Addiction (1996) 91 (Supplement), S211–S229

SECTION III. METHODS FOR ANALYZING LONGITUDINAL DATA ON RELAPSE

Application of random-effects regression models in relapse research

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

[1]*Prevention Research Center and School of Public Health, University of Illinois at Chicago &* [2]*Prevention Research Center and Department of Psychology, University of Illinois at Chicago, USA*

Abstract

This article describes and illustrates use of random-effects regression models (RRM) in relapse research. RRM are useful in longitudinal analysis of relapse data since they allow for the presence of missing data, time-varying or invariant covariates, and subjects measured at different timepoints. Thus, RRM can deal with "unbalanced" longitudinal relapse data, where a sample of subjects are not all measured at each and every timepoint. Also, recent work has extended RRM to handle dichotomous and ordinal outcomes, which are common in relapse research. Two examples are presented from a smoking cessation study to illustrate analysis using RRM. The first illustrates use of a random-effects ordinal logistic regression model, examining longitudinal changes in smoking status, treating status as an ordinal outcome. The second example focuses on changes in motivation scores prior to and following a first relapse to smoking. This latter example illustrates how RRM can be used to examine predictors and consequences of relapse, where relapse can occur at any study timepoint.

Introduction

Longitudinal studies play a prominent role in investigations of alcohol, smoking, and drug relapse. Common issues addressed in these studies include examining the effects of different treatments on relapse rates and identifying predictors of relapse. However, until recently researchers have been limited in their approaches to answering key questions because of difficulties in finding or using analytic techniques that appropriately represent the relapse process. Thankfully, the development of methods to deal with the peculiarities of longitudinal data has been an active area of statistical research. In this regard, random-effects regression models (RRM) may be one approach for broadening researchers' treatment of their data and their conceptualization of the relapse process. This paper will describe the use of RRM, first to analyze outcome data from a smoking cessation study and secondly to explore changes in a psychosocial variable (motivation) both before and after a relapse.

RRM were developed to handle many of the peculiarities of longitudinal data, and variants of

Correspondence to: Donald Hedeker, Ph.D, Division of Epidemiology and Biostatistics (M/C 922), School of Public Health, University of Illinois at Chicago, 2121 West Taylor Street, Room 510, Chicago, IL, 60612-7260, USA. Tel: + 1 (312) 996 4896.

Submitted 29th March 1996; initial review completed 10th June 1996; final version accepted 2nd August 1996.

0965–2140/96/12S211–19 $8.00

Carfax Publishing Company

these models have a variety of names: random-effects models (Laird & Ware, 1982), variance component models (Dempster, Rubin & Tsutakawa, 1981); hierarchical linear models (Bryk & Raudenbush, 1987), multilevel models (Goldstein, 1995), two-stage models (Bock, 1989), random coefficient models (DeLeeuw & Kreft, 1986), mixed models (Longford, 1987), empirical Bayes models (Strenio, Weisberg & Bryk, 1983; Hui & Berger, 1983), unbalanced repeated-measures models (Jennrich & Schlucter, 1986) and random regression models (Bock, 1983a, 1983b). Generalizations of RRM have also been developed for the case of dichotomous response data (Stiratelli, Laird & Ware, 1984; Gibbons & Bock, 1987; Goldstein, 1991) and ordinal response data (Jansen, 1990; Ezzet & Whitehead, 1991; Hedeker & Gibbons, 1994), thus allowing a general framework for analysis of both continuous and categorical outcome variables.

There are several features of RRM that make these methods useful in longitudinal research. First, it is not assumed that all subjects have been measured on the same number of timepoints, thus, in particular, subjects with incomplete data across time are not excluded from the analysis. Subjects contribute as many repeated observations to the analysis that they have data on. The ability to model incomplete data across time is an important advantage relative to procedures that require complete data across time; for example, multivariate repeated measures analysis of variance, since (a) the analysis is more powerful due to the inclusion of more of the data, and (b) complete-case analysis may suffer from selection biases to the degree to which subjects with complete data are not representative of the larger population of subjects. Furthermore, time is treated as a continuous variable in RRM, so that subjects do not have to be measured at the same timepoints. That is, the actual values of time can vary from subject to subject. This is a useful advantage for analysis of longitudinal studies where follow-up times are not uniform across all subjects. Covariates can be included in the model and can be either time-varying or invariant. This flexibility, in terms of model covariates, allows a more reasonable approach for modeling changes in outcomes across time. Namely, changes 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, the traditional approaches to longitudinal data analysis estimate average change (across time) in a population, RRM can also estimate individual change for each subject. These estimates of individual change across time can be particularly useful in longitudinal relapse 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 smoking, there may be a proportion of subjects whose status improves.

Basically, the focus in RRM remains on the individual rather than just on the sample group to which the individual belongs. Specifically, the individual parameters that are estimated, expressing how that particular individual is changing over time, are used at a second step to estimate the more common overall population parameters. This two-step procedure is then repeated, iterating between the estimation of individual and population parameters, until convergence. Also, it is because this technique uses the available data for each individual to estimate his/her parameters, that individuals can be measured at a varying number of timepoints or, in fact, at different timepoints. The underlying assumption is that the individual estimates characterize how each person deviates from the estimated population change during the time-course of the study.

These features of RRM for longitudinal data analysis help to address several limitations that exist in much of the research on relapse to date. First, despite the availability of repeated outcome measurements over time, researchers often do not take full advantage of the possible longitudinal patterns of behavior and, instead, focus on the final data point alone. RRM, however, can allow for a true longitudinal analysis by including all data points, regardless of the pattern of the behavior. Thus, the "problem" of how to deal with subjects who alternate between smoking and abstinence at each time point is not an issue, in and of itself. A second limitation in much of the relapse research is a forced dichotomy of outcome, such as abstinent or smoking/relapsed. The limitation in this case is that this categorization equates all levels of usage, from one slip to a full relapse, as the same. Instead, to better understand the relapse process, it may be useful if researchers considered intermediate outcomes or transitional states as well.

By treating these intermediate outcomes as separate categories, rather than simply recoding them into either the smoking or abstinence categories, a more complete understanding of the relapse process may emerge. In this regard, the generalization of RRM for ordinal outcomes is a promising approach since it allows for the specification of intermediate outcomes. A third problem in longitudinal relapse research is the inevitability of missing data. Researchers have typically opted for one of two approaches to handling this problem: assigning a negative outcome (e.g. smoking) to those who are missing, or analyzing only complete cases. Both these approaches may have significant limitations. Since RRM, however. can include all subjects, even if some data are missing, there is no need from an analysis point-of-view to either discard data (i.e. analyze only complete-cases) or invent data (i.e. recode missing to the negative outcome). Fourth, researchers have often been limited in their analysis of factors that influence relapse by having to select a given variable at one given timepoint to see how it relates to outcome at another point, thus ignoring the fact that many predictor variables change over time. Since RRM can include predictor variables that are time-invariant (e.g. gender) or time-variant (e.g. stress levels), a more dynamic analysis of the relapse process is possible.

In this paper we will describe use of RRM in relapse research. We will first illustrate how RRM can be seen as an extension of an ordinary linear regression model. Starting with a simple linear regression model, the model will slowly be extended and described, in order to guide the reader going from familiar to unfamiliar territory. Generalizations of RRM for longitudinal categorical data will then be described and related to the models for normally-distributed outcomes. Following the descriptions of the statistical models, two examples will be presented from a smoking cessation study to illustrate analysis using RRM. The first examines longitudinal changes in smoking status, treating status as an ordinal outcome. The second example focuses on changes in motivation scores prior to and following a first relapse to smoking, where relapse can occur at any study timepoint.

Other articles illustrating RRM usage for longitudinal psychiatric data include Gibbons *et al.* (1988); Hedeker *et al.* (1989); Sharma *et al.* (1992) and Gibbons *et al.* (1993). Several book-length texts (Bryk & Raudenbush, 1992; Jones, 1993; Longford, 1993; Diggle, Liang & Zeger, 1994; Goldstein, 1995) further describe use of RRM. It should be noted that other approaches besides RRM for longitudinal data analysis have been developed, for example, the generalized estimating equation (GEE) approach (Zeger, Liang & Albert, 1988; Dunlop, 1994), structural equation modeling (McArdle & Hamagami, 1991; Bentler & Newcomb, 1992) survival and event history analysis (Petersen, 1990), latent class modeling (Clogg, Eliason & Grego, 1990) and latent transition analysis (Graham *et al.*, 1991; Velicer, Martin & Collins, this issue), to name a few. In this article, however, we will focus only on RRM in order to more thoroughly describe this particular method and illustrate its usage.

Random-effects regression models (RRM)

To introduce RRM, consider a simple linear regression model for the measurement y of individual i ($i = 1, 2, \ldots, N$ subjects) on occasion k ($k = 1, 2, \ldots, n_i$ occasions):

$$y_{ik} = \beta_0 + \beta_1 t_{ik} + \varepsilon_{ik} \quad (1)$$

Ignoring subscripts, this model represents the regression of the outcome variable y on the independent variable time (denoted t). The subscripts merely help to keep track of the particulars of the data, namely whose observation it is (subscript i) and when was this observation made (the subscript k). The independent variable t gives a value to the level of time and may represent time in weeks, months, etc. Since both y and t carry both i and k subscripts, both the outcome variable and the time variable are allowed to vary both by individuals and occasions. In linear regression models, like (1), the errors ε_{ik} are assumed to be normally and *independently* distributed in the population with zero mean and common variance σ^2. It is the assumption of independence that makes the model given in equation (1) an unreasonable one for longitudinal data. This is because the outcomes y are observed repeatedly from the same individuals, and so it is unreasonable to assume independence of the model errors. For example, smoking status may be repeatedly measured for individuals over a follow-up period. Their status or behavior from one measurement wave to the next is unlikely to be independent. It is much more

likely to assume that errors within an individual are correlated. A solution for dealing with this "challenge" due to data dependency is to introduce into the model individual-specific effects that will account for the data dependency. This is precisely what RRM does. Thus, a simple RRM is given by:

$$y_{ik} = \beta_0 + \beta_1 t_{ik} + \upsilon_{0i} + \varepsilon_{ik} \quad (2)$$

where the additional term υ_{0i} indicates the influence of individual i on his/her repeated observations.

Since individuals in a sample are typically thought to be representative of a larger population of individuals, the individual-specific effects υ_{0i} are treated as random effects. That is, υ_{0i} are considered to be representative of a distribution of individual effects in the population. The most common form for this population distribution is the normal distribution with mean 0 and variance σ_υ^2. In the model given by equation (2), the errors ε_{ik} are now assumed to be normally and *conditionally independently* distributed in the population with zero mean and common variance σ^2. Conditional independence here means conditional on the random individual-specific effects υ_{0i}. Since the errors now have an influence due to individuals removed from them, this conditional independence assumption is much more reasonable than the ordinary independence assumption associated with (1). Since individuals deviate from the regression of y on t in a parallel manner (since there is only one subject effect υ_{0i}) this model is sometimes referred to as a random-intercepts model, with each υ_{0i} indicating how individual i deviates from the model.

A slightly more sophisticated model allows both the intercept and trend due to time to vary by individuals:

$$y_{ik} = \beta_0 + \beta_1 t_{ik} + \upsilon_{0i} + \upsilon_{1i} t_{ik} + \varepsilon_{ik} \quad (3)$$

where, β_0 is the overall population intercept, β_1 is the overall population slope, υ_{0i} is the intercept deviation for subject i, υ_{1i} is the slope deviation for subject i, and ε_{ik} is an independent error term distributed normally with mean 0 and variance σ^2. The assumption regarding the independence of the errors is one of conditional independence, that is, they are independent conditional on υ_{0i} and υ_{1i}. With two random individual-specific effects, the population distribution of intercept and slope deviations is assumed to be a bivariate normal $N(0, \Sigma_\upsilon)$, with

$$\Sigma_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0\upsilon_1} \\ \sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_1} \end{bmatrix}$$

This model can be thought of as a personal trend or change model since it represents the measurements of y as a function of time, both at the individual (υ_{0i} and υ_{1i}) and population (β_0 and β_1) levels. The intercept parameters indicate the starting point, and the slope parameters indicate the degree of change over timepoints. The population intercept and slope parameters represent the trend for the population, while the individual parameters express how the individual deviates from the population trend. Note that the occasions range from $k = 1$ to n_i, with each person being measured on n_i timepoints. Since n carries the i subscript, each subject may vary in terms of the number of measured occasions. Also, since the time variable t carries the i subscript, subjects can be measured on different occasions. The underlying assumption of the model is that the data that are available for a given individual are representative of how that individual deviates from the population trend across the timeframe of the study. If an individual deviates from the population trend in only a random manner, the population trend is the most likely representation of that individual's trend, and the individual parameters will be roughly equal to zero for that person. What is more likely, though, is that each individual will deviate in a personal, and thus systematic, manner from the population trend, and so, the individual parameters characterize these personal trends.

The variability in these individual parameters is assessed by the intercept variance $\sigma^2_{\upsilon_0}$, slope variance $\sigma^2_{\upsilon_1}$ and the covariance of the intercept and the slope $\sigma_{\upsilon_0\upsilon_1}$. To the degree that each individual's deviation from the population trend is only due to random error, these variance terms will approach zero. Alternatively, as each individual's deviation from the population trend is non-random, but characterized by the individual trend parameters υ_{0i} and υ_{1i} as being non-zero, these variance terms will increase from zero. Additionally, the covariance term represents the degree to which the individual intercept and slope parameters covary. For example, a positive covariance term would suggest that individuals with higher initial values have greater positive

slopes, while a negative covariance would suggest the opposite. Finally, the error variance is assessed in the model by σ^2.

We can also include terms in the model for covariates which do not change over time (time invariant) and for covariates which vary across the measured timepoints (time varying). This model can then be written as:

$$y_{ik} = \beta_0 + \beta_1 t_{ik} + \beta_2 x_i + \beta_3 x_{ik} + \upsilon_{0i} + \upsilon_{1i} t_{ik} + \varepsilon_{ik} \quad (4)$$

where β_2 is the coefficient for the time invariant covariate x_i, and β_3 is the coefficient for the time varying covariate x_{ik}. Interactions of the covariates 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_{ik} might be the treatment by time interaction which is obtained as the product of x_i by t_{ik}.

For parameter estimation, a combination of two complementary methods has generally been proposed: empirical Bayes (EB) methods for estimation of the individual effects (υ_{0i} and υ_{1i}), and marginal maximum likelihood (MML) methods for estimation of variance and covariance parameters (σ^2 $\sigma^2_{\upsilon 0}$, $\sigma^2_{\upsilon 1}$ and $\sigma_{\upsilon 0\upsilon 1}$) and covariate effects (β). Details can be found in Laird & Ware (1982), Bock (1989) and Longford (1993).

RRM for longitudinal categorical data

Generalization of RRM for categorical outcomes has been an active area of recent statistical research. In particular, generalizations of RRM have been described for both dichotomous (Stiratelli, Laird & Ware, 1984; Anderson & Aitkin, 1985; Wong & Mason, 1985; Gibbons & Bock, 1987; Goldstein, 1991), and ordinal (Jansen, 1990; Ezzet & Whitehead, 1991; Hedeker & Gibbons, 1994) outcomes. These models generally adopt either a probit or logistic regression model and utilize various methods for incorporating and estimating the influence of the random effects. In general, the models for categorical outcome data are computationally more intensive than the model for continuous outcome data and, therefore, often a single random effect (i.e. a random subject effect, or random intercept) is assumed in these models. Thus, for simplicity, when discussing RRM for categorical data, we will only allow for a single random effect: for a discussion regarding dichotomous and ordinal models with multiple random effects, see Hedeker & Gibbons (1994).

The "threshold concept" (Bock, 1975) is often used in motivating models for dichotomous and ordinal response data. For instance, in probit and logistic regression models it is often assumed that there is an unobservable latent variable (y) which is related to the actual dichotomous or ordinal response (Y) through the threshold concept. For the dichotomous model, one threshold value γ_1 is assumed, while for the ordinal model, a *series* of threshold values $\gamma_1, \gamma_2, \ldots, \gamma_{J-1}$ is assumed, where J equals the number of ordered categories, $\gamma_0 = -\infty$, and $\gamma_J = \infty$. Here, a response occurs in category j ($Y = j$) if the latent response process y exceeds the threshold value γ_{j-1}, but does not exceed the threshold value γ_j. Note that the threshold values are of increasing magnitude.

The random-intercepts regression model for the latent response strength y_{ik} can be written as follows:

$$y_{ik} = \beta_0 + \beta_1 t_{ik} + \beta_3 x_{ik} + \upsilon_i + \varepsilon_{ik} \quad (5)$$

where everything is as before except that y_{ik} is not actually observed (instead the dichotomous or ordinal representation Y_{ik} is observed), and for simplicity we remove the "0" subscript from the random intercept υ_i. With this random-intercepts regression model for the underlying and unobservable variable y_{ik}, the probability, for a given subject i, that $Y_{ik} = j$ (a response occurs for subject i at time k in category j), conditional on β and υ, is given by the following equation:

$$P(Y_{ik} = j|\beta,\upsilon) = \Phi[(\gamma_j - z_{ik})/\sigma] - \Phi[(\gamma_{j-1} - z_{ik})/\sigma] \quad (6)$$

where the response model $z_{ik} = \beta_0 + \beta_1 t_{ik} + \beta_2 x_i + \beta_3 x_{ik} + \upsilon_i$ and $\Phi(\cdot)$ represents the cumulative standard normal density function. The probability of a response less than or equal to j is the first term to the right of the above equality, while the probability of a response less than or equal to $j-1$ is the second term to the right of the equality. The difference between these two then gives the probability of a response in category j. Since y is a latent variable without a fixed scale of measurement, the origin and unit of y, and therefore z, may be chosen arbitrarily. For convenience, it is common to let $\gamma_1 = 0$ and $\sigma = 1$. In this case, z_{ik} is simply a

standard normal deviate, and the conditional probability then simplifies to:

$$P(Y_{ik} = j|\beta,\upsilon) = \Phi[\gamma_j - z_{ik}] - \Phi[\gamma_{j-1} - z_{ik}] \quad (7)$$

Applying Equation (7) for a dichotomous response (coded as 1 and 2) yields:

$$P(Y_{ik} = 1|\beta,\upsilon) = \Phi[-z_{ik}] = 1 - \Phi[z_{ik}] \quad (8)$$

and

$$P(Y_{ik} = 2|\beta,\upsilon) = 1 - \Phi[-z_{ik}] = gF[z_{ik}] \quad (9)$$

since, in the dichotomous case, $\gamma_0 = -\infty$, $\gamma_1 = 0$ and $\gamma_2 = \infty$. This random intercepts (dichotomous) probit model is discussed in Gibbons & Hedeker (1994).

Without the individual-varying effect υ_i, the model given by Equations (5) and (7) is simply either a dichotomous probit regression model (Finney, 1971) or an ordinal probit regression model (McKelvey & Zavoina, 1975). Including the individual-varying effect υ_i, the probability of a response in category j, at a given timepoint, is influenced by either increasing (i.e. positive υ_i) or decreasing (i.e. negative υ_i) the value of the latent response strength y_{ik}, and thus our model z_{ik} of the latent response as well. As in the model for continuous outcomes, these subject-varying effects are assumed to be normally distributed in the population of subjects with mean 0 and variance σ_υ^2. Thus, the subject effects υ_i reflect the influence that subject i has on his/her repeated observations. To the degree that subjects exert little influence on their responses, over and above the influence of the other model terms, values of υ_i, and correspondingly σ_υ^2, will not deviate from zero. However, as subjects do exert influence on their repeated observations, values of υ_i will deviate from zero, resulting in a population variance σ_υ^2 that is greater than zero.

Assuming normally distributed errors ε_{ik} that have zero mean and unit variance (i.e. $\sigma = 1$) leads to the above probit formulation of the model, in which the response probabilities are calculated via the cumulative standard normal density function $\Phi(\cdot)$. To modify this model for the logistic response function, instead of the normal, the logistic function $\Psi(\cdot)$ replaces $\Phi(\cdot)$ in the conditional probability, with $\Psi(\gamma_j - z_{ik}) = 1/(1 + \exp[-(\gamma_j - z_{ik})])$. For the logistic response function, again we let $\gamma_1 = 0$, however, the error variance corresponding to the standard logistic distribution does not equal 1 but equals $\pi^2/3$ instead.

With the logistic response function, the model is written in terms of the log odds of response for a dichotomous outcome (with values, say, 1 and 2) as:

$$\log\left[\frac{P(Y_{ik} = 2)}{1 - P(Y_{ik} = 2)}\right] = \beta_0 + \beta_1 t_{ik} + \beta_2 x_i + \beta_3 x_{ik} + \upsilon_i \quad (10)$$

where, υ_i is the influence of subject i on the log-odds of response across all timepoints. Note that without the subject effect υ_i, this model is an ordinary logistic regression model. This model can be also written as:

$$\log\left[\frac{P(Y_{ik} > 1)}{1 - P(Y_{ik} > 1)}\right] = \beta_0 + \beta_1 t_{ik} + \beta_2 x_i + \beta_3 x_{ik} + \upsilon_i \quad (11)$$

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

$$\log\left[\frac{P(Y_{ik} = > j)}{1 - P(Y_{ik} > j)}\right] = \alpha_j + \beta_1 t_{ik} + \beta_2 x_i + \beta_3 x_{ik} + \upsilon_i \quad j = 1, \ldots, \mathcal{J} - 1, \quad (12)$$

where $\alpha_j = \beta_0 - \gamma_j$. By noting that $P(Y_{ik} > j) = 1 - P(Y_{ik} \leq j)$ and $1 - P(Y_{ik} > j) = P(Y_{ik} \leq j)$, the model can also be written as:

$$\log\left[\frac{P(Y_{ik} \leq j)}{1 - P(Y_{ik} \leq j)}\right] = \gamma_j - (\beta_0 + \beta_1 t_{ik} + \beta_2 x_i + \beta_3 x_{ik} + \upsilon_i \quad j = 1, \ldots, \mathcal{J} - 1, \quad (13)$$

In either representation [i.e. Equation (12) or (13)], a positive value for a regression coefficient β indicates a positive association between Y and the regressor variable, in the sense that large values of Y are relatively more likely to occur at large values of the regressor variable. The latter representation given in Equation (13), however, is more readily seen as a random-intercepts (i.e. υ_i) generalization of the proportional odds model for independent (i.e. not longitudinal) ordinal data described in McCullagh (1980), Agresti (1989) and elsewhere.

The 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 (12) specifies $\mathcal{J} - 1$ cumulative logits each contrasting the combined first j categories to the remaining combined $(\mathcal{J} - j)$ categories. For example, with four possible response categories

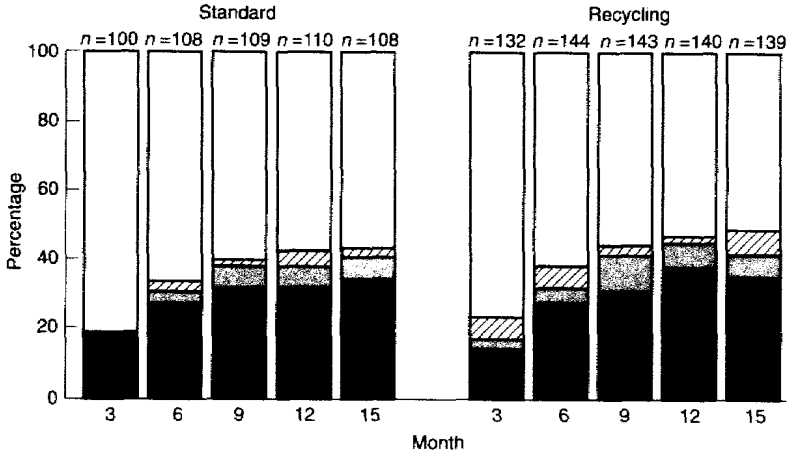

Figure 1. *Days abstinent—end of treatment abstainers by group and time (■ 0 days; ▤ 1–2 days; ▨ 3–5 days; □ 6–7 days).*

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

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

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

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

Since the regression coefficients β do not contain the j subscript, it is assumed that the effect of a regressor variable is the same across these $\mathcal{J}-1$ cumulative logits, or proportional across the cumulative odds. The odds of a response in a category greater than j (for any fixed j) is multiplied by $\exp(\beta_h)$ for every unit change in the regressor variable x_h. In the above case with four ordered categories, the single model simultaneously describes the effect of x_h 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 regressor variable: the simultaneous effect of the regressor variable on the $\mathcal{J}-1$ cumulative logits. If instead, one is interested in separate effects of a regressor variable on each cumulative logit, one can perform $\mathcal{J}-1$ dichotomous analyses by appropriately collapsing the response categories.

Repeated ordinal outcomes: change across time in smoking and abstinence

This first example uses RRM to analyze longitudinal data on smoking behavior. The data for this example come from a study on the use of extended telephone contact in a multi-component smoking cessation program. Subjects were randomized to receive one of two types of treatment (standard or recycling condition) that differed in the content of the phone calls. The goals of the phone calls were: (1) to encourage subsequent quit attempts in subjects still smoking at the end of treatment; (2) to help prevent relapse in subjects who were abstinent at the end of treatment; and (3) to recycle subjects who had relapsed. For the purposes of this example, we will focus on subjects who were abstinent at the end of treatment (with the goal of preventing relapse) and on those who were smoking at the end of treatment (with the goal of encouraging subsequent abstinence). 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. At each follow-up point, the number of days abstinent during that week was obtained and categorized into one of four outcome categories: 0, 1–2, 3–5, or 6–7 days abstinent. Figures 1 and 2 show, separately for the end-of-treatment abstainers and smokers, the observed percentages in these four categories across time for the two intervention groups.

As can be seen in Figs 1 and 2, although the

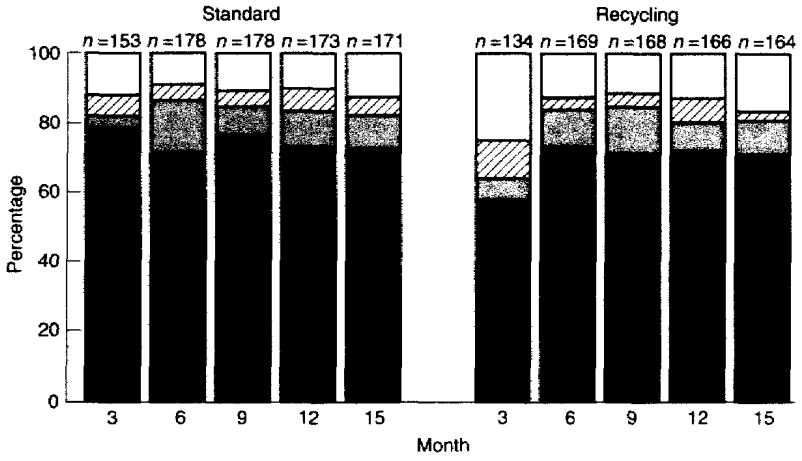

Figure 2. *Days abstinent—end of treatment smokers by group and time (■ 0 days; ▨ 1–2 days; ▩ 3–5 days; □ 6–7 days).*

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 groups. Treating the number of days abstinent as an ordinal outcome, rather than recoding it into a dichotomous outcome, is advantageous since one of the issues with relapse 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 relapse.

Figures 1 and 2 also list 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 timepoint. Also, in terms of an intervention effect, there appears to be no obvious differences between the intervention groups for end of treatment abstainers. Alternatively, for end of treatment smokers, 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, there appears to be a trend of decreased abstinence across time, especially for the end of treatment abstainers.

To more formally examine whether the intervention groups differ in terms of their smoking status across time, the following random-intercepts ordinal logistic regression model was fit separately for end of treatment abstainers and smokers:

$$\log\left[\frac{P(Y_{ik}>j)}{1-P(Y_{ik}>j)}\right] = \alpha_j + \beta_1 T_{ik} + \beta_2 G_i + \beta_3 (G_i \times T_{ik}) + v_i \qquad (14)$$

where T = time in months (coded as month 3 = 0, month 6 = 1, month 9 = 2, month 12 = 3 and month 15 = 4), G = group (0 for standard and 1 for recycling) and, again, $\alpha_j = \beta_0 - \gamma_j$. Since the coding of zero for the first timepoint corresponds to month 3 of the study, T more specifically means time in months *after 3 months*. Also, unit increases of T correspond to 3-month intervals. Using these codings for the variables T and G has the consequence that β_0 and β_1 represent the intercept (i.e. 3-month) and slope across time for the standard group, while β_2 and β_3 represent intervention group differences in the intercept (i.e. at 3 months) and slope, respectively. Also, with the unit coding of T for every 3-month follow-up, the slopes of the time-related terms (β_1 and β_3) represent change per 3 months.

Table 1 lists the parameter estimates, standard errors and p-values for the model parameters, separately for end of treatment abstainers and smokers. The p-values in the table are obtained

Table 1. *Change across time in smoking abstinence. Random-effects ordinal logistic regression estimates (standard errors)*

	End of treatment abstainers		End of treatment smokers	
Parameter	Estimate	$p<$	Estimate	$p<$
STD intercept (at 3 months) β_0	3.652 (0.334)	0.001	− 2.422 (0.302)	0.001
STD slope (per 3 months) β_1	− 0.556 (0.075)	0.001	0.061 (0.066)	0.354
Intervention group differences				
Intercept difference β_2	− 0.124 (0.433)	0.774	0.775 (0.394)	0.049
Slope difference β_3	− 0.028 (0.098)	0.776	− 0.263 (0.089)	0.003
Subject SD σ_v	3.989 (0.311)	0.001	3.034 (0.216)	0.001
Threshold γ_1	0.0		0.0	
Threshold γ_2	0.709 (0.068)	0.001	1.159 (0.074)	0.001
Threshold γ_3	1.196 (0.085)	0.001	1.943 (0.089)	0.001

STD = standard group.

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 χ^2 on 1 degree of freedom. In either case, the p-values are identical.

Inspection of Table 1 reveals that for end of treatment abstainers, there is no evidence of group-related differences either in terms of the intercept or slope. As Fig. 1 illustrates, both groups have relatively similar rates in the four abstinence categories at 3 months, and the trend in these rates across time is also quite similar. There is a significant negative month effect ($\hat{\beta}_1 = -0.556$, $p < 0.001$) indicating that higher values of abstinence (e.g. 6–7 days) are relatively less associated with higher values of time (e.g. 15 months). In other words, abstinence decreases significantly across time.

To interpret the scale of the parameter estimates, consider again the model given by Equation (14). As the model indicates, the cumulative logits depend on the value of the individual effect v_i. Thus, in what follows, the population mean value of 0 will be used in discussing estimated cumulative logit values. Turning to the parameter estimates in Table 1, the intercept β_0 is estimated as positive and as being significantly different from zero. This indicates that the initial (i.e. 3-month) proportion of responses in categories above the first category of 0 days abstinent (i.e. 1–7 days abstinent, or combining categories of 1–2 days, 3–5 days and 6–7 days) is significantly greater than 0.5. In other words, the odds of a 1–7 versus 0 days abstinent response is significantly greater than 1, or the log-odds is significantly greater than 0. This is clearly seen in Fig. 1 which indicates that, overall, approximately 80% of the 3-month responses are in one of the abstinence categories exceeding 0 days. Next, the estimated thresholds in Table 1 reflect the marginal proportion of responses associated with responses in (combined) categories of 1–7 days ($\beta_0 - \gamma_1$), 3–7 days ($\beta_0 - \gamma_2$) and 6–7 days ($\beta_0 - \gamma_3$) abstinence, each relative to the remaining (combined) categories (i.e. 0 days, 0–2 days and 0–5 days, respectively). The estimated cumulative logits at 3 months are then 3.652 (1–7 vs. 0 days), $3.652 - 0.709 = 2.943$ (3–7 vs. 0–2 days) and $3.652 - 1.196 = 2.456$ (6–7 vs. 0–5 days). Since even the last cumulative logit is positive, the estimated proportion of subjects in

even the most stringent abstinence category (i.e. 6–7 days) still exceeds 0.5 at 3 months. Due to the coding of the group variable in the model, these estimated cumulative logits are for the standard condition. For the recycling condition, the estimate of β_2 would be added in (i.e. $3.652 - 0.124 = 3.528$, $2.943 - 0.124 = 2.819$ and $2.456 - 0.124 = 2.332$, respectively), reflecting the very similar 3-month abstinence rates between these two groups (simultaneously contrasting 1–7 vs. 0 days, 3–7 vs. 0–2 days and 6–7 vs. 0–5 days).

To interpret the scale of the time-related parameters, note that since β_1 is estimated as being significantly negative, abstinence decreases significantly across time for the standard condition. Furthermore, since the estimate of β_3 is negative, the trend across time for the recycling condition decreases even more; however, this is not a statistically meaningful difference. To determine the degree of change across time for both groups, estimated cumulative logits based on Equation (14) can be calculated using these estimates of the time-related terms (i.e. $\hat{\beta}_1$and $\hat{\beta}_3$). For example, the estimated logit for 6–7 days abstinent versus 0–5 days abstinent at month 15 (i.e. $T = 4$) for the standard condition is $3.652 + 4 \times (-0.556) - 1.196 = 0.232$. Since this estimated logit is greater than zero, the estimated proportion is still greater than 0.5 for this most stringent abstinence category at month 15. Similarly, for the recycling condition this estimated logit is approximately equal to zero $(3.652 + (-0.124) + 4 \times (-0.556 - 0.028) - 1.196 = -0.004)$, indicating an estimated proportion of roughly 0.5 for this abstinence category. Thus, even at the last timepoint, the estimated proportion of subjects in the 6–7 days abstinence category is greater than or equal to the proportion in the combined category of 0–5 days abstinent.

For the end of treatment smokers Table 1 indicates that, contrary to end of treatment abstainers, group-related differences are observed. First, for the standard condition, the estimated intercept of -2.422 indicates that the 3-month proportion of 1–7 abstinent day responses is significantly less than 0.5 (or the odds of 1–7 versus 0 abstinent days is significantly less than 1). Again, the estimated thresholds simply reflect the marginal response proportions in the four ordered categories. The initial group difference of 0.775 ($p < 0.049$) indicates that the groups differ at 3 months. At this initial timepoint, the recycling group has a greater proportion of abstinence responses (simultaneously contrasting 1–7 vs. 0 days, 3–7 vs. 0–2 days and 6–7 vs. 0–5 days) than the standard condition. The estimate of the month effect for the standard condition is not statistically significant ($p < 0.354$), indicating relatively little change in abstinence rates across time for this condition. However, the group by time effect is significant and negative ($p < 0.003$), indicating that the recycling condition has decreased abstinence across time, relative to the standard condition. Together, the two significant group-related effects represent an initial beneficial effect of the recycling condition, relative to the standard condition, which goes away across time. Inspection of Fig. 2 illustrates this interpretation of the results. Again, to more fully characterize the scale of the parameter estimates, estimated cumulative logits and proportions can be obtained as was done for the end of treatment abstainers.

For both models presented in Table 1, 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 3.989 and 3.034 for end of treatment abstainers and smokers, respectively. 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, the intrasubject correlation is estimated as 0.83 and 0.74 for end of treatment abstainers and smokers, respectively [e.g. $(3.989)^2/((3.989)^2 + \pi^2/3) = 0.83$].

It should be noted that there are some concerns in using the standard errors in constructing a hypothesis test for the random-effect variance term (i.e. the subject standard deviation σ_0), particularly when the variance is near-zero and the number of subjects is small (Bryk & Raudenbush, 1992); in this case, the likelihood-ratio χ^2

Table 2. *Measurement of motivation before and after a relapse for five hypothetical subjects*

	Motivation scores across time						
Subject	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
1	x	x	X	X	X	X	X
2	x	x	x	X	X	X	X
3	x	x	x	x	X	X	X
4	x	x	x	x	x	X	X
5	x	x	x	x	x	x	X

x = scores prior to the (first) relapse.
X = scores following the (first) relapse.

test (Silvey, 1975) may be used by comparing a model ignoring subject variance to a model that includes it. For these data, the likelihood-ratio χ^2 statistic equals 611.8 and 570.8 (each on one degree of freedom) for end of treatment abstainers and smokers, respectively. Thus, again, there is ample evidence that the repeated smoking status assessments are 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. There were no significant overall condition effects for subjects who were abstinent at the end of treatment. For the end-of-treatment smokers, there was an initial benefit to the recycling condition—a greater proportion of these subjects, compared to the standard condition, were able to achieve some abstinent periods following the end of treatment. However, this advantage disappeared by the end of the follow-up period.

A longitudinal continuous outcome: the effects of relapse on motivation

In the previous example we examined smoking status as a time-varying ordinal outcome and related effects of treatment group and time to the longitudinal ordinal outcomes. In this example, we focus on changes that occur both before and after a relapse, defining relapse as a dichotomous event. If we were interested in examining influences on the time to a relapse, survival analysis methods could be used. Here, instead, we will examine the degree to which motivation scores change before and after a first relapse to smoking. The advantage of RRM in this case is that we can take a more dynamic view of the relapse process, examining how motivation changes both before and after the event, rather than selecting one time slice of the process. In addition, RRM allow us to treat relapse in "real" time, as a time-varying event, rather than having to group together all subjects who have relapsed by a set time, regardless of the recency of the event. Thus, RRM provide a more fine-tuned analysis of the relapse process.

Before describing the analysis, we first consider possible patterns of longitudinal data that may occur. Table 2 depicts patterns of data prior to and following a first relapse from five hypothetical subjects. Since the time of relapse varies from individual to individual, the amount of information both before and after the relapse also varies from individual to individual. To contrast what happens to motivation before and after a relapse, for simplicity, we will model the two periods separately and examine trend in motivation before and after the relapse occurs. Another possibility would be to build a combined model that contrasts the two periods, and to test the degree to which the trend in motivation scores varies before and after a relapse occurs.

Data for this example come from the same study described in the previous example. Subjects for this analysis had met at least a 7-day criterion of abstinence at the end of treatment. Motivation was assessed with a one item, 10-point Likert scale question, "How motivated are you to stop smoking/stay quit?" Following the end of treatment, motivation was measured weekly for 3 weeks, then every 2 weeks for 8 weeks, and then every 3 months for 1 year thereafter. Smoking behavior was assessed on the same measurement schedule. Relapse was

Table 3. *Coding of time for data from 8 timepoints and four hypothetical subjects*

	Weeks relative to relapse							
Subject	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
1	−2	−1	0	1	2	3	4	5
2	−3	−2	−1	0	1	2	3	4
3	−4	−3	−2	−1	0	1	2	3
4	−5	−4	−3	−2	−1	0	1	2

Week = 0 corresponds to the week of relapse.

defined as smoking 2 or more days a week. Since our interest was in estimating differences in trends in motivation across time, subjects were included only if they had at least two motivation scores both prior to and following the first relapse. In all, there were 87 subjects who met this criterion. To this group of 87 relapsers, a one-to-one set of matched abstinent subjects (matched on treatment condition, approximate date of entry into the study, and amount of data) was formed. This group was used as a comparison group to the relapsers in terms of changes in motivation over time. The primary question of interest is whether there is differential change across time in both pre-and post-relapse motivation scores for relapsers as compared to continuous abstainers.

In these analyses, "time" was treated as weeks relative to the relapse. For example, based on our selection criteria, Table 3 depicts four codings of time that would be possible with data from eight measured timepoints. As depicted, subject 1 experiences a relapse in week 3, subject 2 in week 4, etc. Thus, the week of relapse "anchors" the value of time for the measurements before and after the relapse. If the motivation score was obtained 2 weeks prior to the relapse, time equals −2, if the score was obtained 2 weeks after the relapse, time equals 2. As such, we will be examining whether there are changes in motivation across time, when time is anchored to the point of relapse, or more specifically, (a) is there trend in motivation in the weeks prior to a relapse? and (b) is there trend in motivation in the weeks following a relapse? Furthermore, since those with a relapse were each matched to a continuous abstainer. we will be able to contrast the trends of relapsers with continuous abstainers over the same time period.

For illustrative purposes, Table 3 lists the coding of time for subjects with data at eight timepoints. For analysis purposes, as mentioned, subjects were included if they had at least two motivation measurements both before and after the relapse, with motivation measurements made at the time the relapse was reported considered as part of the post-relapse period. Thus, at a minimum, a subject could have time codings of any two different negative values in the analysis of the pre-relapse period, and codings of any two different positive values (with a "0" coding as one of the two possible) in the post-relapse analysis. Of the 87 relapsed subjects, only 10 did not have a motivation score during the relapse week. Table 4 lists the number of measurements by group for these two periods.

As can be seen from Table 4, due to the one-to-one matching, the groups are quite similar in terms of the numbers of observations contributed to both the pre-relapse and post-relapse analyses. Also, again due to the matching, the groups are very similar in terms of the values of time. Computing an average value of time for each subject in the pre and post periods, the mean of these average time values equals −4.81 (abstainers) and −4.85 (relapsers) for the pre-relapse period and 6.23 (abstainers) and 6.41 (relapsers) for the post-relapse period. Thus, the matching was successful in ensuring that the groups are comparable in terms of the number of observations and in the timing of those observations.

The following random-effects model for motivation scores across time was used for both the pre- and post-relapse analyses:

$$\begin{aligned} y_{ik} = {} & \beta_0 + \beta_1 \text{Week}_{ik} + \beta_2 \text{Week}_{ik}^2 + \\ & \beta_3 G_i + \beta_4 (G_i \times \text{Week}_{ik}) \\ & + \beta_5 (G_i \times \text{Week}_{ik}^2) + \\ & \upsilon_{0i} + \upsilon_{1i} \text{Week}_{ik} + \upsilon_i \text{Week}_{ik}^2 + \varepsilon_{ik} \end{aligned} \qquad (15)$$

where G_i equals 0 if individual i is a matched continuous abstainer, and equals 1 if individual i

Table 4. *Frequency count: number of pre- and post-relapse motivation measurements by group*

Number of measurements	Pre-relapse		Post-relapse	
	Relapser	Abstainer	Relapser	Abstainer
2	10	14	4	3
3	23	20	16	21
4	16	16	11	13
5	10	13	12	14
6	14	15	20	17
7	12	8	16	13
8	2	1	8	6
Total	87	87	87	87

is a relapser. The inclusion of the $Week^2$ term allows us to examine the degree to which the trend across time is curvilinear. As a result of the above dummy-coding for the group effect, β_0, β_1, and β_2 represent the trend across time for continuous abstainers. In particular, β_0 represents the value of motivation during the "relapse week" (i.e. when $Week = 0$) for continuous abstainers, while β_1 and β_2 characterize the linear and curvilinear components to their trend across time. With the coding of *Week* as described (i.e. anchoring around the relapse week) the group effect β_3 represents differences between groups at the relapse week, while the two group by time interaction terms (β_4 and β_5) represent differential trends between-groups of a linear and curvilinear nature, respectively. Finally, as postulated, the above model allows individuals to deviate from their group curve in terms of the intercept (υ_{0i}), linear trend (υ_{1i}), and curvilinear trend (υ_{2i}). Note that, in this model, individual variation around the intercept represents individual variation in motivation scores during the relapse week. Table 5 lists the results of these analyses for both periods.

Considering results from the pre-relapse period, the estimate of the intercept of 9.412 indicates a very high degree of motivation for the abstainers at the "0" point in time (at the time of relapse). Since both the linear and quadratic terms are not significant, there is no evidence of change across time in motivation levels for the abstainers. In terms of the group-related terms, only the intercept difference is significant ($\hat{\beta}_3 = -1.119$, $p < 0.001$), indicating a lower motivation level for relapsers of approximately 1 point that is consistent across time. Figure 3 depicts the observed and estimated means for these two groups across time.

As the graph and results indicate, relapsers have lower motivation scores consistently across time. There is some suggestion based on the observed means in the graph that motivation scores drop dramatically one week before the relapse; however, the mean at this point in time is based on only 10 of the 87 relapsers. A more conclusive analysis would be possible if more data were available on motivation during the week immediately preceding relapse. Also, as the graph depicts, the means are at the high end of the 10-point Likert scaling of motivation, indicating a skewed distribution of motivation scores in the sample. Although we will analyze these data as continuous outcomes, a further analysis assuming ordinal outcomes might be advisable.

In terms of the random subject-varying terms, for the pre-relapse data, a model with a random intercept and linear term was fit. Although an attempt was made to add the quadratic term as a random effect as well, the data did not support independent estimation of three random effects (intercept, linear and quadratic), thus the simpler model with only two random effects is reasonable.

For the post-relapse period, the intercept was estimated as being approximately 9.6. This corresponds to the estimated motivation level for abstainers at time "0", and agrees closely with the result from the pre-relapse period. Again, there was no evidence of a significant trend across time for abstainers, indicating a generally high level of motivation that was consistent throughout this later time period. Together with

Table 5. *Change in motivation—pre- and post-relapse. Random-effects regression estimates (standard errors)*

Parameter	Pre-relapse		Post-relapse	
	Estimate	$p<$	Estimate	$p<$
ABS intercept β_0	9.412	0.001	9.552	0.001
	(0.232)		(0.200)	
ABS linear β_1	−0.025	0.74	0.0095	0.83
	(0.075)		(0.045)	
ABS quadratic β_2	−0.0037	0.53	−0.0025	0.28
	(0.0057)		(0.0023)	
Group differences				
Intercept difference β_3	−1.119	0.001	−1.962	0.001
	(0.332)		(0.284)	
Linear difference β_4	−0.163	0.13	−0.295	0.001
	(0.108)		(0.064)	
Quadratic difference β_5	−0.0093	0.26	0.013	0.001
	(0.0082)		(0.0032)	
σ^2_{v0}	1.213		2.440	
	(0.242)		(0.381)	
σ_{v0v1}	0.068		−0.088	
	(0.028)		(0.066)	
σ^2_{v1}	0.0076		0.076	
	(0.0036)		(0.019)	
σ_{v0v2}			−0.00031	
			(0.0031)	
σ_{v1v2}			−0.0034	
			(0.00089)	
σ^2_{v2}			0.00018	
			(0.000044)	

p-values not given for variance and covariance parameters (see Bryk & Raudenbush, 1992, p. 55).
ABS = matched abstainer group.

the results from the "pre-relapse" period, these analyses suggest a very high level of motivation that does not vary over time for those subjects who are continuous abstainers. In terms of the group-related effects, there was a significant difference of approximately 2 points at the time of relapse ($\hat{\beta}_3 = -1.962$, $p < 0.001$), indicating a significant drop in motivation scores during the relapse week. The two group by time effects are also highly significant and, based on their magnitude and sign, indicate a significant linear drop in motivation scores across time for the relapsers (the estimated linear trend for relapsers equals $0.0095 - 0.295 = -0.2855$) that eventually levels off and then slowly rises again (the estimated quadratic trend for relapsers equals $-0.0025 + 0.013 = 0.0105$). Based on these estimates for the relapsers (intercept = 7.59, linear slope = −0.2855 and quadratic slope = 0.0105), the average number of weeks that is necessary for the downward descent of motivation scores to stop equals 14 (i.e. the vertex of the quadratic function), at which point the estimated motivation scores have dropped to a level of approximately 5.5 (i.e. the minimum value of the quadratic function). Figure 4 illustrates the observed means across time and the estimated trend lines for these two groups. As can be seen from Fig. 4, means at the last timepoint (23 weeks after relapse) seem somewhat atypical, especially for matched abstainers. However, the amount of data at this timepoint is limited: there are only 10 of 87 subjects in each group. To ensure that the results were not unduly influenced by data from this last timepoint, the analysis was re-run excluding these data with very similar results.

For the post-relapse data, all three terms (intercept, linear and quadratic) were treated as random, and the estimation procedure was able to estimate the population variance covariance matrix of these three subject-varying effects. Not surprisingly, based on these data, the estimate of the variation during the week of relapse ($\hat{\sigma}^2_{v0}$) is

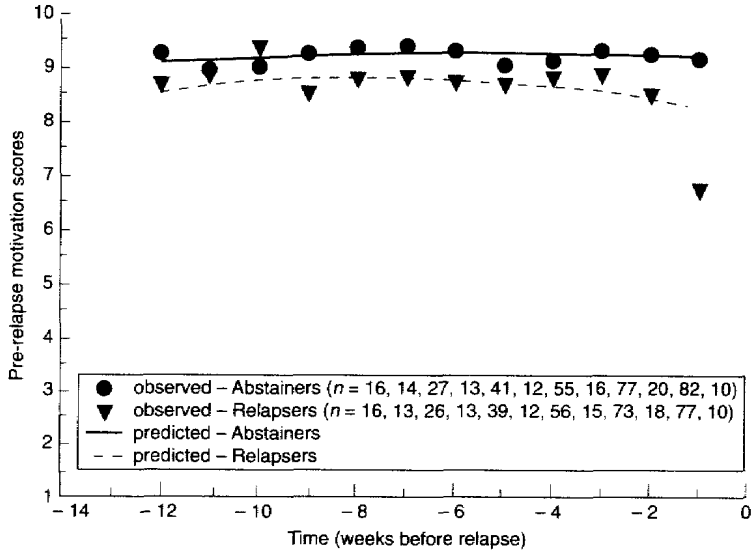

Figure 3. *Pre-relapse motivation scores by time. (Note: not shown are 4 obs. at time = − 14 and 3 obs. at time = − 13).*

appreciably larger during the post-relapse period, as compared to the pre-relapse period. This is most likely due to the fact that the data for the relapse week was considered only in the post-relapse analysis; the pre-relapse analysis only includes data prior to the relapse. As such, the pre-relapse analysis is essentially predicting the amount of variation in motivation scores for the week of relapse using only information before the relapse occurs, whereas the post-relapse analysis has the relapse week information in estimating this variance term. In general, not surprisingly, the amount of individual variation is much greater during the post-relapse period.

In sum, RRM allows a dynamic view of possible precipitants and aftermath of a relapse. As the data in the example show, motivation over time may be a good predictor of impending relapse. Not only did future relapsers have consistently lower motivation levels, prior to their relapse, than did continuous abstainers, but their motivation may have precipitously dropped in the week prior to the relapse. In addition, the results indicate that relapse has a devastating effect on motivation, and on average, subjects needed about three months for their motivation levels to start to climb again. Thus, RRM provide a means of better understanding relapse and recovery processes. Regardless of when in real time the relapse occurred, we can examine trends in motivation both before and after the event. Although the focus here was on changes in motivation before and after the first relapse, the model could be extended in order to investigate the precipitants and sequelae of repeated relapses as well.

Discussion

The study of relapse presents a variety of analytic problems: people relapse at different points during follow-up data collection; the "relapse" itself may be a transitional and not very stable state; there may be different levels of relapse; frequently data are missing; and the relationship between hypothesized predictor variables and relapse may vary by time and by the recency of their joint measurement. Until recently, researchers have had to deal with these analytic dilemmas by reducing their data to more manageable or "cleaner" subsets or by looking only at a limited number of data points at a time. Thus, the richness of many datasets is lost, and the analysis of the relapse process may have been overly simplified. RRM offers a solution to many of these problems with longitudinal data and can

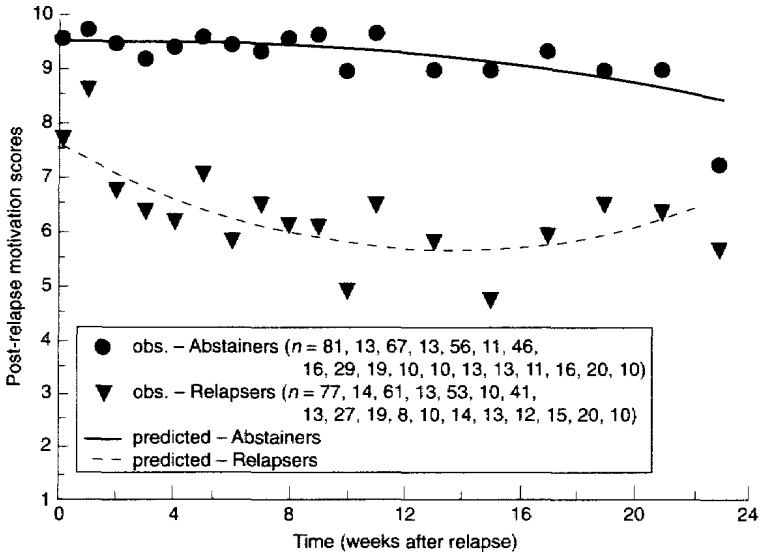

Figure 4. *Post-relapse motivation scores by time.*

be used both for analyzing treatment effects and for examining precipitants and sequelae of the relapse process. Although this paper illustrated the use of RRM with smoking data, it is equally appropriate for use with other substance use data as well.

For longitudinal data analysis, an attractive and important 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 timepoints. Furthermore, as our second example illustrated, subjects can even be measured at different timepoints, which is valuable given that subjects relapse at different timepoints. 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 trend lines that are observed for the whole sample. The model estimates the subject's trend across time based on whatever data that subject has, augmented by the time-trend that is estimated for the sample as a whole and effects of all covariates in the model.

In the examples, we examined repeated observations nested within individuals. In the terminology of multilevel 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 referred to as two-level models. Individuals themselves, though, are often observed clustered within some higher-level unit, for example, a classroom, clinic, or worksite. 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 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) and Longford (1993).

No statistical model is without its limitations, and RRM is no exception. First, the statistical tests typically used in RRM are "large-sample" tests based on asymptotic statistical theory. When applied to small samples, one must be cautious in drawing statistical inferences.

However, in this regard, it should be kept in mind that these large-sample tests (i.e. Wald-tests and likelihood-ratio tests) are also regularly used in logistic regression, log-linear analysis and structural equation modeling. Secondly, the assumption regarding missing data, although flexible, is sometimes unreasonable. This occurs mainly when it is thought that the missing data are related to the value of the outcome variable that would have been observed, and cannot be ade- quately predicted by other non-missing variables (e.g. baseline characteristics of the subject or available outcome measurements). In this case, the missing data are said to be "non-ignorable" (Laird, 1988). Missing smoking status data would be non-ignorable, for example, if subjects who are missing at a particular timepoint are more likely to be smoking than subjects who are assessed at the same timepoint, and this association between missingness and smoking status cannot be explained by past smoking status assessments of the subject or other subject characteristics (e.g. their baseline level of smoking, or their treatment group). In the case of non-ignorable missing data, extended approaches using random-effects analysis are possible (Little, 1995; Hogan & Laird, 1996). Third, the assumption that the individual-specific effects are normally distributed in the population may be questionable in certain circumstances. Some work has been undertaken (Butler & Louis, 1992) to relax this assumption of normality; however, more work is necessary in this area. Finally, the model for the trends across time is a model that is linear in its parameters. While a linear model can fit some forms of non-linear growth through the use of polynomials or variable transformation, more sophisticated methods for more general non-linear growth across time (Vonesh & Carter, 1992) may sometimes be necessary.

Computer programs

Computer software for RRM is becoming increasingly available, especially for normal-theory models (VARCL, Longford, 1986; the BMDP 5V procedure, Schluchter, 1988; HLM, Bryk, Raudenbush & Congdon, 1994; MLn, Rasbash *et al.*, 1995; the SAS procedure MIXED). A detailed comparison of some of these programs is included in Kreft, de Leeuw & van der Leeden (1994). Also, the multilevel homepage (http://www.ioe.ac.uk/multilevel/) provides a wealth of information about the MLn program in particular, and multi-level analysis and RRM in general. For the analyses presented in the current article on changes in motivation scores prior to and following a relapse, the MIXREG program (Hedeker & Gibbons, 1996b) was used. Software programs for RRM are also available for dichotomous (EGRET, Statistics and Epidemiology Research Corporation, 1991) and ordinal (MIXOR, Hedeker & Gibbons, 1996a) response data. The MLn, HLM and VARCL programs additionally have facilities for analysis of dichotomous outcomes. In the current article, MIXOR was used for the analyses presented on changes in smoking status across time. Both MIXREG and MIXOR can be obtained free of charge through the internet from either the multi-level homepage or location http//www.uic.edu/~hedeker/mix.html. Readers interested in the specifications that were required to run MIXREG and MIXOR for the examples in this article, can send a note to the first author at hedeker@uic.edu.

Conclusion

As demonstrated, RRM provides a useful way of analyzing longitudinal relapse data. Specifically, RRM allows 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, dichotomous and ordinal outcomes. Hopefully, this article has demonstrated that RRM represents an important and useful tool for analyzing longitudinal data, and has encouraged use and understanding of these methods.

Acknowledgements

The authors thank Cherry Lowman, PhD for many helpful and constructive comments and for organizing the workshop described below. The authors also thank Siu-Chi Wong for computer and statistical programming assistance. Preparation of this article was supported by National Heart Lung and Blood Institute Grant HL 42485 and National Institutes of Mental Health Grant MH44826-01A2.

Portions of this article were presented at the Methodologies for Relapse Research Workshop sponsored by the Alcohol Research Utilization System (ARUS), National Institute on Alcohol Abuse and Alcoholism, January 1993, Santa Fe, New Mexico.

References

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

ANDERSON, D. A. & AITKIN, M. (1985) Variance component models with binary response: interviewer variability, *Journal of the Royal Statistical Society*, 47, 203–210.

BENTLER, P. M. & NEWCOMB, M. D. (1992) Linear structural equation modeling with nonnormal continuous variables, in: DWYER, J. H., FEINLEIB, M., LIPPERT, P. & HOFFMEISTER, H. (Eds) *Statistical Models for Longitudinal Studies of Health*, pp. 132–160 (New York, Oxford University Press).

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

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. (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 Publications, Inc.).

BRYK, A. S., RAUDENBUSH, S. W. & CONGDON, R. T. (1994) *Hierarchical Linear Modeling with the HLM/2L and HLM/3L Programs* (Chicago, Scientific Software, Inc.).

BUTLER, S. M. & LOUIS, T. A. (1992) Random effects models with non-parametric priors, *Statistics in Medicine*, 11, 1981–2000.

CLOGG, C. C., ELIASON, S. R. & GREGO, J. M. (1990) Models for the analysis of change in discrete variables, in: VON EYE, A. (Ed.) *New Statistical Methods in Longitudinal Research, vol. 2*, pp. 409–442 (San Diego, CA, Academic Press).

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

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., LIANG, K.-Y. & ZEGER, S. L. (1994) *Analysis of Longitudinal Data* (New York, Oxford University Press).

DUNLOP, D. D. (1994) Regression for longitudinal data: a bridge from least squares regression, *The American Statistician*, 48, 299–303.

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

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

FINNEY, D. J. (1971) *Probit Analysis* (Cambridge, England, Cambridge University Press).

GIBBONS, R. D. & BOCK, R. D. (1987) Trend in correlated proportions, *Psychometrika*, 52, 113–124.

GIBBONS, R. D., HEDEKER, D., ELKIN, I. *et al.* (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.

GIBBONS, R. D. & HEDEKER, D. (1994) Application of random-effects probit regression models, *Journal of Consulting and Clinical Psychology*, 62, 285–296.

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).

GRAHAM, J., COLLINS, L. M., WUGALTER, S. E., CHUNG, N. K. & HANSEN, W. B. (1991) Modeling transitions in latent stage-sequential processes: a substance use prevention example, *Journal of Consulting and Clinical Psychology*, 59, 48–57.

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. & 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.

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.

HOGAN, J. W. & LAIRD, N. M. (1996) Mixture models for the joint distribution of repeated measures and event times, *Statistics in Medicine*, in press.

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.

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

JONES, R. H. (1993) *Longitudinal Data Analysis with Serial Correlation: A State–Space Approach* (New York, Chapman and Hall).

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.

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

LONGFORD, N. T. (1986) VARCL—interactive software for variance component analysis, *The 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.

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

MCARDLE, J. J. & HAMAGAMI, F. (1991) Modeling incomplete longitudinal data using latent growth structural equation models, in: COLLINS, L. & HORN, J. L. (Eds) *Best Methods for Analysis of Change*, pp. 276–304 (Washington, DC, APA Press).

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

MCKELVEY, R. D. & ZAVOINA, W. (1975) A statistical model for the analysis of ordinal level dependent variables, *Journal of Mathematical Sociology*, 4, 103–120.

PETERSEN, T. (1990) Analyzing event histories, in: VON EYE, A. (Ed.) *New Statistical Methods in Longitudinal Research, vol. 2*, pp. 258–288 (San Diego, CA, Academic Press).

RASBASH, J., YANG, M., WOODHOUSE, G. & GOLDSTEIN, H. (1995) *MLn: Command Reference Guide* (London, Institute of Education, University of London).

SCHLUCHTER, M. D. (1988) 5V: unbalanced repeated measures models with structured covariance matrices, in: DIXON, W. J. (Chief Ed.) *BMDP Statistical Software Manual*, vol. 2, pp. 1081–1114 (Berkeley, CA, University of California Press).

SHARMA, R. P., HEDEKER, D., PANDEY, G., JANICAK, P. G. & DAVIS, J. M. (1992) A longitudinal study of plasma cortisol and depressive symptomatology by random regression analysis, *Biological Psychiatry*, 31, 304–314.

SILVEY, S. D. (1975) *Statistical Inference* (New York, Chapman & Hall).

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.

VELICER, W. F., MARTIN, R. A. & COLLINS, L. (1996) Latent transition analysis for longitudinal data, *Addiction*, 91: Supplement, S197–S210.

VONESH, E. F. & CARTER, R. L. (1992) Mixed-effects nonlinear regression for unbalanced repeated measures, *Biometrics*, 48, 1–17.

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.

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. & ALBERT, P. S. (1988) Models for longitudinal data: a generalized estimating equation approach, *Biometrics*, 44, 1049–1060.