Published in final edited form as:
Eval Health Prof. 2011 June ; 34(2): 181–200. doi:10.1177/0163278710392982.

Joint Modeling of Longitudinal Data in Multiple Behavioral Change

Richard Charnigo,
Center for Drug Abuse Research and Translation, University of Kentucky; Departments of Statistics and Biostatistics, University of Kentucky

Richard Kryscio,
Center for Drug Abuse Research and Translation, University of Kentucky; Departments of Statistics and Biostatistics, University of Kentucky

Michael T. Bardo,
Center for Drug Abuse Research and Translation, University of Kentucky; Department of Psychology, University of Kentucky

Donald Lynam, and
Center for Drug Abuse Research and Translation, University of Kentucky; Department of Psychological Sciences, Purdue University

Rick S. Zimmerman
Center for Drug Abuse Research and Translation, University of Kentucky; Department of Social and Behavioral Health, Virginia Commonwealth University

Abstract

Multiple behavioral change is an exciting and evolving research area, albeit one that presents analytic challenges to investigators. This manuscript considers the problem of modeling jointly trajectories for two or more possibly non-normally distributed dependent variables, such as marijuana smoking and risky sexual activity, collected longitudinally. Of particular scientific interest is applying such modeling to elucidate the nature of the interaction, if any, between an intervention and personal characteristics, such as sensation seeking and impulsivity. We describe three analytic approaches: generalized linear mixed modeling, group-based trajectory modeling, and latent growth curve modeling. In particular, we identify the strengths and weaknesses of these analytic approaches and assess their impact (or lack thereof) on the psychological and behavioral science literature. We also compare what investigators have been doing analytically versus what they might want to be doing in the future and discuss the implications for basic and translational research.

Public health professionals and psychologists interested in changing behavior across time have, historically, been most often interested in targeting one specific behavior at a time. The general assumption has been that a highly focused intervention or treatment that targets a single health-related behavior will be more effective than a diffuse intervention that may end up “off target”. This strategy is somewhat analogous to research in drug development where a specific disease or psychopathology is targeted, and non-targeted effects of the drug are considered unwanted side effects. However, in contrast to this strategy, recently there has been more interest in developing interventions that effectively produce multiple

Correspondence concerning this paper should be addressed to: Richard Charnigo, Ph.D. 851 Patterson Office Tower, Department of Statistics University of Kentucky, Lexington KY 40506-0027 Phone: 859.218.2072, Fax: 859.257.6430, RJCharn2@aol.com.

behavioral changes, each of which has a positive consequence for health (Morabia and Costanza, 2010; Prochaska, Spring, and Nigg, 2008; Prochaska et al., 2010).

There are at least two potential advantages for examining changes in multiple behaviors in longitudinal studies. First, different behaviors often occur in tandem and studying them in tandem may uncover common etiological or mechanistic risk factors. For example, Stoolmiller and colleagues (2010) found that both exposure to R-rated films and initiation of alcohol increase across the adolescent period. More importantly, sensation seeking status moderated the relation between R-rated films and alcohol use such that exposure to R-rated films was more strongly associated with alcohol use among low sensation seekers than among high sensation seekers. Such results may have important implications for the understanding of phenotypic and environmental interactions for risk-related behaviors across development.

Second, to the extent that two or more health behaviors share common risk factors, designing a single intervention to change multiple behaviors simultaneously would be most efficient and cost effective. For example, after determining that drug use and obesity share some common links in developmental stage, risk factors, and mediators, Pentz and colleagues (Pentz, 2009; Riggs et al., 2007) adapted an evidence-based drug abuse/violence prevention program to an obesity prevention program. While this example illustrates a reduction in two different health risk behaviors, the multiple behavior change can be bidirectional, as would be the case for an intervention designed to *increase* physical activity in order to *decrease* depressive symptoms (Blumenthal et al., 2007).

The present manuscript has two purposes. The first is to describe three potentially useful analytic approaches for longitudinal data in multiple behavioral change: generalized linear mixed modeling, group-based trajectory modeling, and latent growth curve modeling. Our descriptions are non-technical in nature, as this manuscript is not intended to serve as a detailed guide to implementation. Instead, we seek to provide a basic conceptual understanding of each approach, to articulate its strengths and weaknesses, and to illustrate its impact (or lack thereof) on the psychological and behavioral science literature.

The second purpose of this manuscript is to compare what has been done analytically in the psychological and behavioral science literature, particularly when at least one of the dependent variables involved substance use, versus what might desirably be done in the future. Some of the articles referenced herein were located via Google search; however, most were already known to the authors of the present manuscript, who themselves perform research in the field of substance use. The articles referenced herein are not intended as an exhaustive inventory of the literature on multiple behavioral change, even within the field of substance use. However, they may be regarded as representative for the purpose of comparing what has been done analytically versus what might be done in the future. Our comparison is not intended to criticize any previous investigators, who were often working with the best analytic tools that were readily available to them at the time, but rather to provide relevant insights to future investigators. Indeed, we hope that this manuscript will enable future investigators to make informed decisions about data analysis in multiple behavioral change.

Before proceeding to accomplish these purposes, we describe some of the relevant research literature. Assessments of school-based programs for the prevention of substance use or abuse have typically focused on the use of two or more substances; thus, these assessments have, in a narrow sense at least, considered the impacts of the interventions on multiple behaviors. Here we review results of several of the major randomized controlled trials of school-based substance abuse programs published since 1998, with a focus on what

outcomes were assessed and how the data were analyzed. Our review includes studies assessing the 10-year effects of the original *Project D.A.R.E*. program (Lynam et al., 1999), the impact of the more recent version called *Take Charge of Your Life* (Sloboda et al., 2009), the effects of *Project ALERT* (Ringwalt et al., 2009), the results for *All Stars* (Harrington et al., 2001), and the long-term effects of the *Life Skills* program (Botvin et al., 2000). All of these programs were implemented in the 6^{th} or 7^{th} grade, and several had additional subsequent booster components. In each case, schools were randomized to condition.

Every one of these studies presented results for multiple substances (typically including at least cigarettes, alcohol, marijuana, and inhalants) but did so separately and independently for the various substances. Three of the studies used hierarchical linear modeling (HLM) to account for the clustered observations and randomization at the school level (Lynam et al., 1999; Ringwalt et al., 2009; Sloboda et al., 2009), and another used Generalized Estimating Equations (GEE) to control for clustering (Botvin et al., 2000), while the *All Stars* evaluation (Harrington et al., 2001) used neither HLM nor GEE. Conclusions about whether the programs were or were not effective seem to have been based on a review of the significance tests across the independent analyses for the various substances. The studies on the original *Project DARE, Project ALERT, Take Charge of Your Life*, and *All Stars* concluded that these programs were not effective, with only one significant effect (on alcohol) for *Project ALERT* and no significant effects on any substances for the other programs. The study on *Life Skills* found significant intervention effects on heroin and other narcotics, hallucinogens, total illicit substance use, and total illicit substance use other than marijuana. In summary, all five studies assessed impacts of the programs on individual substances separately and then drew conclusions about the overall effectiveness of the programs via some informal aggregation of the results for the individual substances (e.g., one significant result when five substances were considered would not indicate high overall effectiveness).

There has also been some research literature assessing the effects of programs on substance use and at least one other kind of behavioral outcome. The most commonly studied such programs share a focus on substance use or abuse with sexual risk, i.e., either by enrolling individuals with one set of challenges (substance abuse or HIV risk behaviors) into a program addressing the other or by combining intervention components on each topic for a target audience currently or at high risk of being challenged by both issues. We found that most articles describing the first kind of study (e.g., enrolling individuals being treated for substance abuse into a sexual risk reduction program) only discussed the impact of the program on the outcome that was the specific target of the intervention. An exception, in which an HIV risk reduction intervention was incorporated into a substance abuse treatment program and both kinds of behavioral outcomes were assessed, is described below (Copenhaver, Lee, and Margolin, 2007). As noted above, other studies on substance abuse and sexual risk-taking examined an intervention targeted at both kinds of behavioral outcomes. Bryan and colleagues combined elements of an alcohol-related intervention with one for HIV sexual risk reduction, delivering it to adolescents in detention facilities (Bryan, Schmiege, and Broaddus, 2009). Two other interventions combined elements of an alcohol-focused program with an HIV risk reduction program for populations in South Africa, one targeting secondary school students (Cupp et al., 2008) and the other targeting adults who drink in shebeens (similar to taverns, Kalichman et al., 2008). The final study that we reviewed assessed both sexual behaviors and substance use as outcomes for the *Life Skills* school-based substance abuse prevention program whose overall analysis was briefly mentioned above (Griffin, Botvin, and Nichols, 2006).

Analytic methods varied a bit more across these studies than across the multiple substance use papers discussed earlier. Three of these five studies assessed the impacts of the

interventions on behavioral outcomes separately, while the other two used methods that considered outcomes simultaneously. Kalichman and colleagues (2008) assessed the impact of a combined alcohol and sexual risk reduction program using ANCOVA for continuous outcome variables and logistic regression for categorical outcome variables; two sets of behavioral outcomes were examined, each of the outcomes (including several measures of sexual risk-taking and a measure of alcohol use before sex) was evaluated separately, and the program was deemed to impact both sets of behavioral outcomes. Cupp and colleagues (2009) presented separate sets of results for sexual and alcohol-related variables as well, using mixed model repeated measures ANCOVA since schools rather than individual students had been randomly assigned to conditions; the intervention reduced initiation of sexual intercourse but did not affect alcohol initiation or frequency. Bryan and colleagues (2009) used repeated measures ANCOVA and latent growth curve modeling in separate analyses assessing the effects of the program on alcohol problems and sexual risk-taking; the intervention was found to impact both. As noted earlier, two of the studies used methods that evaluated the impacts of the programs on multiple variables simultaneously. Copenhaver and colleagues (2007) employed multivariate ANCOVA to assess four drug-related outcomes in tandem, while the sex-related outcomes were examined together with a separate MANCOVA. Both sets of results suggested that the program was effective. However, knowledge, attitudes, self-efficacy, and behavior were examined together within each MANCOVA, leaving less than clear what to conclude about how the program impacted behavior *per se*. Griffin and colleagues (2006) used latent growth curve modeling to assess the impact of the intervention on alcohol and marijuana use. The authors tested a structural equation model hypothesizing that the prevention program, geared predominantly toward substance use, would have direct effects on substance use (as quantified by the “slope” of substance use in the five years after the intervention) and that, in turn, substance use would affect HIV risk behaviors in young adulthood. The hypothesis was confirmed with an excellent fit of the model to the data.

Generalized Linear Mixed Modeling

A generalized linear mixed model [abbreviated GLMM] (Schall 1991; Breslow and Clayton 1993; Wolfinger and O'Connell 1993; Molenberghs and Verbeke, 2005) is a sort of hybrid between a linear mixed model [LMM] (Searle et al., 1992; Verbeke and Molenberghs, 2000) and a generalized linear model [GLM] (McCullagh and Nelder, 1989; Dobson, 2002). Like a LMM, a GLMM accommodates repeated measures data for which the usual assumption of independent observations is untenable; like a GLM, a GLMM accommodates a non-normally distributed dependent variable.

Although statisticians have been developing GLMM theory and concepts since the 1990's, the recent incorporation of PROC GLIMMIX into Version 9.2 of the SAS statistical software package renders GLMMs far more accessible to behavioral and psychological science researchers than they had been in the past. Yet, even with convenient access to GLMMs, a data analyst faces the question of how to create a single model that accommodates multiple dependent variables simultaneously, as opposed to fitting a separate model for each dependent variable.

To clarify this point, we draw an analogy to a far simpler scenario: two normally distributed dependent variables measured just once, in conjunction with a single categorical independent variable. In this simpler scenario, an investigator may choose to perform two univariate analyses of variance [ANOVAs], one for each dependent variable, or the investigator may opt to perform a single multivariate analysis of variance [MANOVA]. Even putting aside the question of Type I error probability inflation, there is still ample reason to favor the one MANOVA over the two ANOVAs. As Johnson and Wichern (1998)

explain and demonstrate through a numerical example, a MANOVA may have better power to detect group differences than two ANOVAs because the MANOVA takes into account information that the two ANOVAs do not, namely the correlation between the two dependent variables.

Now, returning to the more complicated situation in which we have multiple non-normally distributed dependent variables measured more than once each (hence, a sort of doubly multivariate data structure) and a mix of continuous and categorical independent variables, a data analyst can fit a single GLMM to accommodate all dependent variables simultaneously through the artificial construction of a "pseudo dependent variable" that encompasses all measurements of each dependent variable. For example, if the two dependent variables of marijuana consumption and risky sexual activity are measured at three points in time, then a pseudo dependent variable can be defined that has six instances.

A "pseudo independent variable" is also created to identify the origin of each instance of the pseudo dependent variable. In the present example, the pseudo independent variable can be set at one level for the three instances pertaining to marijuana consumption and at a different level for the three instances corresponding to risky sexual activity. By allowing the pseudo independent variable to interact with the independent variables, the data analyst accounts for the possibility that some measured personal characteristics, environmental factors, and/or interventions may have differential effects on the multiple dependent variables. On the other hand, if the pseudo independent variable is not permitted to interact with the random effects of the GLMM, then the random effects are said to be "shared" across dependent variables. The consequence of shared random effects is that, roughly speaking, a subject inclined toward one risky behavior is presumed more inclined toward another risky behavior, even after adjustment for measured personal characteristics, environmental factors, and/or interventions.

The use of GLMMs and other statistical models with shared random effects is not yet widespread, even outside behavioral and psychological science research, although some examples can be found. Gao (2004) used shared random effects to model jointly the dependent variables of disease and death for dementia patients, while Liu et al. (2007) employed shared random effects to model jointly the dependent variables of medical costs and death for dialysis patients.

Fitting GLMMs without the assumption of shared random effects may impose an immense computational burden when there are many dependent variables. Fieuws and Verbeke (2006) suggested circumventing this difficulty by first fitting a series of preliminary models. Each preliminary model would involve one pair of dependent variables. So, for example, if four illicit drugs were being monitored, then there would be six preliminary models, namely those for drugs 1 & 2, 1 & 3, 1 & 4, 2 & 3, 2 & 4, and 3 & 4 respectively. The results from the preliminary models would then be combined to produce estimated trajectories for all of the dependent variables, almost as if a single GLMM for all of the dependent variables had been fit in the first place.

Fieuws and Verbeke (2006) showed that their procedure yields parameter estimators with the desirable statistical properties of consistency and asymptotic normality. Roughly speaking, consistency means that an estimator is likely to be close to its target when the sample size is large, whereas asymptotic normality means that an estimator is approximately normally distributed when the sample size is large, notwithstanding the possible non-normality of the dependent variables themselves. This latter feature justifies the construction of confidence intervals having the form estimate plus or minus a multiple of the standard error, where the multiple is some number such as 1.96 drawn from the upper tail of a

standard normal distribution. In addition, the procedure of Fieuws and Verbeke (2006) entails only a minor loss of efficiency, which essentially means that the standard errors are not much larger than they would have been if a single GLMM for all dependent variables had been fit in the first place.

The GLMM approach has a number of advantages. First, GLMM coefficients have essentially the same interpretations as GLM coefficients, subject to the caveat that the data analyst is also controlling for random effects; hence, the data analyst may speak in the familiar parlance of estimated odds ratios or rate ratios. Second, if using PROC GLIMMIX, the data analyst may write a CONTRAST statement to test virtually any hypothesis of interest; thus, the data analyst may create his/her own custom hypothesis tests. Third, the GLMM approach can accommodate not just two but three or more dependent variables simultaneously; this is desirable since a given study may monitor more than two different drugs or health behaviors.

However, the GLMM approach is not without its weaknesses. The estimated dependent variable trajectories based on strata defined by the independent variables may not exhibit clinically relevant differences, even if the differences are statistically significant. For example, high sensation seekers may have an estimated marijuana consumption trajectory that is significantly elevated compared to that of low sensation seekers. Yet, such a trajectory is unlikely to be so elevated as to suggest that the typical high sensation seeker is a frequent user of marijuana. In essence, there is no way to prospectively and accurately parse subjects into groups of anticipated frequent, intermittent, experimental, or not-at-all marijuana users based solely on sensation seeking (or, for that matter, any constellation of independent variables assessed prior to the observation of marijuana consumption). Another weakness of GLMMs is that, although missing data can be accommodated, observations not recorded are assumed to be missing completely at random (Little and Rubin, 1987). Such an assumption is generally implausible, although data analysts often live with it because the off-the-shelf procedures available in most statistical software packages do not accommodate sophisticated modeling of missing data patterns.

Group-Based Trajectory Modeling

A group-based trajectory model (Nagin, 1999; Nagin and Tremblay, 2001; Jones, Nagin, and Roeder, 2001; Jones and Nagin, 2007) posits that, for each dependent variable, there exist two or more groups having modest within-group variation but extreme between-group variation. For example, the dependent variable of risky sexual activity measured during adolescence may yield one group characterized by little or no risky sexual activity, a second group for which the typical experience is to succumb to risky sexual activity only late in adolescence, and a third group characterized by pervasive risky sexual activity throughout adolescence.

Since, generally speaking, such groups cannot be accurately established directly from the independent variables, a latent categorical variable is introduced to define group membership. At each level of the latent variable, a typical trajectory for the dependent variable is estimated. The probability of membership in a group is then expressed in terms of the independent variables. Optionally, group membership can also be used to make predictions regarding a subsequent outcome variable distinct from the dependent variables for which trajectories are estimated, such as whether a student graduates from college given his/her group memberships with regard to risky sexual activity and marijuana consumption during adolescence.

As with GLMMs, a data analyst employing group-based trajectory modeling with multiple dependent variables faces the question of how to create a single model that accommodates

all of the dependent variables simultaneously, as opposed to fitting a separate model for each dependent variable. When there are (exactly) two dependent variables and hence two latent variables, the typical solution is to make the probability of membership in a group defined by the second latent variable a function not only of the independent variables but also of the first latent variable. Thus, group memberships for different dependent variables are probabilistically but not deterministically linked. In other words, two subjects belonging to the same group for the first dependent variable can belong to different groups for the second dependent variable. Moreover, there is no requirement that both dependent variables have the same number of groups. This solution for creating a single model to accommodate two dependent variables is implemented in the PROC TRAJ add-on to the SAS statistical software package.

One may consider attempting a similar solution when there are more than two dependent variables, namely to make the probability of group membership defined by the third latent variable a function of the first two latent variables, the probability of group membership defined by the fourth latent variable a function of the first three latent variables, and so forth. However, this is not offered in the PROC TRAJ add-on due to the computational burden. Instead, when there are more than two dependent variables, a data analyst may choose either to handle only two dependent variables at a time or to work with all of the dependent variables simultaneously under the rather restrictive assumption that a single latent variable defines group membership for all of the dependent variables. A consequence of that assumption is that if a data analyst wishes to have three groups for (say) risky sexual activity, then he/she is also forced to have three groups for (say) marijuana consumption and alcohol consumption; moreover, the three groups for risky sexual activity are exactly the same as the three groups for marijuana consumption and the three groups for alcohol consumption.

Group-based trajectory modeling has become well established in the behavioral and psychological science literature during the past several years. We mention here only four examples, which serve to illustrate a few of the applications: (1) Barker et al. (2006) studied trajectories of reactive and proactive aggression; (2) the dependent variables of theft and violence were assessed by Barker et al. (2007); (3) Fontaine et al. (2008) examined patterns of hyperactivity and physical aggression; and, (4) the dependent variables of violence, vandalism, theft, and drug use were studied by van Lier et al. (2009).

The strengths and weaknesses of group-based trajectory modeling are largely complementary to those of generalized linear mixed modeling, except that group-based trajectory modeling also has the limitation of assuming that missing data occur completely at random. On the other hand, unlike generalized linear mixed modeling, group-based trajectory modeling usually creates groups that are so qualitatively different on the dependent variable that there is no question of whether clinically relevant differences exist between groups. For instance, if there are two groups associated with the dependent variable of marijuana consumption, then one group will likely have an estimated trajectory indicative of virtually no marijuana consumption while the other group will likely have an estimated trajectory suggestive of frequent marijuana consumption, at least by the end of the study. Although the latter group may include disproportionately many high sensation seekers, for example, one must keep in mind that neither sensation seeking nor any other independent variable actually defines the groups. Indeed, there will be many high sensation seekers in the former group also. The clinically relevant differences between groups, as well as the typically modest numbers of groups chosen for most data analyses, also ensure that the results of group-based trajectory modeling will be highly amenable to visual interpretation.

At the same time, group-based trajectory modeling does have its limitations. First, there do not exist direct relationships between the independent and dependent variables, so that quantities like odds ratios and rate ratios are not routinely estimated. Second, as noted above, the current version of PROC TRAJ accommodates more than two dependent variables only if all of the dependent variables are constrained to have the same single latent variable defining group membership. Third, the current version of PROC TRAJ does not accommodate custom hypothesis testing. This capability would be useful to assess, for instance, whether an intervention had any association with a particular dependent variable if the model were constructed so that the intervention could interact with a personal characteristic such as sensation seeking; in this case, the appropriate null hypothesis could not be expressed in terms of a single parameter for each group and, as such, would not have test results presented in the PROC TRAJ output. On the other hand, a data analyst could still fit a "reduced" model excluding the intervention and then use the Bayesian Information Criterion available from the PROC TRAJ output to determine whether the original model with the intervention should be deemed superior to the reduced model without it. Hence, this third limitation regarding custom hypothesis testing is mitigated.

Latent Growth Curve Modeling

Whereas GLMMs allow a data analyst to obtain growth curves, these curves pertain to the directly observed dependent variables rather than to any underlying latent constructs of which the directly observed dependent variables may be indicators. To the extent that the underlying latent constructs themselves are of primary interest, as is part and parcel of confirmatory factor analysis and structural equation modeling (Bollen and Long, 1993; Mueller, 1996), a data analyst may prefer to estimate trajectories not for the directly observed dependent variables but for the underlying latent constructs. McArdle and Epstein (1987) developed methodology for this purpose, and since then latent growth curve modeling has become well established in the behavioral and psychological science literature. A couple of applications of this methodology are cited below; see Preacher et al. (2008) for others.

As with GLMMs and group-based trajectory modeling, a data analyst faces the issue of whether to take into account multiple dependent variables simultaneously rather than relying on a one-variable-at-a-time approach. Indeed, some authors have successfully taken into account multiple dependent variables in latent growth curve modeling. For example, Curran et al. (1997) combined structural equations for both self-reported peer alcohol use and personal alcohol use in 363 adolescents. Their data analysis merged features of traditional repeated measures MANOVA, confirmatory factor analysis, and structural equation modeling to fit growth curves simultaneously for each of the two latent constructs under study. More specifically, assuming linear growth for both processes, structural parameters were placed so that the intercept of one factor could influence the growth (or slope) of the other factor. Curran et al. (1997) demonstrated positive linear growth in both latent variables over a three-year period, with consideration for baseline age and parental alcoholism. They demonstrated that growth in adolescent alcohol use was accelerating as a function of peer alcohol use while growth in peer alcohol use was decelerating as a function of adolescent alcohol use. This implied that baseline peer and adolescent alcohol use were predictive of later changes but that these changes differed by construct. Such a conclusion would not have been possible if each construct had been assessed in isolation from the other.

An advantage of latent growth curve modeling over generalized linear mixed modeling is that the former naturally accommodates multiple indicators of the same construct. For instance, suppose that we periodically obtain retrospective self-reports of marijuana consumption and cigarette smoking along with results from laboratory tests administered on

the same days that the self-reports are acquired. Ignoring either the self-reports or the laboratory tests seems wasteful; the self-reports may be susceptible to recall bias, but the laboratory tests capture substance use only in very specific, narrow time intervals. Both a latent growth curve model and a GLMM can accommodate all four dependent variables, but the former naturally tethers together the two dependent variables that indicate marijuana consumption, as well as the two dependent variables that indicate cigarette smoking.

A limitation of latent growth curve modeling is its vulnerability to computational difficulties. For example, McArdle et al. (2009) analyzed records from three cohorts of subjects assessed repeatedly over time for intellectual ability using different instruments at different ages within and between cohorts. They based their analysis on all items in these instruments that measured the specific latent factors of interest. Combining an item response model for estimation of the latent factors with a growth curve model for assessing changes in the latent factors across the lifespan, the investigators encountered computational difficulties when using the standard maximum likelihood approach for estimation due to the large number of items under consideration. The investigators were ultimately able to circumvent these difficulties by adopting a Bayesian framework for statistical inference and constructing a Monte Carlo Markov Chain.

Assessment, Recommendations, and Implications for Basic and Translational Research

The three approaches to analyzing longitudinal data in multiple behavioral change outlined above are state of the art, at least for the year 2011. Of course, these approaches may eventually be superseded by better data analysis schemes in the future. However, at this juncture, we wish to comment on what investigators are actually doing that differs from what we have described above. Our commentary emphasizes studies in which at least one of the dependent variables involved substance use, although our subsequent recommendations apply just as well to other kinds of dependent variables (e.g., sexual risk-taking, delinquency, academic performance) and assorted combinations thereof. This commentary is not intended to criticize any referenced investigators, because many used the best methods that were readily available to them. Rather, we wish to clarify how methodological and computational advances now permit better data analyses.

Generalized linear mixed modeling and its more familiar cousin linear mixed modeling (frequently called “multilevel modeling” or “hierarchical linear modeling”) are actually prominent in the psychological and behavioral science literature, albeit not with multiple dependent variables treated in a single joint model. (We say “joint” to emphasize that the multiple dependent variables are considered all at once rather than in their own separate models; the antonym is “marginal”.) An alternative to fitting a single joint model, decidedly superior to fitting multiple marginal models, was exemplified by Brecht et al. (2008). They studied the use of heroin, cocaine, methamphetamine, alcohol, and marijuana over a ten-year timespan based on data from five longitudinal surveys conducted in California. The goal was to determine how each drug influenced the growth in the use of the other drugs. Their analysis employed a linear growth curve (random intercept and slope) for a primary drug, such as heroin, with the other drugs treated as time-dependent covariates that might impact the intercept and slope. Although this approach entails a series of conditional models rather than a single joint model, these conditional models are at least implicitly connected. (We say “conditional” rather than “marginal” here because, in each model, the expected level of use for one substance is made conditional on the levels of use for the other substances.) Even so, a single joint model is preferable if computationally feasible. With the recent addition of PROC GLIMMIX to SAS, that will typically be the case for an investigator in 2011.

Group-based trajectory modeling also is widespread in the psychological and behavioral science literature, and some authors do in fact fit a single joint model that accommodates all dependent variables. However, many authors fit multiple marginal models, one for each dependent variable, and then proceed to assign subjects to their most likely groups as determined from the marginal models. For instance, this has been done in the context of research on alcohol and marijuana use (Flory et al., 2004), conduct problems and hyperactivity (Shaw et al., 2005), and prosocial behavior and physical aggression (Kokko et al., 2006). A key difficulty is that the probability of belonging to group one for the first dependent variable and group three for the second dependent variable, for example, may not be consistently estimated by the empirical fraction of subjects assigned to group one for the first dependent variable and group three for the second dependent variable, if the assignments are made from multiple marginal models rather than from a single joint model. (A consistent estimator, recall, is one that approaches its target with high probability when the sample size becomes large.) With the current version of PROC TRAJ, an investigator in 2011 does not face this difficulty because he/she can employ a single joint model to estimate the probability of belonging to group one for the first dependent variable and group three for the second dependent variable. Unfortunately, as noted earlier, the current version of PROC TRAJ does not allow for more than two sets of groups in a single joint model. Thus, there is still a computational obstacle to estimating the probability of belonging to group one for the first dependent variable, group three for the second dependent variable, and group two for the third dependent variable.

Latent growth curve modeling is also well established in the psychological and behavioral science literature. However, as with group-based trajectory modeling, some authors construct separate models for different latent constructs. For instance, Farrell and Danish (1993) examined the relationship between peer influence and alcohol use in middle school students using separate longitudinal structural equation models and concluded that earlier peer alcohol use did not predict later adolescent alcohol use. Such an analytic scheme, while perhaps state of the art in 1993, does not take full advantage of the data. Indeed, as Curran et al. (1996) have noted, estimators of model parameters may be biased when one studies the relationship between two co-evolving factors via separate latent growth curve models.

In summary, the principal distinction between what investigators have been doing analytically versus what they might want to be doing in the future is the difference between fitting multiple marginal (or conditional) models within a particular analytic paradigm -- generalized linear mixed modeling, group-based trajectory modeling, or latent growth curve modeling -- and fitting a single joint model within that same analytic paradigm.

Although we have described three different analytic approaches relevant to the assessment of multiple behavior change, investigators need not limit themselves to a single analytic strategy. Most data sets will be amenable to at least two of the three analytic approaches described herein, and there may be instances in which complementary insights are obtained from dual analytic approaches. The complementary insights are not so much about which independent variables are statistically significant predictors or whether intervention interacts with personality, although certainly disagreements over such points from dual analytic approaches warrant attention. Rather, the complementary insights are more about what can be gleaned from one analytic approach that may not be at all obvious from another. For instance, a GLMM can give a more direct estimate of the association between an intervention and a behavior, but a group-based trajectory model can identify archetypal behavioral patterns that will not be evident from a GLMM.

We attach the caveat, however, that our advocacy for using complementary procedures should not be taken as an invitation to data snooping. For example, investigators should not

selectively report the smaller of two p-values obtained for an independent variable from each of two complementary procedures. Rather, we recommend that hypotheses and the primary method of data analysis be stated a priori, and that p-values be reported for independent variables based on the primary data analysis. Results acquired from a second procedure can then be presented in the context of a "sensitivity analysis", and any qualitative disparities with results from the primary data analysis can be discussed, but p-values obtained from a second procedure should not be portrayed as if they had arisen from the primary data analysis.

Generally speaking, we believe that the impact of interventions on multiple behavioral outcomes is best analyzed by treating the outcomes as distinct and employing one or more analytic strategies (such as those described herein) to assess the impact on all of the outcomes simultaneously. However, we also acknowledge that there may be certain contexts or sets of behaviors in which thinking of the outcomes as a collective may make sense (Prochaska et al., 2008). In such a case, the outcomes may be aggregated into a single index or a weighted impact factor, or an over-arching outcome measure may be defined. Investigators pursuing such avenues would do well to identify their weighting/aggregation systems or over-arching outcome measures a priori, as otherwise they could be vulnerable to criticisms that they had engaged in data snooping. Indeed, different choices of weighting/aggregation systems or over-arching outcome measures could well yield different verdicts about significance for the independent variables.

Finally, although we have drawn primarily from the drug abuse field to describe these analytic methods, they represent general approaches that are portable across various substantive research domains, including psychology, public health, and biomedicine. For example, these analytic strategies may be useful for determining if there is a causal relation between two different health risks (e.g., obesity and diabetes) or if the two different health risks are accounted for by some common influence (e.g., sedentary lifestyle as an influence on both cardiovascular and bone disease). In addition, preclinical research with laboratory animals may benefit from these analytic strategies. In the field of developmental teratology, early life events (e.g., maternal separation or stress) are often examined longitudinally for their long-term effects on various neurobehavioral outcomes. Thus, these analytic strategies transcend any specific research domain and may be especially useful for conducting translational research that extends from basic science to real-world application.

Acknowledgments

This research was supported by NIH/NIDA grant P50 DA005312. The authors acknowledge the data management assistance of Yushun Lin and comments from the editor and two anonymous referees.

References

Barker ED, Seguin JR, White HR, Bates M,E, Lacourse E, Carbonneau R, Tremblay RE. Developmental trajectories of male physical violence and theft. Archives of General Psychiatry. 2007; 64:592–599. [PubMed: 17485611]

Barker ED, Tremblay RE, Nagin DS, Vitaro F, Lacourse E. Development of male proactive and reactive physical aggression during adolescence. Journal of Child Psychology and Psychiatry. 2006; 47:783–790. [PubMed: 16898992]

Blumenthal JA, Babyak MA, Doraiswamy PM, Watkins L, Hoffman BM, Barbour KA, Herman S, Craighead WE, Brosse AL, Waugh R, Hinderliter A, Sherwood A. Exercise and Pharmacotherapy in the Treatment of Major Depressive Disorder. Psychosomatic Medicine. 2007; 69:587–596. [PubMed: 17846259]

Bollen, KA.; Long, JS., editors. Testing Structural Equation Models. Sage; Newbury Park: 1993.

Botvin GJ, Griffin KW, Diaz T, Scheier LM, Williams C, Epstein JA. Preventing illicit drug use in adolescents: Long-term follow-up data from a randomized control trial of a school population. Addictive Behaviors. 2000; 25(5):769–774. [PubMed: 11023017]

Brecht M-L, Huang D, Evans E, Hser Y-I. Polydrug use and implications for longitudinal research: ten-year trajectories for heroin, cocaine, and methamphetamine users. Drug and Alcohol Dependence. 2008; 96:193–201. [PubMed: 18329825]

Breslow NE, Clayton DG. Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association. 1993; 88:9–25.

Bryan AD, Schmiege SJ, Broaddus MR. HIV risk reduction among detained adolescents: A randomized, controlled trial. Pediatrics. 2009; 124(6):e1180–e1188. [PubMed: 19901006]

Copenhaver MM, Lee I-C, Margolin A. Successfully integrating an HIV risk reduction intervention into a community-based substance abuse treatment program. The American Journal of Drug and Alcohol Abuse. 2007; 33:109–120. [PubMed: 17366251]

Curran PJ, Harford T, Muthen BO. The relation between heavy alcohol use and bar patronage: a latent growth model. Journal of Studies on Alcohol. 1996; 57:410–418. [PubMed: 8776683]

Cupp PK, Zimmerman RS, Bhana A, Feist-Price S, Dekhtyar O, Karnell A, Ramsoomar L. Combining and adapting American school-based alcohol and HIV prevention programmes in South Africa: The HAPS project. Vulnerable Children and Youth Studies. 2008; 3(2):134–142.

Curran PJ, Stice E, Chassin L. The relation between adolescent alcohol use and peer alcohol use: a longitudinal random coefficients model. Journal of Counseling and Clinical Psychology. 1997; 65:130–140.

Dobson, A. An Introduction to Generalized Linear Models. Second Edition. CRC; Boca Raton: 2002.

Farrell A, Danish S. Peer drug associations and emotional restraint: causes or consequences of adolescents' drug use? Journal of Counseling and Clinical Psychology. 1993; 43:522–537.

Fieuws S, Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics. 2006; 62:424–431. [PubMed: 16918906]

Flory K, Lynam D, Milich R, Leukefeld C, Clayton R. Early Adolescent through Young Adult Alcohol and Marijuana Use Trajectories: Early Predictors, Young Adult Outcomes, and Predictive Utility. Development and Psychopathology. 2004; 16:193–213. [PubMed: 15115071]

Fontaine N, Carbonneau R, Barker ED, Vitaro F, He'bert M, Cote SM, Nagin DS, Zoccolillo M, Tremblay RE. Girls' Hyperactivity and Physical Aggression During Childhood and Adjustment Problems in Early Adulthood. Archives of General Psychiatry. 2008; 65:320–328. [PubMed: 18316678]

Gao S. A Shared Random Effect Parameter Approach for Longitudinal Dementia Data with Non-Ignorable Missing Data. Statistics in Medicine. 2004; 23:211–219. [PubMed: 14716723]

Griffin KW, Botvin GJ, Nichols TR. Effects of a school-based drug abuse prevention program for adolescents on HIV risk behavior in young adulthood. Prevention Science. 2006; 7(1):103–112. [PubMed: 16604429]

Harrington NG, Giles SM, Hoyle RH, Feeney GJ, Yungbluth SC. Evaluation of the All Stars character education and problem behavior prevention program: Effects on mediator and outcome variables for middle school students. Health Education and Behavior. 2001; 28:533–542. [PubMed: 11575684]

Johnson, RA.; Wichern, DW. Applied Multivariate Statistical Analysis. Fourth Edition. Prentice Hall; Upper Saddle River: 1998.

Jones BL, Nagin DS. Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. Sociological Methods and Research. 2007; 35:542–571.

Jones BL, Nagin D, Roeder K. A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. Sociological Methods and Research. 2001; 29:374–393.

Kalichman SC, Simbayi LC, Vermaak MA, Cain D, Smith G, Mthebu J, Jooste S. Randomized trial of a community-based alcohol-related HIV risk-reduction intervention for men and women in Cape Town South Africa. Annals of Behavioral Medicine. 2008; 36:270–279. [PubMed: 18836789]

Kokko K, Tremblay RE, Lacourse E, Nagin DS, Vitaro F. Trajectories of prosocial behavior and physical aggression in middle childhood: Links to adolescent school dropout and physical violence. Journal of Research on Adolescence. 2006; 16:403–428.

Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. Wiley; New York: 1987.

Liu L, Wolfe RA, Kalbfleisch JD. A Shared Random Effects Model for Censored Medical Costs and Mortality. Statistics in Medicine. 2007; 26:139–155. [PubMed: 16526006]

Lynam DR, Milich R, Zimmerman R, Novak SP, Logan TK, Martin C, Leukefeld C, Clayton R. Project DARE: No effects at 10-year follow-up. Journal of Consulting and Clinical Psychology. 1999; 67(4):590–593. [PubMed: 10450631]

McArdle JJ, Epstein D. Latent growth curves within developmental structural equation models. Child Development. 1987; 58:110–133. [PubMed: 3816341]

McArdle JJ, Grimm KJ, Hamagami F, Bowles RP, Meredith W. Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. Psychological Methods. 2009; 14:126–149. [PubMed: 19485625]

McCullagh, P.; Nelder, JA. Generalized Linear Models. Second Edition. CRC; Boca Raton: 1989.

Molenberghs, G.; Verbeke, G. Models for Discrete Longitudinal Data. Springer; New York: 2005.

Morabia A, Costanza MC. Multiple health behavior change interventions: Tell us what you see. Preventive Medicine. 2010; 50:1–2. [PubMed: 20109761]

Mueller, RO. Basic Principles of Structural Equation Modeling. Springer-Verlag; New York: 1996.

Nagin D. Analyzing Developmental Trajectories: Semi-Parametric, Group-Based Approach. Psychological Methods. 1999; 4:39–177.

Nagin DS, Tremblay RE. Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method. Psychological Methods. 2001; 6:18–34. [PubMed: 11285809]

Pentz, MA. Understanding and Preventing Risks for Obesity. In: DiClemente, R.; Crosby, R.; Santelli, J., editors. Adolescent Health: Understanding and Preventing Risk. Wiley; Hoboken: 2009. p. 147-164.

Preacher, KJ.; Wichman, AL.; MacCallum, RC.; Briggs, NE. Latent Growth Curve Modeling. Sage; Los Angeles: 2008.

Prochaska JJ, Nigg CR, Spring B, Velicer WF, Prochaska JO. The benefits and challenges of multiple health change in research and in practice. Preventive Medicine. 2010; 50:26–29. [PubMed: 19948184]

Prochaska JJ, Spring B, Nigg CR. Multiple health behavior change research: An introduction and overview. Preventive Medicine. 2008; 46:181–188. [PubMed: 18319098]

Prochaska JJ, Velicer WF, Nigg CR, Prochaska JO. Methods of quantifying change in multiple risk factor interventions. Preventive Medicine. 2008; 26:260–265. [PubMed: 18319099]

Ringwalt CL, Clark HK, Hanley S, Shamblen SR, Flewelling RL. *Project ALERT:* A randomized trial. Archives of Pediatric and Adolescent Medicine. 2009; 163(7):625–632.

Riggs NR, Sakuma KL, Pentz MA. Preventing Risk for Obesity by Promoting Self-Regulation and Decision-Making Skills: Pilot Results from the PATHWAYS to Health Program. Evaluation Reviews. 2007; 31:287–310.

Schall R. Estimation in Generalized Linear Models with Random Effects. Biometrika. 1991; 78:719–727.

Searle, SR.; Casella, G.; McCulloch, CE. Variance Components. Wiley; New York: 1992.

Shaw DS, Lacourse E, Nagin DS. Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. Journal of Child Psychology and Psychiatry. 2005; 46:931–942. [PubMed: 16108996]

Sloboda Z, Stephens RC, Stephens PC, Grey SF, Teasdale B, Hawthorne RD, Williams J, Marquette JF. The Adolescent Substance Abuse Prevention Study: A randomized field study of a universal substance abuse prevention program. Drug and Alcohol Dependence. 2009; 102:1–10. [PubMed: 19332365]

Stoolmiller M, Gerrard M, Sargent JD, Worth KA, Gibbons FX. R-rated Movie Viewing, Growth in Sensation Seeking, and Alcohol Initiation: Reciprocal and Moderation Effects. Prevention Science. 2010; 11:1–13. [PubMed: 19655251]

Van Lier PAC, Vitaro F, Barker ED, Koot HM, Tremblay RE. Developmental links between trajectories of physical violence, vandalism, theft, and alcohol-drug use from childhood to adolescence. Journal of Abnormal Child Psychology. 2009; 37:481–492. [PubMed: 19089610]

Verbeke, G.; Molenberghs, G. Linear Mixed Models for Longitudinal Data. Springer; New York: 2000.

Wolfinger R, O'Connell M. Generalized Linear Mixed Models: A Pseudo-likelihood Approach. Journal of Statistical Computation and Simulation. 1993; 4:233–243.