.K, Winn D.M., and Pierce J.P. 1993. Health conditions among the currently employed. In: Vital and Health Statistics. National Center for Health Statistics. DHHS Publication No. (PHS) 93–1514.

11. Richardson, D., Loomis, D., Wolf, S.H., and Gregory, E. 1997. Fatal agricultural injuries in North Carolina by race and occupation, 1977–1991. *Am. J. Ind. Med.* **31,** 452–458.
12. Robinson, J.C. 1989. Exposure to occupational hazards among Hispanics, blacks, and non-Hispanic whites in California. *Am. J. Public Health* **79,** 629–630.
13. Rowe, M.L. 1983. *Backache at Work.* Fairport, NY: Perrington Press.
14. SAS Institute Inc. 1989. SAS/STAT User's Guide, Vol. 1 and Vol. 2, Version 6, Cary, NC: SAS Institute Inc.
15. Shah, B., Barnwell, B.G., and Bieler, G.S. 1996. SUDAAN, Software for the statistical analysis of correlated data. RTI User's Manual, Release 7.0 April, 1996. Research Triangle Park, NC.
16. U.S. Bureau of Labor Statistics. 1997. BLS News: Workplace injuries and illnesses in 1995.
17. U.S. Bureau of the Census.1978. Technical report No. 40.
18. U.S. Bureau of the Census. 1980. Classified Index of Industries and Occupations. Washington, DC: U.S. Department of Commerce.
19. U.S. Department of Health and Human Service. 1996. Healthy people 2000 review 1995-96. National Center for Health Statistics. DHHS Publication No. (PHS) 96–1256.
20. Wagener, D.K. and Winn, D.W. 1991. Injuries in working population: Black-White differences. *Am. J. Public Health* **81,** 1408–1414.
21. Zwerling, C., Sprince, N.L., Ryan, J., and Jones M.P. 1993. Occupational injuries: comparing the rates of male and female. *Am. J. Epidemiol.* **138,** 46–55.

## APPENDIX

A sample S is taken from the population U. Clusters were selected by pps design and the persons in the cluster were selected by equal probability. The population U included the units $\{U_{oij}\}$ for i = (1,A), j = (1,$B_i$), and o = (1,O).

Defined the relative risk as $R_{o/r}$ for relative risk of injury at work comparing occupation "o" to reference occupation "r".

The sample S = $\{(i,j)\colon i \in S^*;\ j \in S_i\}$ where S* is a set of segments and $S_i$ is the number of persons in segment i. The sample segments are indexed by i = (1, a) and the persons by j = (1, $b_i$). The subscripts i and j for the population are not the same as the i and j for the sample.

The sample S included a set of sample persons $\{u_{oij}\}$.$\{u_{oij}\}$ are measured by $\{x_{oij}\}$.

$$x_o = \Sigma_{ij} w_{oij} x_{oij}$$

where $w_{oij}$ is the weight of the (oij)-th person, and $x_{oij}$ = 1 if the person (oij) is injured and $x_{oij}$ = 0 otherwise,

$$w_{oij} = (1/\Pi_{oij})(C_{aij}/S_{aij})$$

where $\Pi_{oij}$ is the probability of person (oij) included in the sample, and $C_{aij}$ is the known population count in cell a where the ij-th person belongs and $S_{aij}$ is the corresponding sample estimate of $C_{aij}$.

$$n_o = \Sigma_{ij} w_{oij}, \quad x_o = \Sigma_{ij} \delta_{oij}$$

where $\delta_{oij} = 1$ if the oij = the person injured, and $\delta_{oij} = 0$, otherwise.

$x_o$ and $n_o$ are the estimated numbers of the injured persons and all persons for the occupation "o". Similarly, $x_r$ and $n_r$ are the numbers of injured persons and all persons, respectively, for the reference occupation "r". Denote the estimator of a sample relative risk by $r_{o/r}$, 0 = 1, . . . , 0.

$$r_{o/r} = \frac{\dfrac{x_o}{n_o}}{\dfrac{x_r}{n_r}}$$

Let the expected values of $x_o$, $n_o$, $x_r$, $n_r$ be $\mu_o$, $\mu_{no}$, $\mu_r$, $\mu_{nr}$, respectively, and the variances be $\sigma_o^2$, $\sigma_{no}^2$, $\sigma_r^2$, $\sigma_{nr}^2$, respectively. Also denote the ratio of percent by $R_{o/r} = R_o/R_r$ in the population, where $R_o = \mu_o/\mu_{no}$ and $R_r = \mu_r/\mu_{nr}$. Taylor expansion of $r_{o/r}$ is

$$r_{o/r} = R_{o/r}\left(1 + \frac{x_o}{\mu_o} - \frac{n_o}{\mu_{no}} - \frac{x_r}{\mu_r} + \frac{n_r}{\mu_{nr}}\right) + O\left(n^{-1}\right)$$

When the occupations are independent, it can be shown that the variance of $r_{o/r}$ is given by

$$\mathrm{Var}\left(r_{o/r}\right) = R^2\left[\frac{\sigma_o^2}{\mu_o^2} + \frac{\sigma_{no}^2}{\mu_{no}^2} - 2\frac{\mathrm{cov}\left(x_o, n_o\right)}{\mu_o\mu_{no}} + \frac{\sigma_r^2}{\mu_r^2} + \frac{\sigma_{nr}^2}{\mu_{nr}^2} - 2\frac{\mathrm{cov}\left(x_r, n_r\right)}{\mu_r\mu_{nr}}\right]$$

where $\mathrm{var}(x_o) = \sigma_o^2$.

Since $x_o$, $n_o$, $x_r$, $n_r$ are the weighted and correlated counts as shown before, and $II_{ij} = II$ for a pps sample S, the variances, $\sigma_o^2$, $\sigma_{no}^2$ $\sigma_r^2$, and $\sigma_{nr}^2$ and covariances, $\mathrm{cov}(x_o, n_o)$ and $\mathrm{cov}(x_r, n_r)$, are obtained by Taylor expansion and correlation models. Under correlation assumptions, it can be shown

$$\mathrm{var}\left(x_o\right) = s_{do}^2\, n\left[1 + (b-1)\theta\right] / \Pi^2$$

$$s_{do}^2 = \frac{1}{\left(N_o - 1\right)} \Sigma_{i=1,A\,;\,j=1,Bi}\left(d_{oij} - \overline{D}_o\right)^2$$

where $s_{do}^2$ is the variance of $d_{oij} = (x_{oij} - R_o\, s_{oij})$, $s_{oij}$ is the estimated number of persons falling in the age-sex-race cell where the (ij)-th person belongs. $\theta$ is the common intracluster correlation between $d_{oij}$ and $d_{oij'}$ for $j \neq j'$ in cluster i. $\overline{D}_o$ is the average of all $d_{oij}$, and $\Pi$ the common selection probability of (ij) element for pps sample, $N_o$ is the number of persons with occupation "o" in the population, n is the number of sample units n = ab, and b is the number of units in the cluster when $b_i = b$.

The ratio $R_o = E(x_{oij})/E(s_{oij})$ and $E(d_{oij}) = 0$.

$var(x_o)$ reflects two facts: Taylor expansion of ratio and the intracluster correlation.

Denote the number of persons in the age-sex-race cell where (ij) person belongs by $U_{oij}$ and its sample estimate by $S_{oij}$. Write the difference by $d^*_{oij} = (U_{oij} - S_{oij})$ and $\overline{D}^*_o$ for the average of all $d^*_{oij}$. $E(d^*_{oij}) = 0$. $\theta^*$ is intracluster correlation between $d^*_{oij}$ and $d^*_{oij'}$ for $j \neq j'$ in cluster i. It can be shown

$$var(n_o) = \mathbf{S}^2_{d*o} n\left[1 + (b-1)\theta^*\right]$$

$$\mathbf{S}^2_{do} = \frac{1}{(N_o - 1)} \Sigma_{i=1,A\,;\,j=1,Bi}\left(d_{oij} - \overline{D}_o\right)^2$$

Similarly, the covariance between $x_o$ and $n_o$ can be written as, if injured and not injured are independent,

$$Cov(x_o, n_o) = ns = var(x_o)$$

$n_o = x_o + x_{o1}$. (Injured plus not injured.)

Similarly derive the variances $var(x_r)$ and $var(n_r)$, and covariance $cov(x_r, n_r)$ of the reference occupation "r". Replacing these variances and covariances and correlations with respective sample estimates, we can obtain a sample estimate of $var(r_{o/r})$. Methods of correlation are suggested in a reference (Choi and Landis, 1997).

If the correlations $\theta$ and $\theta^*$ are all negligible, then the variances and covariances retain only the first term, and we can use SUDAAN to obtain approximations.

The hypothesis $H_o : R_{o/r} = R$ (*e.g.*, R = 1) may be tested with a test statistic

$$Z = \left(r_{o/r} - R\right) \Big/ \left[var\left(r_{o/r}\right)\right]^{1/2}$$

$$\text{or interval} \qquad \left(r_{o/r} - 2s\right) \leq R \leq \left(r_{o/r} + 2s\right)$$

If the r's are unstable, scattered around or skewed, taking log may stabilize r's.

Then $H_o$: log $R_{o/r}$ = log R (*e.g.*, log 1 = 0). Test statistics are

$$Z' = \left(\log r_{o/r} - \log R\right) \Big/ \left[\mathrm{var}\left(\log r_{o/r}\right)\right]^{1/2}$$

and $$\left(\log r_{o/r} - 2s*\right) \le \log R \le \left(\log r_{o/r} + 2s*\right)$$

where var(log $r_{o/r}$) is the same as the var($r_{o/r}$) except R is replaced by log R in the formula.