Internet Interventions 2 (2015) 110-120

Contents lists available at ScienceDirect



Internet Interventions

journal homepage: www.invent-journal.com/



CrossMark

## Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions

### Hugo Hesser \*

Department of Behavioural Sciences and Learning, Linköping University, Sweden

#### ARTICLE INFO

Article history: Received 17 November 2014 Received in revised form 20 February 2015 Accepted 22 February 2015 Available online 4 March 2015

Keywords: Growth models Structural equation modeling Multilevel modeling Internet interventions Mediation Moderation

#### ABSTRACT

Growth models (also known as linear mixed effects models, multilevel models, and random coefficients models) have the capability of studying change at the group as well as the individual level. In addition, these methods have documented advantages over traditional data analytic approaches in the analysis of repeated-measures data. These advantages include, but are not limited to, the ability to incorporate time-varying predictors, handle dependence among repeated observations in a very flexible manner, and to provide accurate estimates with missing data under fairly unrestrictive missing data assumptions. The flexibility of the growth curve modeling approach to the analysis of change makes it the preferred choice in the evaluation of direct, indirect and moderated intervention effects. Although offering many benefits, growth models present challenges in terms of design, analysis and reporting of results. This paper provides a nontechnical overview of growth models in the analysis of change in randomized experiments and advocates for their use in the field of internet interventions. Practical recommendations for design, analysis and reporting of results from growth models are provided.

© 2015 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND licenses (http://creativecommons.org/licenses/by-nc-nd/4.0/).

### 1. Introduction

Despite considerable advances in applied statistics and methodology over the years, the randomized experiment is still the only established way to test cause–effect associations (Pearl, 2009). By randomizing participants to conditions and holding everything, save one variable (i.e., experimental manipulation), constant across conditions, the average difference between participants treated and controls is an unbiased estimate of the average casual effect of the treated (Rubin, 1974). A well-designed randomized experiment also allows for the exploration of how this direct effect between treatment and outcome arises (i.e., mediation/indirect effects), and under what circumstances or for whom (i.e., moderation) this effect is most pronounced (Baron and Kenny, 1986).

A number of factors can bias the estimate of the direct, indirect, and moderated effects in a randomized experiment. Modern statistical analyses of repeated-measures data offer potential remedies to common problems encountered in field experiments, for example, addressing problems with attrition/missing data, dependence among repeated observations, and statistical power. Methodological advances in the analysis of change also provide researchers with excellent analytic tools to answer important research questions regarding individual differences. These advances are most relevant for the statistical analysis of mediation and moderation in randomized experiments. Despite the benefits of these techniques, however, these methods also present specific and real challenges for researchers in terms of designing, analyzing and reporting results of treatment trials.

The field of internet interventions has grown rapidly over the past decade (Andersson, 2009; Andrews et al., 2010). Numerous randomized controlled trials have been conducted to evaluate whether internetdelivered interventions produce beneficial outcomes relative to control conditions, other treatments, or delivery-formats for a wide variety of health issues (see for reviews, Andersson et al., 2011, 2014; Andrews et al., 2010; Hedman, 2014; Spek et al., 2007). There is clear heterogeneity, however, among studies on internet interventions in terms of the scientific methodology used (Barak et al., 2009; Danaher and Seeley, 2009; Ritterband and Tate, 2009). Recently, in an attempt to raise the quality of research in the area, general research guidelines were developed that provide researchers with recommendations for how to best report studies within the field (Proudfoot et al., 2011); yet, to my knowledge, no guidelines exist that provide recommendations for how to appropriately design and analyze randomized experiments of internet interventions. Such guidelines can be of critical importance for the scientific progression of the field at large, as an appropriate methodology of experiments is at the heart of scientific claims and

2214-7829/© 2015 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>\*</sup> Corresponding author at: Department of Behavioural Sciences and Learning, Linköping University, SE-581 83 Linköping, Sweden. Tel.: +46 13 28 58 45; fax: +46 13 28 21 45.

E-mail address: hugo.hesser@liu.se.

inferences (Mayo, 1996). Specifically, no guidelines have been developed that provide recommendations for how to make use of modern statistical techniques in the analysis of change in experiments of internet interventions. Perhaps as a consequence, coupled with the rapid growth of research in the area, the data analytic procedures used in clinical trials on internet interventions have varied between studies and some studies have used out-of date methods to analyze clinical trial data. This could lead to incorrect conclusions and in the long run be problematic for the field. In addition, studies within the field have yet to make full use of the advantages of modern data analytic techniques offer in the study of individual differences in change and correlates of individual change in experimental designs.

The overall purpose of this paper is to increase awareness of the advantages of modern data analytic procedures for repeated-measures data and to encourage their use over traditional methods in the field of internet interventions. The paper provides a nontechnical overview of growth curve models in the analysis of direct, indirect and moderated treatment effects and summarizes some recommendations from lead-ing methodological authorities within the field of longitudinal data analysis. The objective is not to provide a detailed technical account of how to conduct these analyses. For this, readers are referred to the many accessible didactic articles and books available (Bollen and Curran, 2006; Hedeker and Gibbons, 2006; Kwok et al., 2008; Peugh and Enders, 2005; Singer, 1998; Singer and Willett, 2003; Snijders and Bookker, 2012).

The structure of the paper is as follows. The article begins with a brief review of growth models in the analysis of change in randomized experiments. Here, I also focus on design issues of relevance when using growth models. An example is provided to illustrate how to implement and interpret a linear growth model in the context of a randomized controlled trial. I then provide some recommendations for modeling building issues concerning functional form of change, coding of time, and variance components. Next, I focus on a common problem in randomized experiments that can seriously bias inferences: missing data. In the context of Rubin's (1976) missing data theory, I discuss maximum likelihood estimation in growth models and argue for its use in clinical trials with incomplete data over other common ad hoc missing data handling techniques. In the subsequent section, I discuss how to take full advantage of the modeling capabilities of growth models in the statistical analysis of mediation and moderation in randomized experiments. The paper concludes with some recommendations for the reporting of results from growth models in studies on internet interventions.

#### 2. Measure and model change in randomized experiments

When researchers conduct a randomized experiment to evaluate an outcome of an intervention they are, implicitly or explicitly, concerned with how that outcome changes over time. They might be interested in how a specific set of behaviors or symptomatology expressed by the individual at certain time point has changed in response to the intervention. Consequently, when evaluating the outcome of an intervention, they need to consider how to model and measure change. Historically, change has often been examined using one or two assessment points, such as in a pre- and post-treatment design (Francis et al., 1991). For such research designs, repeated measures analysis of variance (RM-ANOVA) or multiple regression analysis (including analysis of covariance which is a form of multiple regression) has frequently been employed. The primary purpose here is to study mean group differences. Individual deviations from these averages are treated as error variance, rather than reflecting the true amount of change that take place at the individual level (Duncan and Duncan, 2004). Although this design and data analytic choice can be defended, concerns from a conceptual and statistical standpoint have been raised by methodologists, in particular, related to the study of correlates of change (Rogosa and Willett, 1985). On a conceptual level, these models do not explicitly model individual change and change is viewed as incremental rather than as a continuous developmental process that unfolds over time (Francis et al., 1991; Rogosa, 1988; Rogosa and Willett, 1985). On a statistical level, these models are concerning because of the (un-) reliability of the difference score, inability to adequately handle timevarying predictors, missing data as well as unbalanced data (Duncan and Duncan, 2004; Francis et al., 1991; Gueorguieva and Krystal, 2004; Kwok et al., 2008; Rogosa and Willett, 1985; Willett and Sayer, 1994). Moreover, our data analytic approach must be able to handle nesteddata structures and dependence among observations in repeatedmeasures data (i.e., observations are nested within individuals over time). Traditional data analytic approaches (e.g., RM-ANOVA) rest on strong data analytic assumptions regarding covariance structures for handling dependence among repeated-measurements (e.g., sphericity assumption) and if these assumptions are violated it will lead to incorrect decisions (Francis et al., 1991; Gueorguieva and Krystal, 2004; Kwok et al., 2008).

An alternative approach is to view change as a continuous process that occurs between any two-time periods. Change is captured by an individual's underlying growth trajectory. Growth models (also known as random coefficients models, multilevel models, and mixed effect models) are well suited for the purpose of studying change at the group as well as the individual level (Bryk and Raudenbush, 1987; Hedeker and Gibbons, 2006; Meredith and Tisak, 1990; Muthén and Curran, 1997; Willett and Sayer, 1994). This makes the data analytic approach highly relevant for the analysis of change in randomized experiments. These models have several clear advantages over traditional data analytic approaches for the analysis of repeated-measures data. These advantages include, but are not limited to, the ability to incorporate time-varying predictors, handle dependence among repeated observations in a very flexible manner, and to provide accurate estimates with missing data under fairly unrestrictive missing data assumptions (Duncan and Duncan, 2004; Gueorguieva and Krystal, 2004; Muthén and Curran, 1997). The data analytic approach also allows researchers to handle other types of dependence among observations due to clustering. For example, effects may, in certain intervention studies, be due to variability among therapists providing the treatment and random effects modeling can be one way to control for clustering due to therapists in experimental designs (Wampold and Serlin, 2000). Finally, although it is largely an unstudied topic, statistical power to detect nonzero parameters can be increased in certain circumstances when using growth models. For example, compared to traditional data analytic approaches (e.g., RM-ANOVA) in the analysis of repeated-measures data from between-group designs, growth models have shown to produce greater statistical power in detecting a difference in a linear slope estimate (i.e., time by group interaction) in small to moderate sample sizes (Fan, 2003).

#### 2.1. Recommendations for design

Although it is possible to apply growth models to two-wave data, the traditional pre–post-treatment design is not optimal, especially if the researcher intends to model individual heterogeneity in growth (Rogosa and Willett, 1985).<sup>1</sup> That is, when only two time points are available one cannot separate error variance from individual heterogeneity in change. This precludes the study of individual differences in change and of correlates of individual change. In addition, if the researcher suspects that the functional form of change (time trend) cannot be adequately described with a straight line, two data points are not sufficient. Also, inclusion of multiple measurements during the

<sup>&</sup>lt;sup>1</sup> Even if the aim of the study is not to examine individual growth, there are situations in which the researcher still might want to consider growth models over traditional approaches (e.g., RM-ANOVA) because they offer other clear advantages, such as the ability to include time-varying covariates, handle missing and unbalanced data, and to fit various forms of variance-covariance structures to the data.

trial is beneficial when there is missing data. Having more observations on the outcome between pre- and post-treatment assessment can provide valuable information about the individual who fails to return the last assessment point in the study; information that can be used to obtain accurate parameter estimates with incomplete data (I will return to this last point when I discuss missing data in Sections 5.1 and 5.2). Finally, increasing the number of measurements can positively affect the statistical power and reliability of the assessment of individual change (Muthén and Curran, 1997; Raudenbush and Bryk, 2002).

Thus, more than two time points is often preferred and too few measurements can in certain situations be detrimental to a longitudinal study (Collins, 2006). Several factors can be used to determine the number of time points to be assessed. Consideration needs to be given to the underlying theory of change such as when change is most likely to occur and measure relevant constructs frequently and tightly spaced during these time periods (Collins, 2006). The researcher need also consider the nature of the measures used to assess change, such as whether the measures are sensitive enough to detect a significant change during the time period, whether there is a risk of floor or ceiling effects, and whether repeated testing changes the meaning of the construct being measured (Jackson, 2010). The test-retest reliability of the instrument as well as power issues also factor into decisions about number of measurement occasions. As a general recommendation, researchers should aim to include, at the minimum, four measurement points in a trial because the flexibility in the modeling approach is increased substantially when more than three time points are available. However, whenever possible, the aim should be to include more measurements during the active intervention phase, for example, weekly or biweekly measures of primary outcomes (and process variables) can be beneficial, especially when examining moderators and mediators in a trial.

#### 3. Implementing and interpreting growth models

Growth models can be estimated within a multilevel, mixed model or structural equation modeling framework. Although there are some differences in modeling capabilities, different frameworks often yield identical results (Bollen and Curran, 2006; Enders, 2011b). In fact, Preacher et al. (2008) point to the fact that differences between frameworks to a large extent are products of the software and as new versions of software are released, differences between frameworks become more difficult to identify. In this paper, I use the generic term growth models (or growth curve models), but readers should be aware that there are some differences between frameworks that will affect results in certain types of models (e.g., models with time-varying predictors). Some models can also be more readily estimated within one framework as compared to another (e.g., parallel process growth models and not missing at random growth models are easier to implement in structural equation modeling framework as compared to linear mixed or multilevel framework). Jackson (2010) has provided some general recommendations when one framework may be preferred over another, but as stated earlier, the software may play an even more important role here.

# 3.1. An example of a linear growth model in the context of a randomized trial

To illustrate how to implement and interpret a linear growth model, consider the following hypothetical randomized depression trial in which participants were randomly assigned to either an internetdelivered intervention or an active control condition and the primary outcome was measured at four time points weekly throughout the active treatment phase. Fig. 1 illustrates a randomly selected sample of observed individual trajectories and the average mean trend for the entire sample (i.e., the treatment group and control group combined). As can be seen in the figure, individuals started at different levels and changed at different rates over time, but, on average, they seemed to decrease in depression scores over time. Now consider if the researchers wish to fit a linear growth trajectory to the observed data and evaluate whether individuals assigned to the intervention decrease at faster linear rate in depression symptoms as compared to those assigned to the control condition (and hence ending up with lower depression scores at the end of the treatment period).

Growth models can be expressed in the form of a linear regression equation in which the outcome is modeled as function of a temporal predictor variable that captures how the individual changes on the outcome over time. A linear growth model is,

$$Y_{ti} = \beta_0 + \beta_1(TIME_{ti}) + b_{0i} + b_{1i}(TIME_{ti}) + \varepsilon_t$$

where  $Y_{ti}$  is the outcome at time point t for individual i,  $TIME_{ti}$  is the value of the predictor for individual i at time point t (e.g., the time passed since the onset of the study),  $\beta_0$  is the mean intercept,  $\beta_1$  is the mean growth rate (i.e., expected change in the outcome as function of 1 unit change on the time predictor variable),  $b_{0i}$  and  $b_{1i}$  are so-called random effects that allow the intercepts and growth rates to vary across individuals, and  $\varepsilon_{ii}$  is a time-specific residual that expresses the difference between an individual's fitted linear trajectory and the observed data. Thus, the model includes both fixed effects (i.e., averages across individuals) and random effects (i.e., individual deviations from these averages); hence, the use of the term mixed effects models.

The most noteworthy aspects of the model are the random effects,  $b_{0i}$  and  $b_{1i}$ , that capture individual heterogeneity (normally assumed to be multivariate normally distributed). Thus, each individual would get his or her own intercept,  $b_{0i}$  (here the value on the outcome at initial assessment) and slope,  $b_{1i}$  (expected change in the outcome as function of 1 unit change, e.g. one week, on the time predictor variable). In addition, by estimating the covariance between random effects we can obtain information on whether individuals' intercepts and slopes are correlated; that is, in this case, whether there is a relationship between individuals' starting values and individuals' rates of change on the outcome. For example, the researcher may expect that individuals with high initial depression scores will change at faster rate over time (decrease more in depression symptoms) as compared to those with low scores at baseline. This covariance between intercepts and slopes is estimated in the covariance structure associated with random effects (error variance also has an associated covariance structure; I return to this covariance structure in Section 4.1). When only two random effects are included in the model, we can either estimate this covariance (i.e., unstructured covariance structure) or constrain it to be zero (i.e., diagonal covariance structure).



Fig. 1. Observed mean trend and randomly selected individual trajectories with artificial depression trial data.

Up until now, the model only examines change over time in the entire sample of individuals and does not reflect the fact the individuals were randomized to different groups. The primary goal of a treatment trial is, of course, to determine whether individuals randomized to the intervention change at a faster (or slower) rate over time as compared to those in the control group. To accomplish this, our linear growth model needs to include a binary coded predictor variable representing conditions (1 = intervention, 0 = control). The model then becomes,

$$\begin{array}{l} Y_{ti} = \beta_{0} + \beta_{1}(TIME_{ti}) + \beta_{2}(CONDITION_{i}) + \beta_{3}(CONDITION_{i})(TIME_{ti}) \\ + b_{0i} + b_{1i}(TIME_{ti}) + \varepsilon_{ti} \end{array}$$

where  $\beta_0$  is the population estimate of the intercept for the control group,  $\beta_1$  is the population estimate of the linear slope for the control group (i.e., control = 0),  $\beta_2$  and  $\beta_3$  capture the estimates of the mean difference in intercept and slope between conditions,  $b_{0i}$  and  $b_{1i}$  are random effects that allow the intercepts and growth rates to vary across individuals, and  $\varepsilon_{ti}$  is a time-specific residual that expresses the difference between an individual's fitted linear trajectory and the observed data. Most relevant, of course, for the primary goal of the study is to test whether  $\beta_3$  (i.e., condition by time interaction; mean difference in slopes between conditions) is nonzero. What is important to note, however, if this coefficient is significantly different from zero it will explain some of the variance in linear slope trajectories,  $b_{1i}$ . That is, by including the condition variable in the model we aim to account for the individual heterogeneity and random effects now act as residuals (i.e., unexplained individual heterogeneity not accounted for by condition). In fact, this model is sometimes described as a conditional growth model in which random effects (i.e., intercepts and slopes) are conditioned on the predictor variable, whereas the former model is often referred to as an unconditional growth model because it only includes the temporal predictor of time without other explanatory variables (Bollen and Curran, 2006). Thus, by allowing intercepts and slopes to vary across individuals we can add predictor variables to account for this heterogeneity in growth trajectories.

To aid in the interpretation of this linear growth model, I used artificial data to mimic results of a depression trial and implemented the growth models (unconditional and conditional model) in the linear mixed effects framework using SPSS version 21 (full information maximum likelihood estimation with an unstructured covariance structure for random effects and an identity covariance structure for error variance). The results can be found in Table 1. As can be seen in the table, there were significant individual heterogeneity in initial levels (i.e., intercepts) and slopes. Participants, regardless of assignment, decreased, on average, in depression scores by -0.71 ( $\beta_1$ ) points per week. Most noteworthy, the slope heterogeneity could in part be accounted for by the inclusion of the condition predictor variable in the growth model (i.e., the variance associated with random slope was reduced from 0.14 to 0.11, see Table 1). Indeed, there was a significant condition by time interaction,  $\beta_3 = -0.40$ , indicating that participants assigned to the intervention decreased, on average, by 0.4 points more in depression scores per week as compared to those assigned to control. Thus, condition could systematically explain some of the variance in individual slope estimates. To aid in the interpretation of this effect,  $Pseudo-R^2$  can be calculated as the proportion of explained variance in the random effect by the condition predictor variable (Singer and Willett, 2003). By subtracting the estimate of the random effect variance in the conditional growth model (i.e., the model with the predictor) from the estimate of the variance in the unconditional growth model (i.e., the model without the predictor) and by dividing this difference with the unconditional random effects variance, we obtain the proportion of explained variance by the predictor. Based on the information in Table 1, it can be concluded that the condition predictor variable accounted for about 21% of the slope variance ((0.14-0.11) / 0.14 = 21.4%).

#### 3.2. Growth models within structural equation modeling framework

As previously stated, growth models can be implemented in different frameworks. One useful and flexible framework is structural equation modeling. Although a review of structural equation modeling is beyond the scope of this article, here I briefly show how the linear growth model can be cast within this framework. In structural equation modeling, the linear growth model is a two-factor confirmatory factor analysis, where the random intercept and slope are captured by the two latent factors. Fig. 2 depicts a path diagram of the conditional linear growth model. The ellipses are latent variable (with means and variances), rectangles are observed measured variables, single-headed straight arrows represent regression coefficients, and the doubleheaded arrow denotes covariance. Specifically, the variances associated with the latent variables capture the heterogeneity in growth trajectories (intercept and slopes;  $b_{0i}$ , and,  $b_{1i}$ ) and the mean of the latent intercept and latent slope variable corresponds to the mean of the intercept and slope ( $\beta_0$ , and  $\beta_1$ ; here the average initial level and rate of change in the control condition, as the condition status variable is included in the model). The coefficients for the single-headed arrows that connect the condition status variable (binary coded variable) to the latent variables correspond with the mean differences between conditions ( $\beta_2$  and  $\beta_3$ ; mean difference in initial level and slope). Finally, the factor loadings for the intercept (1, 1, 1, and 1) and slope (0, 1, 2, and 3) stipulate the functional form of change (i.e., a linear model with equally spaced measurements).

#### 3.3. Type of predictors in growth models

One of the strengths of this modeling approach is that it can incorporate various forms of predictors in the model. The predictor values may vary over time within individuals, so-called time-varying covariates (e.g., change in medication use over time), or values may be fixed across time within individuals but vary across individuals, so-called timeinvariant covariates (e.g., condition in between-group design) (Singer and Willett, 2003). In other words, predictors may either covary with the outcome over time or represent stable characteristics that do not change over time. Time-varying covariates generally account for the variance within individuals (i.e., error variance), whereas time-invariant covariates account for variance between individuals (i.e., random effects; although these variance components are related within a general covariance structure so both within and between variance may be affected with inclusion of both types of predictors, Singer and Willett, 2003; Wu et al., 2009). An applied example of the use of time-varying covariates can be found in a study that examined associations between the frequency of patient's use of particular words in their text communication with therapists and outcomes in an

Results from a linear growth model with artificial depression trial data.

	Unconditional		Conditional	
	Estimate	SE	Estimate	SE
Fixed effects Initial level, $\beta_0$ Time, $\beta_1$ Condition, $\beta_2$	5.33 <sup>*</sup> - 0.71 <sup>*</sup> -	0.06 0.04 -	$5.39^{*}$ - 0.39^{*} - 0.08 - 0.08	0.13 0.08 0.15
Random effects Variance initial level, $b_{0i}$ Variance slope, $b_{1i}$ Covariance initial level-slope Residual, $\varepsilon_{ti}$	- 0.33 <sup>*</sup> 0.14 <sup>*</sup> 0.05 0.69 <sup>*</sup>	- 0.09 0.03 0.04 0.05	-0.40 0.33 <sup>*</sup> 0.11 <sup>*</sup> 0.05 0.69 <sup>*</sup>	0.09 0.03 0.04 0.05

Note. Unconditional growth model does not include condition, whereas the conditional model does. Condition is binary coded variable (1 = intervention, 0 = control). SE = standard error.

\* p < .001.

internet-delivered psychological intervention for generalized anxiety disorder (Dirkse et al., 2014). Researchers can also incorporate regression among random effects in the growth model; in such models, the random effects (intercepts and/or slopes) serve as predictors of other random effects in the model (I return to this type of growth model when I discuss moderation and mediation in Sections 6.1 and 6.2).

#### 3.4. The coding of time and interpretation of the intercept

The interpretation of both the average and individuals' intercepts and slopes changes with the coding of the predictor variables in the equation. Thus, it is necessary to know how time is coded in order to interpret the coefficients. Similar to standard regression analysis, the intercept represents the value at the time point when the predictor variable(s) in the model is zero. For example, if individuals are measured weekly from pre- to post-treatment (4 weeks), a possible coding of time could be 0, 1, 2, and 3. As the first time point is coded as zero, we would interpret the intercept as the average level at the first assessment and individual heterogeneity at that time point (in an unconditional growth model). We could also recode the variable so that the intercept would be at the end of the assessment period, as in, -3, -2, -1, and 0. This would only be a reparameterization of the same model, but it would change the interpretation of the intercept and associated variance (i.e., random effects) (Biesanz et al., 2004). The inclusion of other predictors in the model, besides time, would also change the interpretation of the intercept, so it is import to know how each predictor variable is coded in the model.

Although it is possible to use coding schemes common in standard regression or ANOVA, such as centering or polynomials contrasts, it is generally not recommended (Biesanz et al., 2004). The coding of time should produce readily interpretable estimates that reflect the substantive research question (Biesanz et al., 2004). For example, if researchers wish to evaluate the outcome at the end of the treatment period, making the last measurement point the intercept makes sense. Of course, if one is interested in the time point in the middle of all assessment points (e.g., as a transition point between two phases) centering the time variable is an option (Raudenbush and Bryk, 2002). It is also important to observe that coding of time should reflect the time



**Fig. 2.** Schematic figure of a linear growth model with a condition status variable (1 = intervention; 0 = control) predicting individual intercepts and slopes.

intervals between measurement occasions; for instance, if measurements were taken every other week, rather than every week, the coding of time would be 0, 2, 4, and 6 (for a linear model).

#### 4. Model building strategies for growth models

In order not to obtain biased parameters estimates, we need to fit an accurate model to the observed data. As with most model building strategies, the goal is to identify a parsimonious statistical model that describes the observed data to satisfactory extent. To identify an appropriate growth model it is important to consider, the time trend, as well as variance within individuals (i.e., error variance) and between individuals (i.e., random effects) (Wu et al., 2009). Given the flexibility of the approach and the many models available to choose from, it is recommended to start the analysis phase by plotting both the average trend and individual trends over time (Kwok et al., 2008). In a randomized controlled trial, plotting these trends as a function of condition can help to identify the functional form of change and allow for the examination of individual heterogeneity. This visual inspection serves to limit the number of models that need to be tested.

#### 4.1. Modifying the time trend and variance components

Several different time trends or functional forms of change can be specified in growth models and this is accomplished by altering how time is coded in the model (Bollen and Curran, 2006). Growth models can incorporate polynomial terms (as standard regression) to capture non-linear change. For example, to obtain a quadratic function we could simply square the linear term (using the aforementioned depression trial example, the quadratic term is 0, 1, 4, and 9) and then include both terms in the model. As before, we can incorporate individual heterogeneity in the trajectories by including random effects associated with the intercept, linear and quadratic terms in the model. As the quadratic function captures the curvature from the straight line, it may be suited for a trial with follow-up data; that is, individuals may change rapidly during the active treatment phase and then change may stabilize during the follow-up phase of the trial. However, an even better data analytic approach in this situation would be to model the treatment trial as qualitatively distinct phases by estimating a separate regression coefficient for each phase of the trial (Duncan and Duncan, 2004; Muthén and Curran, 1997). This can be accomplished with a piecewise function. With a sufficient number of data points (at least 3 per phase, assuming a linear change model in each phase), researchers may wish to entertain such a model because this would allow them to include all data points in the same analysis and also model change in a way that fits nicely with how change is likely to occur in most trials that include follow-up assessments. In addition, using a piecewise growth model, a set of predictors may be used to predict individual growth in one phase (e.g., pre- to post-treatment) and a separate set may be used during the other phase of the trial (e.g., follow-up phase) (Duncan and Duncan, 2004; Muthén and Curran, 1997). An example of this model can be found in a study that compared internet-delivered cognitive behavior therapy with face-toface group therapy for depression over a time period of three years using a randomized design (Andersson et al., 2013).

The choice of functional form of change should be based on theory of change and previous examination of the particular phenomenon under study. Indeed, methodologists have emphasized the importance of theory in model building and design in longitudinal data analysis (Collins, 2006). Sometimes theoretical assumptions about change are nonexistent (Jackson, 2010). In these situations, researchers may choose to test a number of different models to determine the model that best fits the observed data. Several different relative and global fit measures can be used here and some practical guidance for model building is also available (Bollen and Curran, 2006; Wu et al., 2009).

It is often appropriate to assess the statistical (and practical) significance of model parameters (e.g., means, variance and covariance) when evaluating model fit. Assuming a large enough sample, a common and general way to test the significance of a parameter is the Wald test, in which the point estimate is divided by its standard error; if the ratio exceeds 1.96, it is said to be significantly different from zero at the .05 level. When models are nested, researchers may use a likelihood ratio test to determine whether the inclusion of additional parameters (one or more) in the model contributes significantly to the model (Snijders and Bosker, 2012). The test is a deviance score, which is the difference in -2 Log Likelihood values between a less restrictive model and a more restrictive one. The significance of the deviance test value is evaluated against the chi-square distribution, where degrees of freedom is equal to the number of parameters that are set to zero in the more restrictive model. Likelihood ratio tests may be useful when examining individual heterogeneity in the form of the inclusion of random effects and associated covariance structure, error variance covariance structures as well as fixed effects (Snijders and Bosker, 2012).<sup>2</sup> It should be noted, however, that under certain circumstances, the Wald-test and the likelihood ratio test are too conservative and can produce biased standard errors (which is especially true for variance components as these are bounded; they are by definition nonnegative, making the standard normal symmetric sampling distribution inapplicable) (Snijders and Bosker, 2012). Various corrections and modifications to these commonly used tests can reduce such bias and create more powerful tests for parameters (Molenberghs and Verbeke, 2007). More detailed information on hypothesis tests of fixed and random effects in multilevel models can be found in the text by Snijders and Bosker (2012).

Once the number of random effects has been analytically determined, researchers may also decide to modify the error variance (Kwok et al., 2007). Longitudinal data analysis has historically often incorporated correlated error terms in the model to account for dependence among repeated measurements (e.g., autoregressive structure) (Willett and Sayer, 1994). Dependence is now, at least in part, accounted for by the inclusion of random effects in the model (i.e., dependence associated with that the same individuals are measured repeatedly over time); still, it may be meaningful to test whether the assumption of independence of error variance is tenable. It may also be reasonable to test whether error variance changes over time (i.e., diagonal covariance structure) rather than assuming that variance is homogenous across measurement points (i.e., identity covariance structure) (Willett and Sayer, 1994), which is the default in most multilevel or linear mixed software.

#### 4.2. Overall approach to model building

Unfortunately, the methodological literature offers little clear-cut guidance when it comes to the order in which components should be added to the model and where to start modifying the growth model (Wu et al., 2009). Some recommend estimating all parts in the model and then removing components that do not contribute significantly to the model (e.g., Verbeke and Molenberghs, 2000). Others argue for a bottom-up approach in which one starts with a simple model and evaluates whether the inclusion of additional components adds to the model, working upward from level 1 in multilevel repeated-measures terminology (e.g., Raudenbush and Bryk, 2002; Snijders and Bosker, 2012).

Wu et al. (2009) provide a baseline model as a starting point for evaluating and re-specifying growth models in both multilevel and structural equation framework. It is often advisable to begin by modifying and evaluating the functional form of change and then adding random effects as needed (see also a recent article that argued that random effects structure should be kept maximal in confirmatory hypothesis testing, Barr et al., 2013). After considering the time trend and taking into account heterogeneity at the individual level, researchers may want to modify the error variance. In a randomized trial, it is often useful to do this process for each condition separately before including all groups in the same model. Finally, predictors (e.g., condition) could be added to account for individual heterogeneity in growth trajectories.

There are no clear and fixed rules to follow when it comes to model specification and both substantive and statistical considerations should be taken into account when choosing among data models (Snijders and Bosker, 2012). As stated earlier, theory should guide researchers and it is often advisable to not test models in isolation, but to compare competing theoretically derived models (Preacher et al., 2008). It is also critical that the specified model accurately reflects the design of the study and the theoretical predictions. This also means that researchers should not drop or include variables in the model solely based on statistical criteria (e.g., significance level). Snijders and Bosker (2012) distinguish between two parts of the data analytic model that can be treated distinctly in the model building phase: the part that involves a priori hypotheses of parameters and the part that is required to get the model to fit the data well and is a prerequisite for valid tests of hypotheses. For the latter part, they argued, an inductive data driven-approach can sometimes be adequate, whereas for the former part of the model a data-driven approach to model specification is not suitable. However, see also Barr et al. (2013) for specific concerns regarding data-driven approaches to the specification of random effect structure in hypothesis testing.

#### 5. Statistical inferences in growth models with missing data

Missing data is arguably one of the biggest data analytic concerns in a randomized controlled trial. For example, in an intervention study for depression, it is reasonable to expect that participants who do not respond to treatment, and subsequently are the most depressed, are more likely to drop out of the study and not return to complete the follow-up assessment. This is an example of when there is a relationship between outcome and the probability for missing data, so-called nonignorable missing data (Enders, 2010, 2011b). To understand how missing data, and the extent to which missing data, influence statistical inferences, Rubin's (1976) missing data theory is of critical importance. Although the framework has been long established in the methodological literature, it has not always spread to the applied sciences. One reason is probably that the terminology is somewhat confusing and the practical implications of the theory are often elusive to most researchers (Graham, 2009). Three so-called missing data mechanisms are central to the theory: Missing Completely at Random (MCAR), Missing at Random (MAR), and Not Missing at Random (NMAR) (Little and Rubin, 2002). Each one of these mechanisms describes the association between the probability of missing data and other variables and the outcome variable itself. Missing mechanisms are important to consider as they act as assumptions for missing data handling techniques and determine the performance of these techniques (Enders, 2010).

#### 5.1. Ignorable missing data: MCAR and MAR

In the case of MCAR, the probability of missing data on a variable should be unrelated to other measured variables and to the would-be values of that variable (Enders, 2011a). For example, if we have measured participants at four measurement points in a randomized control trial (such as in the depression trial) and we only have missing data at the final time point, MCAR requires that missing, at this time point, is unrelated to treatment group status, to scores on the outcome variable at previous assessments, and to the would-be scores on the

<sup>&</sup>lt;sup>2</sup> When using restricted maximum likelihood estimation (REML), likelihood ratio tests cannot be used to test fixed effects nor to compare models with different covariates, as the method adjusts the likelihood for number of covariates in the model (Hedeker and Gibbons, 2006).

outcome variable at the final time point. In other words, if missing data is a random sample of the whole data set then MCAR holds. This is often an unrealistic assumption. Yet, most simple forms of strategies for handling missing data, such as eliminating cases with missing data (e.g., pairwise or list wise deletion) or single regression or mean imputation, assume MCAR. Indeed, numerous simulations studies have shown that these techniques produce biased estimates when data are missing under MAR or NMAR mechanisms (see for reviews, Enders, 2010; Graham, 2009).

A more lenient missing data assumption is MAR, in which the probability of missing data on a variable is allowed to be related to other observed variables, but not to the would-be values of that variable (Enders, 2011a). Despite what the name implies (random missingness), this is, in fact, systematic missing, because it allows the propensity for missing data to be related to observed variables included in the data analytic model. Graham (2009), for example, proposed that conditional missing at random might be a better word than MAR. Using the aforementioned example, MAR allows the probability of missing data at final assessment to be related to treatment group status, and to scores of previous assessments, but not to the would-be values at the final assessment. Thus, MAR is a much more reasonable assumption than MCAR because it is unlikely that other observed variables are unrelated to the outcome and the propensity for missing data. MAR allows for such associations, assuming that these variables are included in the data analytic model (Enders, 2010).

#### 5.2. Maximum likelihood estimation in growth models

Full information maximum likelihood estimation, as implemented in most software routines that estimate growth models, assumes MAR. Nontechnically, full information maximum likelihood estimation is an iterative process with the aim to identify population parameter values that have the highest probability of reproducing the sample data (readers interested in the technical details of maximum likelihood can consult, Enders, 2010). The technique does not strictly impute missing values, but borrows information from all of the observed data. Using the depression example, individuals who had not completed the final assessment, provided observed data to the estimation process on the first three measurement points and this information was used when estimating all parameters in the model. Assuming a multivariate normal distribution, the information from the incomplete cases steers the estimation process toward more accurate parameter estimates (Enders, 2010, 2011a). Thus, rather than disregarding incomplete cases, full information likelihood estimation uses all available data in the estimation process of parameter values. This method of incorporating all individuals in the analysis is most relevant when it comes to a randomized controlled trial for three primary reasons. First, to obtain an unbiased estimate of the average casual effect of the treated requires that all individuals randomized are included in the analysis (i.e., intention-to-treat). Second, if individuals are deleted due to partial missing data, statistical power loss is to be expected. Third, missing data is now handled under MAR rather than MCAR.

Indeed, full information maximum likelihood estimation is one of two recommended methods for handling missing data (the other is multiple imputation) (Schafer and Graham, 2002) and the technique has shown to outperform other methods commonly used in clinical trials. Most noteworthy are direct comparisons between this technique and an ad hoc missing data strategy frequently employed in the clinical trial literature: last observation carried forward. Methodological studies have repeatedly demonstrated that full information maximum likelihood estimation is a better alternative than last observation carried forward because it provides more accurate estimates and standard errors as well as increased statistical power (e.g., Lane, 2008; Mallinckrodt et al., 2001; Salim et al., 2008). In fact, despite its widespread use, last observation carried forward has shown to produce substantial bias, even under an MCAR mechanism (Enders, 2011a). Thus, full information maximum likelihood estimation has several advantages as compared to traditional remedies for missing data in clinical trials. There are also several examples of studies in the literature of internet interventions that use maximum likelihood estimation to cope efficiently with data loss under MAR (e.g., Hesser et al., 2012; Ljótsson et al., 2014; Newby et al., 2013).

#### 5.3. Non-ignorable missing data: NMAR

There are situations when MAR-based techniques (e.g., full information maximum likelihood estimation) also yield biased results. NMAR, outcome-dependent missing data or non-ignorable missing, occurs when the probability of missing data on a variable is related to the would-be value of that variable (Enders, 2011b). So returning to our example, NMAR occurs when the likelihood of dropping out of the study before completing the final assessment is directly related to the would-be values of the outcome at that time point of assessment, even after controlling for scores on previous assessments and treatment group status. This is the most problematic situation to handle and unfortunately there is no formal test of MAR (as it is based on an untestable assumption i.e., unobserved data or the would-be values of the outcome), so we cannot determine conclusively whether an NMAR mechanism is at play (Enders, 2011b). Even more problematic is that although several so-called NMAR analyses have been developed over the years (and continue to be developed), these analyses come with their own set of untestable assumptions and even minor violations of these assumptions can result in bias (Enders, 2011b). This has led to methodologists sometimes recommending either that MAR-based techniques be performed even if one suspects NMAR (e.g., Graham, 2009; Molenberghs et al., 2004), or that a NMAR analysis be conducted as a form of sensitivity analysis to examine whether results hold under various data assumptions and conditions (e.g., Enders, 2011b).

As stated above, statisticians have devoted much energy to develop different forms of NMAR models to handle non-ignorable data (for an overview see, Enders, 2011b), but to date they are not in widespread use in the behavioral sciences. One reason for this is that they have historically been difficult to implement, but new software developments have made these models much more accessible to researchers in applied sciences (Enders, 2011b). One prominent NMAR model for longitudinal data is the pattern mixture model (Hedeker and Gibbons, 1997, 2006), in which individuals are stratified according to their missing data patterns and a separate growth model is estimated for each pattern (or the patterns are included as dummy predictor variables in the model). The model allows one to examine whether individuals with a certain missing data pattern (e.g., missing data due to attrition) differ in their growth trajectory from those with a complete set of data, and, more importantly, whether these differences affect overall estimates and standard errors. To obtain corrected full sample parameter estimates (e.g., a mean difference in slope between treatment and control), proportion-weighted pattern-specific estimates are averaged across subgroups with different missing data patterns (Enders, 2011b). One can compare this combined overall estimate with the estimate obtained in a traditional MAR-based growth model to determine whether outcome-dependent missing data are likely to have influenced the findings. An applied example of a pattern-mixture model is provided in an open study that examined internet-delivered cognitive behavioral therapy for depression in routine psychiatric care (Hedman et al., 2014; for similar data analytic model in applied setting see also, Hadjistavropoulos et al., 2014).

# 6. Making full use of growth models: implications for mediation and moderation

Up until now we have considered the strengths of growth models to evaluate the direct effect of intervention on outcome. Yet to realize the full potential and possibilities of these models we should consider how we could apply them to answer substantive research questions related to differences between individuals within intervention studies. Several methodologists have pointed out that these methods respond to central questions in intervention studies, namely who responds best to what intervention and why (e.g., Duncan and Duncan, 2004; Muthén and Curran, 1997). Questions such as these are commonly answered by the means of statistical analysis of moderation and mediation. A moderator is a variable that affects the strength between the independent (e.g., intervention) and dependent variable (e.g., outcome), whereas a mediator is a variable that is part of casual chain in which the mediator acts as link between the independent variable and dependent variable (Baron and Kenny, 1986).

The importance of the study of moderators and mediators has also been emphasized in the field of internet interventions. For example, Ritterband et al. (2006) made the observation that, "The chief goal of any internet intervention is to produce cognitive and behavior change that leads to symptom improvement. Examining and testing this process using theories and models of behavior change is critical to furthering the understanding of how Internet interventions, and even treatments in general, work." (p. 2). They further argued that, "An advantage of conducting randomized controlled trials through the internet is the ease of obtaining large sample sizes (no geographical limitations), making it possible to better examine mediators and moderators of treatment." (Ritterband et al., 2006, p. 3). To adequately answer questions related to mediation and moderation, researchers need to model change at the individual level. Growth models are therefore of high value when it comes to testing mediators and moderators in intervention studies.

#### 6.1. Examining moderators using growth models

Moderation has often been evaluated using standard regression analysis with an interaction term (MacKinnon, 2008; MacKinnon et al., 2007). A nonzero interaction effect between moderator and condition in explaining variance in the outcome is interpreted as a moderated treatment effect. That is, the effect of treatment on outcome differs as function of values on the moderator variable. The simplest way to accomplish this with growth models would be to model individual growth on the outcome and then include a potential moderator (continuous or as a dummy variable) as a predictor in the model. A nonzero interaction effect between moderator, group and time would be interpreted as a moderated treatment effect. It is important to remember that by including this new variable in the model we aim to account for individual heterogeneity in slope trajectories. In other words, the question we aim to answer is whether individuals' rate of change varies systematically as a function of both condition and another third variable (i.e., moderator). An applied example of this model can be found in a study that examined moderators in an internet-based prevention trial for eating disorders (Völker et al., 2014). An extension of this moderator growth model would be using the random effect intercept to model the interaction effect between "true" initial status (i.e., intercepts vary over individuals at first assessment point) and condition in their influence on the growth rate. Thus, this model corresponds to a substantive research question, namely whether the intervention effect varies as a function of individuals' initial symptomatology. This can be accomplished by regression among random effects in the structural equation modeling framework (for variations of this model see, Muthén and Curran, 1997).

We can also use more sophisticated approaches to test moderated effects in intervention studies. One potentially useful technique is to empirically classify individuals based on their growth trajectory (intercepts and slopes), so-called group-based trajectory modeling or growth mixture modeling (Nagin and Odgers, 2010). We could add predictors in these types of growth models with the aim of trying to explore relationships between individual baseline characteristics and the empirically based sub-populations of trajectories of change and also examine associations with distal outcomes. Growth mixture models have been specifically developed for the purpose of evaluating intervention effects in sub-populations (Muthen et al., 2002). Although growth mixture models could potentially be very useful when examining moderators in clinical trials, up to date, I am aware of no applied example in the field of internet interventions that has implemented this type of growth model.

#### 6.2. Examining mediators using growth models

In randomized designs, mediation is often evaluated by examining whether the intervention (relative to control) changes the mediator and, if this change, in turn, is correlated with change in the outcome (MacKinnon, 2008). Growth modeling is most relevant when it comes to test mediators in intervention studies. One important reason is that these models allow for the examination of individual differences in change in both the outcome and mediator. Indeed, if one is concerned with how two constructs are related over time, such as a mediator and outcome, the investigation of how change comes about takes precedence. That is, we need an appropriate model for change for both the mediator and the outcome before we can consider if, and how, they are correlated (Cheong et al., 2003). Once an appropriate growth model has been developed for the mediator and outcome, multivariate extensions of growth models then allow researchers to examine whether individual change in the mediator is associated with individual change in the outcome over time in one combined model (Cheong et al., 2003; for an alternative multivariate mediation approach within the multilevel framework see, Bauer et al., 2006). This can be accomplished by regression among random effects in which the random slope of the outcome is regressed on the random slope of the mediator. Support for mediation is found when the treatment group significantly accounts for individual change in the mediator, which, in turn, is associated with individual change in the outcome. Appropriate tests of indirect effects (mediated effects) can then be applied (Cheong et al., 2003; MacKinnon et al., 2002). A good applied example of this parallel process mediator growth model can be found in a study that compared two forms of internet-delivered psychological treatments for irritable bowel syndrome using weekly measurements of potential mediators and primary outcome (Ljótsson et al., 2013).

One important methodological obstacle for any researcher who aims to evaluate mediation in the context of a randomized controlled trial is the correlational nature of the association between mediator and outcome (MacKinnon, 2008; MacKinnon et al., 2007). That is, we cannot conclusively determine whether the mediator has a casual effect on the outcome as we do not have experimental control over the mediator. One way to perform a more stringent test of causality in mediator models is to establish the timeline between mediator and outcome; that is, whether change in the mediator preceded and contributed to subsequent change in the outcome (Kazdin, 2007). One way to test this assumption is to include the mediator as lagged time-varying covariate in the growth model. That is, we could predict outcome at time point *t* using the mediator at time point t - 1 (or any other lag that is relevant) and then reverse the order between the outcome and mediator to compare strength of association. More sophisticated growth models have also been developed with the specific aim to model temporal order between two (or more) variables over time in one combined model, such as the bivariate latent difference scores analysis (McArdle, 2009) and the autoregressive latent trajectory model (Bollen and Curran, 2004).

Finally, growth models can be very useful when researchers test various forms of moderated mediation and mediated moderation models (MacKinnon, 2008; MacKinnon et al., 2007). An applied example of a moderated mediator model is found in a study that compared two internet-delivered psychological treatments for tinnitus distress (Hesser et al., 2013). Using the mediator as a time-varying covariate in linear mixed effects growth model, the study showed that the treatments produced similar results on the primary outcome, but

that different processes in part explained the effects. That is, the mediated effect varied as function of a moderator variable, that is, treatment condition (for technical details see, Bauer et al., 2006). Another interesting model to consider in intervention studies is one in which the mediated effect depends on the baseline level of the mediator (MacKinnon et al., 2007). Similar to the model testing baseline moderated treatment effects, the true initial level on the mediator can be captured by the random intercept in a growth model and the change in mediator and outcome can be modeled with individual growth trajectories (i.e., parallel process growth modeling).

#### 7. Reporting results from growth models

The overall guiding principle when reporting results is that the analyses should be described in sufficient detail so other researchers are able to replicate them. This guiding principle, of course, also applies to the reporting of results from growth models. Given the flexibility of the approach, this means that researchers need to report more information about the data analytic procedures than they generally do when traditional forms of data models are used (e.g., ANOVA). In other words, it will not suffice to only report, for example, that a linear mixed effects model was conducted to evaluate the outcome of an intervention. Some researchers might find the correct reporting of growth models to be tedious and it is often a tradeoff between clarity, target group and journal page limit/reporting requirements. Still, to report too little information can be a serious problem for the particular study and the field in the long run (Jackson, 2010).

What needs to be reported will to a large extent depend on the growth model that is employed and the particular phenomenon under study. Below, I summarize some general recommendations for reporting of growth models. For more information, Jackson (2010) has provided a guide for the reporting of results from growth models in the multilevel and structural equation modeling framework.

#### 7.1. Recommendations for the reporting of results from growth models

One important aspect to report is the framework used to implement the model, that is, multilevel modeling/linear mixed effects or structural equation framework. As stated earlier, different frameworks often yield identical results; nevertheless it is important to report this so that readers can interpret the findings and replicate the analysis. It is also important to report the software and type of estimation method used. How missing data were handled and under what assumption (MCAR, MAR, or NMAR) should also be communicated clearly in the paper.

Regarding final model(s) interpreted in the results, researchers need to report the functional form of change and describe the coding of time. It is also valuable if researchers provide arguments for these choices in relation to theory of change and/or a substantive research question(s) (Jackson, 2010). The inclusion of random effects (with associated covariance structure) and the choice of error structure (e.g., whether homogenous and independent) are also important to communicate. It is often very useful to include the regression equation for the model(s) in the paper. When estimating models in structural equation modeling framework, researchers may also consider including a schematic figure of the model. Equations and figures can be very informative and be an efficient way to communicate advanced data analytic models (Jackson, 2010).

It is important to describe not only the final model(s), but also the process of model evaluation and re-specification of models during the analysis phase. This includes reporting the procedures and decisions criteria for evaluating and re-specifying models, the type of relative and global fit measures used, and the various models that were tested before selecting the final model(s). It is also important to specify whether models that were tested were determined a priori before the data were collected or were analytically determined post hoc in an exploratory fashion (Jackson, 2010).

In terms of reporting the results of the models, useful summary statistics should be included (e.g., observed and model-implied means, standard deviations, variance). If individual heterogeneity is of particular importance for the study (e.g., in a moderation/mediation study) variance associated with random effects should also be described in the paper, along with  $Pseudo-R^2$  (when appropriate) (Singer and Willett, 2003). In such studies, I would also recommend a graphical visualization of individual growth trajectories (Carrig et al., 2004), as this would aid in the interpretation of the results concerning individual differences in change and correlates of individual differences in change. It is often appropriate and useful to report the unstandardized beta coefficients of fixed effects, as this allows readers to interpret the results in the metric of the original outcome scale (e.g., mean difference between groups as function of one unit change in time). Standardized effect sizes for growth models have been developed (Feingold, 2009) and relevant effect sizes (when available) ought to be reported along with test statistics (e.g., z, t or F statistics and degrees of freedom) and confidence intervals for fixed effects. In addition, statistical power to reject a nonzero parameter of primary interest for study (e.g., parameter estimate of group difference at the end of the treatment period) needs to be reported. Power will generally depend on effect size, number of repeated-measurements, sample size (with expected attrition), and the type of data analytic model (e.g., covariance structure, number of random effects, time trend, contrasts) (Hedeker et al., 1999). These aspects, along with alpha level, should be presented when reporting power calculations for growth models.

#### 8. Concluding remarks

The primary aim of this paper was to highlight some recent methods in the analysis of change and encourage their use in the field of internet interventions. The particular details of specific advanced models were not given here. There are many excellent books that cover in great detail various forms of growth models, both within the framework of linear mixed effects models/multilevel modeling (Hedeker and Gibbons, 2006; Singer and Willett, 2003) and structural equation modeling (Bollen and Curran, 2006; Preacher et al., 2008). Readers who wish to gain a better understanding of these models can do so by consulting these books. In addition, there are specific and important topics that were not fully addressed in this overview. There are several texts that provide in-depth treatments of specific issues, such as estimation and statistical tests of parameters (Snijders and Bosker, 2012; Verbeke and Molenberghs, 2000), the evaluation of model fit (Wu et al., 2009), the statistical power of rejecting nonzero parameters in between groupdesigns with attrition (Hedeker et al., 1999), the coding and interpretation of time (Biesanz et al., 2004), and the choice of error structure (Kwok et al., 2007) and random effects structure (Barr et al., 2013).

In this paper, I have argued for the use of growth models in the analysis of change in experiments of internet interventions. These models offer the potential of answering substantive research questions regarding change, individual differences in change and correlates of individual change. The methods also have clear documented advantages for handling common problems in longitudinal data analysis, most significantly, missing data. As such, they have much to offer researchers with respect to treatment evaluation and I believe that these techniques should be seriously considered in studies on internet interventions.

The methods use, however, is also associated with some specific challenges. Growth models require that researchers reconsider the use of the traditional pre-post-treatment design, which is not optimal for these types of models. Given that this general approach to the analysis of change is very flexible, researchers are now also challenged in terms of model building, as they need to choose and argue for an appropriate data analytic model for change. This includes specifying an appropriate time trend and modeling correctly variance within individuals (i.e., error covariance structure) as well as between individuals (i.e., random effects with covariance structure). Finally, these models also require researchers to report sufficient information about models (and model evaluation) so as to be able to replicate the analyses and to interpret the results. Despite these challenges, I do believe that the advantages outweigh the potential drawbacks of using these models in applied research (e.g., underreporting, misspecifying models) and, if implemented correctly, these methods can have an immense impact for the evaluation of direct, indirect and moderated intervention effects in the field of internet interventions.

#### Acknowledgments

I would like to thank Brjánn Ljótsson and Heather Hadjistavropoulos who provided valuable comments on an early draft of this article.

#### References

- Andersson, G., 2009. Using the internet to provide cognitive behaviour therapy. Behav. Res. Ther. 47 (3), 175–180.
- Andersson, G., Ljotsson, B., Weise, C., 2011. Internet-delivered treatment to promote health. Curr. Opin. Psychiatr. 24 (2), 168–172.
- Andersson, G., Hesser, H., Veilord, A., Svedling, L., Andersson, F., Sleman, O., Carlbring, P., 2013. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J. Affect. Disord. 151 (3), 986–994.
- Andersson, G., Cuijpers, P., Carlbring, P., Riper, H., Hedman, E., 2014. Guided internetbased vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 13 (3), 288–295.
- Andrews, G., Cuijpers, P., Craske, M.G., McEvoy, P., Titov, N., 2010. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 5 (10), e13196.
- Barak, A., Klein, B., Proudfoot, J.G., 2009. Defining internet-supported therapeutic interventions. Ann. Behav. Med. 38 (1), 4–17.
- Baron, R.M., Kenny, D., 1986. The moderator-mediator variable distinction in social psychological research. Conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol. 51 (6), 1173–1182.
- Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68 (3).
- Bauer, D.J., Preacher, K.J., Gil, K.M., 2006. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. Psychol. Methods 11 (2), 142–163.
   Biesanz, J.C., Deeb-Sossa, N., Papadakis, A.A., Bollen, K.A., Curran, P.J., 2004. The role of
- Biesanz, J.C., Deeb-Sossa, N., Papadakis, A.A., Bollen, K.A., Curran, P.J., 2004. The role of coding time in estimating and interpreting growth curve models. Psychol. Methods 9 (1), 30–52.
- Bollen, K., Curran, P., 2004. Autoregressive latent trajectory (ALT) models: a synthesis of two traditions. Sociol. Methods Res. 32, 336–383.
- Bollen, K.A., Curran, P.J., 2006. Latent Curve Models: A Structural Equation Perspective. Wiley, Hoboken, NJ.
- Bryk, A.S., Raudenbush, S.W., 1987. Application of hierarchical linear models to assessing change. Psychol. Bull. 101 (1), 147–158.
- Carrig, M.M., Wirth, R.J., Curran, P.J., 2004. A SAS macro for estimating and visualizing individual growth curves. Struct. Equ. Model. 11 (1), 132–149.
- Cheong, J., MacKinnon, D.P., Khoo, S.T., 2003. Investigation of mediational processes using parallel process latent growth curve modeling. Struct. Equ. Model. 10, 238–262.
- Collins, L.M., 2006. Analysis of longitudinal data: the integration of theoretical model, temporal design, and statistical model. Annu. Rev. Psychol. 57, 505–528.
- Danaher, B.G., Seeley, J.R., 2009. Methodological issues in research on web-based behavioral interventions. Ann. Behav. Med. 38 (1), 28–39.
- Dirkse, D., Hadjistavropoulos, H.D., Hesser, H., Barak, A., 2014. Linguistic analysis of communication in therapist-assisted internet-delivered cognitive behavior therapy for generalized anxiety disorder. Cogn. Behav. Ther. 1–12.
- Duncan, T.E., Duncan, S.C., 2004. An introduction to latent growth curve modeling. Behav. Ther. 35 (2), 333–363.
- Enders, C.K., 2010. Applied Missing Data Analysis. Guilford Press, New York.
- Enders, C.K., 2011a. Analyzing longitudinal data with missing values. Rehabil. Psychol. 56 (4), 267–288.
  Enders, C.K., 2011b. Missing not at random models for latent growth curve analyses.
- Psychol. Methods 16 (1), 1–16. Fan, X., 2003. Power of latent growth modeling for detecting group differences in linear
- growth trajectory parameters. Struct. Equ. Model. 10 (3), 380–400. Feingold, A., 2009. Effect sizes for growth-modeling analysis for controlled clinical trials in
- the same metric as for classical analysis. Psychol. Methods 14 (1), 43–53. Francis, D.J., Fletcher, J.M., Stuebing, K.K., Davidson, K.C., Thompson, N.M., 1991. Analysis
- of change: modeling individual growth. J. Consult. Clin. Psychol. 59 (1), 27–37. Graham, J.W., 2009. Missing data analysis: making it work in the real world. Annu. Rev.
- Psychol. 60 (1), 549–576. 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. Arch. Gen. Psychiatry 61 (3), 310–317.
- Hadjistavropoulos, H.D., Pugh, N.E., Nugent, M.N., Hesser, H., Andersson, G., Ivanov, M., Austin, D.W., 2014. Therapist-assisted internet-delivered cognitive behavior therapy

for depression and anxiety: translating evidence into clinical practice. J. Anxiety Disord. 28, 884–893.

- Hedeker, D., Gibbons, R.D., 1997. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol. Methods 2 (1), 64–78.
- Hedeker, D., Gibbons, R.D., 2006. Longitudinal Data Analysis. Wiley, Hoboken, NJ.
- Hedeker, D., Gibbons, R.D., Waternaux, C., 1999. Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. J. Educ. Behav. Stat. 24 (1), 70–93.
- Hedman, E., 2014. Therapist guided internet delivered cognitive behavioural therapy. BMJ 348, g1977.
- Hedman, E., Ljotsson, B., Kaldo, V., Hesser, H., El Alaoui, S., Kraepelien, M., Lindefors, N., 2014. Effectiveness of internet-based cognitive behaviour therapy for depression in routine psychiatric care. J. Affect. Disord. 155, 49–58.
- Hesser, H., Gustafsson, T., Lunden, C., Henrikson, O., Fattahi, K., Johnsson, E., Andersson, G., 2012. A randomized controlled trial of internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. J. Consult. Clin. Psychol. 80 (4), 649–661.
- Hesser, H., Zetterqvist Westin, V., Andersson, G., 2013. Acceptance as a mediator in internet-delivered acceptance and commitment therapy and cognitive behavior therapy for tinnitus. J. Behav. Med. 37 (4), 756–767.
- Jackson, D.L., 2010. Reporting results of latent growth modeling and multilevel modeling analyses: some recommendations for rehabilitation psychology. Rehabil. Psychol. 55 (3), 272–285.
- Kazdin, A.E., 2007. Mediators and mechanisms of change in psychotherapy research. Annu. Rev. Clin. Psychol. 3, 1–27.
- Kwok, O.M., West, S.G., Green, S.B., 2007. The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: a Monte Carlo study. Multivar. Behav. Res. 42 (3), 557–592.
- Kwok, O.M., Underhill, A.T., Berry, J.W., Luo, W., Elliott, T.R., Yoon, M., 2008. Analyzing longitudinal data with multilevel models: an example with individuals living with lower extremity intra-articular fractures. Rehabil. Psychol. 53 (3), 370–386.
- Lane, P., 2008. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm. Stat. 7 (2), 93–106.
- Little, R., Rubin, D., 2002. Statistical Analysis With Missing Data. 2nd ed. John Wiley & Sons, Inc., New Jersey.
- Ljótsson, B., Hesser, H., Andersson, E., Lindfors, P., Hursti, T., Ruck, C., Hedman, E., 2013. Mechanisms of change in an exposure-based treatment for irritable bowel syndrome. J. Consult. Clin. Psychol. 81 (6), 1113–1126.
- Ljótsson, B., Hesser, H., Andersson, E., Lackner, J.M., El Alaoui, S., Falk, L., Hedman, E., 2014. Provoking symptoms to relieve symptoms: a randomized controlled dismantling study of exposure therapy in irritable bowel syndrome. Behav. Res. Ther. 55, 27–39.
- MacKinnon, D.P., 2008. Introduction to Statistical Mediation Analysis. Lawrence Erlbaum Associates, New York.
- MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., Sheeis, V., 2002. A comparison of methods to test mediation and other intervening variable effects. Psychol. Methods 7 (1), 83–104.
- MacKinnon, D.P., Fairchild, A.J., Fritz, M.S., 2007. Mediation analysis. Annu. Rev. Psychol. 58, 593–614.
- Mallinckrodt, C.H., Clark, W.S., David, S.R., 2001. Accounting for dropout bias using mixedeffects models. J. Biopharm. Stat. 11 (1–2), 9–21.
- Mayo, D.G., 1996. Error and the Growth of Experimental Knowledge. University of Chicago Press, Chicago.
- McArdle, JJ., 2009. Latent variable modeling of differences and changes with longitudinal data. Annu. Rev. Psychol. 60, 577–605.
- Meredith, W., Tisak, J., 1990. Latent curve analysis. Psychometrika 55 (1), 107-122.
- Molenberghs, G., Verbeke, G., 2007. Likelihood ratio, score, and Wald tests in a constrained parameter space. Am. Stat. 61 (1), 22–27.
- Molenberghs, C., Thijs, H., Jansen, I., Beunckens, C., Kenward, M.G., Mallinckrodt, C., Carroll, R.J., 2004. Analyzing incomplete longitudinal clinical trial data. Biostatistics 5 (3), 445–464.
- Muthén, B.O., Curran, P.J., 1997. General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. Psychol. Methods 2 (4), 371–402.
- Muthen, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Liao, J., 2002. General growth mixture modeling for randomized preventive interventions. Biostatistics 3 (4), 459–475.
- Nagin, D.S., Odgers, Candice L., 2010. Group-based trajectory modeling in clinical research. Annu. Rev. Clin. Psychol. 6, 109–138.
- Newby, J.M., Mackenzie, A., Williams, A.D., McIntyre, K., Watts, S., Wong, N., Andrews, G., 2013. Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. Psychol. Med. 43 (12), 2635–2648.
- Pearl, J., 2009. Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge University Press, New York.
- Peugh, J.L., Enders, C.K., 2005. Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. Educ. Psychol. Meas. 65 (5), 811.
- Preacher, K.J., Wichman, A.L., MacCallum, R.C., Briggs, N.E., 2008. Latent Growth Curve Modeling. Sage, Thousand Oaks, CA.
- Proudfoot, J., Klein, B., Barak, A., Carlbring, P., Cuijpers, P., Lange, A., Andersson, G., 2011. Establishing guidelines for executing and reporting internet intervention research. Cogn. Behav. Ther. 40 (2), 82–97.
- Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage, Newbury Park, CA.
- Ritterband, L.M., Tate, D.F., 2009. The science of internet interventions. Introduction. Ann. Behav. Med. 38 (1), 1–3.

Ritterband, L.M., Andersson, G., Christensen, H.M., Carlbring, P., Cuijpers, P., 2006. Directions for the International Society for Research on Internet Interventions (ISRII). J. Med. Internet Res. 8 (3), e23.

Rogosa, D., 1988. Myths about longitudinal research. Methodol. Issues Aging Res. 171-209.

Rogosa, D., Willett, J., 1985. Understanding correlates of change by modeling individual

Rubin, D., 1974. Estimating cusal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66 (5), 688–701.

Rubin, D.B., 1976. Inference and missing data. Biometrika 63 (3), 581–590.

Salim, A., Mackinnon, A., Christensen, H., Griffiths, K., 2008. Comparison of data analysis strategies for intent-to-treat analysis in pre-test-post-test designs with substantial dropout rates. Psychiatry Res. 160 (3), 335-345.

Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. Psychol. Methods 7 (2), 147-177.

Singer, J.D., 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J. Educ. Behav. Stat. 23 (4), 323–355.

Singer, J.D., Willett, J.B., 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York.

Sniiders, T., Bosker, R., 2012, Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, 2nd ed. Sage, London.

Spek, V., Cuijpers, P., Nyklicek, I., Riper, H., Keyzer, J., Pop, V., 2007. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol. Med. 37 (3), 319–328.

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

Völker, U., Jacobi, C., Trockel, M.T., Taylor, C.B., 2014. Moderators and mediators of outcome in Internet-based indicated prevention for eating disorders. Behav. Res. Ther. 63, 114-121.

Wampold, B.E., Serlin, R.C., 2000. The consequence of ignoring a nested factor on measures of effect size in analysis of variance. Psychol. Methods 5 (4), 425–433.

Willett, J.B., Sayer, A.G., 1994. Using covariance structure analysis to detect correlates and predictors of individual change over time. Psychol. Bull. 116, 363.

Wu, W., West, S.G., Taylor, A.B., 2009. Evaluating model fit for growth curve models: integration of fit indices from SEM and MLM frameworks. Psychol. Methods 14 (3), 183-201.