A Procedure for Linking Psychosocial Job Characteristics Data to Health Surveys

JOSEPH E. SCHWARTZ, PHD, CARL F. PIEPER, MPH, AND ROBERT A. KARASEK, PHD

Abstract: A system is presented for linking information about psychosocial characteristics of job situations to national health surveys. Job information can be imputed to individuals on surveys that contain three-digit US Census occupation codes. Occupational mean scores on psychosocial job characteristics—control over task situation (decision latitude), psychological work load, physical exertion, and other measures—for the linkage system are derived from US national surveys of working conditions (Quality of Employment Surveys 1969, 1972, and 1977). This paper discusses a new method for reducing the biases in multivariate analyses that are likely to arise when utilizing linkage systems based on mean scores. Such biases are reduced by modifying the linkage system to adjust imputed individual scores for demographic factors such as age, education, race, marital status and, implicitly, sex (since men and women have separate linkage data bases). Statistics on the linkage system's efficiency and reliability are reported. All dimensions have high inter-survey reproducibility. Despite their psychosocial nature, decision latitude and physical exertion can be more efficiently imputed with the linkage system than earnings (a non-psychosocial job characteristic). The linkage system presented here is a useful tool for initial epidemiological studies of the consequences of psychosocial job characteristics and constitutes the methodological basis for the subsequent paper. (*Am J Public Health* 1988; 78:904–909.)

Introduction

Investigations into the adverse psychosocial health consequences of various occupations are most informative when the job situation of each individual is measured directly. However, information on health status and work conditions are rarely available in the same data base. This has spurred us to develop linkage systems for estimating the job situations of individuals in health data bases. This paper will discuss the rationale for, as well as the potential risks of, linking data bases and will describe one such system that uses US Census occupation codes to link to health data bases psychosocial characteristics of work structure that have previously been shown to be associated with job stress.[1]

There is a substantial history, particularly within stress research and sociology, of linking events and situations at the macro or ecological level with data on health status and well-being—for example, contextual analyses of the association of a plant closing,[2] socio-ecologic stress,[3] or working in a stressful occupation[4] with individuals' levels of blood pressure or the association of community-level unemployment rate with individuals' mental health.[5] In these studies, an explanatory variable that is defined (or at least measured) only at the macro level is linked to individual level health data. Several other studies have examined aggregate or ecological level associations between social environment and health, such as between unemployment and mortality rates.[6–8] In these latter studies, aggregate-level data on two or more types of events for the same units (e.g., neighborhood, state, or year) are merged. Historically, these contextual and ecological analyses, despite potential methodological problems,[9] have provided the first suggestion of effects of societal or stressful events on individual health and well-being.

The tradition of data base linkage systems which specifically focus on occupational risks and individual level health status is more recent. It includes a Swedish version of our own system[10,11] associating psychosocial job stressors and cardiovascular illness, a job-title based study relating chronic diseases to diverse work requirements,[12] a linkage system for toxic exposures and carcinogenesis,[13] and studies relating exposure to both physical and psychosocial stressors with colon and rectal cancers.[14,15] While these studies report contextual analyses, the authors realize that their use of occupational-level data for the hypothesized risk factor(s) is a weak, albeit necessary, substitute for individual-level data.

Previous research[1,16] has suggested that a job's psychological work load and the work's latitude for decision making on the job may be associated with psychological stress and coronary heart disease. Unfortunately, no American data base simultaneously contains information on these aspects of an individual's job situation and coronary heart disease status. This motivated us to develop a system that utilizes job information from comprehensive national surveys of working conditions to impute job psychosocial characteristics to surveys of health status.

Methods

Data Linkage

One general, albeit simplified, model for estimating the effect of a job characteristic on the risk of coronary heart disease, for person i, is

$$Y_i = a + B_1 W_i + B_2 A_i + B_3 C_i + U_i \qquad [1]$$

where Y is the risk of coronary heart disease,
W is a work condition (e.g. job strain),
A is age,
C is a control variable (e.g. cholesterol), and
U is a residual, all for person i.

All variables except job strain exist on the health data bases; all but heart disease and cholesterol exist on the work data bases. The primary linking variable—three-digit Census occupation code—exists on both data bases.

Mean Job Strain Scores

Since the job strain variable does not exist in the health data bases, the most straightforward approach for estimating an individual's job strain is to compute the mean value of job strain for all those having the same occupation (as defined by the Census classification system) in the work data base and

Address reprint requests to Joseph E. Schwartz, PhD, Department of Psychiatry and Behavioral Medicine, State University of New York at Stony Brook, Stony Brook, NY 11794. Mr. Pieper is with the Columbia University School of Public Health; Dr. Karasek is with the Department of Industrial and Systems Engineering, University of Southern California. This paper, submitted to the Journal August 29, 1986, was revised and accepted for publication September 30, 1987.
Editor's Note: See also related article p 910 this issue.

© 1988 American Journal of Public Health 0090-0036/88$1.50

impute this mean to all individuals with this occupation in the health data bases. This strategy for imputing scores attributes to the individual an estimate of the score for his occupation. Assuming that both survey samples are drawn from the same population (our work and health data bases are national samples from approximately the same time period), this method captures the between-occupation variance of job strain, but not the within-occupation variance. If the jobs within each occupation are relatively homogeneous with respect to job strain, most of the variance is between occupations and is not lost by using the occupation score rather than the individual's actual job score. For example, this approach can only estimate what it is like to be a baker; it cannot predict differences between bakers.

Adjusted-Mean Job Strain Scores

Within occupations, age generally is negatively related to job strain. Since age is available in both the health status and working conditions data bases, we can obtain a more precise estimate of the job strain of individual i, who is in occupation k, using the equation:

$$\hat{W}_i = D(A_i - \overline{A}) + \overline{W}_{adj(k)} \qquad [2]$$

where W = job strain and A = age.

Both the coefficient D, measuring the relationship between job strain and age (controlling for occupation) and the occupation means of job strain (adjusted for age) are estimated from the working conditions surveys and used to generate an estimate of job strain for each subject in the health surveys. These adjusted-mean estimates ($\hat{W}_i$) fully capture the relationship between A and W, a condition which is important for obtaining good estimates of Equation 1. Thus, if job strain decreases with age, an older baker will be estimated to have a lower score than a younger baker.

Efficacy of the Mean and Adjusted-Mean Estimates

The primary weakness of both the mean and adjusted-mean estimates is that the imputed scores capture only a fraction of the actual variance of the job characteristic. This loss of variance results in a proportionate decline in the estimated strength of association between heart disease and job strain and a corresponding drop in the explanatory power (contribution to R^2) of job strain. Nevertheless, it can be shown (see Appendix, Case 1, for details) that in the *simple* regression of heart disease on either type of estimates of job strain, the unstandardized coefficient is unbiased.

The Appendix describes in some detail the biases that result when the mean and adjusted-mean estimates are used in multiple regression analyses. To summarize, if heart disease were regressed on *mean* job strain estimates and age, the estimated effect of job strain would be attenuated (biased toward zero). Because job strain and age are negatively correlated, the effect of age would also be attenuated. These biases are due to the mean job strain estimates' failure to reproduce the covariation between age and job strain within occupations. If *adjusted-mean* job strain estimates were used in this multiple regression, neither coefficient would be biased because the job strain estimates would incorporate the association of job strain with age.

Often, however, there are other explanatory variables, such as cholesterol in Equation 1, which cannot be incorporated into the adjusted-mean estimates of job strain because of their unavailability in the working conditions surveys. To the degree that these variables are correlated with a job characteristic, controlling for occupation and the demographic factors already used to predict the job characteristic, the estimated coefficients of Equation 1 will tend to be biased, and the anticipated result will be a conservative deflation of the estimated effect of job strain on heart disease in Equation 1.

Other linkage systems with which we are familiar, such as the Dictionary of Occupational Titles,[17] impute scores on the basis of occupation that are analogous to our mean estimates. While the simple regression of a dependent variable on these scores should yield an unbiased coefficient, the coefficients from a multiple regression are likely to be biased. Our adjusted-mean estimates of job characteristics, by replicating the within-occupation covariation of the job characteristic with basic demographic variables, can eliminate or at least substantially reduce these biases. This is, we believe, an important advantage of this linkage system.

Application (Results)

The Health Examination Survey (HES) conducted in the early 1960s and the first National Health and Nutrition Examination Survey (NHANES) conducted in the first half of the 1970s each contain appropriate information on heart disease and its conventional risk factors. They also contain basic demographic data which can be used, when appropriate, in the equations that generate the adjusted-mean estimates of job characterisrtics.

Our source of data for estimating job characteristics consists of three nationally representative Surveys of Working Conditions, also known as the Quality of Employment Surveys, sponsored by the US Department of Labor and conducted by the Institute for Survey Research at the University of Michigan in 1969, 1972, and 1977.[18] These cross-sectional surveys (with sample sizes of 993, 985, and 968 males,* respectively) ask respondents (persons ages 20 to 65 who worked at least 20 hours in the paid civilian labor force during the week prior to the interview) to report on a wide range of characteristics about their jobs. Based on factor analyses of a subset of these rather detailed characteristics, it is possible to construct the following scales of job decision latitude and psychological demands.

Decision Latitude (an equally weighted sum of two subscales)**

- **Skill Discretion:** 1) Keep learning new things; 2) Requires high level of skill; 3) Requires creativity; 4) Not repetitious; 5) Can develop skills; 6) Job has variety.
- **Decision Authority:** 1) Can choose how to perform work; 2) Can make decisions on my own; 3) Have a lot of say on the job; 4) Take part in decisions that affect me.

Psychological Job Demands: 1) Work fast; 2) Work hard; 3) Work is excessive; 4) Not enough time to do work; 5) Face conflicting demands at work.

Since we are investigating cardiovascular illness, we also want to estimate the physical demands of the job, a potentially important control variable. Only one indicator exists in the working conditions surveys.

Physical Exertion: 1) Job requires lots of physical effort.

Table 1 summarizes our assessment of the reliability of these scales in the working conditions surveys. The within-survey reliabilities of the multiple-item scales are estimated

*The female subsamples have also been analyzed, but the results are not reported here.

**The skill discretion and decision authority scales are highly correlated at the occupation level (r = .77) and theoretically linked in organizational theory literature.[1]

TABLE 1—Reliability Estimates of Job Characteristics in the Three Surveys of Working Conditions

Job Characteristics	Within-Survey Reliability[a]				Between-Survey Consistency[b]	Total
	1969	1972	1977	Pooled		
Decision Latitude	.703	.851	.847	.805	.965	.776
Skill Discretion	.738	.780	.771	.766	.963	.737
Decision Authority	.580	.787	.793	.724	.970	.703
Psychological Demands	.633	.592	.616	.614	.961	.591
Physical Exertion	—	—	—	—	1.000	1.000
Annual Earnings	—	—	—	—	1.000	1.000

[a]Cronbach's alpha measure of reliability
[b]Weighted intra-class correlation of occupation means for three surveys

TABLE 2—Proportion of Variance Captured by Occupation Means

Job Characteristics	Decomposition of Reliable Variance[a]			% Ratio of Predicted to Total Reliable Variance[b]
	Between Occupations %	Within Occupations %	Total %	
Decision Latitude	34.7	42.9	77.6	44.7
Skill Discretion	34.9	38.8	73.7	47.4
Decision Authority	25.0	45.3	70.3	35.6
Psychological Demands	4.2	54.9	59.1	7.1
Physical Exertion	25.9	74.1	100.0	25.9
Annual Earnings	20.2	79.8	100.0	20.2

[a]as percentage of total variance
[b]ratio of first to third column

by Cronbach's alpha coefficient (columns 1-3). A pooled measure of within-survey reliability is reported in the fourth column. (The within-survey reliability cannot be estimated for physical exertion since there is only one indicator of this construct.) Because of some inter-year disparities between questions in the nominally similar job scales and because of the possibility that the job conditions of occupations might have actually changed between the 1969 and 1977 working conditions surveys, it is desirable to assess the comparability of the scales across the three surveys. The fifth column of Table 1 shows the estimated consistency of the occupation means across surveys. These are reassuringly high (≥.96) and justify pooling the job characteristic scores from the three surveys before estimating the parameters of the linkage system. This high inter-survey repeatability of occupation-level job scores increases our confidence in the estimated job scores on the national health surveys. The last column of Table 1 presents a summary measure of reliability for the pooled combination of the three surveys; it is the product of the pooled within-survey reliability and between-survey consistency. This last column represents an upper bound on the estimated per cent of variance that can be reliably estimated. The last row of this and subsequent tables presents the same information for self-reported annual earnings (not one of our psychosocial job characteristics) to provide a benchmark for the other potentially more subjective job characteristics.

As an external check on the content validity of these scales, the 1970 Census occupation scores for decision latitude and its components were compared to analogous scales from the US Dictionary of Occupational Titles (DOT) aggregated into 1970 Census occupation categories using a bridge tape.[19,20] The DOT data are based on direct observation by US Employment Service employees rather than self-report survey data, but scores for 45 per cent of the 12,099 detailed DOT occupations are based on observing *one* or *zero* jobs. Other aspects of the DOT, including the apparent pooling of observations made as many as 30 years apart, are discussed by Miller *et al.*[21] The correlations of our decision latitude scale with the DOT measures of occupational self-direction, routineness, closeness of supervision, and functional complexity with people and with data are 0.76, −0.53, −0.71, 0.75, and 0.51, respectively. The correlation of our physical exertion measure with the DOTs is 0.62. Given the entirely independent methodologies and somewhat differing scale content, these correlations seem acceptably high. The imputed job dimension scores also differentiate occupations in a plausible manner. The difference, for instance, in mean job decision latitude between managers and assembly workers (in the transportation and electrical equipment industry) is 2.6 standard deviations (see also Figure 1 in the subsequent paper).[22]

The 1970 three-digit Census occupation codes include about 440 categories. Those occupations with fewer than three male respondents in the pooled sample of the working conditions surveys were either merged into another substantively similar category or omitted from the analysis. This resulted in 221 occupations into which 98.5 per cent of the work force could be classified.

Table 2 presents summary statistics for the *mean* estimates. The first column reports the estimated percentage of total variance (adjusted for degrees of freedom) in the working conditions surveys that is between occupations. The remaining reliable variance is necessarily due to differences among persons having the same occupation. Given the goal of predicting the reliable variance, the best summary measure of how well the occupation means estimate individuals' actual job characteristics is the ratio (column 4) of the between-occupation variance to the total *reliable* variance (column 3).

This last statistic indicates that the mean estimates capture about two-fifths of the reliable variance of decision latitude (and its two subscales) and about one-fourth of the variance of physical exertion. These estimates compare favorably with the analogous result for annual earnings (20 per cent). However, we only capture about 7 per cent of the reliable variance of the psychological demands scale.

In deriving the adjusted-mean equations for the same job characteristics, we used those demographic covariates that are generally available in most surveys: race, years of education, age, marital status, region of residence (south versus other), urban versus rural residence, and self-employment status. In addition, efforts were made to model any non-linear relationships between these covariates (especially age and education) and the job characteristics. Table 3 shows the number of covariates that provide a statistically significant improvement in the predictability of each job characteristic and the proportion of additional variance that they predict.

As an example, the covariates that improve the predictive power of the physical exertion equation are age, education, and a dummy variable for whether the person is self-employed. The resulting equation for physical exertion, obtained from the working conditions surveys, is:

TABLE 3—Proportion of Reliable Variance Captured by Adjusted-Mean Estimates

Job Characteristics	Number of Covariates	Increment to R-SQ Due to Covariates %	Ratio of Predicted to Total Reliable Variance[a] %	Improvement[b] of Adjusted Over Unadjusted Mean Estimates %
Decision Latitude	5	3.9	49.7	11.2
Skill Discretion	6	2.7	51.0	7.6
Decision Authority	4	3.5	40.5	13.8
Psychological Demands	6	4.6	14.9	109.9
Physical Exertion	3	1.2	27.1	4.6
Annual Earnings	10	14.5	34.7	71.8

[a]comparable to last column of Table 2
[b]increase in Column 3 relative to last column of Table 2

$$\hat{W}_i = (-.004)\ Age_i + (-.035)\ Educ_i + (.253)\ self\text{-}employed_i + \bar{W}_{adj(k)} \quad [3]$$

The effects of these covariates imply that *within* occupations, there is variation in physical exertion (1.2 per cent, as shown in column 2 of Table 3) which is systematically related to three demographic variables. It should not be surprising that men who are older or have more education tend to have less physically demanding jobs. It is this type of relationship that the adjusted-mean estimates will correctly replicate on the health surveys, in contrast to the (unadjusted) mean estimates. The third column of Table 3 shows the percentage of reliable variance that the adjusted-mean estimates capture while the final column reports the relative improvement over the comparable statistic (last column of Table 2) for the mean estimates. This completes the development of the linkage system from the working conditions surveys.***

We can now begin the actual linkage process (Figure 1). The occupation-level job score data base, which is the input to the linkage system, consists of a set of simple (unadjusted) means of each job dimension for the 211 occupations and a parallel set of *adjusted* occupational means. Also, the system includes a coefficient for each covariate of each job dimension to later compute the adjusted-mean estimates. The first step of the process of imputing job condition scores to subjects in the health data bases is to collapse the detailed Census occupation variable in these surveys into the same 211 categories that were used in the work surveys.† Next, the unadjusted and adjusted occupation means are merged onto each subject's record by occupation. Last, the adjusted-means need to have the effects of the covariates added to them in order to create the adjusted-mean estimates; this is achieved with a single compute statement (having the form of Equation 3) for each job dimension. Empirical results using the imputed job scores to test the association of job strain with coronary heart disease are reported in the subsequent paper.[22]

***Our linkage system includes several other job scales, not used in the test of the job strain model for coronary heart disease, tapping such dimensions as job security, social support, and hazardous conditions at the worksite. These and further details on the above scales are described in [23].

†Since the Health Examination Survey classifies subjects according to the 1960 three-digit Census codes, we have also prepared a linkage data base for the job dimensions using a slightly collapsed version of 1960 occupational categories.

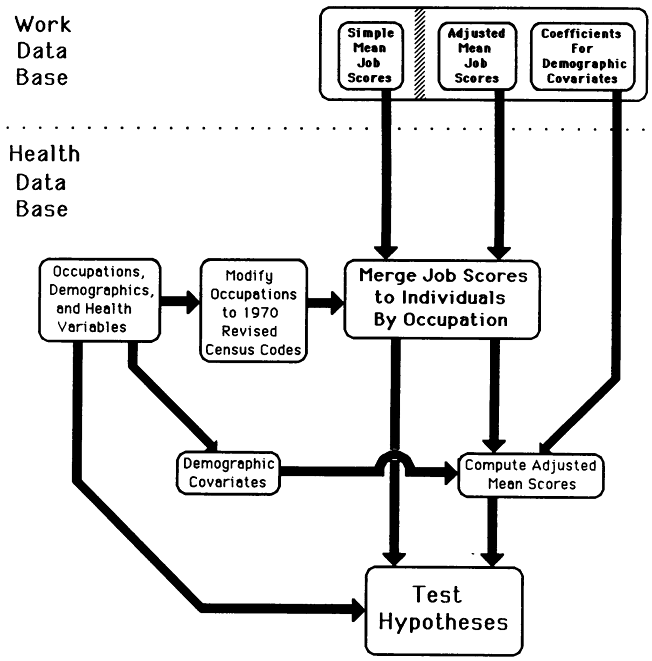

FIGURE 1—Flow Chart of Job Characteristic Linkage System

Discussion

Our occupation-based job characteristic linkage system has both advantages and disadvantages over directly measured job data from questionnaires. The fact that the job characteristics are not reported by the subjects themselves, but are estimated from average scores (from an independent sample) for occupational groups, greatly diminishes the potential impact of self-report bias in job measurement. Although it cannot protect against the possibility of a systematic tendency for persons within particular occupations to collectively over- or under-report their status with respect to some dimension, the vast majority of "subjective variance" (self-report bias) is usually thought to vary across individuals within occupations and is therefore averaged out by the estimation process. This is a substantial advantage for studies of psychosocial risk factors where it is often impossible to determine whether results are genuinely valid or attributable to self-report bias.

However, our method also has the disadvantage of almost certainly underestimating the true strength of associations (correlations) between job characteristics and potential outcome variables because of the lack of more detailed information about respondents' jobs beyond that provided by the three-digit Census occupational classification system. This causes the standard error of the unstandardized coefficient to be larger than if no variance were lost in the imputation process which, in turn, causes confidence interval estimates of regression coefficients to be wider. This problem is especially severe for a dimension like psychological demands which has relatively little between-occupation variance.

Our job dimensions linkage system certainly does not eliminate the desirability and need to collect detailed information about the jobs of health survey subjects.†† Rather,

††The job dimensions discussed in this paper, and related scales, are available for use in questionnaire form for direct data gathering.[24] Mean scores on the questionnaire scales can be compared with the national mean scores

the proper function of this system is to facilitate preliminary analyses of the possible effects of working conditions when the optimal data are unavailable. Given the substantial cost of collecting detailed data on individuals' job situations and the scarcity of resources, we hope that such preliminary analyses using the job dimensions linkage system will aid decision-makers in determining when the additional investment is likely to justify the cost.

APPENDIX

Analysis of Bias when using Mean and Adjusted-Mean Estimates in Regression Analyses

Case 1: Simple regression of Y on W

Suppose we want to estimate the equation

$$Y_i = a + B\,W_i + U_i \qquad \text{[A-1]}$$

and the only obstacle is the absence of data on W (a working condition such as job strain). In particular, the usual assumptions about the error term are presumed to hold: $E(U_i) = E(U_i\,W_i) = 0$ and $E(U_i^2) = \sigma_u^2$. Finally, using either occupation means or adjusted-mean estimates (Equation 2, in text), we have generated predicted scores W such that

$$W_i = \hat{W}_i + V_i \qquad \text{[A-2]}$$

and it is reasonable to assume that $E(V_i) = E(V_i\,\hat{W}_i) = 0$. Substituting into Equation A-1 yields

$$\begin{aligned} Y_i &= a + B(\hat{W}_i + V_i) + U_i \\ &= a + B\,\hat{W}_i + B\,V_i + U_i \\ &= a + B\,\hat{W}_i + U_i'. \end{aligned} \qquad \text{[A-3]}$$

The assumption that $E(U_i\,W_i) = 0$ implies that $E(U_i\,\hat{W}_i) + E(U_i\,V_i) = 0$, but we need to introduce the common assumption§ that $E(U_i\,V_i) = 0$ which implies that $E(U_i\,\hat{W}_i) = 0$ also. From the above, it follows that $U_i' = B\,V_i + U_i$ has an expectation of zero and is uncorrelated with $\hat{W}$ which implies that the ordinary least-squares (OLS) estimates of the regression of Y on $\hat{W}$ will be unbiased estimates of the parameters of Equation A-1.§§

Although fairly reasonable assumptions enable one to obtain unbiased estimates for Equation A-1, the R^2 for Equation A-3 grossly underestimates the corresponding R^2 for Equation A-1. However, if one is prepared to use estimates of σ_v^2 from the working conditions survey to infer what proportion of the sum of squares for U_i' is due to $(B\,V_i)$, then one can estimate the sum of squares for U_i and the R^2 for Equation A-1.

The preceeding analysis is unaffected by whether we use the unadjusted means or the adjusted-mean estimates to generate $\hat{W}_i$, except in so far as the latter (1) may be more likely to satisfy the assumption that $E(U_i\,V_i) = 0$ and (2) marginally increase the variance of $\hat{W}$, thus reducing the standard error of B.

Case 2: Regression of Y on W and A

It is this situation, when estimating

$$Y_i = a + B_1\,W_i + B_2\,A_i + U_i, \qquad \text{[A-4]}$$

where the difference between using the means and adjusted-mean estimates to predict W becomes consequential. When, in place of W, we use the adjusted-mean estimates which are conditioned on occupation and A (age), the V_i in Equation A-2 should be uncorrelated with both $\hat{W}$ and A (since the unexplained variance in W is uncorrelated with all variables used to predict W). Following the same logic as Case 1, it follows that the OLS regression of Y on $\hat{W}$ and A produces unbiased estimates of the parameters of Equation A-4.

Problems arise when the unadjusted occupation means are used to estimate W and A is correlated with W controlling for occupation. In this situation, V and A are necessarily correlated and therefore so are U' and A. If the remaining assumptions hold, it can be shown that the OLS estimates of B_2 will certainly be biased and those of B_1 will usually be biased. More precisely,

$$E(B_1) = \beta_1\left(1 - \frac{\sigma_{VA}\,\sigma_{A\hat{W}}}{\sigma_A^2\,\sigma_{\hat{W}}^2\,(1 - r_{A\hat{W}}^2)}\right) \qquad \text{[A-5]}$$

$$E(B_2) = \beta_2 + \beta_1\,\frac{\sigma_{VA}}{\sigma_A^2\,(1 - r_{A\hat{W}}^2)} \qquad \text{[A-6]}$$

Equation A-5 implies that in the usual case where A will be correlated with $\hat{W}$ in the same direction as its correlation with V—for example, the average job discretion of an occupation tends to be higher in those occupations, like managers, for which the average age is higher (job discretion and age are positively correlated at the aggregate level of occupations) while within an occupation, older individuals tend to have more job discretion—OLS estimates of B_1 will tend to be biased toward zero. If the correlations are in opposite directions, then B_1 will be exaggerated. Only if A is uncorrelated with $\hat{W}$ at the aggregate occupational level will estimates of B_1 be unbiased. Equation A-6 indicates that estimates of B_2 will be biased whenever B_1 does not equal zero. The direction of the bias corresponds to the sign of the product of B_1 times the correlation of A with V. The stronger the association between A and $\hat{W}$, the greater the bias of both coefficients will be.

The results of this section clearly demonstrate that if one wants an unbiased estimate of the effect of W on Y controlling for A, it is important that the imputed scores, $\hat{W}$, be conditioned on A so that the V_i will be uncorrelated with A.

Case 3: Regression of Y on W, A, and C

Having considered the two simplified situations above, it is relatively clear what happens when one uses $\hat{W}$ to estimate the general equation with which we began this paper,

$$Y_i = a + B_1\,W_i + B_2\,A_i + B_3\,C_i + U_i \qquad \text{[A-7]}$$

With the adjusted-mean estimates of W, one can assume that $E(V_i\,A_i) = 0$, but not that $E(V_i\,C_i) = 0$: the behavior of C (cholesterol level) in Equation A-7 is similar to that of A in Equation A-4 when using the occupation means to estimate W. The only difference is that the relationship between A and W could be determined in the working conditions survey (and incorporated into the adjusted-mean estimates of W) while variable C is only available in the health survey. Hence, there are no data with which to assess the potential correlation of C with V. Generalizing from the results of Equations A-5 and A-6, the existence of a correlation between C and V will directly bias the estimate of B_3 while bias in the other two coefficients depends on both this correlation *and* the correlation of C with A and $\hat{W}$. If the latter (observable) correlations are small, then the bias of B_1 and B_2 should be correspondingly small.

discussed above, for males, for females, by occupation, or by industry.

§This assumption will be invalid to the degree that there are omitted variables that are related to *both* V_i in Equation A-2 and U_i in Equation A-1. Such variables would have to be able to improve the prediction equation for W and be related to Y after controlling for $\hat{W}$.

§§If it is further true that $E(V_i^2) = \sigma_v^2$ is a constant—that is, that the heterogeneity of jobs with respect to W is constant across occupations—then the U_i' will be homoscedastic and the estimate of Equation A-3 will be BLUE. It is possible to use the working conditions surveys to test if the within-occupation variance of W varies across occupations and, if so, estimates of the variances of the V_i could be used to generate statistically more efficient *weighted* least-squares estimates of Equation A-3.

ACKNOWLEDGMENTS

The research reported in this paper was supported by NIOSH grant R01-OH-00906 to Columbia University. We would also like to gratefully acknowledge assistance in programming and manuscript preparation provided by Laurie Beck, Chris Carlin, Frank Chemely, Linda Cranor, Joanne Factor, Louise Fry, and Chris Schafter at Columbia University.

REFERENCES

1. Karasek RA: Job demands, job decision latitude, and mental strain: Implications for job redesign. Admin Sci Q 1979; 24:285–307.
2. Kasl SV, Cobb S: Blood pressure changes in men undergoing job loss: A preliminary report. Psychosom Med 1970; 2:19–38.
3. Harburg E, Erfurt JC, Havenstein LS, Chape C, Schull WJ, Schork MA: Socio-ecological stress, suppressed hostility, skin color, and Black-White male blood pressure: Detroit. Psychosom Med 1973; 35:276–296.
4. Rose RM, Jenkins CD, Hurst MW: Air Traffic Controller Health Change Study. Report prepared for US Dept of Transportation, FAA Office of Aviation Medicine. Springfield, VA: Natl Tech Info Serv (FAA-Am-78-39) 1979.
5. Catalano RA, Dooley CD: Economic predictors of depressed mood and stressful life events. J Health Soc Behav 1977: 18:292–307.
6. James SA, Kleinbaum DG: Socioecologic stress and hypertension related mortality rates in North Carolina. Am J Public Health 1976; 66:354–358.
7. Brenner HM: Economic changes and heart disease mortality. Am J Public Health 1971; 59:1154–1168.
8. Eyer J: Prosperity as a cause of death. Int J Health Serv 1977; 7:125–150.
9. Morgenstern H: Uses of ecologic analysis in epidemiologic research. Am J Public Health 1982; 72:1336–1344.
10. Alfredsson L, Karasek RA, Theorell T: Myocardial infarction risk and psychosocial work environment: An analysis of the male Swedish working

force. Soc Sci Med 1982; 16:463–467.
11. Alfredsson L, Spetz C, Theorell T: Type of occupation and near-future hospitalization for myocardial infarction and some other diagnoses. Int J Epidemiol 1985; 14:378–388.
12. Shaw JB, Riskind JH: Predicting job stress using data from the position analysis questionnaire. J Appl Psychol 1983; 68:253–261.
13. Hoar SK, Morrison AS, Cole P, Silverman DT: An occupational and exposure linkage system for the study of occupational carcinogenesis. JOM 1980; 22:722–726.
14. Garabrandt DH, Peters JM, Mack TM, Bernstein L: Job activity and colon cancer risk. Am J Epidemiol 1984; 119:1005–1014.
15. Spiegelman D, Wegman DH: Occupation-related risks for colorectal cancer. JNCI 1985; 75:813–821.
16. Karasek RA, Baker D, Marxer F, Ahlbom A, Theorell T: Job decision latitude, job demands and cardiovascular disease: a prospective study of Swedish men. Am J Public Health 1981; 71:694–705.
17. US Dept of Labor: Selected Characteristics of Occupations Defined in the Dictionary of Occupational Titles. Washington, DC: DOL, 1981.
18. Quinn RP, Staines GL: The 1977 Quality of Employment Survey: Descriptive Statistics with Comparison Data from the 1960–70 and the 1972–73 Surveys. Ann Arbor, MI: Survey Research Center, Institute for Social Research, 1979.
19. Temme T: Occupation: Meanings and Measures. Washington, DC: Bureau of Social Science Research, 1975.
20. Spenner K: Occupational characteristics and classification systems: New uses of the Dictionary of Occupational Titles in social research. Soc Meth Res 1980; 9:239–264.
21. Miller A, Treiman D, Cain P, Roos P: Work, Jobs, and Occupations: A Critical Review of the Dictionary of Occupational Titles. Washington, DC: National Academy Press, 1980.
22. Karasek RA, Theorell T, Schwartz JE, Schnall PL, Pieper CF, Michela JL: Job characteristics in relation to the prevalence of myocardial infarction in the US Health Examination Survey and Health and Nutrition Examination Survey. Am J Public Health 1988; 78:910–918.
23. Karasek RA, Schwartz JE, Theorell T: Job Characteristics, Occupation, and Coronary Heart Disease. Final Report, Phase I to NIOSH (Grant R01-OH-00906). 1982 (mimeo) New York: Columbia University.
24. Karasek RA: The Job Content Questionnaire and User's Guide, Version 1.1. Los Angeles: Department of Industrial and Systems Engineering, University of Southern California, March 1985.

NCHSR's Hospital Studies Program Using New Data Base

Increased understanding of how hospitals provide health care services and what determines patterns of hospital use and costs is the goal of the Hospital Studies Program (HSP) being conducted by staff of the National Center for Health Services Research (NCHSR) and Health Care Technology Assessment.

The Hospital Cost and Utilization Project (HCUP) is studying how hospitals react to changes in their environment, especially with respect to technology, market conditions, and third-party payers. Until recently HCUP analyses were based on information from its original data base—a powerful and unique combination of interlinked facts about hospitals, patients, physicians, and communities from 1970 to 1977.

The HCUP data base has now been expanded to include similar data for 1980 through 1985: data for 1986 and 1987 will be added as soon as they become available. In addition, the original sample of 384 short-term, general, non-federal hospitals is being enlarged to 600 hospitals. To expedite research on these new data, HCUP analysts are using about 400 of the hospitals and roughly two-thirds of the discharge records that will be in the full data base. By legal agreements, the identities of HCUP sample hospitals, providers, and patients are removed; the remaining data are encrypted and protected with special computer technology before delivery to the federal government. HSP studies present data and finding for use by the public in summary form only.

With the expansion of the HCUP data base, NCHSR researchers have a unique window on major changes in health care delivery during the 1970s and 1980s and can analyze the effect of these changes on patterns of hospital use and costs. The new HCUP data are now being used to support policy research on:

- hospital case mix issues including diagnosis related groups (DRGs) and alternative classifications;
- hospital financial position;
- effects of Medicare's prospective payment system (PPS);
- hospital utilization patterns in general and in relation to specific types of hospitals, patients, payers, or policies; and
- variations in patients outcomes and in medical practice patterns.

Publication of findings begin during summer 1988. Additional information on HCUP activities is available from Rosanna M. Coffey, PhD, Director, Hospital Studies Program, NCHSR, Parklawn Building, Room 18A-55, Rockville, MD 20857; telephone 301/443-5706.