

Analysis of Clustered Data in Community Psychology: With an Example from a Worksite Smoking Cessation Project¹

Donald Hedeker²

University of Illinois at Chicago

Susan D. McMahon and Leonard A. Jason

DePaul University

Doreen Salina

Northwestern University

Although it is common in community psychology research to have data at both the community, or cluster, and individual level, the analysis of such clustered data often presents difficulties for many researchers. Since the individuals within the cluster cannot be assumed to be independent, the use of many traditional statistical techniques that assumes independence of observations is problematic. Further, there is often interest in assessing the degree of dependence in the data resulting from the clustering of individuals within communities. In this paper, a random-effects regression model is described for analysis of clustered data. Unlike ordinary regression analysis of clustered data, random-effects regression models do not assume that each observation is independent, but do assume data within clusters are dependent to some degree. The degree of this dependency is estimated along with estimates of the usual model parameters, thus adjusting these effects for the dependency resulting from the clustering of the data. Models are described for both continuous and dichotomous outcome variables, and available statistical software for these

¹Preparation of this article was supported by National Heart, Lung, and Blood Institute Grant R18 HL42987-01A1, National Institutes of Mental Health Grant MH44826-01A2, and University of Illinois at Chicago Prevention Research Center Developmental Project CDC Grant R48/CCR505025.

²All correspondence should be addressed to Donald Hedeker, Division of Epidemiology and Biostatistics (M/C 922), School of Public Health, University of Illinois at Chicago, 2121 West Taylor Street, Room 510, Chicago, Illinois 60612-7260.

models is discussed. An analysis of a data set where individuals are clustered within firms is used to illustrate features of random-effects regression analysis, relative to both individual-level analysis which ignores the clustering of the data, and cluster-level analysis which aggregates the individual data.

KEY WORDS: intraclass correlation; unit of analysis; nested design; correlated data; random-effects model.

Community psychology has long recognized the importance of context and the interaction between person and environment (e.g., Barker, 1968; Kelly, 1966; Moos, 1976). The ecological perspective, particularly the principle of interdependence, which assumes that all parts of systems are interrelated, guides us to attend to the relationships within the system and utilize the community as the level of analysis (Levine & Perkins, 1987). Yet, as Shinn (1990) asserts, community research has had little success in understanding patterns of reciprocal relations that involve higher level units. Assessments of the individual, organizational, and community constructs are often considered in interventions (Linney & Reppucci, 1982); however, most traditional statistical methods utilized by community psychologists (i.e., fixed-effects ANOVA or regression of individual or aggregate scores) do not adequately account for the person-environment relationship. The question of what level of analysis is appropriate and representative of individuals in an ecological context thus presents an important methodological issue that has proved difficult for community psychologists to address.

Although community psychologists recognize the importance of context in assessing individuals, either the individual or the context (e.g., the community, neighborhood, firm, hospital clinic, school, classroom, or family) is examined, often to the exclusion of the other. This is unfortunate since neither analysis at the individual or aggregate level tells us very much about the degree to which the context and its variables influence the response of the individual. One of the difficulties in analyzing individuals that are part of a large group, or cluster, is that many traditional statistical procedures assume independence of observations. However, since the individuals within the cluster cannot necessarily be assumed to be independent, the use of these techniques is problematic. In fact, it is this dependency that is of interest in informing us about the influence of the cluster on the individual, and so to assume that it does not exist (individual-level analysis) or to remove its influence by aggregating the individual outcomes (cluster-level analysis) misses the point that this influence is of interest, and as such, should be examined and estimated. Furthermore, there is often interest in assessing effects of both cluster-level variables (e.g., intervention, classroom size) and individual-level variables (e.g., sex,

age, SES), and additionally, interaction of the two (e.g., sex by intervention), on the individual-level outcome.

Clustered designs, where individuals are members of a larger group or context, are common in community psychology and can take a variety of forms. A common example of a clustered design is from school-based interventions. Here, children are typically assessed within various grades and schools, and the culture of the classroom, as well as the school (Sarason, 1972) may have an impact on student responses. Another example in community research is familial studies, where for example, one might be interested in the responses of both parents and children to interventions delivered to the families. It is unlikely that the individuals within a family respond independently, in fact, estimating the degree of within-family dependency is often of interest, in and of itself. Physician- or clinic-based studies are another example of a clustered design where individuals are observed within a broader context that may influence the response of the individuals. Other examples include studies of individuals within social clubs, neighborhoods, wards, households, or worksites. In this article we focus on a smoking cessation program within worksites to demonstrate the utility of the random-effects regression model for clustered designs, but it is important to realize that this approach has many other potential areas of application in community research.

The potential of using worksite health programs as a method of promoting health behavior changes in the community is evident (Fielding, 1984; Glasgow & Terborg, 1988). Smoking cessation programs implemented at worksites have shown encouraging results and appear more effective than clinic-based programs regarding long-term abstinence and participant recruitment (Klesges, Cigrang, & Glasgow, 1987; Schwartz, 1987). The economic and health benefits derived from nonsmoking employees have led many businesses to establish smoking cessation programs. Yet, findings regarding the types of interventions (i.e., social support, incentives, self-help) that are effective have differed. Differences between worksites in attitudes and norms regarding smoking cessation have suggested that the culture of the workplace influences its employees (Sorensen, Pechacek, & Pallonen, 1986). Unfortunately, as is the case with many clustered designs, the impact of the worksite on the employee is often not considered statistically. As noted by Klesges et al. (1987) assignment to treatment conditions is often done at the company level, while the results are analyzed at the individual level. In terms of worksite health interventions, this unit-of-analysis problem may lead to inappropriate conclusions about the effectiveness of such interventions.

Kenny and La Voie (1985) pointed out some of the problems with traditional methods of analysis when individuals within clusters are not in-

dependent, and they developed a technique that adjusts and estimates the correlations at both individual and cluster levels to control for the influence of the other level. Using an example of organizational climate, 25% of the correlations between independent and dependent variables changed from nonsignificant to significant or vice versa when this technique was used, rather than unadjusted correlations (Florin, Giamartino, Kenny, & Wandersman, 1990). These findings suggest that misinterpretations may result when cluster or community-level influences are not accounted for.

If the sample size within clusters is balanced, one might analyze individual data using analysis of variance (ANOVA) models which account for nesting of individuals within clusters, employing a randomized block or split-plots model, for example. These models are termed mixed models since they include both fixed (i.e., intervention group) and random (i.e., cluster) effects. As Hopkins (1982) indicated, when sample size within clusters is the same, proper use of mixed-model ANOVA yields identical results for the test regarding the effect of a cluster-level intervention, regardless of whether cluster means or individual observations are analyzed. Thus, for balanced designs, the mixed-model ANOVA resolves the issue concerning the level of analysis. Unfortunately, the sample size within clusters is often unequal, and further, the inclusion of multiple independent variables or covariates into the model at either the cluster or individual level can quickly render the design unbalanced. As noted by Searle (1987), for unbalanced data the use of the mixed-model ANOVA approach using least squares estimation is problematic.

An alternative way of statistically handling the problem of clustering is to test the cluster as an additional *fixed* factor in the ANOVA model, or similarly, to use dummy-coded variables in a linear regression model to account for the influence of the cluster on the individual outcomes. This approach does not suffer from the computational difficulties that result from using least squares estimation for an unbalanced mixed model, however, it is conceptually unsatisfying since it treats the clusters as representing *all* clusters of interest in the population. Instead, it is more natural to conceptualize the cluster effect as a random effect, meaning that one is interested in characterizing the population of clusters from which the current sample of clusters was drawn. Furthermore, treating the cluster as a fixed effect does not allow one to estimate the population variance associated with the cluster effect, and thus one cannot partition the total variance in the population into that which is due to the individual and that which is due to the cluster.

Another complication in the analysis of clustered data is that often the outcome variable is not a measurement, but instead, a dichotomous (e.g., yes/no or true/false) or ordinal (e.g., none, low, high) classification.

While, again, traditional statistical techniques can be used for individual-level analysis (e.g., logistic regression or log-linear models) or cluster-level analysis (e.g., regression or ANOVA of the proportion of individuals in a cluster responding yes), as noted above, analysis at either level is incomplete and potentially misleading. In sum, traditional methods of analysis often do not adequately account for the impact of the cluster upon the individual, a relationship that warrants examination in community psychology-related studies.

Random-effects regression models (RRM) have been developed and proposed for analysis of unbalanced clustered data in educational (Aitkin & Longford, 1896; Bock, 1989; DeLeeuw & Kreft, 1986; Goldstein, 1987; Raudenbush & Bryk, 1986), psychological (Bock, 1983; Bryk & Raudenbush, 1987), biological (Laird & Ware, 1982; Longford, 1987), epidemiological (Donner, 1985), and psychiatric (Gibbons, Hedeker, Waternaux, & Davis, 1988; Hedeker, Gibbons, & Davis, 1991; Hedeker, Gibbons, Waternaux, & Davis, 1989), literatures. RRM are useful in the analysis of clustered data since outcomes at the individual level are modeled in terms of both individual- and cluster-level variables, while concurrently estimating and adjusting for the amount of intraclass correlation present in the data. Conceptually, RRM can be thought of as representing an extension of the balanced mixed analysis of variance models, like those described by Hopkins (1982), for unbalanced data; as such, RRM make no assumption regarding cluster sample size, allowing for a varying number of individuals within each cluster. Some work has generalized RRM to the case of dichotomous (Stratelli, Laird, & Ware, 1984; Anderson & Aitkin, 1985; Conaway, 1989; Gibbons & Bock, 1987; Goldstein, 1991; Wong & Mason, 1985; Zeger, Liang, & Self, 1985) and ordinal (Ezzet & Whitehead, 1991; Harville & Mee, 1984; Hedeker & Gibbons, 1994; Jansen, 1990) outcome variables.

Hedeker, Gibbons, and Flay (1994) described the use of RRM for clustered continuous response data, where students were observed within classrooms and schools. That paper focused on describing the model terms and their estimation. In this paper we apply RRM to data from a worksite smoking cessation project and compare its use to more traditional statistical techniques. First, we examine a continuous response variable (change in cigarettes per day following the intervention) and compare RRM to ordinary linear regression analysis at both the individual level and the aggregated firm level. This outcome variable is considered first not because it is the primary outcome variable of the study (which it is not), but because researchers are typically more familiar with methods for statistical analysis of continuous response variables (e.g., linear regression and ANOVA models). Since the primary outcome variable from this study is dichotomous

(smoking or nonsmoking), we further illustrate the use of the generalized RRM for noncontinuous response variables to examine the influences upon an individual's smoking status, comparing it to analysis at both the individual and firm level.

WORKSITE STUDY EXAMPLE

Conditions

Employees from 61 worksites participated in one of three conditions that received (a) self-help manuals only (SH), (b) self-help manuals and expectations of incentives (I), or (c) maintenance manuals, expectation of incentives, support groups, and cognitive-behavioral strategies (G).³ As this paper aims primarily to illustrate the use of RRM, results are reported from the posttest, which immediately followed a 3-week intervention.

Subjects

Approximately 400 companies, varying in size and type, were selected from the Chicago Metropolitan Area. Directors of personnel received a letter and follow-up phone call requesting their participation in this study. To participate, companies were required to accept any of the three conditions and hold all group meetings on company time. This prerequisite was intended to identify companies that were similarly interested and invested in the program prior to random assignment. The 63 companies that agreed to participate were matched in terms of size and type, and then randomly assigned to the three conditions. The companies ranged in size from 100–400 employees, and consisted of a variety of types of business (manufacturing, communications, legal, banking, retail, health, and hospitality). One company dropped out immediately prior to the intervention, and another dropped out during the intervention, leaving 61 companies participating in this intervention study.

Measures

Participants in all three conditions were assessed at pretest (immediately prior to the intervention) and posttest (immediately following the 3-week intervention). Individuals were eligible to participate if they were

³For more information regarding the overall design of this longitudinal study, refer to Jason et al., 1991.

classified as cigarette smokers at pretest (i.e., they were not cigar or pipe smokers). The primary study outcome was the smoking status (smoking/nonsmoking) of these individuals following the intervention. To provide an example using RRM for a continuous outcome variable, we also present analysis of the pre- to postintervention change in the number of cigarettes smoked per day. In all, there were 605 individuals from 61 firms who had available data at the postintervention time point in terms of their smoking status (and 577 in terms of their cigarettes per day measurements). The resulting data set was unbalanced with a range of 1 to 29 individuals per firm. Postintervention quit rates, broken down by condition, at both the individual and firm level are given in Table I, as are means for cigarettes per day at preintervention, postintervention, and change between pre and post. As Table I illustrates, the average level of smoking at preintervention was approximately 20 cigarettes per day (i.e., a pack a day). These descriptive statistics are provided to provide an overview of the data and to aid in interpretation of the subsequent RRM results.

RANDOM-EFFECTS REGRESSION MODEL

In the present context, individuals are observed clustered within firms with the intervention being performed in the firms. Here, firms represent

Table I. Smoking Status and Cigarettes Per Day—Descriptive Statistics by Condition^a

| | Employee level | | | Firm level | | |
|-----------------------------|----------------|-------|-------|------------|-------|-------|
| | SH | I | G | SH | I | G |
| Smoking status (<i>n</i>) | 191 | 221 | 193 | 20 | 21 | 20 |
| Proportion quit | | | | | | |
| <i>M</i> | .052 | .122 | .477 | .059 | .119 | .551 |
| <i>SD</i> | — | — | — | .087 | .136 | .251 |
| Cigarettes/day | | | | | | |
| <i>n</i> | 188 | 216 | 187 | | | |
| Preintervention | | | | | | |
| <i>M</i> | 21.95 | 21.47 | 19.64 | 21.61 | 21.65 | 19.82 |
| <i>SD</i> | 9.97 | 11.21 | 10.94 | 5.26 | 5.11 | 6.63 |
| Postintervention | | | | | | |
| <i>M</i> | 16.17 | 13.97 | 5.89 | 15.84 | 14.24 | 6.30 |
| <i>SD</i> | 10.68 | 11.68 | 9.81 | 4.53 | 5.53 | 6.04 |
| Difference | | | | | | |
| <i>M</i> | 5.66 | 7.50 | 13.75 | 5.76 | 7.41 | 13.52 |
| <i>SD</i> | 8.29 | 10.10 | 11.98 | 2.50 | 3.39 | 4.75 |

^aSH = Self-help manuals; I = SH and expectation of incentives; G = maintenance manuals, I, support groups, and cognitive-behavioral strategies.

the clusters of interest, however, in different research settings the clusters (within which the individuals are observed) could be, for example, families, households, classrooms, schools, clinics, neighborhoods, or social clubs. Characteristics of both the individuals and the firms were obtained prior to the start of the intervention, and the individuals were reassessed following the delivery of the intervention. Our interest is in modeling the postintervention outcomes of the individuals, and in particular, examining the effect of the intervention, which is cluster-level variable, on these outcomes. Suppose that we are also interested in controlling for the preintervention level of the outcome variable observed for each individual. Schematically, the regression model for the postintervention outcome may be expressed as follows:

$$\left[\begin{array}{c} \text{Post-Int} \\ \text{Outcome} \end{array} \right] = [\text{Constant}] + \left[\begin{array}{c} \text{Pre-Int} \\ \text{Level} \end{array} \right] + \left[\begin{array}{c} \text{Firm} \\ \text{Condition} \end{array} \right] + \left[\begin{array}{c} \text{Cluster} \\ \text{Effect} \end{array} \right]$$

As represented, there are four independent components which constitute the above model of postintervention outcomes. The first term is a constant term which represents the scale of the outcome variable, much like the grand mean in an analysis of variance, and does not vary by firms or individuals. The second term is at the individual level and represents the influence of preintervention levels. This term reflects the influence a person's preintervention score has on their postintervention score. The third term is at the firm level and represents the influence of the intervention on the outcome. This term is assumed to be same for all firms within the same condition, and thus for all individuals within the same condition. Finally, the cluster effect represents the environmental influence of the firm on the individual outcomes, over and above the influence of the intervention. This final term is assumed to be the same for all individuals within a firm, but varies from firm to firm.

As sampled firms are thought representative of a larger population of firms, the cluster effects are typically considered random effects, and what is of interest is the estimation of the parameters of the population from which these random firms were drawn. In particular, it is usually assumed that the population of cluster effects is a normal distribution and interest is in estimating the degree of variation, or the variance associated with that normal distribution (the mean of the normal distribution is usually assumed to be zero, as the cluster effects are thought to be deviations from an average firm in a given condition). It is the inclusion of the random cluster effect that separates this model from, say, the usual (fixed-effects) multiple regression model, and it is this term that represents the effect of the clustering within firms on the outcome variable. To the degree that clustering within firms has little effect on the outcome data, estimates of

the cluster effects will all be near zero and the estimate of cluster variance in the population will approach zero. In this case the individuals within firms can be considered to be independent, conditional on the effects due to the intervention and preintervention scores, and analysis at the individual level is reasonable since the data within firms are statistically independent. If, on the other hand, clustering within firms has a strong effect on the outcome data, estimates of the cluster effects will deviate from zero and differ for each firm, and thus the population cluster variance will increase in value. While clustering within firms may not have a strong effect on the outcome data, it is likely that the data within firms will be dependent to some degree, the result being that individual within some firms do better than average (the average being based on the firm's condition and the individual's preintervention score) and individuals within other firms do worse than average. Notice that the above formulation of the model illustrates the relationship between the random-effects regression model and ordinary regression analysis at the individual level (ignoring the clustering). Namely, ordinary regression at the individual level is a special case of the random-effects model when all of the cluster effects equal zero.

In addition to the above model terms, there is a residual error term associated with the model that represents the portion of the outcome variable that is unexplained by the other model terms. As the residuals vary from individual to individual, they are at the individual level and are typically assumed to be normally distributed in the population of individuals with zero mean and an unknown, but estimable, variance. As a result, there are two sources of random variation in this model: the residual error terms at the level of the individual, and the cluster effects at the firm level. These two random sources of variation together equal the total amount of variation that is not explained by the independent variables in the model (e.g., preintervention level of the individual and firm condition). The ratio of the cluster variance to this total variance is often called the *intraclass correlation*, which indicates proportion of (unexplained) variance in the data attributable to the cluster. The intraclass correlation then reveals the degree of dependency in the data resulting from the clustering of individuals within firms, and thus indicates the degree to which analysis at the individual level ignoring the clustering effect is problematic. Further, the intraclass correlation is of interest to community psychologists since it quantifies the degree of the person-cluster relationship.

For a dichotomous outcome variable (e.g., smoking/nonsmoking) the above model is usually reframed in terms of a model of the *probability* of smoking versus not smoking. The idea here is that the independent variables in the model influence the degree to which an individual is likely to smoke or not. For this, a useful approach has been to use either a probit

or logistic regression model (Agresti, 1990) to estimate the influence of variables on an individual's probability of smoking. Again, since the firm environment may influence the degree to which individuals within the firm may be likely to smoke or not, it is important to include a cluster effect in the model to assess this potential influence. Schematically, the model for the probability of smoking may be as follows:

$$\left[\begin{array}{c} \text{Probability} \\ \text{of Smoking} \end{array} \right] = F \left\{ \left[\text{Constant} \right] + \left[\begin{array}{c} \text{Firm} \\ \text{Condition} \end{array} \right] + \left[\begin{array}{c} \text{Cluster} \\ \text{Effect} \end{array} \right] \right\}$$

Here, F represents a function that translates the influences of the independent variables to the probability scale (values between 0 and 1) associated with smoking. As mentioned, common choices for this are the probit and logistic transforms. Since all individuals are presumably smoking prior to the intervention, we have taken out of the model the individual's pre-intervention level on the dependent variable (i.e., smoking status). Other individual-level influences, for example, the number of years an individual smoked, could be included into the model instead. Again, as discussed above, the cluster effect should be considered a random effect in this model, leading to a random-effects probit or logistic regression model for these dichotomous outcome data.

Details regarding estimation of the RRM terms are beyond the scope of the present article. Hedeker et al. (1994) discussed parameter estimation for the case of a continuous outcome measurement in a relatively nontechnical way. Other articles dealing with parameter estimation for RRM at a more sophisticated level include Laird and Ware (1982), Longford (1987), Bock (1989), and Bryk and Raudenbush (1992) for continuous outcomes, and Hedeker and Gibbons (1994) for dichotomous and ordinal outcomes. A few points are worth mentioning with regard to parameter estimation. The solution involves estimating two general sets of parameter values: (a) the fixed regression coefficients, the cluster variance (the population variance associated with the random cluster effects), and the residual variance; and (b) each of the random cluster effects. The estimation of each of these two set of parameters depends on parameter values from the other set; if one had values for the parameters in set (a) the parameters of set (b) could be estimated, and vice versa. As a result, methods proposed for the estimation of these two sets of parameters are of an iterative nature, that is, the exact solution is not obtained directly but is obtained following a sequence of approximate solutions (alternating between the two sets of parameters) which ultimately converge on the exact solution for both sets of parameters. Due to the iterative nature of the solution, the amount of computation required is much greater than for common statistical techniques like ANOVA, though this is less of an issue with ever-increasing compu-

tational capacities. Statistical tests of the model parameters are generally of two types: likelihood-ratio tests (Wilks, 1938) and so-called "Wald statistics" (Wald, 1943). The likelihood-ratio test is useful for testing the significance of alternative "nested" models, that is, when a model, say Model B, includes all the parameters of another model, say Model A, plus some additional terms. In this case the likelihood-ratio test provides a way of assessing the statistical significance of the additional terms in Model B not included in Model A. The Wald statistics are useful for testing the statistical significance of individual model parameters. Note that these tests (i.e., the likelihood-ratio tests and tests based on the Wald statistics) are not unique to RRM but are used in other statistical techniques as well, for example, logistic regression (Agresti, 1990).

Changes in Cigarettes Per Day

Since most researchers are more accustomed to statistical models for continuous measurements, we present the results for the analysis of the secondary outcome variable, the change in cigarettes per day, first. These change scores were computed as the difference in numbers of cigarettes per day between pre- and postintervention (i.e., pre minus post). While all individuals were smoking at preintervention, at postintervention some individuals were not smoking (cigarettes per day equal to 0); these change scores then merely reflect the difference in cigarettes smoked per day in preintervention smokers; the actual smoking status (smoking or not smoking) at postintervention is examined later. Prior to analysis of these change scores, histograms were generated and indicated a positively skewed distribution of scores. As a result, a square root transformation of these change scores was used in all of the following statistical analyses. Several regression models were fit to these data. Results from these analyses are given in Table II. In all cases, the change in cigarettes per day was modeled in terms of baseline cigarettes per day, and two condition-related comparisons. The first of these two compared the SH versus I, and the second I versus G. The first two columns of Table II list results for the usual regression analyses at firm level and individual level. Firm-level analysis aggregated individual scores and so modeled the *average* within the firm of change in cigarettes per day in terms of firm-level condition effects, controlling for the *average* preintervention level of cigarettes per day within the firm. The individual-level analysis treated all individuals as independent observations, ignoring the nesting within firms. While both analyses indicate the positive effect of G relative to I, it is interesting to note that different conclusions would be drawn regarding the merit of the I condition relative

to the SH condition: This condition difference is not statistically significant by firm-level analysis but is significant by individual-level analysis. Thus, individual-level analysis indicates that the use of self-help manuals and expectation of incentives is superior to the use of self-help manuals alone, while firm-level analysis indicates no statistically significant difference between these two smoking cessation interventions.

The third column of Table II lists results obtained from analysis using the two-level random-effects regression model. Results from this are similar to those obtained from ordinary regression analysis at the individual level, at least in terms of the parameter estimates. However, a difference does emerge which could influence the interpretation drawn from the data. Notice that, unlike ordinary individual-level analysis, random-effects regression analysis indicates that the difference between the I and SH conditions is only marginally significant, and is not statistically significant at the $p < .05$ level. The main reason for this difference is not due to differences in terms of the parameter estimates (since these are very similar), but rather to differences in the standard errors associated with the parameters. This is not surprising since the standard errors reflect the uncertainty in the parameter estimates, and this uncertainty decreases as the amount of independent

Table II. Change in Cigarettes Per Day (Square Root of Pre-Post Intervention Scores)
Linear Regression Estimates (and Standard Errors)

| | Ordinary regression models | | Random-effects regression model |
|----------------------|--------------------------------|--------------------------------|---------------------------------|
| | Firm level | Individual level | Individuals in firms |
| <i>n</i> | 61 | 577 | 577 |
| Constant | 6.689 ^c (0.119) | 6.400 ^c (0.061) | 6.401 ^c (0.065) |
| Baseline Cigs/day | 0.016 ^c (0.005) | 0.029 ^c (0.003) | 0.029 ^c (0.003) |
| SH vs. I | -0.054 (0.037) | -0.069 ^b (0.032) | -0.068 ^a (0.040) |
| I vs. G | -0.229 ^c (0.037) | -0.243 ^c (0.032) | -0.247 ^c (0.041) |
| Residual variance | 0.056 | 0.399 | 0.374 ^c (0.023) |
| Cluster variance | | | 0.021 ^b (0.011) |

^a $p < .10$.

^b $p < .05$.

^c $p < .01$.

data used in the calculations is increased. Thus, whereas the ordinary individual-level analysis bases its standard errors on the assumption that the 577 observations are independent, the random-effects analysis bases its standard errors on having 577 observations with a certain degree of dependency (and therefore less independent information).

Smoking Status

Turning to the analysis of the main outcome measure, the individual's smoking status at postintervention, we turn to statistical models for categorical, rather than continuous, data. Thus, instead of ordinary linear regression we must use regression procedures for dichotomous outcomes, most notably, either probit or logistic regression models. However, although this is the case for analysis at the individual-level and for the random-effects analysis, analysis at the firm level can still use ordinary linear regression techniques. The reason for this is that at the firm level the outcome variable is the proportion of individuals within the firm who are abstinent, and so by aggregating a dichotomous outcome variable (individual smoking status) at the level of the firm, a continuous variable (the proportion of abstinent individuals) is obtained.

Results from these analyses are given in Table III. In these analyses, the same two condition-related effects were examined: the SH versus I difference, and the I versus G difference. The first two columns of Table III give results for the regression analysis at the firm level of the proportion of abstainers, and the probit regression analysis of abstinence versus smoking at the individual level, respectively. As can be seen, the results are similar to those observed in the analysis of cigarettes per day change scores. Namely, both analyses indicate the positive effect of the G condition relative to the I condition, and yield different conclusions regarding the I versus SH difference. Again, individual-level analysis indicates that the use of self-help manuals and expectation of incentives is superior to the use of self-help manuals alone, while firm-level analysis indicates no statistically significant difference between these two.

The third column of Table III lists results obtained from analysis using a two-level random-effects probit regression model (as described in Gibbons & Hedeker, 1994). Again, parameter estimates based on this analysis are similar to those obtained from ordinary probit regression analysis at the individual level. However, as opposed to the analysis of changes in cigarettes per day, here the same conclusion is reached in terms of the I versus SH difference. Namely, both approaches yield a statistically significant difference at the $\alpha = .05$ level, although the level of significance is clearly

Table III. Postintervention Smoking Status Linear and Probit Regression Estimates (and Standard Errors)

| | Ordinary regression model ^a | Probit regression model | Random-effects probit regression model |
|------------|---|--------------------------------|--|
| | (Firm level) | (Individual level) | (Individuals in firms) |
| n | 61 | 605 | 605 |
| Constant | 0.305 ^c (0.027) | -0.841 ^c (0.088) | -0.860 ^c (0.116) |
| SH vs. I | -0.030 (0.027) | -0.229 ^b (0.093) | -0.224 ^b (0.111) |
| I vs. G | -0.216 ^c (0.027) | -0.553 ^c (0.071) | -0.593 ^c (0.088) |
| Cluster SD | | | 0.263 ^b (0.140) |

^aNote that the dependent variable for the firm-level analysis is the *proportion* of abstinent individuals within the firm, whereas the individual-level and random-effects probit analyses model the status of the individual (abstinent or smoking). As a result, the scale of the effects is quite different between models; comparisons with the firm-level analysis should focus instead on the significance tests associated with the model effects.

^b $p < .05$.

^c $p < .01$.

more in favor of rejection of the null hypothesis for the ordinary individual-level analysis ($p < .014$ for the individual-level analysis, and $p < .042$ for the random-effects analysis). It should be emphasized that it was somewhat fortuitous that the p value associated with the RRM analysis ($p < .042$) did not exceed the .05 criterion, and thus that the same *general* conclusion was obtained for both individual-level and RRM analysis. Clearly, as our previous example using changes in cigarettes per day indicated, this is not always the case.

Intraclass Correlation

The amount of dependency present in the data as a result of the clustering of individuals within firms was statistically significant, as evidenced by significant cluster variance terms for both models of changes in cigarettes per day and smoking status. For the changes in cigarettes per day, intraclass correlation (which equals the cluster variance divided by the sum of the cluster variance plus the residual variance) was equal to 0.053, indicating that clustering of individuals within firms accounted for roughly

5% of the variability in the data not explained by the model covariates. Similarly, for smoking status the intraclass correlation was estimated as 0.065 (in the case of a probit regression the residual variance is assumed to be equal to 1, see Finney, 1971), indicating that roughly 7% of the unexplained variance was attributable to the clustering of individuals within firms. Thus, even with rather low levels of intraclass correlation, it is important to note that conclusions regarding other model parameters can change, as did, for example, the conclusion regarding the significance of the self-help versus incentives difference. In addition, the size of the intraclass correlation is meaningful for community psychologists, as it enables us to better understand the extent to which individuals are influenced by their specific environments.

DISCUSSION

This example illustrates that choice of statistical models can influence conclusions drawn from a given data set. In the current context, all of the statistical models assessed significance of two condition-related comparisons: SH versus I, and I versus G. While all models examined in this paper indicated a significant positive effect of the G condition relative to the I condition, results concerning significance of the I condition compared to the SH condition were mixed. Using an aggregate approach at the firm level, no statistically significant difference was observed between these two conditions. On the other hand, ordinary regression analysis at the individual level indicated a positive effect of the I relative to the SH condition. Based on random-effects regression analysis, this difference was not statistically significant when examining changes in cigarettes per day, and just statistically significant when examining smoking status. Thus, the choice of statistical analysis performed clearly has an impact on the conclusions drawn from this study.

Regarding parameter estimates, the estimates from the individual-level analyses were reasonably similar to the estimates obtained from random-effects regression models. However, the estimates of the corresponding standard errors were much smaller for ordinary regression analysis compared to RRM analyses. Underestimation of standard errors by ordinary regression analysis was observed for nearly all the model parameters, and is expected since ordinary regression assumes independence of observations. Since the observations were not independent, the amount of independent information available in parameter estimation is inflated.

Levels of intraclass correlation observed in these data (between 5 and 7%) were of moderate size and are consistent with reported levels from

the literature. For example, Donner (1982) observed intraclass correlation levels from 5 to 12% for data from spouse pairs, 0.16 to 1.26% for data from physician practices, and 0.05 to 0.85% for data clustered by counties. In a worksite smoking-cessation study similar to the one reported here, Jacobs, Jeffery, and Hannan (1989) observed intraclass correlation levels between 0.9 to 2.4% for various smoking outcome variables. The intraclass correlation levels we observed were larger than those of Jacobs et al., (1989) even though these two studies were similar in design and in terms of the outcome variables. This is reasonable, though, since the average worksite size in their study was 420, whereas our range was between 100 and 400, and as noted by Kish (1965), intraclass correlation tends to be larger for smaller clusters. Thus, the size and type of cluster can influence the level of intraclass correlation.

The example given illustrates the use of RRM for the analysis of data from a worksite study, additionally, this model can be useful in the analysis of other types of clustered data sets, for example, data clustered by families, households, classrooms, schools, clinics, neighborhoods, or social clubs. Another common clustering in psychological research occurs when data are clustered by raters. In this case, each of, say, N raters provide measures (perhaps a measure of psychological functioning) on different individuals, with the raters typically measuring a different number of individuals as well. One might then be interested in assessing how much of the total variance in the measured variable is attributable to the raters, as well as the impact any rater-level covariates might have on the measures (perhaps the experience of the rater was thought to have an influence on the rating). RRM would be useful for data of this type, since it provides an estimate of the variability attributable to the cluster (in this case, the rater) as well as examining the effect of cluster- or individual-level covariates on the dependent measure of the model (which is measured at the level of the individual).

Another context for clustered data is in the area of repeated measures data. Here, the repeated measures are observed clustered within individuals. Notice that in the present example focusing on changes in cigarettes per day, there were two repeated observations clustered within individuals (pre- and postintervention scores). With only two repeated observations, the longitudinal nature of the data can be modeled by analyzing change scores or by analyzing the postscore covarying out the prescore, however, with more repeated observations per individual a longitudinal approach is more appropriate and informative. In general, random-effects models for longitudinal data extend the model presented above by including multiple random effects to account for the clustering of observations within individuals. For instance, one might allow individuals to vary in terms of both their intercept, or starting point, and their trend across time, yielding a

model with two random effects at the individual level. As others have noted, these more general random-effects models provide a powerful approach to the analysis of longitudinal data (Bock, 1983; Bryk & Raudenbush, 1987; Gibbons et al., 1993; Laird & Ware, 1982).

Computer Programs

There has been an increasing development of computer software for the random-effects regression model. Several commercially available software programs exist for computing RRM, although these programs exist primarily for models of continuous dependent variables. The BMDP3V (Jennrich & Sampson, 1988) procedure is available as part of the general BMDP statistical package for both mainframe and DOS-based personal computers (PCs). SAS has introduced the MIXED procedure in version 6.07 which runs only on mainframe computers, however this procedure will soon be available for PC versions of SAS as well. Other software programs—ML3 (Prosser, Rasbash, & Goldstein, 1991), HLM (Bryk, Raudenbush, Seltzer, & Congdon, 1989), VARCL (Longford, 1986), MIXREG (Hedeker, 1992b)—are stand-alone programs which are primarily suited for use on DOS-based personal computers, though available for some types of mainframe computers. All of these programs were written by researchers with published articles on random-effects models, and the manuals refer to these articles as examples and to clarify the use of RRM. Though all of the aforementioned programs are designed primarily for analysis of a continuous dependent variable, for analysis of a categorical dependent variable (i.e., dichotomous or ordinal) the amount of available software is limited. ML3 and the software package EGRET have facilities for performing some types of (dichotomous) logistic regression models with random effects. An alternative program, MIXOR (Hedeker, 1992a, available from its author), has been developed as a DOS-based PC program to estimate an ordinal random-effects (logistic or probit) regression model. MIXOR can handle and estimate the influence of multiple random-effects, individual-level and cluster-level covariates, and dichotomous or ordinal outcomes.

Conclusion

Our example highlights features of analysis using RRM, and how analysis by more traditional regression models can lead to biased estimates of uncertainty and different conclusions. As demonstrated, RRM provides a powerful tool for analysis of clustered data where the number of individuals within clusters varies. Alternatively, as noted by many researchers

(Burstein, 1980; Jacobs et al., 1989; Koepke & Flay, 1989; McKinlay, Stone, & Zucker, 1989; Murray, Hannan & Zucker, 1989; Raudenbush & Bryk, 1988-1989), traditional analysis at the individual or cluster level is problematic. Traditional analysis at the individual level ignores dependency in the data that results from clustering, while analysis at the cluster level does not permit a straightforward analysis of individual-level characteristics. By controlling for and estimating the degree of dependency resulting from the clustering of data, RRM allows testing of relationships at the individual level in a manner consistent with data collection, namely, from individuals within clusters. Furthermore, RRM can include independent variables at either the cluster or individual levels and estimate their effects as well as interactions.

The usefulness of RRM is particularly noteworthy when dealing with research designs like those utilized in worksite smoking cessation programs. Since the impact of the worksite on the individual can be identified and accounted for, we can begin to investigate what smoking interventions are most effective in specific types of worksites. If smoking interventions can be tailored to worksites, long-term abstinence rates may improve. The possibility of examining both effects of the individual and the worksite on individual smoking outcomes in a cohesive and comprehensive manner should thus help to overcome the fragmented results that necessarily emerge when the focus is on either of the two levels of data to the exclusion of the other.

Since community psychologists often develop and evaluate interventions that involve natural clustering within larger groups (i.e., schools, neighborhoods, families, organizations), RRM can be particularly useful. Through the examination of the influence of context on individuals, and the relationships that evolve within communities, RRM can facilitate a better understanding of communities. Further, mixing and matching levels of statistical analysis, measurement, and conceptualization can result in inaccurate inferences (Shinn, 1990). RRM provides an example of a statistical technique that fits the theory, principles, and values of community psychology, as it enables us to better understand higher levels of aggregation and their impact on individual well-being.

REFERENCES

- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society, Series A*, 149, 1-43.
- Anderson, D., & Aitken, M. (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society, Series B*, 47, 203-210.

- Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.
- Bock, R. D. (1983). The discrete Bayesian. In H. Wainer & S. Messick (Eds.), *Modern advances in psychometric research* (pp. 103-115). Hillsdale, NJ: Erlbaum.
- Bock, R. D. (1989). Measurement of human variation: A two stage model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 319-342). New York: Academic Press.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147-158.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. (1989). *An introduction to HLM: Computer program and users' guide*. Chicago: Scientific Software.
- Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. Berliner (Ed.), *Review of research in education* Vol. 8, pp. 158-233. Washington, DC: American Educational Research Association.
- Conaway, M. R. (1989). Analysis of repeated categorical measurements with conditional likelihood methods. *Journal of the American Statistical Association*, *84*, 53-61.
- DeLeeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*, 57-85.
- Donner, A. (1982). An empirical study of cluster randomization. *International Journal of Epidemiology*, *11*, 283-286.
- Donner, A. (1985). A regression approach to the analysis of data arising from cluster randomization. *International Journal of Epidemiology*, *14*, 322-326.
- Ezzet, F., & Whitehead, J. (1991). A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine*, *10*, 901-907.
- Fielding, J. (1984). Health promotion and disease prevention at the worksite. *Annual Review of Public Health*, *5*, 237-265.
- Finney, D. J. (1971). *Probit analysis* (3rd ed.). New York: Cambridge University Press.
- Florin, P., Giamartino, G. A., Kenny, D. A., & Wandersman, A. (1990). Levels of analysis and effects: Clarifying group influence and climate by separating individual and group effects. *Journal of Applied Social Psychology*, *20*, 881-900.
- Gibbons, R. D., & Bock, R. D. (1987). Trend in correlated proportions. *Psychometrika*, *52*, 113-124.
- Gibbons, R. D., & Hedeker, D. (1994). Application of random-effects probit regression models. *Journal of Consulting and Clinical Psychology*, *62*, 285-296.
- Gibbons, R. D., Hedeker, D., Elkin, I., Waternaux, C., Kraemer, H. C., Greenhouse, J. B., Shea, M. T., Imber, S. D., Sotsky, S. M., & Watkins, J. T. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry*, *50*, 739-750.
- Gibbons, R. D., Hedeker, D., Waternaux, C., & Davis, J. M. (1988). Random regression models: a comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacology Bulletin*, *24*, 438-443.
- Glasgow, R., & Terborg, J. (1988). Occupational health promotion programs to reduce cardiovascular risk. *Journal of Consulting and Clinical Psychology*, *56*, 365-373.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.
- Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, *78*, 45-51.
- Harville, D. A., & Mee, R. W. (1984). A mixed-model procedure for analyzing ordered categorical data. *Biometrics*, *40*, 393-408.
- Hedeker, D. (1992a). *MIXOR: A Fortran program for mixed-effects ordinal probit and logistic regression*. Technical Report, School of Public Health, University of Illinois at Chicago.
- Hedeker, D. (1992b). *MIXREG: A Fortran program for mixed-effects linear regression with auto-correlated errors*. Technical Report, School of Public Health, University of Illinois at Chicago.

- Hedeker, D., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel data. *Biometrics*, 50.
- Hedeker, D., Gibbons, R. D., & Davis, J. M. (1991). Random regression models for multicenter clinical trials data. *Psychopharmacology Bulletin*, 27, 73-77.
- Hedeker, D., Gibbons, R. D. & Flay, B. R. (1994). Random-effects regression models for clustered data: With an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, 62, 757-765.
- Hedeker, D., Gibbons, R. D., Waternaux, C., & Davis, J. M. (1989). Investigating drug plasma levels and clinical response using random regression models. *Psychopharmacology Bulletin*, 25, 227-231.
- Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal*, 19, 5-18.
- Jacobs, D. R., Jeffery, R. W., & Hannan, P. J. (1989). Methodological issues in worksite health intervention research: II. Computation of variance in worksite data: Unit of analysis. In K. Johnson, J. H. LaRosa, C. J. Scheirer, et al. (Eds.), *Proceedings of the 1988 methodological issues in worksite research conference* (pp. 77-88). Airlie, VA: United States Department of Health and Human Services.
- Jansen, J. (1990). On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, 39, 75-84.
- Jason, L., Salina, D., Hedeker, D., Kimball, P., Kaufman, J., Bennett, P., Bernstein, R., & Lesondak, L. (1991). Designing an effective worksite smoking cessation program using self-help manuals, incentives, groups and media. *Journal of Business and Psychology*, 6, 155-166.
- Jennrich, R. I., & Sampson, P. F. (1988). 3V: General mixed model analysis of variance. In W. J. Dixon (Chief Ed.), *BMDP statistical software manual* (Vol. 2, pp. 1025-1043). Berkeley: University of California Press.
- Kelly, J. G. (1966). Ecological constraints on mental health services. *American Psychologist*, 21, 535-539.
- Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology*, 48, 339-348.
- Kish, L. (1965). *Survey sampling*. New York: Wiley.
- Klesges, R. C., Cigrang, J., & Glasgow, R. E. (1987). Worksite smoking modification programs: A state-of-the-art review and directions for future research. *Current Psychological Research & Reviews*, 6(1), 26-56.
- Koepke, D., & Flay, B. R. (1989). Levels of analysis. In M. T. Braverman (Ed.), *Evaluating health promotion programs. New directions for program evaluation* (No. 43, pp. 75-87). San Francisco: Jossey-Bass.
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Levine, M., & Perkins, D. V. (1987). *Principles of community psychology: Perspectives and applications*. New York: Oxford University Press.
- Linney, J. A., & Reppucci, N. D. (1982). Research design and methods in community psychology. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 535-566). New York: Wiley.
- Longford, N. T. (1986). VARCL—Interactive software for variance component analysis. *Professional Statistician*, 74, 817-827.
- Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817-827.
- McKinlay, S. M., Stone, E. J., & Zucker, D. M. (1989). Research design and analysis issues. *Health Education Quarterly*, 16, 307-313.
- Moos, R. H. (1976). *The human context: Environmental determinants of behavior*. New York: Wiley.
- Murray, D. M., Hannan, P. J., & Zucker, D. M. (1989). Analysis issues in school-based health promotion studies. *Health Education Quarterly*, 16, 315-320.
- Prosser, R., Rasbash, J., & Goldstein, H. (1991). *ML3 software for three-level analysis, users' guide for v.2*. London: Institute of Education, University of London.

- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Raudenbush, S. W., & Bryk, A. S. (1988-89). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 423-475). Washington, DC: American Educational Research Association.
- Sarason, S. B. (1972). *The culture of the school and the problem of change*. Boston: Allyn & Bacon.
- Schwartz, J. L. (1987, April). *Review and evaluation of smoking cessation methods: The United States and Canada, 1975-1985* (DHHS No. 87-2940). Washington, DC: National Cancer Institute.
- Searle, S. R. (1987). *Linear models for unbalanced data*. New York: Wiley.
- Shinn, M. (1990). Mixing and matching: Levels of conceptualization, measurement, and statistical analysis in community research. In P. Tolan, C. Keys, F. Chertok, & L. Jason (Eds.) *Researching community psychology: Issues of theory and methods* (pp. 111-126). Washington, DC: American Psychological Association.
- Stirateli, R., Laird, N. M., & Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics*, 40, 961-971.
- Sorensen, G., Pechacek, & Pallonen, U. (1986). Occupational and worksite norms and attitudes about smoking cessation. *American Journal of Public Health*, 76, 544-549.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426-482.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9, 60-62.
- Wong, G. Y., & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80, 513-524.
- Zeger, S. L., Liang, K.-Y., & Self, S. G. (1985). The analysis of binary longitudinal data with time independent covariates. *Biometrika*, 72, 31-38.