# Three-Level Multilevel Growth Models for Nested Change Data: A Guide for Group Treatment Researchers

**Article** *in* Psychotherapy Research · July 2009

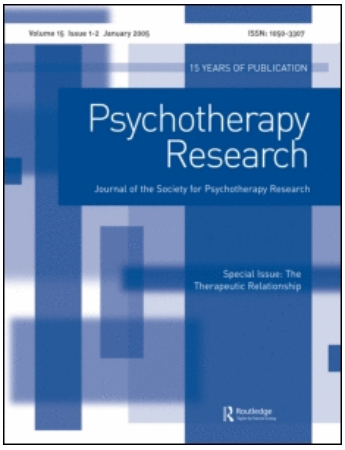
DOI: 10.1080/10503300902933188 · Source: PubMed



**4 authors**, including:

Giorgio A. Tasca
University of Ottawa
**156** PUBLICATIONS **3,086** CITATIONS
SEE PROFILE

Anthony Joyce
University of Alberta
**151** PUBLICATIONS **4,579** CITATIONS
SEE PROFILE




**Psychotherapy Research**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713663589

# Three-level multilevel growth models for nested change data: A guide for group treatment researchers

Giorgio A. Tasca[a]; Vanessa Illing[b]; Anthony S. Joyce[c]; John S. Ogrodniczuk[d]
[a] Department of Psychology, The Ottawa Hospital-General, Ottawa, Ontario [b] Department of Psychology, University of Ottawa, Ottawa, Ontario [c] Department of Psychiatry, University of Alberta, Edmonton, Alberta [d] Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada

**To cite this Article** Tasca, Giorgio A. , Illing, Vanessa , Joyce, Anthony S. and Ogrodniczuk, John S.(2009) 'Three-level multilevel growth models for nested change data: A guide for group treatment researchers', Psychotherapy Research, 19: 4, 453 — 461
**To link to this Article: DOI:** 10.1080/10503300902933188
**URL:** http://dx.doi.org/10.1080/10503300902933188

*Psychotherapy Research*, July–September 2009; 19(4–5): 453–461



PSYCHOTHERAPY RESEARCH METHODS

# Three-level multilevel growth models for nested change data: A guide for group treatment researchers

GIORGIO A. TASCA[1], VANESSA ILLING[2], ANTHONY S. JOYCE[3], & JOHN S. OGRODNICZUK[4]

[1]*Department of Psychology, The Ottawa Hospital–General, Ottawa, Ontario;* [2]*Department of Psychology, University of Ottawa, Ottawa, Ontario;* [3]*Department of Psychiatry, University of Alberta, Edmonton, Alberta &* [4]*Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada*

*(Received 14 November 2008; revised 27 March 2009; accepted 30 March 2009)*



**Abstract**
Researchers have known for years about the negative impact on Type I error rates caused by dependencies in hierarchically nested and longitudinal data. Despite this, group treatment researchers do not consistently use methods such as multilevel models (MLMs) to assess dependence and appropriately analyse their nested data. The goals of this study are to review some of the study design issues with regard to hierarchically nested and longitudinal data, discuss MLMs for assessing and handling dependence in data, and present a guide for developing a three-level growth MLM that is appropriate for group treatment data, design, and research questions. The authors present an example from group treatment research to illustrate these issues and methods.

**Keywords:** group psychotherapy; outcome research; statistical methodology; multilevel modeling

The defining design element of group treatment research is that clients are organized into groups, and so the client data are "nested" within groups. Nesting refers to data with a multilevel or hierarchical structure wherein units are clustered into a hierarchy. Another common design feature in group treatment research is that outcomes are often repeatedly measured. In these longitudinal designs, one may conceptualize "time" (i.e., data measured at multiple time points) as nested within individuals (Singer & Willett, 2003). Thus, in group treatment research, nesting occurs in both hierarchically structured data and longitudinal data. Nesting may result in nonindependence in the group treatment data.

Researchers have long been aware that even a moderate amount of dependence in hierarchically nested data has a dramatic effect on Type I error, possibly resulting in spurious "significant" results (Kenny & Judd, 1986). Dependence in one's data is measured by the intraclass correlation coefficient ($\rho$), and a $\rho > 0$ indicates dependence that inflates Type I error ($\rho$ and its components are discussed in more detail later). For example, Kreft and de Leeuw (1998) reported that in a study with 10 participants per group and $\rho = .05$, the actual Type I error rate was .11. Hence, the group effect (i.e., the nesting factor) that accounted for 5% of the variance in the dependent variable resulted in an actual Type I error rate that was more than doubled. Despite these problems with dependence in hierarchically nested data, group treatment researchers typically have not used methods such as multilevel models (MLMs) to address these issues. For example, in a study outlining this problem, Baldwin, Murray, and Shadish (2005) found that none of the 33 studies of group interventions on the list of empirically supported treatments (Task Force on Promotion and Dissemination of Psychological Procedures, 1998) assessed and adjusted for dependence in their data.

A common method for assessing longitudinal data is the repeated measurement analysis of variance ANOVA (rANOVA) model, in which mean

Correspondence concerning this article should be addressed to Giorgio A. Tasca, Department of Psychology, The Ottawa Hospital–General, 501 Smyth Road, Room 4428, Ottawa, Ontario K1H 8L6, Canada. E-mail: gtasca@ottawahospital.on.ca

ISSN 1050-3307 print/ISSN 1468-4381 online 
DOI: 10.1080/10503300902933188

differences across time are compared. A restriction of rANOVA is the requirement of having complete data for each individual at each time point. This is often not achievable with longitudinal designs (especially in the context of clinical research) and may result in loss of data through listwise deletion or imputation of data by a variety of methods, and both of these could bias the results. Further, the rANOVA model has a restrictive assumption of sphericity in which variability of measurements across time is required to be equal, and correlations between any two measurements taken across multiple time points on the same individual are assumed to be equivalent. This assumption is often untenable for group treatment research change data because (a) the variability one sees in initial scores might shrink at subsequent assessments and (b) correlations of consecutive measurements tend to be higher than correlations of measurements taken further apart from each other in time.

Violations of the sphericity assumption in rANOVA result in inflated Type I error rates. One method of correcting for this problem is to reduce the numerator degrees of freedom of the significance tests of rANOVA by applying the Greenhouse–Geisser or Huyn–Feldt corrections. However, the Greenhouse–Geisser correction tends to be very conservative, and the Huyn–Feldt correction less so, and both reduce power (Gueorguieva & Krystal, 2004). Further, these methods are subject to the same problems and limitations related to missing data outlined previously (Holden, Kelley, & Argarwal, 2008). Despite these issues, rANOVA continues to be used widely and more frequently than MLM growth models for longitudinal data in research appearing in some publications with high-impact factors (Gueorguieva & Krystal, 2004).

Arguably, the most common design in group treatment research is that in which repeatedly measured outcomes occur within individuals who are, in turn, nested within groups. Yet group treatment researchers almost exclusively prefer to use analytic techniques based on the general linear model that do not address problems associated with dependence in grouped and longitudinal data (Baldwin et al., 2005).

Currently, there are a few emerging examples in group treatment research of the use of MLMs that account for dependence in hierarchically nested data. Imel, Baldwin, Bonus, and MacCoon (2008), for example, used two-level hierarchically nested random-effects MLMs to study group effects on outcomes. Although such two-level MLMs address problems associated with dependencies in grouped data, these models are not longitudinal in that they evaluate outcomes at a single time point, such as at posttreatment only. However, group treatment researchers typically are interested in the amount, rate, and shape (linear or nonlinear) of change in outcomes across multiple sessions of therapy or across several posttreatment time points. Three-level growth MLMs are parsimonious models for hierarchically nested longitudinal designs. Despite this, we are aware of no examples of the use of a three-level growth MLM in a published group treatment study. Our view is that one of the main reasons why group therapy researchers do not frequently use MLMs, especially three-level growth MLMs, is that most articles and chapters on the topic do not focus on the practical steps for developing these MLMs for the design challenges posed by group treatment research.

Further, there are very few examples of three-level growth MLMs in the statistical and methodology literatures. Examples that do exist are not easily accessible or applicable to group treatment researchers. Some of these examples require knowledge of matrix notation and the role of matrix vectors in algebraic equations (e.g., Hedeker & Gibbons, 2006). Other examples make assumptions that are not often appropriate to the group treatment research context, such as assuming that there is no variation in the initial status that cannot be accounted for by the baseline score itself (Elkin, Falconnier, Martinovich, & Mahoney, 2006). Raudenbush and Bryk (2002) and Atkins (2005) do provide more accessible examples of a three-level MLM in educational and in marital research, respectively. However, their examples are not easily translated to the group treatment research context. For example, group treatment researchers often wish to (a) control for initial scores when evaluating change and (b) test nonlinear change in process and outcome variables at the level of the groups. Neither of these basic design elements of group treatment research is currently discussed in the three-level growth MLM literature.

Our goal in this presentation is to provide a guide for the use of three-level growth MLMs specific to group treatment research design. By doing so, we hope that group treatment researchers will be less intimidated by the apparent complexities of MLMs and more likely to appropriately use a three-level growth MLM when designing studies and analysing their data. The purposes of this article are (a) to provide a guide to group treatment researchers for research design that is informed by MLMs of nested (i.e., grouped) change data and (b) to provide a technical primer for developing and using three-level MLM equations specific to group treatment research contexts.

## Study Design Issues with Nested Data

The primary issue that arises in research design with hierarchically nested data is the unit of analysis

problem (Pollack, 1998). That is, researchers must decide on the appropriate level on which to focus their study design, data collection, and data analyses (i.e., the individual or the group). Generally, when one is designing a study with a hierarchically nested data structure, one should consider the highest level of nesting as defining the unit of analysis and degrees of freedom. For example, if participants are nested within a group, it is the number of groups, not the number of individuals, that determines the salient unit of analysis. Hence, when designing a study, group treatment researchers should think in terms of creating more total groups as opposed to focusing primarily on recruiting more individuals, because it is the group *N*, not necessarily the individual *N,* that will have greater relevance for statistical power (Maas & Hox, 2005). However, these statistical issues must be considered within the clinical requirements of the minimum number of individuals per group necessary to achieve desired group processes and clinical outcomes (Yalom & Leszcz, 2005). In any case, accruing an adequate number of groups creates a greater demand on resources, and this can be challenging for group treatment researchers. So far, it has not been typical for randomized comparative group treatment studies to have large numbers of therapists or groups (e.g., Piper, Ogrodniczuk, Weideman, Joyce, & Rosie, 2007). However, although degrees of freedom associated with the nesting factor tend to be small, parameters (i.e., estimates of intercepts and slopes) associated with the nesting factor tend to be very reliable. These parameters are based on aggregated data from the individuals, thus resulting in more accurate standard errors. Stevens (2002) argued that the smaller degrees of freedom from aggregated data "will not cause as drastic a drop in power as some have feared" (p. 260).

With regard to repeated measurement or longitudinal data, an important design issue is the number of times and points at which data are collected. An attractive feature of MLMs for longitudinal data is that "time" can be treated flexibly, allowing the researcher to ask unique questions about the rate and shape of change over time (Singer & Willett, 2003). For example, outcomes may be assessed at several sessions during and after treatment in order to identify early responders, identify shifts that occur at substantively interesting moments as predicted by the treatment model, or test a hypothesis of immediate change at posttreatment that is maintained at follow-up assessments. When designing group treatment studies, generally, assessing outcomes at more time points—and thus making use of more data—will result in more reliable estimates of parameters (i.e., of the intercepts and slopes; Crits-Christoph & Gallop, 2006).

## MLM Terminology

Before considering the equations, we review some MLM terms. MLMs go by different names in the literature, including hierarchical linear models, mixed-effects models, random-effects models, and growth curve models for longitudinal data. Random effects may include the initial status (intercept) or the rate of change (slope) for an individual or group whose data are repeatedly measured. These effects are random in that they are allowed to vary within or between individuals or groups. They also reflect population parameter estimates (i.e., average values for the population), and so have error terms related to them. Fixed effects have a single value for all units of the sample and do not vary, such as treatment condition to which a group is assigned.

MLMs are multilevel and hierarchical models because parameters at a lower level of a model are nested within the parameters at a higher level. That is, for the longitudinal part of a three-level growth model, the data at repeated measurement time points modeled at Level 1 become the data by which individual initial status (intercepts) and growth (slopes) at Level 2 are calculated. For the hierarchically nested part of the model, the intercepts and slopes of an individual modeled at Level 2 become the data by which the intercepts and slopes of the group are calculated at Level 3.

## Choosing an MLM Analytic Strategy

The structure of the data, amount of dependence in the data, and the research design and questions will define the researcher's choice of the multilevel model. Figure 1 is a decision tree in which three fundamental questions are presented to guide initial decisions on whether to use an MLM and on which MLM to use.

The first question, related to structure of the data, is, are individuals nested within groups? By definition, the answer to this question for group treatment research is "yes." If the data are hierarchically structured, then the second question in Figure 1 asks, is there evidence of dependence in the data, and if so, what is the extent of dependence in the data? Evidence of dependence in hierarchically nested data would then require that some procedure be used to mitigate its effects on Type I error. As we illustrate later, one such set of procedures to assess dependence, as well as to appropriately analyse hierarchically nested change data, is the MLM.

The third question asks whether the data are repeatedly measured, which leads one to consider which MLM is appropriate for one's research design and research questions. If the design involves a single end-point analysis, in which groups are compared at

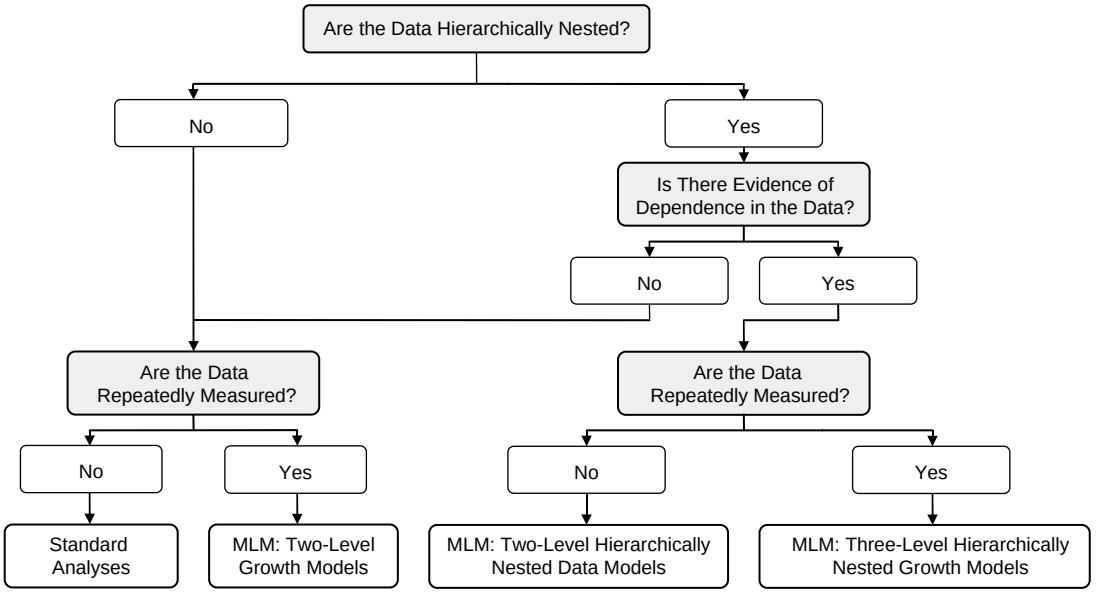



Figure 1. Decision tree to guide design and analyses of hierarchically nested change data with multilevel models (MLMs). Note: Shaded boxes indicate three fundamental questions when evaluating the need for an MLM. This decision tree applies to normally distributed continuous outcomes.



one time point only (i.e., if the data are not repeatedly measured), then the researcher may use two-level hierarchically nested data models (e.g., Imel et al., 2008). However, if the data are within a repeated measurement design but not nested within groups, then two-level growth curve MLMs may be considered (Tasca & Gallop, 2009).

Group treatment research designs represent particular analytic challenges because they often have (a) repeated measurement data that (b) are also hierarchically structured, which may result in dependence in the change data. MLMs are flexible enough to accommodate these complex group treatment designs within three-level hierarchically structured growth models. These three-level growth models address questions related to the hierarchically nested design, such as what is the effect of the nesting factor (e.g., group) on the rate of change? Further, these three-level nested growth MLMs address the longitudinal design research questions, such as what is the group rate of change, what is the shape of this change (i.e., linear or curvilinear), and was there an effect of treatment conditions or other predictors on group rate or shape of change?

## Intraclass Correlation

As indicated earlier, the decision to use MLMs for hierarchically structured data is based partly on answering the question, is there evidence for dependence in the data? (See Figure 1) The statistic to assess dependence in data is the intraclass correlation coefficient ($\rho$). A method commonly used to derive $\rho$ is taken from the random ANOVA model in which the nesting factor is a random effect (Stevens, 2002). Larger between-group variance relative to within- plus between-group variance indicates a group effect (i.e., an effect of the nesting factor), so that $\rho > 0$. A group effect suggests that participants within groups are more similar to each other and less similar to participants from independent groups. Hence, the individual participant data are not independent.

Although the random ANOVA model is generally satisfactory for assessing $\rho$ at individual time points, the method is less satisfactory for change data. That is, in a group treatment study, for example, it is the change from pretreatment that is most relevant when assessing for nonindependence of grouped data. What may complicate matters further is if one were interested in the rate of change over time that is characteristic of repeated measurement designs. In such cases, the static time point assessment of $\rho$ offered by random ANOVA is inadequate to evaluate dependence of the dynamic nature of change data. As indicated later, three-level growth MLMs provide a means of assessing $\rho$ for nested change data.

## A Guide to Developing Three-Level Multilevel Models to Assess Dependence and Change

To assess change, we suggest a plan of developing a sequence of models from simplest to most complex by starting with (a) a base model (i.e., a completely unconditional or intercept-only model), (b) an unconditional three-level growth model while

controlling for prescores, and finally (c) a conditional three-level growth model while controlling for prescores.

*Base Model.* Base models are not longitudinal in the sense that time is not modeled. However, base models do allow the researcher to estimate within- and between-person variance by which subsequent longitudinal models may be compared.

$$\begin{aligned}&\text{Level 1: } Y_{tij} = \pi_{0ij} + e_{tij}\\ &\text{Level 2: } \pi_{0ij} = \beta_{00j} + r_{0ij}\\ &\text{Level 3: } \beta_{00j} = \gamma_{000} + u_{00j}\end{aligned} \quad (1)$$

At Level 1, $Y_{tij}$ is the dependent variable score repeatedly measured at time $t$ for individual $i$ nested within group $j$, $\pi_{0ij}$ is the individual mean score across all time points, and $e_{tij}$ is the residual or error term indicating the deviation of each individual score from their mean score. At Level 2, each individual's intercept (i.e., individual mean score across all time points), $\pi_{0ij}$, is modeled as the group mean of individual scores, $\beta_{00j}$, plus each individual's deviation from their respective group mean, $r_{0ij}$. At Level 3, the group mean score, $\beta_{00j}$, is a function of the grand mean across all groups, $\gamma_{000}$, and each group mean's deviation from that grand mean, $u_{00j}$.

*Unconditional Three-Level Growth Model Controlling for Prescores.*

$$\begin{aligned}&\text{Level 1: } Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{LOGTIME}_{tij}) + e_{tij}\\ &\text{Level 2: } \pi_{0ij} = \beta_{00j} + \beta_{01j}\ (\text{individual prescore}) + r_{0ij}\\ &\qquad\quad \pi_{1ij} = \beta_{10j} + \beta_{11j}\ (\text{individual prescore}) + r_{1ij}\\ &\text{Level 3: } \beta_{00j} = \gamma_{000} + \gamma_{001}\ (\text{group prescore}) + u_{00j}\\ &\qquad\quad \beta_{01j} = \gamma_{010} + u_{01j}\\ &\qquad\quad \beta_{10j} = \gamma_{100} + \gamma_{101}\ (\text{group prescore}) + u_{10j}\\ &\qquad\quad \beta_{11j} = \gamma_{110} + u_{11j}\end{aligned} \quad (2)$$

*Conditional Three-Level Growth Model Controlling for Prescores.*

$$\begin{aligned}&\text{Level 1: } Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{LOGTIME}_{tij}) + e_{tij}\\ &\text{Level 2: } \pi_{0ij} = \beta_{00j} + \beta_{01j}\ (\text{individual prescore}) + r_{0ij}\\ &\qquad\quad \pi_{1ij} = \beta_{10j} + \beta_{11j}\ (\text{individual prescore}) + r_{1ij}\\ &\text{Level 3: } \beta_{00j} = \gamma_{000} + \gamma_{001}(\text{group prescore})\\ &\qquad\qquad + \gamma_{002}(\text{treatment condition}) + u_{00j}\\ &\qquad\quad \beta_{01j} = \gamma_{010} + u_{01j}\\ &\qquad\quad \beta_{10j} = \gamma_{100} + \gamma_{101}(\text{group prescore})\\ &\qquad\qquad + \gamma_{102}(\text{treatment condition}) + u_{10j}\\ &\qquad\quad \beta_{11j} = \gamma_{110} + u_{11j}\end{aligned} \quad (3)$$

A model for repeated measurement data becomes a growth model when "time" is specified at Level 1. A nested three-level growth model is unconditional when there is no fixed treatment or experimental condition factor at Level 3 of the model (i.e., parameters $\gamma_{002}$ and $\gamma_{102}$). In these models, as is typical in treatment outcome research, prescores are controlled when evaluating growth (i.e., change).

At Level 1, within-person growth or rate of change is modeled. Note that the intercept in growth models, $\pi_{0ij}$, now represents the initial status (a latent variable or predicted value representing the prescore in this example) for the individual. $\text{LOGTIME}_{tij}$ is the log transformation of the time metric 1, 2, 3 (representing pre-, post-, and follow-up assessments, respectively) into .00, .30, and .48. The log transformation reflects the curvilinear nature of change from pre- to post- to 6 months posttreatment based on the assumption or theory that change caused by group treatment occurs most rapidly at posttreatment and less rapidly after post- to 6 months posttreatment. Alternatively, a linear model could be posited depending on the shape of the individual data. Singer and Willett (2003) suggest a visual inspection of the individual ordinary least squares (OLS) growth plots to help determine whether a linear or nonlinear model would best fit the data. Finally, $\pi_{1ij}$ represents the growth rate for the individual across each time point, and $e_{tij}$ is the residual or error term indicating the deviation of each individual's score from his or her own modeled line.

The Level 2 equations represent variability in between-person growth or rate of change. At Level 2, each individual's intercept, $\pi_{0ij}$, is modeled as the group mean initial status, $\beta_{00j}$, the interaction with the individual's group centered prescore, $\beta_{01j}$, and the residual, $r_{0ij}$. Each individual's slope (i.e., rate of change), $\pi_{1ij}$, is a function of the average rate of change for his or her group, $\beta_{10j}$, the interaction between rate of change and the individual group centered prescore, $\beta_{11j}$, plus each individual's growth parameters' deviation from the average rate of change, $r_{1ij}$. Note that individual prescores are centered at the group mean prescore. Because of this, $\beta_{10j}$ is the expected rate of change for an individual prescore equal to the group mean prescore. This allows one to assess variance in individual growth rate after taking into account differences in prescores.

At Level 3, between-group growth is modeled. The group mean initial status, $\beta_{00j}$, is a function of the grand mean for initial status, $\gamma_{000}$, the interaction with the group's grand mean centered prescore, $\gamma_{001}$, and a residual, $u_{00j}$. The interaction between group initial status and group grand mean centered prescores, $\beta_{01j}$, is a function of the grand mean of this interaction, $\gamma_{010}$, and a residual, $u_{01j}$. A group's growth parameter or slope representing rate of change, $\beta_{10j}$, is a function of the mean rate of change for all groups, $\gamma_{100}$, and the group grand mean

centered prescore, $\gamma_{101}$, the effect of treatment condition, $\gamma_{102}$, and the group residual, $u_{10j}$. Group prescores are grand mean centered. Because of this, $\gamma_{100}$ represents the expected rate of change when the group mean prescore is equal to the grand mean prescore. Finally, the group mean prescore × rate of change interaction, $\beta_{11j}$, is a function of the mean group prescore/rate of change interaction, $\gamma_{110}$, and $u_{11j}$ is the group random effect associated with this parameter. The statistical significance of the $\gamma_{100}$ parameter indicates if the group slope or rate of change across time is greater than zero while controlling for prescores, and the statistical significance of the $\gamma_{102}$ parameter indicates reliable differences between treatment conditions on this rate of change across time.

Each equation in the three-level growth models outlined previously has a variance component associated with the error term (i.e., residual). Error terms have a mean of 0 and are assumed to be normally distributed. The Level 1 within-person variance component is $\sigma^2$ associated with the $e_{tij}$ error term. The Level 2 between-person variance components include $\tau_0$ associated with the intercept error term, $r_{0ij}$, and $\tau_1$ associated with the individual growth or slope error term, $r_{1ij}$. Level 3 between-group variance components include $\tau_{00}$ associated with the intercept error term, $u_{00j}$, and $\tau_{10}$ associated with the group growth or slope parameter error term, $u_{10j}$.

The process of adding predictors in subsequent models is aided by evaluating variance components from Level 3 equations with a chi-square test. For example, if the chi-square statistics from the unconditional growth model are significant, this would indicate that there is enough variance in the Level 3 slope (i.e., group growth) remaining to be explained by the addition of treatment condition as a predictor.

*Assessing Dependence, Explained Variance, and Deviance Statistics.* As indicated in Figure 1, assessing for dependence is an important early step to deciding whether data must be modeled as nested within groups. To calculate ρ from growth models, the following equation may be used:

$$\rho = \tau_{10j}(\text{conditional})/(\tau_{10j}(\text{unconditional}) + \tau_{1ij}). \quad (4)$$

$\tau_{10j}$(conditional) and $\tau_{10j}$(unconditional) refer to the Level 3 variance components associated with the group slope parameter from the conditional and unconditional growth models, $u_{10j}$, respectively. $\tau_{1ij}$ refers to the unconditional Level 2 variance component associated with the individual slope parameter, $r_{1ij}$. In this model, ρ estimates the effect of the nesting factor (e.g., group) on the individual and group growth for which prescores are controlled and the effect of treatment condition is removed (Elkin et al., 2006).

In addition, one can assess "pseudo $R^2$" for the amount of within-person variance explained by the addition of the time variable. To do so, the within-person variance, $\sigma^2$, from the unconditional growth model is assessed relative to $\sigma^2$ from the base model.

$$\sim R^2 = (\sigma^2(\text{base}) - \sigma^2(\text{unconditional}))/\sigma^2(\text{base}) \quad (5)$$

An alternative approach that has superior statistical properties for comparing models is the deviance statistic (Singer & Willett, 2003). The deviance statistic is based on the sample log-likelihood statistic (LL) generated from maximum likelihood estimation. The deviance statistic compares the LL for the currently specified model with a saturated model that is presumed to fit the data perfectly. The smaller the deviance statistic, the better is the fit of the model to the data. To compare two models, the models must be based on identical data, and one model must be nested within the other. For example, the unconditional growth model specified previously (Equation 2) is nested within the conditional model (Equation 3). The difference in deviance statistics ($\Delta D$) between two models is tested with a chi-square distribution, with degrees of freedom defined by the difference in the number of parameters tested between the two models.

## A Group Treatment Research Example

The following example is presented to illustrate the prior discussion. The data were drawn from a previously published randomized comparative group treatment trial. The reader is referred to that publication for more information about the participants, methods, and all results (Piper et al., 2007). MLM analyses were done with Hierarchical Linear Modeling software version 6.04 using a full maximum likelihood estimation method (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004).

*Participants and Procedures.* Briefly, for the current presentation, data were drawn from 82 participants referred to a psychiatry department group outpatient treatment program at two university hospitals who were assessed to have complicated grief. Piper et al. (2007) created four conditions in their study. Two conditions involved groups composed to be homogeneous based on a client characteristic, quality of object relations (QOR; Azim, Piper, Segal, Nixon, & Duncan, 1991). That is, high-QOR clients were in interpretive short-term group therapy (STG; Piper, McCallum, & Joyce, 1995), and low-QOR clients were in supportive STG (McCallum, Piper, & Joyce, 1995). The other two treatment conditions involved

matching clients in pairs on QOR and randomly assigning the members of each pair to interpretive or supportive STG. These groups were characterized by heterogeneity on the QOR variable. For this presentation, participants were nested within 17 groups: eight interpretive STGs ($n=36$) and nine supportive STGs ($n=46$). Treatment condition, interpretive versus supportive STG, was the fixed effect. The dependent variable presented here is the State-Trait Anxiety Inventory (STAI; Spielberger, 1983) administered at pre-, post-, and 6 months posttreatment.

## Results and Discussion

Is there evidence of dependence in these nested data? (See Figure 1) Variance components from the three-level growth models were used to calculate ρ using Equation 4: 10.51/(21.56+518.37). The ρ value from the three-level growth model indicated that the group that participants attended (the nesting factor) accounted for 2.0% of the growth variance. These results suggested that a hierarchically nested MLM approach is a useful strategy to avoid modestly inflated Type I error caused by dependence in the individual outcome or growth data. Following our decision tree (see Figure 1), the next question to consider was, which MLM is most appropriate for these data? Given our interest in longer term follow-up of outcomes, the three-level growth MLM approach was selected.

The unconditional growth model (Equation 2) showed that the group slope parameter, $\gamma_{100}$, was significantly different from zero (Table I). This suggested that for all groups STAI scores decreased rapidly from pre- to posttreatment, and this decrease was maintained to 6 months posttreatment. This modeling of change in STAI scores accounted for 73.21%, of the within-person variance, that is, (80.47–21.56)/80.47 (Equation 5). To assess whether there was variance remaining to be accounted for by a predictor, the $\tau_{10j}$ variance component from the unconditional model was examined using a chi-square test. The results indicated that the growth variance component, $\tau_{10j}$, $\chi^2(15,\ N=82)=16.29$, $p=.36$, did not have significant variance remaining to be modeled. For illustrative purposes, we specified a conditional growth model in which treatment condition was added as a predictor at Level 3 (Equation 3). The conditional growth model assessed the difference between the treatment conditions on the significant decreasing slope of STAI scores. There was no significant difference between the treatment conditions on rate of change, $\gamma_{102}$, from pre- to post- to 6 months posttreatment (see Table I). To illustrate the use of the deviance statistic to assess the models, we compared the deviance statistic of the unconditional growth model that had 20 parameters with the deviance statistic of the conditional growth model that had 22 parameters, $\Delta D=1629.88–1629.48=\chi^2(2,\ N=82)=0.40$, $p>.05$. The conditional growth model was not a better fit to the data compared with the unconditional growth model.

## Conclusion

The first question in Figure 1 asks whether the data are hierarchically nested. The prior example was drawn from group treatment research, and so by definition the data were hierarchically nested. Based on the second question of our decision tree, the next task was to estimate the amount of variance in the change data that was associated with the nesting factor (i.e., the amount of dependence in the data). We found that 2.0% of STAI rate of change outcome variance was associated with differences between groups. Some authors consider this level of dependence to be small (Hox, 2002), and based on values in the table provided by Kreft and de Leeuw (1998), the actual Type I error rate for these data is somewhere between .06 and .11. Thus, there was evidence of modest dependence in the data, and common OLS analyses of the individual data likely would have resulted in some inflation of Type I error. Hence, we opted to use an MLM to analyse the change data.

Table I. Selected Fixed and Random Effects Results from Three Multilevel Models

| Model | Parameter | Coefficient | *t* | *df* | *p* | $\sigma^2$ | $\tau$ | Deviance[a] |
|---|---|---|---|---|---|---|---|---|
| Base | | | | | | 80.47 | | 1880.17 |
| Level 3: intercept | $\gamma_{000}$ | 50.11 | 40.58 | 16 | <.001 | | 71.27 | |
| Unconditional growth | | | | | | 21.56 | | 1629.88 |
| Level 3: slope | $\gamma_{100}$ | −17.88 | 6.14 | 15 | <.001 | | 12.71 | |
| Conditional growth | | | | | | 21.55 | | 1629.48 |
| Level 3: Slope × Treatment | $\gamma_{102}$ | 2.88 | 0.46 | 14 | .660 | | 10.51 | |

*Note.* Individual $N=82$; group $N=17$.

[a]Number of parameters: base model = 4, unconditional growth model = 20, conditional growth model = 22. The variance component associated with the Level 2 growth parameter from the unconditional model was $\tau_{1ij}=518.37$.

The next question in Figure 1 asks, are the data repeatedly measured? As we indicated earlier, the choice of the appropriate MLM for assessing dependence and for analysing change data is contingent on the research design and questions. In the group treatment research example presented previously (Piper et al., 2007), the data were nested and the research design and questions addressed longitudinal hypotheses; hence, a three-level nested growth model was appropriate.

We argue that the three-level growth MLM is a useful model for many group treatment research designs for a number of reasons. First, three-level growth models allow the researcher to assess variance related to rates of change for all participants in addition to testing whether rates of change differ by treatment condition. In the group treatment example presented previously, the unconditional growth model indicated that all participants had reduced STAI scores up to 6 months posttreatment, but there were no differences between conditions. This provided more information than was possible with a single end-point analysis using a two-level random-effects model, which would have only compared conditions on postscores or follow-up scores while controlling for prescores. Second, growth MLMs allow one to specify linear or nonlinear models depending on one's concept of change. In our example, we hypothesized a nonlinear model of rapid change at posttreatment and then maintenance to 6 months posttreatment. Third, as indicated earlier, growth parameters (i.e., slopes) tend to be based on several waves of data, and so these parameters are likely more reliable than those based on two data points, as would have been the case with the use of a two-level random-effects model (Crits-Christoph & Gallop, 2006). This leads to a fourth advantage: Growth parameters are based on multiple data points and so the estimates of rates of change are less affected by missing data, when the data are missing at random (Gueorguieva & Krystal, 2004).

Despite these advantages, three-level growth MLMs present some challenges and limitations. First, these are complex models, and it is very easy to misspecify the models or misinterpret the results. Second, calculating and understanding estimates of variance accounted for by adding predictors at upper levels by pseudo-$R^2$ are very complex. Pseudo-$R^2$ for variables at Levels 2 and 3 of growth models is composed of variances related to different parts or levels of the model (Singer & Willett, 2003, pp. 103–104). As a result, unlike traditional $R^2$ statistics, pseudo-$R^2$ can have a negative value. For this reason, Hox (2002) argued that pseudo-$R^2$ may not be appropriate for MLMs. Kreft and de Leeuw (1998), however, suggested that pseudo-$R^2$ is appropriate when variance components of subsequent models are compared with variance components of base models only. For these reasons, some prefer the use of the deviance statistic to assess best fitting nested models (Singer & Willett, 2003). Third, as indicated, acquiring an adequate number of groups at Level 3 of these models requires greater resources. Not having enough groups relative to parameters may result in problems with convergence of the model, less than ideally reliable estimates of parameters at Level 3, and underpowered analyses at Level 3. When designing studies, group treatment researchers would do well to measure outcomes more often and to accumulate larger numbers of groups.

There are some potential limitations to the particular model presented in Equations 2 and 3. For example, in this model, we assumed an imperfect relationship between the observed prescore and the latent intercept. Hence, we chose to allow the intercept error terms to vary at Levels 2 and 3 even after covarying the prescores. One could argue that the observed prescore and the latent intercept will be nearly equivalent, and so the error terms in this model, $r_{0ij}$ and $u_{00j}$, may be unreliable. Some three-level models do constrain variability in the intercept, and as a result all participants' predicted baselines are made to be equal (e.g., Elkin et al., 2006). This might be done to enhance the reliability of slope parameters, which is an important consideration for change research (Lutz, Leon, Martinovich, Lyons, & Stiles, 2007). However, this modeling that constrains intercept variability has caused some debate partly because it assumes that the initial scores have no measurement error, and this may not be realistic in the psychotherapy research context (Wampold & Bolt, 2006). As this and other such debates suggest, MLMs are not "correct" in and of themselves, but each model reflects a variety of decisions and a particular conceptualization of the data and of the phenomenon being studied. The use of MLM in group treatment research is relatively new, with few agreed-upon conventions. As they make their modeling decisions, group treatment researchers are encouraged to consider such issues as whether to control for prescores and whether to allow intercepts to vary when prescores are controlled. To aid in such decisions, deviance statistics of nested models (e.g., MLMs that allow intercept error terms to vary vs. those that do not) can be compared or chi-square distributions of the intercept variance components can be examined.

We encourage researchers to systematically approach their data and research design questions by evaluating whether (a) the data are nested by nature, (b) there is dependency in the data, and (c) the data are repeatedly measured (see Figure 1). Further, we

encourage researchers to build their models in a sequence from the least to the most complex (Hox, 2002; Singer & Willett, 2003). Start with a base model that allows one to estimate overall variance in the data and provides a basis for estimating variance accounted for by adding parameters (Hox, 2002; Kreft & de Leeuw, 1998). Then add predictors based on examination of the chi-square statistic associated with Level 2 and Level 3 variance components. For example, in growth models, adding the time parameter at Level 1 of a second iteration of the model and then adding treatment condition as a fixed effect at Level 3 in a third iteration of the model allows one to assess variance explained by these added factors and the relative fit of the nested models to the data.

MLMs represent a powerful set of tools for group treatment researchers to evaluate and work with nested change data. Three-level growth MLMs are appropriate for the design and data analysis challenges posed by some group treatment research. We have presented some basic issues related to research design and data analysis with three-level MLMs as a way of encouraging group treatment researchers to move beyond traditional and inadequate methods of modeling nested change data. Adopting an MLM addresses some problems with hierarchically nested and longitudinal data analysis. However, MLMs are more than just techniques for addressing problems. MLMs also allow the group treatment researcher to examine important issues of group dynamics in new ways. For example, MLMs now afford the opportunity to appropriately model grouped change data and estimate how much variance in the outcome or process is attributed to group relative to individual effects (e.g., Imel et al., 2008).

## Acknowledgements

We thank Kimberly Corace, Robert Gallop, and Tim Ramsay for their comments.

## References

Atkins, D. C. (2005). Using multilevel models to analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology*, *19*, 98–110.

Azim, H. F. A., Piper, W. E., Segal, P. M., Nixon, G. W. H., & Duncan, S. (1991). The Quality of Object Relations Scale. *Bulletin of the Menninger Clinic*, *55*, 323–343.

Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or Type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology*, *73*, 924–935.

Crits-Christoph, P., & Gallop, R. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program and other psychotherapy studies. *Psychotherapy Research*, *16*, 178–181.

Elkin, I., Falconnier, L., Martinovich, Z., & Mahoney, C. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Research Program. *Psychotherapy Research*, *16*, 144–160.

Gueorguieva, R., & Krystal, J. H. (2004). Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in papers published in the *Archives of General Psychiatry*. *Archives of General Psychiatry*, *61*, 310–317.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. New York: Wiley.

Holden, J. E., Kelley, K., & Argawal, R. (2008). Analysing change: A primer on multilevel models with applications to nephrology. *American Journal of Nephrology*, *28*, 792–801.

Hox, J. (2002). *Multilevel analysis: Techniques and applications* (2nd ed.). Mahwah, NJ: Erlbaum.

Imel, Z., Baldwin, S., Bonus, K., & MacCoon, D. (2008). Beyond the individual: Group effects in mindfulness-based stress reduction. *Psychotherapy Research*, *18*, 735–742.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, *99*, 422–431.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology*, *54*, 32–39.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92.

McCallum, M., Piper, W. E., & Joyce, A. S. (1995). *Manual for time-limited, short term, supportive group therapy for patients experiencing pathological bereavement*. Unpublished manuscript.

Piper, W. E., McCallum, M., & Joyce, A. S. (1995). *Manual for time-limited, short-term, interpretive group therapy for patients experiencing pathological bereavement*. Unpublished manuscript.

Piper, W. E., Ogrodniczuk, J. S., Weideman, R., Joyce, A. S., & Rosie, J. S. (2007). Group composition and group therapy for complicated grief. *Journal of Consulting and Clinical Psychology*, *75*, 116–125.

Pollack, B. (1998). Hierarchical linear modeling and the "unit of analysis" problem: A solution for analyzing responses of intact group members. *Group Dynamics*, *2*, 299–312.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM6: Hierarchical linear and non-linear modeling*. Lincolnwood, IL: Scientific Software International.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis; Modeling change and event occurrence*. New York: Oxford University Press.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.

Tasca, G. A., & Gallop, R. (2009). Multilevel modeling of longitudinal data for psychotherapy researchers: I. The basics. *Psychotherapy Research*, *19*, 429–437.

Task Force on Promotion and Dissemination of Psychological Procedures. (1998). Update on empirically supported therapies: II. *Clinical Psychologist*, *48*, 3–23.

Wampold, B. E., & Bolt, D. M. (2006). Therapist effects: Clever ways to make them (and everything else) disappear. *Psychotherapy Research*, *16*, 184–187.

Yalom, I. D., & Leszcz, M. (2005). *The theory and practice of group psychotherapy* (5th ed.). New York: Basic Books.