Addiction (2000) 95(Supplement 3), S381–S394

MAKING THE MOST OUT OF DATA ANALYSIS AND INTERPRETATION

Analysis of longitudinal substance use outcomes using ordinal random-effects regression models

DONALD HEDEKER[1] & ROBIN J. MERMELSTEIN[2]

[1]*Division of Epidemiology and Biostatistics, and Health Research and Policy Centers, University of Illinois at Chicago &* [2]*Health Research and Policy Centers, and Department of Psychology, University of Illinois at Chicago, IL, USA*

Abstract

In this paper we describe analysis of longitudinal substance use outcomes using random-effects regression models (RRM). Some of the advantages of this approach is that these models allow for incomplete data across time, time-invariant and time-varying covariates, and can estimate individual change across time. Because substance use outcomes are often measured in terms of dichotomous or ordinal categories, our presentation focuses on categorical versions of RRM. Specifically, we present and describe an ordinal RRM that includes the possibility that covariate effects vary across the cutpoints of the ordinal outcome. This latter feature is particularly useful because a treatment can have varying effects on full versus partial abstinence, for example. Data from a smoking cessation study are used to illustrate application of this model for analysis of longitudinal substance use data.

Introduction

Longitudinal studies play an important role in substance use research. In these studies, levels of substance use are measured repeatedly across a series of timepoints, and the goal is often to examine the effects of different treatments and/or predictors on usage levels across time. In some cases, there is interest only in the final measurement, or the repeated usage levels can be aggregated to provide a single outcome per subject, for example, an average substance use level or a simple difference in substance use (e.g. pre- to post-change). In these cases, standard statistical analysis procedures can be readily applied. However, these approaches are limited because they either ignore change across time (i.e. end-point or averaged value analysis), or they only consider within-subjects change that is linear (i.e. pre- to post-change analysis). Furthermore, if subjects from different treatment groups drop out differentially across time, then treatment group and time are confounded in these analyses. Inclusion of time-varying covariates presents another problem for these approaches. Finally, from a statistical point of view these approaches are inefficient because much of the data that are collected are

Correspondence 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 525, Chicago, IL 60612–7260, USA. e-mail: hedeker@uic.edu

Submitted 22nd February 2000; initial review completed 25th May 2000; final version accepted 10th July 2000.

ISSN 0965–2140 print/ISSN 1360–0443 online/00/11S381–14

Carfax Publishing, Taylor & Francis Limited

DOI: 10.1080/09652140020004296

ignored. For these reasons, development of more general statistical methods for longitudinal data analysis has been an active area of statistical research in recent years. In particular, random-effects regression models (RRM) have been developed to overcome many of these limitations. With the concurrent development of software, RRM represent a popular approach for analysis of longitudinal data.

Variants of RRM have been developed under a variety of names: random-effects models (Laird & Ware, 1982), variance component models (Dempster, Rubin & Tsutakawa, 1981); hierarchical linear models (Bryk & Raudenbush, 1992), multilevel models (Goldstein, 1995), two-stage models (Bock, 1989), random coefficient models (de Leeuw & Kreft, 1986), mixed models (Longford, 1987; Wolfinger, 1993), empirical Bayes models (Hui & Berger, 1983; Strenio, Weisberg & Bryk, 1983), unbalanced repeated-measures models (Jennrich & Schlucter, 1986) and random regression models (Bock, 1983a, 1983b). A basic characteristic of these models is the inclusion of random subject effects into regression models in order to account for the influence of subjects on their repeated observations. These random subject effects thus describe and explain the correlational structure of the longitudinal data. Additionally, they indicate the degree of subject variation that exists in the population of subjects.

There are several features that make RRM especially useful in longitudinal research. First, subjects are not assumed to be measured on the same number of timepoints, thus, subjects with incomplete data across time are included in the analysis. The ability to include subjects with incomplete data across time is an important advantage relative to procedures that require complete data across time because (a) by including all data, the analysis has increased statistical power, and (b) complete-case analysis may suffer from biases to the extent that subjects with complete data are not representative of the larger population of subjects. Because time is treated as a continuous variable in RRM, subjects do not have to be measured at the same timepoints. This is useful for analysis of longitudinal studies where follow-up times are not uniform across all subjects. Both time-invariant and time-varying covariates can be included in the model. Thus, changes in the outcome variable may be due to both stable characteristics of the subject (e.g. their gender or race) as well as characteristics that change across time (e.g. life-events). Finally, whereas traditional approaches estimate average change (across time) in a population, RRM can also estimate change for each subject. These estimates of individual change across time can be particularly useful in longitudinal studies where a proportion of subjects exhibit change across time that deviates from the average trend. For example, while most subjects over time may show an increase in drinking or smoking, there may be a proportion of subjects who do not.

In terms of measurement scale, substance use outcomes are often categorical in nature. For example, subjects may simply be dichotomously classified as being abstinent or not at each timepoint. Alternatively, subjects might be classified in terms of degrees of use as none, mild, moderate or severe. Because substance use classifications like these are often the primary outcome variables in a study, statistical models for dichotomous or ordinal responses are particularly germane for analysis of alcohol and smoking data. In this regard, generalizations of RRM have been developed for dichotomous response data (Wong & Mason, 1985; Gibbons & Bock, 1987; Goldstein, 1991; Rosner, 1992; Stiratelli, Laird & Ware, 1984; Wolfinger & Lin, 1997) and ordinal response data (Jansen, 1990; Ezzet & Whitehead, 1991; Hedeker & Gibbons, 1994; Ten Have, 1996), thus allowing a general framework for analysis of both continuous and categorical longitudinal outcomes.

As these methods have been developed and used more widely, application of RRM for substance use data has grown. For longitudinal smoking data, Hu *et al.* (1998) reviews and compares RRM to the generalized estimating equations (GEE) method for analysis of longitudinal dichotomous data. For longitudinal ordinal smoking outcomes, Hedeker & Mermelstein (1996) describe and illustrate use of ordinal RRM. Nich & Carroll (1997) compare RRM to traditional repeated measures analysis of variance for analysis of continuous drug composite scores. In addition to these papers focusing on RRM description and dissemination, several outcomes-orientated papers using RRM for analysis of longitudinal substance use data include Bernstein *et al.* (1994), Carroll *et al.* (1994, 1998), Chassin *et al.* (1996), Gallagher *et al.* (1997), Gruder *et al.* (1993), Halikas *et al.* (1997), Jason *et al.* (1997), O'Malley *et al.* (1996)

and Salina *et al.* (1994). These latter papers are especially useful for illustrating how results from RRM can be described and reported.

More generally, several textbooks describing RRM for longitudinal data analysis have recently been published (Bryk & Raudenbush, 1992; Longford, 1993; Diggle, Liang & Zeger, 1994; Goldstein, 1995; Hand & Crowder, 1996) and review, comparison, and/or tutorial articles on longitudinal data analysis treating RRM have proliferated (Gibbons *et al.*, 1993; Burchinal, Bailey & Snyder, 1994; Manor & Kark, 1996; Cnaan, Laird & Slasor, 1997; Everitt, 1998; Albert, 1999; Delucchi & Bostrom, 1999; Keselman *et al.*, 1999; Lesaffre, Asefa & Verbeke, 1999; Omar *et al.*, 1999; Sullivan, Dukes & Losina, 1999). Most of these papers concern continuous response variables, although ones dealing specifically with dichotomous outcomes have also appeared (Zeger & Liang, 1992; Fitzmaurice, Laird & Rotnitzky, 1993; Gibbons & Hedeker, 1994; Pendergast *et al.*, 1996), although most of these are somewhat technical.

Statistical software to perform RRM analysis has also proliferated, especially for continuous outcomes (SAS PROC MIXED; BMDP 5V; HLM: Bryk, Raudenbush & Congdon, 1996; MIXREG: Hedeker & Gibbons, 1996b; MLwiN: Goldstein *et al.*, 1998). For categorical data, software has become available for dichotomous (EGRET, Statistics and Epidemiology Research Corporation, 1991) and ordinal or nominal outcomes (SAS PROC NLMIXED; HLM; MLwiN; MIXOR: Hedeker & Gibbons, 1996a; MIXNO: Hedeker, 1999). Of course, software for nominal and ordinal outcomes can be used to fit models for dichotomous outcomes. Review articles comparing some of these software programs include Kreft *et al.*, (1994), van der Leeden, Vrijburg & de Leeuw (1996) and de Leeuw & Kreft (1999).

Because substance use outcomes are primarily categorical, we will focus on RRM for categorical outcomes in this paper. Also, since a dichotomous outcome is simply a special case of an ordinal outcome with only two categories, we will consider ordinal outcomes only. For longitudinal ordinal outcomes, most of the models include an assumption called the proportional odds assumption (McCullagh, 1980). A proportional odds model, for an ordinal response with K categories, assumes that the effect of a covariate is proportional across the model's $K-1$ cumulative odds, or homogeneous across the corresponding $K-1$ cumulative logits (i.e. the log odds). For alcohol and smoking outcomes this assumption may not be reasonable. For example, suppose that there are three categories (abstinence, mild use, severe use) and suppose that an intervention is not successful in increasing the proportion of individuals in the abstinence category but is successful in moving individuals from severe to mild use. In this case, the effect of intervention group (i.e. the covariate) would not be observed on the first cumulative odds (i.e. comparing abstinence versus the two use categories combined), but would be observed on the second cumulative odds (i.e. comparing abstinence and mild use combined versus the severe use category).

Recently, Hedeker & Mermelstein (1998) described an extension of the random-effects proportional odds model to allow for non-proportional odds for the covariate effects. In this extended model, covariates can be specified either requiring or relaxing the proportional odds assumption. As applied to longitudinal substance use data, this model allows the influence of covariates (e.g. time and intervention group) to vary, for example, across the levels of alcohol or smoking usage. Thus, variables can have different, or heterogeneous, effects on mild and severe usage. In this paper, data from a longitudinal smoking cessation study will be used to illustrate application of this more general model.

RRM for longitudinal categorical data

RRM for categorical outcomes generally adopt either a probit or logistic regression model and utilize various methods for incorporating and estimating the influence of the random effects. A recent review paper (Pendergast *et al.*, 1996) presents and describes many of these approaches. In general, parameter estimation is computationally more intensive for these models than for models of continuous outcome data. In this article, we will focus on application of the model. Interested readers should consult the Pendergast *et al.* review paper for details on estimation. Also, because the logistic regression model is probably more commonly used than the probit regression model, we will present the model in terms of the logistic response function.

For an introduction on the use of the probit model for longitudinal dichotomous outcomes, see Gibbons & Hedeker (1994).

With the logistic response function, the model for subject i ($i = 1, 2, \ldots, N$ subjects) on occasion j ($j = 1, 2, \ldots, n_i$ occasions) can be written in terms of the log odds of response for a dichotomous outcome Y (with values, for instance 1 and 2) as:

$$\log\left[\frac{P(\mathrm{Y}_{ij}=1)}{1-P(\mathrm{Y}_{ij}=1)}\right] = \alpha_0 + \beta_1 t_{ij} +$$

log odds of 1 response = constant factor + time effect +

$\beta_2 x_i$ + $\beta_3 x_{ij}$ + v_i

subject-varying covariate + subject- and time-varying covariate + subject specific effect

The numerator is the probability of a 1 response, and the denominator $1 - P(\mathrm{Y}_{ij}=1)$ equals the probability of a 2 response. The ratio of these probabilities is the odds of a 1 response, and the log of this ratio is the log odds of a 1 response (sometimes called the logit of P). Notice that the log odds is equal to 0 when the probability of a 1 response equals 0.5 (i.e. equal odds of a response in category 1 and category 2), is negative when the probability is less than 0.5 (i.e. odds favoring a response in category 2), and is positive when the probability is greater than 0.5 (i.e. odds favoring a response in category 1).

In terms of the regression parameters, α_0 is the intercept, β_1 is the coefficient for the effect of time t_{ij}, β_2 is the coefficient for the time-invariant covariate x_i, and β_3 is the coefficient for the time-varying covariate x_{ij}. Notice that the subscripts for the covariates indicate whether the variable varies by subjects (i) or across subjects and time (ij). Covariate interactions can be included in the same way as interactions are included into the usual multiple regression model. For example, in the above model x_i might represent the type of treatment that a subject is assigned to for the course of the study, while x_{ij} might be the treatment by time interaction which is obtained as the product of x_i by t_{ij}.

The remaining term, v_i, which represents the influence of subject i on the log-odds of response across all time-points, is what separates the above model from an ordinary (fixed-effects) logistic regression model. This term indicates the influence of individual i on his/her repeated observations, and because individuals in a sample are thought typically to be representative of a larger population of individuals, the individual-specific effects v_i are treated as random effects. That is, v_i are considered to be representative of a distribution of individual effects in the population. The most common assumed form for this population distribution is the normal distribution with mean 0 and variance σ^2_v. Because there is only one random individual effect in this model, it is sometimes referred to as a random-intercepts model, with each v_i indicating how individual i deviates from the (fixed-effects part of the) model.

This model for the dichotomous outcome can also be written as:

$$\log\left[\frac{P(\mathrm{Y}_{ij}\leq 1)}{1-P(\mathrm{Y}_{ij}\leq 1)}\right] = \alpha_0 + \beta_1 t_{ij} + \beta_2 x_i + \beta_3 x_{ij} + v_i$$

which then generalizes to the random-intercepts ordinal logistic regression model:

$$\log\left[\frac{P(\mathrm{Y}_{ij}\leq k)}{1-P(\mathrm{Y}_{ij}\leq k)}\right] = \alpha_{0k} + \beta_1 t_{ij} + \beta_2 x_i + \beta_3 x_{ij} + v_{ij}, \quad k = 1, \ldots, K-1, \tag{1}$$

where α_{0k} are the $K - 1$ intercept terms to model the marginal frequencies in the K ordered categories. In this representation, a positive value for a regression coefficient β indicates a negative association between Y and the covariate, in the sense that large values of Y are relatively less likely to occur for larger values of the covariate. Thus, if the outcome is coded as 1 = no use, 2 = minimal use, and 3 = heavy use, a covariate with a positive regression coefficient would imply less use with higher values of the covariate.

The above model makes what is called the "proportional odds assumption". This proportional odds characterization for ordinal response models, discussed in detail by Agresti (1989), has some important features. The left-hand side of the equality in Equation (1) specifies $K - 1$ cumulative logits each contrasting the combined first k categories to the remaining combined $(K - k)$ categories. For example, with four possible response categories (coded as

1, 2, 3 or 4), the following three cumulative logits are indicated by the model:

$$\log\left[\frac{P(Y_{ij}\leq 1)}{1-P(Y_{ij}\leq 1)}\right]=\log\left[\frac{P(Y_{ij}=1)}{P(Y_{ij}=2,3 \text{ or } 4)}\right]$$

$$\log\left[\frac{P(Y_{ij}\leq 2)}{1-P(Y_{ij}\leq 2)}\right]=\log\left[\frac{P(Y_{ij}=1 \text{ or } 2)}{P(Y_{ij}=3 \text{ or } 4)}\right]$$

$$\log\left[\frac{P(Y_{ij}\leq 3)}{1-P(Y_{ij}\leq 3)}\right]=\log\left[\frac{P(Y_{ij}=1,2 \text{ or } 3)}{P(Y_{ij}=4)}\right]$$

As the regression coefficients β do not carry the k subscript, it is assumed that the effect of a covariate is homogeneous across these $K-1$ cumulative logits, or proportional across the cumulative odds. The odds of a response in a category greater than K (for any fixed K) is multiplied by exp (β) for a unit change of the given covariate x. With four ordered categories, the model simultaneously describes the effect of x on all three cumulative comparisons between the probabilities (i.e. 1 vs. 2, 3 or 4; 1 or 2 vs. 3 or 4; and 1, 2 or 3 vs. 4). Thus, a single effect is estimated for each covariate: the homogeneous effect of the covariate on the $K-1$ cumulative logits.

The proportional odds assumption is not always reasonable, and examples violating this assumption are not hard to find (Peterson & Harrell, 1990). For example, an intervention might be successful at reducing alcohol use from heavy to intermediate usage categories (i.e. mild usage) but not total abstinence. For such heterogeneous effects across the ordered response categories, a model that relaxes the proportional odds assumption is necessary. For cross-sectional data, Peterson & Harrell (1990) and Terza (1985) describe ordinal logistic and probit models, respectively, relaxing this assumption. Applying this model to stages of change data, Hedeker, Mermelstein & Weeks (1999) described it as a "thresholds of change" model. For longitudinal data, Hedeker & Mermelstein (1998) further developed this model and termed it a multilevel thresholds of change model. This model can be written as:

$$\log\left[\frac{P(Y_{ij}\leq k)}{1-P(Y_{ij}\leq k)}\right]=\alpha_{0k}+\mathbf{w}_{ij}'\alpha k+\mathbf{x}_{ij}'\mathrm{b}+v_i,$$

$$k=1,\ldots,K-1, \qquad (2)$$

where $\mathbf{w}_{ij}$ and $\mathbf{x}_{ij}$ denote the covariate vectors with heterogeneous and homogeneous effects, respectively. Because the regression coefficient vector $\mathrm{a_k}$ carries the k subscript, each of the covariates in $\mathbf{w}_{ij}$ have $K-1$ effects, one for each of the $K-1$ cumulative logits. In this way, a covariate can have no effect on complete cessation (e.g. the first category, thus the first logit), but can have positive effects on reducing the proportions of subjects in the highest usage categories (e.g. the second and third logits). Additionally, by comparing model fit assuming homogeneous versus heterogeneous effects for a covariate or set of covariates, a test of the proportional odds assumption can be performed.

It is worth nothing that ordinal regression models are often motivated and described using the "threshold concept" (Bock, 1975). Here, it is assumed that a continuous latent variable underlies the observed ordinal response. Specifically, with K ordered response categories, $K-1$ strictly increasing thresholds separate values of the unobserved continuous variable into the observed ordinal responses. This concept, and its description via thresholds, is what leads to the "thresholds of change" formulation of the above heterogeneous effects model described in Hedeker, Mermelstein & Weeks (1999) and Hedeker & Mermelstein (1998). As noted by McCullagh & Nelder (1989), the assumption of a continuous latent distribution, while providing a useful motivating concept, is not a strict requirement for use of ordinal regression models like the kind presented in this paper.

Repeated ordinal outcomes: change across time in smoking abstinence

In our example, we present and describe use of RRM to analyze longitudinal smoking data, but these same methods can be applied to alcohol or other substance use data. The data for our example come from a study on the use of extended telephone contact in a multi-component smoking cessation program. All subjects participated in a 7-week group treatment program. Following the group program, subjects were randomized to receive one of two types of telephone treatment (Standard or Recycling condition) that differed in the content of the phone calls. Seven counselor-initiated telephone calls were scheduled in both conditions. The calls started the week following the last group meeting and were spaced

Table 1. *Distribution of participants by group and days abstinent across time*

	Standard group			Recycling group		
	0 days	1–5 days	6–7 days	0 days	1–5 days	6–7 days
3-month	55.3% (140)	5.5% (14)	39.1% (99)	36.5% (97)	13.2% (35)	50.4% (134)
9-month	59.2% (170)	11.5% (33)	29.3% (84)	52.7% (164)	15.8% (49)	31.5% (98)
15-month	57.7% (161)	12.9% (36)	29.4% (82)	54.8% (166)	12.9% (39)	32.3% (98)

approximately every other week. In the Standard condition, counselors only gave subjects words of encouragement without specific guidance (e.g. "keep trying" if the subject had relapsed, or "congratulations" if the subject was abstinent). In the Recycling condition, the telephone call treatment protocol varied depending on the subject's smoking status: still smoking; abstinent; "slipped;" or relapsed. The goals of the Recycling calls were: (1) to encourage subsequent quit attempts in subjects still smoking at the end of treatment by helping them to problem solve barriers and reset quit dates; (2) to help prevent relapse in subjects who were abstinent by providing continued encouragement and planning for high-risk situations; and (3) to help Recycle subjects who had relapsed (to quit again) by debriefing the relapse episode, overcoming barriers and resetting quit dates.

Following the end of the group treatment, subjects were interviewed every 3 months for 15 months and asked to recall retrospectively their daily smoking behavior. In addition, reports of abstinence were verified biochemically through a combination of expired-air carbon monoxide and saliva cotinine at the post 1 (end of groups), 3-, 6- and 15-month follow-up points. Carbon monoxide values less than 8 p.p.m. and cotinine values less than 10 ng/ml verified abstinence. At each follow-up point, the number of days abstinent during that week was obtained and categorized into one of three outcome categories: 0, 1–5 or 6–7 days abstinent. For this example, for simplicity, we will focus on only three of the follow-up time points: 3−, 9− and 15-month follow-up points. Table 1 lists the observed percentages in these three categories across time for the two intervention groups.

As can be seen in Table 1, although the majority of subjects fell into the two extreme categories (either 0 days abstinent per week or 6–7 days abstinent per week), there were still noticeable proportions of subjects who fell in the middle outcome group. Treating the number of days abstinent as an ordinal outcome, rather than recording it into a dichotomous outcome, is advantageous since one of the issues with alcohol and smoking research is documenting "transitional states" that are not represented by dichotomous outcomes. Furthermore, the cutpoint for "abstinence" or success is often debatable. The advantage of treating the outcome as more than a dichotomy is that it allows researchers to consider more fine grained patterns of behavior.

Table 1 also lists the group sample sizes across time and, as can be seen, the amount of data per subject varied. Clearly, not all subjects were measured at each and every time-point. Also, in terms of an intervention effect, there appears to be an initial advantage for the Recycling group. However, by the last follow-up the groups appear to be similar. In general, a trend of decreased abstinence across time is observed.

To prepare for the analysis, Table 2 lists the observed cumulative odds and logits (corresponding to the three outcome categories) across the three time-points broken down by group. Since the first cumulative logit separates 0 days abstinent from the 1–5 and 6–7 days abstinent category, it is termed the partial abstinence threshold. Similarly, the second cumulative logit is termed the full abstinence threshold because it separates those below 6–7 days to those 6–7 days abstinent. As can be seen from Table 2, essentially all of these logits increase in value across time indicating less abstinence. However, comparing the 9- and 15- to the 3-month logits, it appears that the time effect is more pronounced in terms of the full abstinence threshold than the

Table 2. *Days abstinent by group across time: observed cumulative odds (and logits)*

	Standard group		Recycling group	
	1 vs. 2–3* Partial abstinence threshold	1–2 vs. 3 Full abstinence threshold	1 vs. 2–3 Partial abstinence threshold	1–2 vs. 3 Full abstinence threshold
3-month	140/113 = 1.24 (0.21)	154/99 = 1.56 (0.44)	97/169 = 0.57 (−0.56)	132/134 = 0.99 (−0.02)
9-month	170/117 = 1.45 (0.37)	203/84 = 2.42 (0.88)	164/147 = 1.12 (0.11)	213/98 = 2.17 (0.78)
15-month	161/118 = 1.36 (0.31)	197/82 = 2.40 (0.88)	166/137 = 1.21 (0.19)	205/98 = 2.09 (0.74)

*1 = 0 days abstinent, 2 = 1–5 days abstinent, 3 = 6–7 days abstinent.

partial abstinence threshold. In other words, over time, the proportion of subjects in the 0 days abstinent category is more stable than in the 6–7 days abstinent category.

To examine these observations more formally, the following random-intercepts ordinal logistic regression model was fit:

$$\log\left[\frac{P(Y_{ij}\leq k)}{1-P(Y_{ij}\leq k)}\right]=\alpha_{0k}+\beta_1 Time1_j+\beta_2 Time\ 2_j + \beta_3 Group_i+\beta_4(Group_i\times Time1_j) + \beta_5\ (Group_i\times Time2_j)+v_i \quad (3)$$

where *Time*1 = contrasts 9- to 3- months, *Time*2 contrasts 15- to 3-months, and *Group* is coded as 0 for Standard and 1 for Recycling. Because of these codings and because of the interactions in the model, β_1 and β_2 represent the time effects for the Standard group, β_3 represents the group difference at 3-months and β_4 and β_5 represent differences in the time effects between groups. Also, because subjects were measured at the same time-points in this study, the time variables only carry the occasion subscript j and not the additional subject subscript i.

The left-hand side of Table 3 lists the parameter estimates and standard errors for this model assuming homogeneous covariate effects. All analyses were performed using MIXOR (Hedeker & Gibbons, 1996a). This shareware program and its manual are available at http://www.uic.edu/~hedeker/mix.html. Readers interested in the program specifications for the examples in this article can send a note to the first author at hedeker@uic.edu. The *p*-values indicated next to the parameter estimates are obtained using the so-called "Wald test" (Wald, 1943), which uses the ratio of the parameter estimate to its standard error to determine statistical significance. These test statistics (i.e. Z = ratio of the parameter estimate to its standard error) are compared to a standard normal frequency table to test the null hypothesis that a given parameter equals 0. Alternatively, these Z-statistics are sometimes squared, in which case the resulting test statistic is distributed as chi-square on 1 degree of freedom. In either case, the *p*-values are identical.

Inspection of Table 3 reveals significant time, group and group × time interaction terms. The significant time effects indicate the increasing levels of smoking at 9- and 15-, relative to 3 months, for the Standard group. Since the group effect is negative and significant, the Recycling group had increased abstinence at 3 months, relative to the Standard group. However, since the group × time interaction terms are positive and significant, this beneficial effect of the Recycling group goes away at the 9- and 15-month follow-ups.

The next set of estimates, presented in the middle columns of Table 3, is for a model allowing heterogeneous effects for time and group. To represent this change, the regression coefficients for the time and group effects carry the k subscript and are denoted with α rather than β:

$$\log\left[\frac{P(Y_{ij}\leq k)}{1-P(Y_{ij}\leq k)}\right]=\alpha_{0k}+\alpha_{1k}\ Time1_j+\alpha_{2k}\ Time2_j + \alpha_{3k}\ Group_i+\beta_4\ (Group_i\times Time1_j) + \beta_5\ (Group_i\times Time2_j)+v_i \quad (4)$$

Table 3. *Ordinal logistic random-effects regression model estimates (standard errors): homogeneous and heterogeneous time, group and group by time effects*

	Homogeneous effects		Some heterogeneous effects		All heterogeneous effects	
Parameter	Partial abs	Full abs	Partial abs	Full abs	Partial abs	Full abs
Intercept	0.388 (0.301)	1.711*** (0.306)	0.696** (0.324)	1.427*** (0.325)	0.769** (0.330)	1.360*** (0.334)
9-month (vs. 3-month)	0.621** (0.265)		0.298 (0.287)	0.950*** (0.318)	0.234 (0.298)	1.025*** (0.370)
15-month (vs. 3-month)	0.560** (0.242)		0.288 (0.273)	0.816*** (0.275)	0.124 (0.290)	1.026** (0.298)
Group (0 = std, 1 = recy)	−1.324*** (0.402)		−1.632*** (0.433)	−1.261*** (0.415)	−1.751*** (0.445)	−1.052** (0.434)
Group × 9-month	0.989*** (0.365)		1.031*** (0.374)		1.191*** (0.417)	0.853* (0.468)
Group × 15-month	1.039*** (0.324)		1.085*** (0.336)		1.456*** (0.393)	0.642 (0.399)
Subject SD σ_v	3.831*** (0.250)		3.815*** (0.250)		3.830*** (0.252)	
−2log L	2611.8		2596.8		2592.4	

***$p < 0.01$; **$p < 0.05$; *$p < 0.10$.

Comparing this model to the previous one by a likelihood-ratio χ^2 test indicates an improvement in model fit; the difference in deviance (-2 log L) equals 2611.8–2596.8 = 15, which is significant at $p < 0.005$ for this 3 degrees of freedom χ^2 test. The degrees of freedom for this test is equal to the difference in number of estimated parameters between the two models. Additionally, the null hypothesis of equal effects across the two cumulative logits is rejected for each of the three parameters ($p < 0.004$, 0.02 and 0.02 for the 9 months, 15 months and group terms, respectively). As can be seen, the time effects are not significant in terms of the first cumulative logit, they are only significant in terms of the second cumulative logit. Alternatively, the group effect is appreciable on both, but is more pronounced for the first cumulative logit. Because the first cumulative logit compares 0 days abstinent to 1–5 and 6–7 days abstinent, the non-significance of the time effects indicates that the proportion of subjects with 0 days abstinent is relatively similar across time for the Standard group. This is seen in Tables 1 and 2. The heterogeneous group effect indicates that there is a larger group difference at 3 months for the 0 days abstinent category than the 6–7 days abstinent category. This is clearly seen in Table 2, where the difference at 3 months is -0.77 and -0.46 for the first and second cumulative logits, respectively.

The final model presented in Table 3 relaxes the homogeneity (i.e. proportional odds) assumption for all regression coefficients.

$$\log\left[\frac{P(\mathrm{Y}_{ij} \le k)}{1 - P(\mathrm{Y}_{ij} \le k)}\right] = \alpha_{0k} + \alpha_{1k}\, Time1_j + \alpha_{2k}\, Time2_j + \alpha_{3k}\, Group_i + \alpha_{4k}\,(Group_i \times Time1_j) + \alpha_{5k}\,(Group_i \times Time2_j) + v_i \quad (5)$$

The likelihood-ratio statistic for comparing this model to the previous one equals 4.4, which is not significant on 2 degrees of freedom. Thus, there is not sufficient evidence to reject the assumption of homogenous interaction effects across the two cumulative logits. The overall diminishing across time of the group effect that is seen in the first two models remains. Descriptively, it is interesting to note that in this final model, the interaction terms are more pronounced in terms of the first, rather than the second cumulative logit.

Figure 1 plots the observed and estimated proportions for the first two models in Table 3.

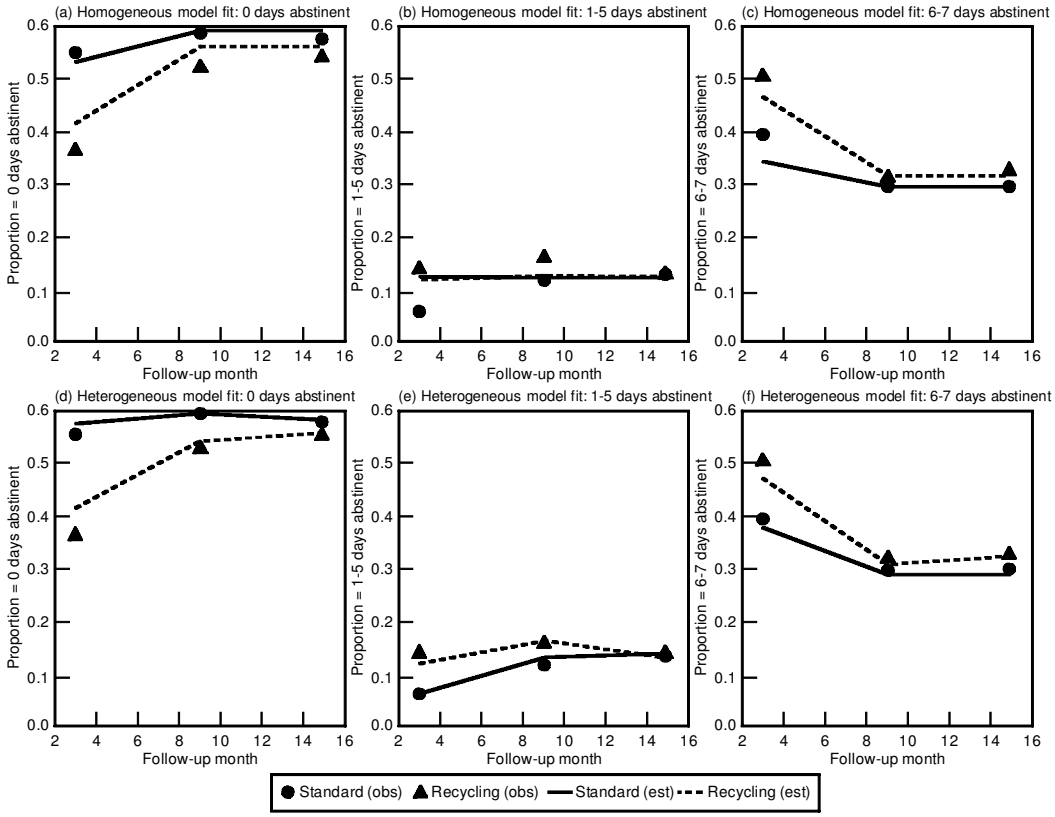

Figure 1. *The observed and estimated proportions for the first two models in Table 3.*

The top half of Fig. 1 is for the model assuming homogeneous effects, and the bottom half is for the model allowing heterogeneous effects due to time and group (but not group × time). As can be seen, the second model fits the data better, especially for the middle response category. Also, the estimates based on the homogeneous model deviate more from the observed proportions in terms of the initial group differences for the two extreme categories (0 and 6–7 days abstinent). This is due to the larger group effect for 0 days abstinent compared to 6–7 days abstinent.

For all models presented in Table 3, there is a considerable effect of the subject on their repeated observations. The population standard deviation associated with the random subject-varying effects is estimated as approximately equal to 3.8 for all three models. These estimates greatly exceed their standard errors, although this is not surprising since it would be unreasonable to assume that a subject's repeated smoking status assessments are independent. The degree of dependency attributable to subjects in their repeated observations is sometimes expressed as an intraclass, or more appropriately intrasubject, correlation. This correlation represents the average correlation between any two observations within the same subject. It also measures the proportion of total variance which is between-subjects. In the present case, since the error variance is assumed to be equal to $\pi^2/3$ for the logistic model (see Long, 1997, p. 119), the intrasubject correlation is estimated as 0.82 [e.g. $(3.8)^2/((3.8)^2 + \pi^2/3) = 0.82$]. Thus, the smoking measures are highly correlated within individuals.

In sum, with the present example, ordinal RRM allows us to examine time trends in outcome, to include all subjects regardless of missing data, and to examine several ordered levels of outcome. Relaxing the proportional odds assumption allows an examination of whether the covariate effects vary across the levels of the ordinal outcome. For these data, the time effects

were more important in distinguishing full abstainers from partial and non-abstainers (i.e. second cumulative logit), than in distinguishing non-abstainers from partial and full abstainers (i.e. first cumulative logit). Alternatively, the initial group difference was larger in terms of non-abstainers than in terms of full abstainers. The decrease across time in the group effect was relatively similar for both of these comparisons of the ordinal outcome.

Discussion

As demonstrated, RRM provide a useful way of analyzing longitudinal outcomes data. Specifically, RRM allow for the presence of missing data, irregularly spaced measurements across time, time-varying and invariant covariates, accommodation of individual-specific deviations from the average time trend and estimation of the population variance associated with these individual effects. Additionally, methods and software exist for analysis of continuous and categorical outcomes.

Perhaps the most useful feature of RRM is its treatment of missing data. As has been illustrated, subjects are not assumed to be measured at the same number of time-points. Since there are no restrictions on the number of observations per individual, subjects who are missing at a given interview wave are not excluded from the analysis. The assumption of the model is that the data that are available for a given subject are representative of that subject's deviation from the average population trend across time (which is estimated based on the whole sample). This is termed "ignorable" missingness in the statistical literature (Laird, 1988) and falls under Rubin's (1976) "missing at random" (MAR) assumption, in which the missingness depends only on observed data. That is, the probability of missingness is dependent on observed covariates and previous values of the outcome variable from subjects with missing data. The notion here is that if subject attrition is related to previous performance, in addition to other observable subject characteristics, then RRM provides valid statistical inference.

In some cases, the assumption of ignorable missingness may not be reasonable for longitudinal substance use data. For example, in the substance use literature, it is often thought that missingness equals substances use, regardless of subject covariate values or prior observed substance use levels. For this reason, researchers sometimes recode missing observations as the highest substance use level. Because it is unlikely that missingness and substance use are completely correlated (as this recoding assumes), a more statistical approach to this problem is desirable. One such approach for dealing with non-ignorable missingness is based on use of "pattern-mixture modeling" (Little, 1995). For this, subjects are first grouped based on their available data pattern across time. For example, in the simplest case, subjects can be classified as complete-data subjects or incomplete-data subjects. The between-subjects classification variables that are formed based on these missing data patterns are then included in the model to examine their effect on the outcome variable. Interactions can also be included in order to determine, for example, if treatment group-related effects vary by missing data pattern. Utilizing this pattern-mixture approach within RRM, Hedeker & Gibbons (1997) illustrate its application to psychiatric clinical trials data, and Hedeker & Rose (2000) describe its use for longitudinal smoking data. For longitudinal substance use data, pattern-mixture modeling is particularly useful because subjects with missing data across time often have increased baseline levels and/or increased trajectories across time (Tebes, Snow & Arthur, 1992). Other approaches for dealing with non-ignorable missingness in longitudinal settings are described by Conaway (1992) and Diggle & Kenward (1994).

Due to the categorical nature of alcohol and smoking outcomes, recent extensions of RRM for categorical data are particularly important. In this paper, we have presented use of the ordinal RRM. A common characteristic of the ordinal model is the proportional odds assumption. This assumption specifies that the covariate effects are homogeneous (on the logit scale) across the (comparisons of the) ordinal outcome categories. For smoking and alcohol data, this assumption does not always hold. For example, an intervention might be able to reduce use in the middle outcome categories, but not at the highest level of use. The model presented in this article allows for relaxation of the proportional odds assumption, which helps identify cutpoints among the ordinal categories where variables have the strongest (and weakest) effects.

In dealing with ordinal outcomes in practice,

researchers sometimes dichotomize the ordinal variable. This practice is often performed more for convenience than for any substantive considerations. An issue that emerges is the location on the scale that should be used to dichotomize the variable. For example, should abstinence be defined as 0 days abstinent or 0–1 days abstinent. The model presented in this paper overcomes this issue because it can estimate the effects of covariates for all $K - 1$ dichotomizations of an ordinal outcome with K categories. Thus, the dilemma of where to dichotomize is effectively solved.

While it is important to consider the sample size for application of the ordinal random-effects regression model, especially with heterogeneous effects, it is not easy to provide global recommendations for what the required sample size should be. One consideration in estimating heterogeneous effects is the numbers of observations in the K response categories as they are broken down by the covariates and covariate interactions of the model. Consider the simplest case of one covariate with two categories (e.g. gender); the data then form a $2 \times K$ cross-tabulation table. Estimating heterogeneous threshold effects for gender then requires observations in both gender groups at each of the $K - 1$ comparisons across the response variable (i.e. category 1 vs. 2 to K combined, categories 1 and 2 combined vs. 3 to K combined, …, categories 1 to K–1 combined vs. K). Because allowing interactions (e.g. gender by treatment group) to have heterogeneous effects splits the data up even further, it may not always be possible or may require collapsing of covariate or response categories. A further point regarding sample size is that the significance tests that are formed by taking the ratio of the parameter estimate to its standard error are based on asymptotic statistical theory. However, many other statistical techniques using maximum likelihood estimation also invoke asymptotic theory for hypothesis testing (e.g. logistic regression, log-linear models and structural equation models). For more details on asymptotic theory as applied to random-effects models, see Longford (1993).

In the example, repeated observations were observed nested within individuals. In the terminology of multi-level analysis (Goldstein, 1995) and hierarchical linear models (Bryk & Raudenbush, 1992), this is termed a two-level data structure with individuals representing level 2 and the nested repeated observations level 1. The models that we have presented are thus referred to as two-level models. Individuals themselves, however, are often observed clustered within some higher-level unit; for example, a classroom, clinic or work-site. Cross-sectional clustered data can also be considered as two-level data, with the clusters representing level 2 and the clustered subjects level 1. Analysis of cross-sectional clustered (substance use) data using RRM is discussed by Hedeker, Gibbons & Flay (1994) and Hedeker *et al.* (1994). In some studies, subjects are clustered and also repeatedly measured, resulting in three levels of data: the cluster (level 3), individual (level 2) and repeated observation (level 1). Analysis of three-level data is described in Goldstein (1995), Bryk & Raudenbush (1992), Longford (1993) and Gibbons & Hedeker (1997).

Since longitudinal designs are increasingly used to study alcohol, smoking and other substance use patterns across time, it is important that statistical methods are developed and used to extract the most out of these longitudinal datasets. RRM provide an attractive approach for addressing some key questions that emerge from longitudinal designs. Hopefully, this paper has helped in increasing the understanding of these methods and their potential for use in analyzing longitudinal substance use outcomes.

Acknowledgements

The authors thank Siu-Chi Wong for computer and statistical programming assistance. Preparation of this paper was supported by National Heart Lung and Blood Institute Grant HL 42485 and National Institutes of Mental Health Grant MH 44826.

References

AGRESTI, A. (1989) Tutorial on modeling ordered categorical response data *Psychological Bulletin*, 105, 290–301.

ALBERT, P. S. (1999) Longitudinal data analysis (repeated measures) in clinical trials, *Statistics in Medicine*, 18, 1707–1732.

BERNSTEIN, G. A., CARROLL, M. E., CROSBY, R. D., PERWIEN, A. R., GO, F. S. & BENOWITZ, N. L. (1994) Caffeine effects on learning, performance, and anxiety in normal school-age children, *Journal of the American Academy of Adolescent Psychiatry*, 33, 407–415.

BOCK, R. D. (1975) *Multivariate Statistical Methods in Behavioural Research* (New York, McGraw-Hill).

BOCK, R. D. (1983a) Within-subject experimentation in psychiatric research, in: GIBBONS, R. D. & DYSKEN, M. W. (Eds) *Statistical and Methodological Advances in Psychiatric Research*, pp. 59–90 (New York, Spectrum).

BOCK, R. D. (1983b) The discrete Bayesian, in: WAINER, H. & MESSICK S. (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: BOCK, R. D. (Ed.) *Multilevel Analysis of Educational Data*, pp. 319–342 (New York, Academic Press).

BRYK, A. S. & RAUDENBUSH, S. W. (1992) *Hierarchical Linear Models: applications and data analysis methods* (Newbury Park, CA, Sage Publications, Inc.).

BRYK, A. S., RAUDENBUSH, S. W. & CONGDON, R. T. (1996) *HLM. Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs* (Chicago, IL, Scientific Software, Inc.).

BURCHINAL, M. R., BAILEY, D. B. & SNYDER, P. (1994) Using growth curve analysis to evaluate child change in longitudinal investigations, *Journal of Early Intervention*, 18, 403–423.

CARROLL, K. M., ROUNSAVILLE, B. J., NICH, C., GORDON, L. T., WIRTZ, P. W. & GAWIN, F. (1994) One-year follow-up of psychotherapy and pharmacotherapy for cocaine dependence, *Archives of General Psychiatry*, 51, 989–997.

CARROLL, K. M., NICH, C., BALL, S. A., MCCANCE, E. & ROUNSAVILLE, B. J. (1998) Treatment of cocaine and alcohol dependence with psychotherapy and disulfiram, *Addiction*, 93, 713–728.

CHASSIN, L., PRESSON, C. C., ROSE, J. S. & SHERMAN, S. J. (1996) The natural history of cigarette smoking from adolescence to adulthood: differential stability as a function of age, educational attainment, and parental smoking, *Health Psychology*, 15, 478–484.

CNAAN, A., LAIRD, N. M. & SLASOR, P. (1997) Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data, *Statistics in Medicine*, 16, 2349–2380.

CONAWAY, M. R. (1992) The analysis of repeated categorical measurements subject to nonignorable nonresponse, *Journal of the American Statistical Association*, 87, 817–824.

DE LEEUW, J. & KREFT, I. (1986) Random coefficient models for multilevel analysis, *Journal of Educational Statistics*, 11, 57–85.

DE LEEUW, J. & KREFT, I. (1999) *Software for Multilevel Analysis*, preprint 239 Los Angeles, CA, (UCLA Statistics).

DELUCCHI, K. & BOSTROM, A. (1999) Small sample longitudinal clinical trials with missing data: a comparison of methods, *Psychological Methods*, 4, 158–172.

DEMPSTER, A. P., RUBIN, D. B. & TSUTAKAWA, R. K. (1981) Estimation in covariance component models, *Journal of the American Statistical Society*, 76, 341–353.

DIGGLE, P. & KENWARD, M. G. (1994) Informative dropout in longitudinal data analysis (with discussion), *Applied Statistics*, 43, 49–94.

DIGGLE, P. J., LIANG, K. Y. & ZEGER, S. L. (1994) *Analysis of Longitudinal Data* (New York, Oxford).

EGRET (1991) *Reference Manual* (Seattle, WA, Statistics & Epidemiology Research Corporation).

EVERITT B. S. (1998) Analysis of longitudinal data: beyond MANOVA, *British Journal of Psychiatry*, 172, 7–10.

EZZET, F. & WHITEHEAD, J. (1991) A random effects model for ordinal responses from a crossover trial, *Statistics in Medicine*, 10, 901–907.

FITZMAURICE, G. M., LAIRD, N. M. & ROTNITZKY, A. G. (1993) Regression models for discrete longitudinal responses, *Statistical Science*, 8, 284–309.

GALLAGHER, T. J., COTTLER, L. B., COMPTON, W. M. & SPITZNAGEL, E. (1997) Changes in HIV/AIDS risk behaviors in drug users in St. Louis: application of random regression models, *Journal of Drug Issues*, 27, 399–416.

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. (1997) Random-effects probit and logistic regression models for three-level data, *Biometrics*, 53, 1527–1537.

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.

GOLDSTEIN, H. (1991) Nonlinear multilevel models, with an application to discrete response data, *Biometrika*, 78, 45–51.

GOLDSTEIN, H. (1995) *Multilevel Statistical Models*, 2nd edn (New York, Halsted Press).

GOLDSTEIN, H., RASBASH, J., PLEWIS, I., DRAPER, D., BROWNE, W., YANG, M., WOODHOUSE, G. & HEALY, M. (1998) *A User's Guide to MLwiN* (London, Institute of Education, University of London).

GRUDER, C. L., MERMELSTEIN, R. J., KIRKENDOL, S., HEDEKER, D., WONG, S. C., SCHRECKENGOST , J., WARNECKE, R. B., BURZETTE, R. & MILLER, T. Q. (1993) Effects of social support and relapse prevention training as adjuncts to a televised smoking cessation intervention, *Journal of Consulting and Clinical Psychology*, 61, 113–120.

HALIKAS, J. A., CROSBY, R. D., PEARSON, V. L. & GRAVES, N. M. (1997) A randomized double-blind study of carbamazepine in the treatment of cocaine abuse, *Clinical Pharmacology and Therapeutics*, 62, 89–105.

HAND, D. & CROWDER, M. (1996) *Practical Longitudinal Data Analysis* (New York, Chapman & Hall).

HEDEKER, D. (1999) MIXNO: a computer program for mixed-effects nominal logistic regression, *Journal of Statistical Software*, 4, 1–92.

HEDEKER, D. & GIBBONS, R. D. (1994) A random-effects ordinal regression model for multilevel analysis, *Biometrics*, 50, 933–944.

HEDEKER, D. & GIBBONS, R. D. (1996a) MIXOR: a

computer program for mixed-effects ordinal regression analysis, *Computer Methods and Programs in Biomedicine*, 49, 157–176.

HEDEKER, D. & GIBBONS, R. D. (1996b) MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors, *Computer Methods and Programs in Biomedicine*, 49, 229–252.

HEDEKER, D. & GIBBONS, R. D. (1997) Application of random-effects pattern-mixture models for missing data in longitudinal studies, *Psychological Methods*, 2, 64–78.

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., MCMAHON, S. D., JASON, L. A. & SALINA, D. (1994) Analysis of clustered data in community psychology: with an example from a worksite smoking cessation project, *American Journal of Community Psychology*, 22, 595–615.

HEDEKER, D. & MERMELSTEIN, R. J. (1996) Application of random-effects regression models in relapse research, *Addiction*, 91 (suppl.), S211–S229.

HEDEKER, D. & MERMELSTEIN, R. J. (1998) A multilevel thresholds of change model for analysis of stages of change data, *Multivariate Behavioral Research*, 33, 427–455.

HEDEKER, D., MERMELSTEIN, R. J. & WEEKS, K. A. (1999) The thresholds of change model: an approach to analyzing stages of change data, *Annals of Behavioral Medicine*, 21, 61–70.

HEDEKER, D. & ROSE, J. S. (2000) The natural history of smoking: a pattern-mixture random-effects regression model, in: ROSE, J. S., CHASSIN, L., PRESSON, C. C. & SHERMAN, S. J. (Eds) *Multivariate Applications in Substance Use Research*, pp. 79–112 (Hillsdale, NJ, Lawrence Erlbaum Associates).

HU, F. B., GOLDBERG, J., HEDEKER, D., FLAY, B. R. & PENTZ, M. A. (1998) A comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes, *American Journal of Epidemiology*, 147, 694–703.

HUI, S. L. & BERGER, J. O. (1983) Empirical Bayes estimation of rates in longitudinal studies, *Journal of the American Statistical Association*, 78, 753–759.

JANSEN, J. (1990) On the statistical analysis of ordinal data when extravariation is present, *Applied Statistics*, 39, 75–84.

JASON, L., SALINA, D., MCMAHON, S. D., HEDEKER, D. & STOCKTON, M. (1997) A worksite smoking intervention: a 2 year assessment of groups, incentives, and self-help, *Health Education Research*, 12, 129–138.

JENNRICH, R. I. & SCHLUCHTER, M. D. (1986) Unbalanced repeated-measures models with structured covariance matrices, *Biometrics*, 42, 805–820.

KESELMAN, H. J., ALGINA, J., KOWALCHUK, R. K. & WOLFINGER, R. D. (1999) A comparison of recent approaches to the analysis of repeated measurements, *British Journal of Mathematical and Statistical Psychology*, 52, 63–78.

KREFT, I. G., DE LEEUW, J. & VAN DER LEEDEN, R. (1994) Comparing five different statistical packages for hierarchical linear regression: BMDP-5V, GENMOD, HLM, ML3, and VARCL, *American Statistician*, 48, 324–335.

LAIRD, N. M. (1988) Missing data in longitudinal studies, *Statistics in Medicine*, 7, 305–315.

LAIRD, N. M. & WARE, J. H. (1982) Random-effects models for longitudinal data, *Biometrics*, 38, 963–974.

LESAFFRE, E., ASEFA, M. & VERBEKE, G. (1999) Assessing the goodness-of-fit of the Laird and Ware model—an example: the Jimma infant survival differential longitudinal study, *Statistics in Medicine*, 18, 835–854.

LITTLE, R. J. A. (1995) Modeling the drop-out mechanism in repeated-measures studies, *Journal of the American Statistical Association*, 90, 1112–1121.

LONG, J. S. (1997) *Regression Models for Categorical and Limited Dependent Variables* (Thousand Oaks, CA, Sage Publications).

LONGFORD, N. T. (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects, *Biometrika*, 74, 817–827.

LONGFORD, N. T. (1993) *Random Coefficient Models* (New York, Oxford University Press).

MANOR, O. & KARK, J. D. (1996) A comparative study of four methods for analysing repeated measures data, *Statistics in Medicine*, 15, 1143–1159.

MCCULLAGH, P. (1980) Regression models for ordinal data (with discussion), *Journal of the Royal Statistical Society, Series B*, 42, 109–142.

MCCULLAGH, P. & NELDER, J. A. (1989) *Generalized Linear Models*, 2nd edn (London, Chapman and Hall).

NICH, C. & CARROLL, K. (1997) Now you see it, now you don't: a comparison of traditional versus random-effects regression models in the analysis of longitudinal follow-up data from a clinical trial, *Journal of Consulting and Clinical Psychology*, 65, 252–261.

O'MALLEY, S. S., JAFFE, A. J., CHANG, G., RODE, S., SCHOTTENFELD , R., MEYER, R. E. & ROUNSAVILLE, B. (1996) Six month follow-up of naltrexone and psychotherapy for alcohol dependence, *Archives of General Psychiatry*, 53, 217–224.

OMAR, R. Z., WRIGHT, E. M., TURNER, R. M. & THOMPSON, S. G. (1999) Analysing repeated measures data: a practical comparison of methods, *Statistics in Medicine*, 18, 1587–1603.

PENDERGAST, J. F., GANGE, S. J., NEWTON, M. A., LINDSTROM, M. J., PALTA, M. & FISHER, M. R. (1996) A survey of methods for analyzing clustered binary response data, *International Statistical Review*, 64, 89–118.

PETERSON, B. & HARRELL, F. E. (1990) Partial proportional odds models for ordinal response variables, *Applied Statistics*, 39, 205–217.

ROSNER, B. (1992) Multivariate methods for binary longitudinal data with heterogeneous correlation over time, *Statistics in Medicine*, 11, 1915–1928.

RUBIN, D. B. (1976) Inference and missing data, *Biometrika*, 63, 581–592.

SALINA, D., JASON, L. A., HEDEKER, D., KAUFMAN, J., LESONDAK, L., MCMAHON, S. D., TAYLOR, S. &

KIMBALL, P. (1994) A follow-up of a media-based, worksite smoking cessation program, *American Journal of Community Psychology*, 22, 257–271.

STIRATELLI, R., LAIRD, N. M. & WARE, J. H. (1984) Random-effects models for serial observations with binary response, *Biometrics*, 40, 961–971.

STRENIO, J. F., WEISBERG, H. I. & BRYK, A. S. (1983) Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates, *Biometrics*, 39, 71–86.

SULLIVAN, L. M., DUKES, K. A. & LOSINA, E. (1999) An introduction to hierarchical linear modelling, *Statistics in Medicine*, 18, 855–888.

TEBES, J. K., SNOW, D. L. & ARTHUR, M. W. (1992) Panel attrition and external validity in the short-term follow-up study of adolescent substance use, *Evaluation Review*, 16, 151–170.

TEN HAVE, T. R. (1996) A mixed effects model for multivariate ordinal response data including correlated failure times with ordinal responses, *Biometrics*, 52, 473–491.

TERZA, J. (1985) Ordinal probit: a generalization, *Communications in Statistics*, 14, 1–12.

VAN DER LEEDEN, R., VRIJBURG, K. & DE LEEUW, J. (1996) A review of two different approaches for the analysis of growth data using longitudinal mixed linear models, *Computational Statistics and Data Analysis*, 21, 583–605.

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.

WOLFINGER, R. D. (1993) Covariance structure selection in general mixed models, *Communications in Statistics, Simulation and Computation*, 22, 1079–1106.

WOLFINGER, R. D. & LIN, X. (1997) Two Taylor-series approximation methods for nonlinear mixed models, *Computational Statistics and Data Analysis*, 25, 465–490.

WONG, G. Y. & MASON, W. M. (1985) The hiearchical logistic regression model for multilevel analysis, *Journal of the American Statistical Association*, 80, 513–524.

ZEGER, S. L. & LIANG, K.-Y. (1992) An overview of methods for the analysis of longitudinal data, *Statistics in Medicine*, 11, 1825–1839.