Psychotherapy Research, 2017
Vol. 27, No. 4, 425–436, http://dx.doi.org/10.1080/10503307.2015.1114688

METHOD PAPER

Partially nested designs in psychotherapy trials: A review of modeling developments

SONYA K. STERBA

Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA

(Received 26 May 2015; revised 17 September 2015; accepted 16 October 2015)

Abstract
Objective: Individually-randomized psychotherapy trials are often partially nested. For instance, individuals assigned to a treatment arm may be clustered into therapy groups for purposes of treatment administration, whereas individuals assigned to a wait-list control are unclustered. The past several years have seen rapid expansion and investigation of methods for analyzing partially nested data. Yet partial nesting often remains ignored in psychotherapy trials. **Methods**: This review integrates and disseminates developments in the analysis of partially nested data that are particularly relevant for psychotherapy researchers. **Results**: First, we differentiate among alternative partially nested designs. Then, we present adaptations of multilevel model specifications that accommodate each design. Next, we address how moderation by treatment as well as mediation of the treatment effect can be investigated in partially nested designs. Model fitting results, annotated software syntax, and illustrative data sets are provided and key methodological issues are discussed. **Conclusions**: We emphasize that cluster-level variability in the treatment arm need not be considered a nuisance; it can be modeled to yield insights about the treatment process.

Keywords: partial nesting; individually-randomized psychotherapy trials; partial clustering; group therapy; multilevel modeling

Many psychotherapy treatment outcome studies compare group therapy treatment with a wait-list or information-only control. In such studies, after individuals are randomly assigned to study arm, only individuals assigned to the treatment arm are clustered into therapy groups for the purpose of treatment administration. This type of design has been termed a *partially nested individually-randomized trial* because outcomes are clustered in a particular manner in at least one study arm but not in other study arm(s) (Bauer, Sterba, & Hallfors, 2008). In the treatment arm, outcomes (e.g., depression scores) for patients within the same therapy group can be dependent because of shared group experiences or mutual influence (e.g., Dishion, Poulin, & Burraston, 2001). In contrast, in the control arm, patients' scores on the depression outcome may be independent. Partially nested designs pose unique methodological challenges because different model-implied variances and/or effects are theoretically anticipated in the nested versus non-nested arms.

Over the past several years there has been rapid growth in methodological attention to strategies and issues involving analyzing partially nested data (e.g., Baldwin, Bauer, Stice, & Rohde, 2011; Baldwin & Fellingham, 2013; Bauer et al., 2008; Candel & van Breukelen, 2009, 2010; Chu, 2013; Hedges & Citkowicz, 2015; Korendijk, Maas, Hox, & Moerbeek, 2012; Lachowicz, Sterba, & Preacher, 2015; Lai & Kwok, 2014; Lee & Thompson, 2005a, 2005b; Lohr, Schochet, & Sanders, 2014; Luo, Cappaert, & Ning, 2015; Mehta, 2015; Moerbeek & Wong, 2008; Pals et al., 2008; Roberts & Roberts, 2005; Roberts & Walwyn, 2013; Sanders, 2011; Sterba et al., 2014; Talley, 2013; Tessler, 2014; Walwyn & Roberts, 2010). However, recent developments have not been synthesized and furthermore have appeared in the methodological, education, or medical literatures, rather than the

Correspondence concerning this article should be addressed to Sonya K. Sterba, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203, USA. Email: Sonya.Sterba@Vanderbilt.edu

psychotherapy literature. Although a monograph on partial nesting in education research was recently commissioned by the US Department of Education (Lohr et al., 2014), it did not focus on issues[1] of particular relevance to psychotherapy researchers—such as assessing moderation and mediation of treatment effects in partially nested individually-randomized trials.

In the psychotherapy literature, it is still routine for partial nesting to be ignored in empirical data analysis (e.g., Christensen, Frostholm, Ornbol, & Schroder, 2015; Conrad et al., 2015; Hedman et al., 2014; Jewell, Malone, Rose, Sturgeon, & Owens, 2015; Liu et al., 2015; Marshall et al., 2015; Sundquist et al., 2015; Teismann et al., 2014; Wesner et al., 2014; Wolff et al., 2015; Zhang, Yan, Du, & Liu, 2014). Simulations have shown that treating partially nested data as non-nested in data analyses leads to a biased standard error for the treatment effect (e.g., Baldwin et al., 2011; Hedges & Citkowicz, 2015; Korendijk et al., 2012; Lohr et al., 2014; Roberts & Roberts, 2005). It is also routine for partial nesting to be ignored in the methodological literature on testing mediation and moderation of treatment effects; for instance, Emsley, Dunn, and White's (2010) pedagogical paper on the topic states "like those [analyses] of previous authors, we make no attempt to allow for the clustering of the data" (p. 241). The purpose of this review is to motivate, integrate, and disseminate developments in the analysis of partially nested data that are particularly relevant for a psychotherapy treatment outcome audience.

The remainder of this paper proceeds as follows. First, we differentiate partially nested designs from related designs arising in individually-randomized psychotherapy trials. Second, we describe a basic model for analyzing partially nested data, and how it can be extended when partial nesting at one level occurs in conjunction with full nesting at another level. Third, we describe developments involving assessing moderation and mediation in the context of partially nested designs. Fourth, we address key methodological issues relevant to the context of partially nested designs. We limit attention to two-arm designs for simplicity, although additional nested or non-nested arms could be included (e.g., Bauer et al., 2008; Sterba et al., 2014; Walwyn & Roberts, 2010). In the Online Appendix we provide annotated *Mplus* 7.31 input syntax (Muthén & Muthén, 1998–2015) for each of the four models presented here. At the website http://www.vanderbilt.edu/peabody/sterba/ an Online Supplement will be made available containing four illustrative simulated data sets that can be used to fit these four models, using the aforementioned syntax, to produce the estimates in manuscript Tables I–IV and the test results described in sections below.

Table I. Example 1 results from fitting the Equation (1) multiple-arm partial nesting (MA-PN) model to data generated from the group therapy arm versus wait-list control arm design, in Figure 2 **Panel A**.

Parameter	Estimate	SE
$\gamma_{00}^{(t)}$	1.107	.203
$\sigma^{2(t)}$	1.254	.229
$\tau^{(t)}$	0.367	.230
$\gamma_{00}^{(c)}$	2.666	.204
$\sigma^{2(c)}$	1.865	.393

Table II. Example 2 results from fitting the Equation (2) multiple-arm partial nesting (MA-PN) model to data generated from the group therapy arm versus individual therapy arm design, in Figure 2 **Panel D**.

Parameter	Estimate	SE
$\gamma_{000}^{(t)}$	0.899	0.184
$\sigma^{2(t)}$	1.362	0.144
$\tau^{(t)}$	0.252	0.127
$\phi^{(t)}$	0.201	0.146
$\gamma_{000}^{(c)}$	2.991	0.203
$\sigma^{2(c)}$	1.520	0.358
$\phi^{(c)}$	0.066	0.189

Table III. Example 3 results from fitting the Equation (3) conditional multiple-arm partial nesting (MA-PN) model allowing for moderation by treatment, in the Figure 2 **Panel D** design.

Parameter	Estimate	SE
$\gamma_{000}^{(t)}$	1.135	0.177
$\gamma_{001}^{(t)}$	1.259	0.231
$\gamma_{010}^{(t)}$	0.447	0.202
$\gamma_{100}^{(t)}$	0.849	0.095
$\sigma^{2(t)}$	1.187	0.125
$\tau^{(t)}$	0.306	0.131
$\phi^{(t)}$	0.172	0.135
$\gamma_{000}^{(c)}$	3.198	0.236
$\gamma_{001}^{(c)}$	0.394	0.182
$\gamma_{100}^{(c)}$	1.099	0.245
$\sigma^{2(c)}$	1.279	0.303
$\phi^{(c)}$	0.209	0.235

Individually-Randomized Fully Nested Designs

Sources in the psychotherapy literature have historically recognized that individually-randomized trials can involve *fully nested designs* (e.g., Martindale, 1978). In fully nested designs the same nesting structure exists in all study arms. Data from such designs can be analyzed with conventional multilevel models. In one fully nested design, depicted in Figure 1 **Panel A**, two kinds of individual therapies are compared. In each arm, patient outcomes may be

Table IV. Example 4 results from fitting the Equation (4) multivariate multiple-arm partial nesting (MA-PN) model allowing for mediation of the treatment effect, in the Figure 2 **Panel A** design.

Parameter	Estimate	SE
$\gamma_{00}^{(yt)}$	1.753	.313
$\gamma_{01}^{(y)}$	0.469	.136
$\gamma_{10}^{(yt)}$	−0.546	.187
$\gamma_{00}^{(mt)}$	1.545	.271
$\sigma^{2(yt)}$	1.620	.296
$\sigma^{2(mt)}$	0.770	.141
$\tau^{(yt)}$	0.477	.298
$\tau^{(mt)}$	0.947	.403
$\gamma_{00}^{(yc)}$	2.748	.230
$\gamma_{01}^{(y)}$	0.469	.136
$\gamma_{00}^{(mc)}$	0.789	.175
$\sigma^{2(yc)}$	1.856	.391
$\sigma^{2(mc)}$	1.374	.290

nested[2] (i.e., correlated) within therapist because therapists have varying amounts of experience or therapists are not using a manual, for instance (Crits-Christoph & Mintz, 1991; Wampold & Brown, 2005). Viewing therapists as representative of a wider population of therapists (e.g., Kim, Wampold, & Bolt, 2006; Roberts, 1999; Serlin, Wampold, & Levin, 2003), therapist could be specified as a random effect in a conventional two-level model for data from this design (i.e., a random intercept varying across therapists and a fixed slope of treatment). Despite a history of attention to fully nested designs in individually-randomized psychotherapy trials, Walwyn and Roberts (2010) state that conventional multilevel models are still underapplied in this context, because "while psychotherapy researchers readily recognize that average patient

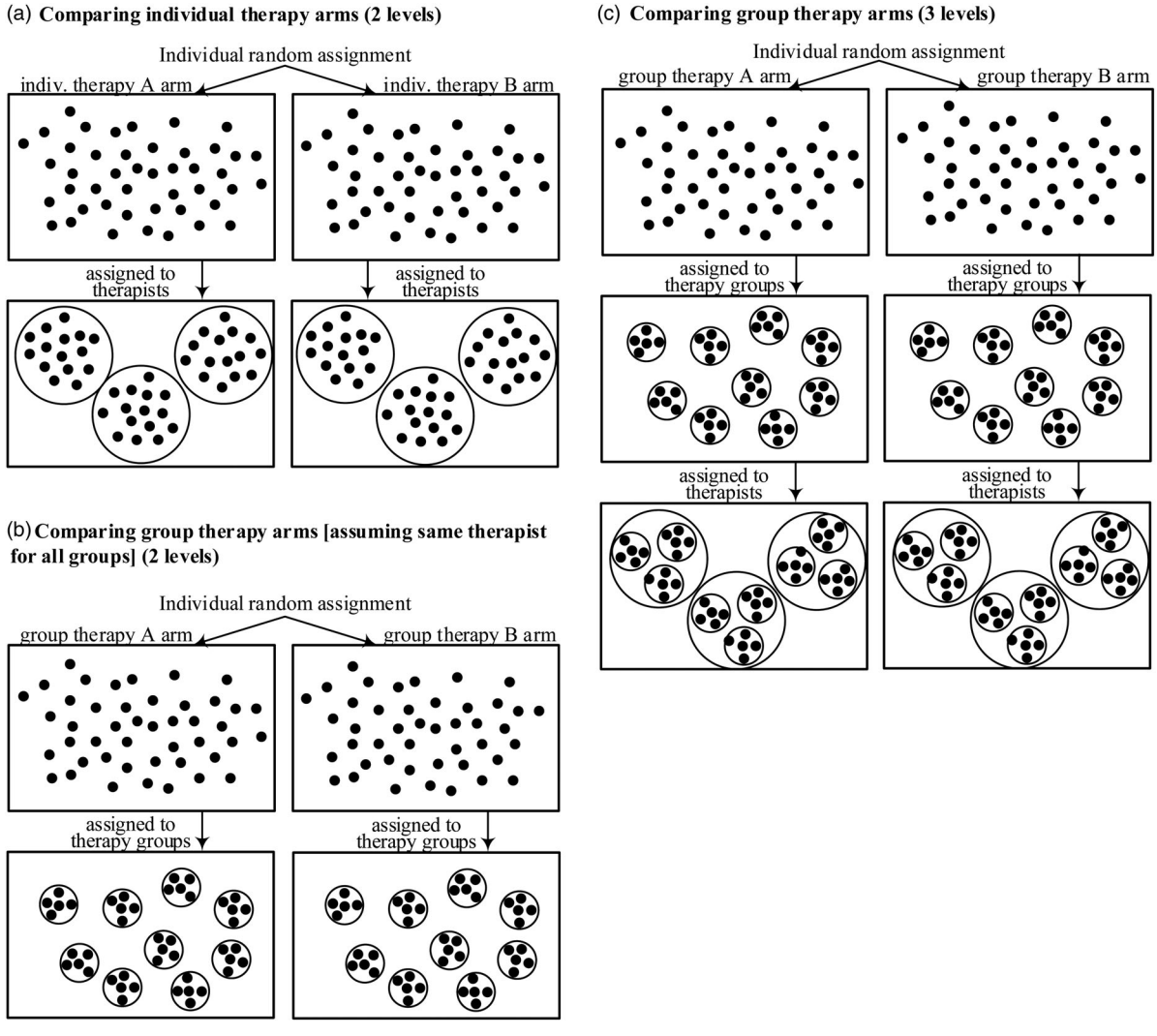

Figure 1. Alternative fully nested designs for individually-randomized trials.

outcomes may vary between therapists, they infrequently appreciate that this is equivalent to patient outcomes being *clustered* within individual therapists leading to intra-therapist correlation" (p. 292).

In another fully nested design, individuals are clustered into therapy groups in *each* study arm in order to compare group therapies (Figure 1 **Panel B**, e.g., Lee & Thompson, 2005a; Pals et al., 2008). Viewing therapy groups as representative of a wider population of therapy groups (e.g., Baldwin, Murray, & Shadish, 2005; Rohde, 2008), therapy group could be considered a random effect in a conventional two-level model for data from this design (i.e., a random intercept varying across therapy groups and a fixed slope of treatment). If certain therapy groups share the same therapist (as in Figure 1 **Panel C**), a conventional three-level model may be considered (e.g., a random intercept varying across therapists and therapy groups; a fixed slope of treatment).

Individually-Randomized Partially Nested Designs

Here we adapt the conventional multilevel model to accommodate partially nested design variants of the fully nested designs presented in Figure 1. First we consider the most basic partially nested design where there is dependence among outcomes within cluster in one arm, but independence among outcomes in the other arm. The *same* model can be used whether we are comparing an unclustered (e.g., wait-list only) control arm to the treatment arm in Figure 2 **Panel A** or to the treatment arm in Figure 2 **Panel B**. The treatment arm in Figure 2 **Panel A** involves group therapy, with patients (level 1 units) clustered within therapy groups (level 2 units). The treatment arm in Figure 2 **Panel B** involves individual therapy, with patients (level 1 units) clustered within therapists (level 2 units).

The adapted multilevel model initially used to accommodate these designs employed a *fixed* intercept but *random* slope of treatment (where treatment was coded 1 for treated and 0 for control) along with heteroscedastic level 1 residual variances across study arm (e.g., Bauer et al., 2008; Lee & Thompson, 2005b; Moerbeek & Wong, 2008; Roberts & Roberts, 2005). This specification implies that a random effect is toggled into the reduced form model for the treatment arm (where treatment = 1) but toggled out of the reduced form model for the control arm (where treatment = 0). As such, the random effect accommodates dependency among scores within cluster *only* in the treatment arm.

Instead, here we specify an *equivalent* model using a different approach: a *multiple-arm*[3] *multilevel model for partial nesting* (MA-PN) (Lachowicz et al., 2015; Lohr et al., 2014; Sterba et al., 2014). We will later see that this approach poses benefits for ease of extension to more advanced specifications (including investigating moderation of predictor effects by treatment in partially nested designs or investigating mediators of treatment effects in partially nested designs). This approach also readily accommodates heteroscedastic variances across study arm, which is a typical consequence of partially nested designs.

Multiple-Arm Partial Nesting (MA-PN) Models

MA-PN Model for a Two/One Level Combination Design

The MA-PN model for a two-level treatment arm and single-level control arm design is represented in Equation (1). Superscript t indicates treatment arm, superscript c indicates control arm, subscript i indicates individual, and subscript j indicates cluster. Note that, although it is not necessary to have a j subscript for the outcome y_{ij} in the control arm, the j subscript can be retained if one conceptually considers each person to be their own cluster.

$$\begin{array}{lll} & \text{Group therapy arm:} & \text{Control (wait} - \text{list) arm:} \\ \text{Level 1:} & y_{ij} = \beta_{0j}^{(t)} + r_{ij}^{(t)} & y_{ij} = \gamma_{00}^{(c)} + r_{ij}^{(c)} \\ \text{Level 2:} & \beta_{0j}^{(t)} = \gamma_{00}^{(t)} + u_{0j}^{(t)} & \\ \text{Where:} & r_{ij}^{(t)} \sim N(0, \sigma^{2(t)}) & r_{ij}^{(c)} \sim N(0, \sigma^{2(c)}) \\ & u_{0j}^{(t)} \sim N(0, \tau^{(t)}) & \end{array} \quad (1)$$

In Equation (1), $\gamma_{00}^{(t)}$ is the treatment-arm mean and $\gamma_{00}^{(c)}$ is the control-arm mean. In the treatment arm, the mean for cluster j, $\beta_{0j}^{(t)}$, deviates from the treatment arm mean, $\gamma_{00}^{(t)}$, by the cluster-specific residual, $u_{0j}^{(t)}$. In the treatment arm, the outcome variance is decomposed into a sum of between-cluster variance, $\tau^{(t)}$ (allowing for variation in residuals $u_{0j}^{(t)}$ across clusters) and within-cluster variance, $\sigma^{2(t)}$ (allowing for variation in residuals $r_{ij}^{(t)}$ within-cluster). In the control arm, the variance is simply $\sigma^{2(c)}$ (allowing for variation in residuals $r_{ij}^{(c)}$ across individuals). Hence, the model-implied variances are heteroscedastic across study arm. Table I provides estimates of these parameters for our first simulated example; the associated dataset is available in the Online Supplement [equation1.dat] and the associated syntax is available on p. 1 of the Online Appendix. This first example involves a clustered

depression treatment arm (with 15 therapy groups of size 5) and an unclustered wait-list control arm (with 45 singletons). The intra-class correlation (ICC) in the treatment arm is computed as $\tau^{(t)}/(\tau^{(t)} + \sigma^{2(t)})$. For our first example, in Table I, the ICC in the treatment arm is .367/(.367 + 1.254) = .226. The treatment effect is $\gamma_{00}^{(t)} - \gamma_{00}^{(c)}$. This effect can be tested using a likelihood ratio test (LRT). The LRT statistic is $-2(\text{LL}_{\text{restricted}} - \text{LL}_{\text{unrestricted}})$ where $\text{LL}_{\text{unrestricted}}$ is the log-likelihood for the Equation (1) model and $\text{LL}_{\text{restricted}}$ is the log-likelihood for a model imposing the restriction $\gamma_{00}^{(t)} = \gamma_{00}^{(c)}$. This test statistic is compared to a reference $\chi^2(df = 1)$ distribution. In our first example, this LRT statistic = $-2(-210.068 - -199.555) = 21.026(df = 1)$, $p < .001$, indicating a significant treatment effect.

MA-PN Model for a Three/Two Level Combination Design

Now we consider modeling data from more complex partially nested designs. Suppose a group therapy treatment is being compared to an individual therapy condition. Further suppose that, in the group therapy treatment arm, multiple therapy groups are led by the same therapist. Also suppose that, in the individual therapy arm, each therapist has a case load of multiple individual patients. As depicted in Figure 2 **Panel D** this design has three levels of nesting in the group therapy arm and two levels of nesting in the individual therapy condition (as in Compas et al., 2015; Walwyn & Roberts, 2010).[4] The MA-PN model to accommodate this design (e.g., Lohr et al., 2014) is given in Equation (2), where the t superscript designates the group therapy arm and the c superscript now designates the individual therapy arm. The k subscript designates therapist.

$$\begin{array}{lll} & \text{Group therapy arm:} & \text{Individual therapy arm:} \\ \text{Level 1:} & y_{ijk} = \beta_{0jk}^{(t)} + r_{ijk}^{(t)} & y_{ijk} = \beta_{00k}^{(c)} + r_{ijk}^{(c)} \\ \text{Level 2:} & \beta_{0jk}^{(t)} = \beta_{00k}^{(t)} + u_{0jk}^{(t)} & \\ \text{Level 3:} & \beta_{00k}^{(t)} = \gamma_{000}^{(t)} + u_{00k}^{(t)} & \beta_{00k}^{(c)} = \gamma_{000}^{(c)} + u_{00k}^{(c)} \\ \text{Where:} & r_{ijk}^{(t)} \sim N(0,\, \sigma^{2(t)}) & r_{ijk}^{(c)} \sim N(0,\, \sigma^{2(c)}) \\ & u_{0jk}^{(t)} \sim N(0,\, \tau^{(t)}) & \\ & u_{00k}^{(t)} \sim N(0,\, \phi^{(t)}) & u_{00k}^{(c)} \sim N(0,\, \phi^{(c)}) \end{array} \quad (2)$$

Note that individual-specific residuals are still heteroscedastic across arms ($\sigma^{2(t)}$ vs. $\sigma^{2(c)}$). Now variability across therapists is estimated and is also allowed to be heteroscedastic across study arm ($\phi^{(t)}$ vs. $\phi^{(c)}$). Table II provides estimates of the Equation (2) parameters for our second simulated example; the associated dataset is available in the Online Supplement [equation2.dat] and the associated syntax is on p. 2 of the Online Appendix. In this second example, nine therapists in the treatment arm each lead five therapy groups (with five patients each) and nine therapists in the individual treatment arm each have five individual patients. Though it happens to be the case in this simulated example, the MA-PN model does not require an equal number of patients per therapist, patients per therapy group, or therapy groups per therapist. It is possible to test whether the amount of therapist-level variability is equal across group therapy vs. individual therapy arms using a LRT comparing the Equation (2) model with a restricted model that constrains $\phi^{(t)} = \phi^{(c)}$. In our second example, this test indicates that the amount of therapist-level variability does not differ across arms: LRT statistic = $-2(-447.886 - -447.744) = .284$ $(df = 1)$, $p > .05$.

Note that if *different* therapists administer treatment across arms then the covariance between therapist-level residuals $u_{00k}^{(t)}$ and $u_{00k}^{(c)}$ can be constrained to 0. But if therapists are "crossed" with treatments (see Kazdin, 1986) covariance among these level-3 residuals would need to be estimated (this covariance is here denoted $\phi^{(t,c)}$) (Walwyn & Roberts, 2010). In Equation (2), the therapy-group level variance (level-2 variance), $\tau_{00}^{(t)}$, is still estimated only in the group therapy treatment arm. The ICC for therapy groups (proportion of outcome variability attributable to therapy groups) is now calculated as $\tau^{(t)}/(\tau^{(t)} + \phi^{(t)} + \sigma^{2(t)})$, which is .252/(.201 + .252 + 1.362) = .139 for our second example, from Table II. The ICC for therapists (proportion of outcome variability attributable to therapists) in the group therapy arm is $\phi^{(t)}/(\tau^{(t)} + \phi^{(t)} + \sigma^{2(t)})$ and in the individual therapy arm is $\phi^{(c)}/(\phi^{(c)} + \sigma^{2(c)})$; these ICCs are .201/(.201 + .252 + 1.362) = .111 and .066/(.066 + 1.520) = .042, respectively in our second example, from Table II. Though we did not consider cluster randomized trials in this review (see Footnote 1), the Equation (2) MA-PN could be employed if, for instance, intact schools (level 3 units) had been randomly assigned, study/learning groups (level 2 units) were constructed in only one arm of the design, and groups were led by the same leader.

MA-PN for a Three/One Level Combination Design

To accommodate a design where this group-therapy treatment is instead compared to a wait-list only control condition (i.e., unclustered control arm) as in Figure 2 **Panel C**, we need to make only one

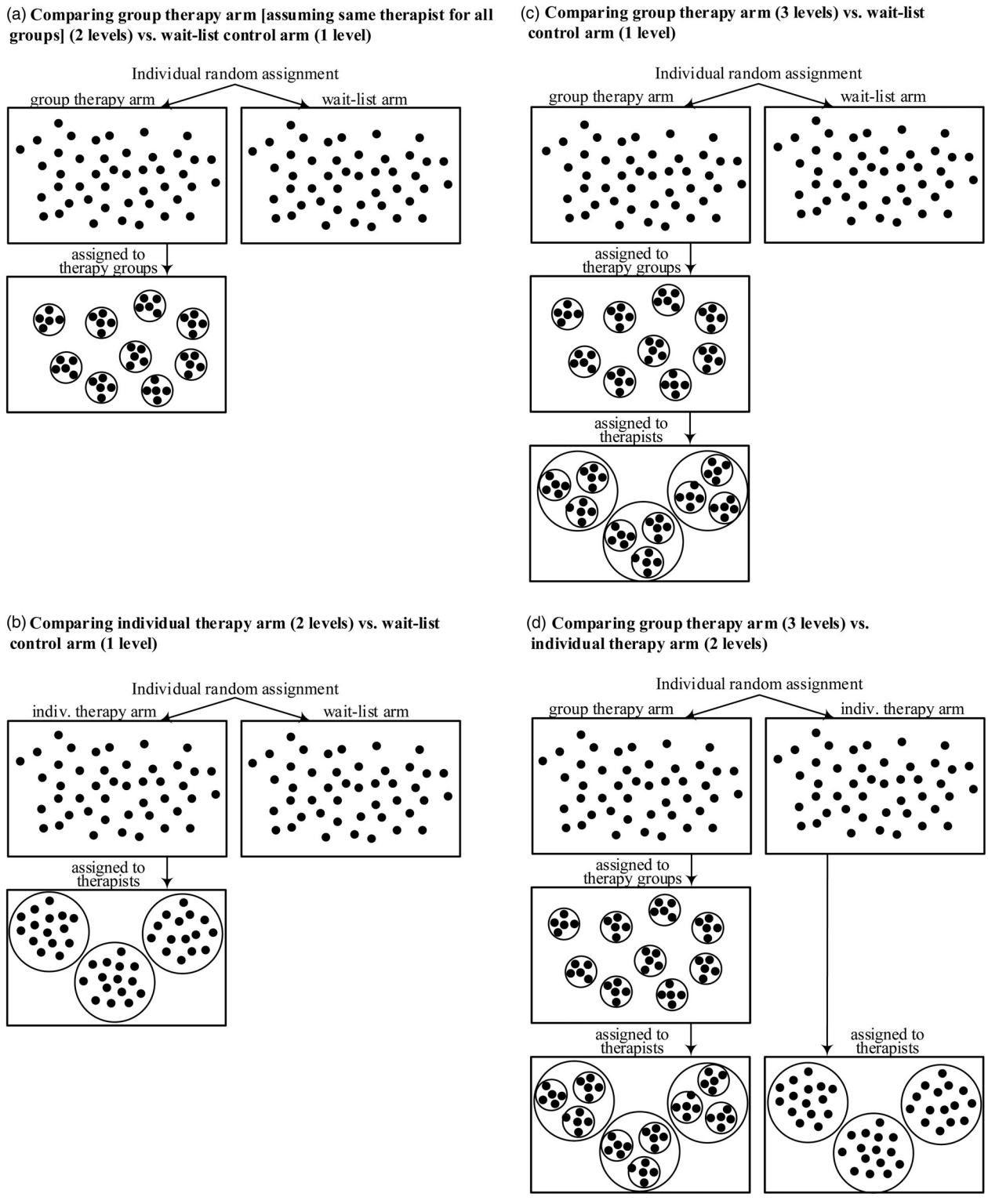

Figure 2. Alternative partially nested designs for individually-randomized trials.

modification to the Equation (2) model. Specifically, we fix $\phi^{(c)} = 0$. With this modification, the MA-PN model would then have three levels in the treatment arm and one level in the control arm (as in Tessler, 2014). Simulation results indicate that ignoring therapist-level variation in a partially nested design (i.e., fitting the Equation (1) model when the design is consistent with the Equation (2) model) biases the standard error of the treatment effect, artificially inflates power, as well as inflates estimates of

group-therapy-level variability in the treatment arm and individual-level variability in the control arm (Tessler, 2014).

Effect Sizes

There are two main options for constructing a standardized effect size for the treatment effect in partially nested designs (Hedges & Citkowicz, 2015).[5] If a researcher wants the treatment effect size to be measured in terms of within-cluster standard deviations as computed under no manipulation, this leads to $\left|\gamma_{00}^{(t)} - \gamma_{00}^{(c)}\right| / \sqrt{\sigma_c^2}$. This option yields a treatment effect size of $1.559/\sqrt{1.865} = 1.142$ in the context of our first example, from Table I. This option is sensible when the control arm is considered the "natural" comparison situation, and clustering is induced by treatment administration (e.g., Fuchs, Sterba, Fuchs, & Malone, under review; Hedges & Citkowicz, 2015). Another option is to instead divide by the standard deviation in the group therapy treatment arm. For instance, using the Equation (2) model in conjunction with the design in Figure 2 **Panel D**, the effect size is $\left|\gamma_{000}^{(t)} - \gamma_{000}^{(c)}\right| / \sqrt{\tau^{(t)} + \phi^{(t)} + \sigma_t^2}$. According to Hedges and Citkowicz (2015) this option is sensible when the group therapy treatment administration is the typical or established approach, and it is being compared to an alternative individual administration of treatment.

Assessing Moderation by Treatment in Partially Nested Designs

Psychotherapy researchers may be interested in including predictors of within-cluster or between-cluster variation in outcomes in the treatment arm, and predictors of outcome variation in the control arm (e.g., Compas et al., 2009). In this section, we expand the Equation (2) MA-PN model to include such predictors. We also explain how, using the MA-PN model, it is straightforward to examine interactions of therapist-level predictors with treatment and interactions of individual-level predictors with treatment; doing so does not require constructing and including product terms (see Sterba et al., 2014). We do not discuss interactions of therapy-group-level predictors with treatment because therapy-group-level predictors will often be missing by design in the individual therapy arm (see later discussion).

In Equation (3) we expand the Equation (2) MA-PN to include fixed slopes of therapist-level, therapy-group level, and individual-level predictors.[6] Below, we discuss each in turn. Table III provides estimates of the Equation (3) parameters for our third simulated example (the associated dataset is available in the Online Supplement [equation3.dat] and the associated syntax is available on p. 3 of the Online Appendix). This Table III example uses the same sample sizes and design as in the Table II example.

$$
\begin{array}{lll}
 & \text{Group therapy arm:} & \text{Individual therapy arm:} \\
\text{Level 1:} & y_{ijk} = \beta_{0jk}^{(t)} + \gamma_{100}^{(t)}(x_{ijk} - \bar{x}_{.jk}) + r_{ijk}^{(t)} & y_{ijk} = \beta_{00k}^{(c)} + \gamma_{100}^{(c)} x_{ijk} + r_{ijk}^{(c)} \\
\text{Level 2:} & \beta_{0jk}^{(t)} = \beta_{00k}^{(t)} + \gamma_{010}^{(t)} \bar{x}_{.jk} + u_{0jk}^{(t)} & \\
\text{Level 3:} & \beta_{00k}^{(t)} = \gamma_{000}^{(t)} + \gamma_{001}^{(t)} w_k + u_{00k}^{(t)} & \beta_{00k}^{(c)} = \gamma_{000}^{(c)} + \gamma_{001}^{(c)} w_k + u_{00k}^{(c)} \\
 & \text{Where: } r_{ijk}^{(t)} \sim N(0, \sigma^{2(t)}) & r_{ijk}^{(c)} \sim N(0, \sigma^{2(c)}) \\
 & u_{0jk}^{(t)} \sim N(0, \tau^{(t)}) & \\
 & u_{00k}^{(t)} \sim N(0, \phi^{(t)}) & u_{00k}^{(c)} \sim N(0, \phi^{(c)})
\end{array}
\tag{3}
$$

Equation (3) includes a therapist-level predictor, therapist experience, denoted w_k. Researchers may want to include therapist-level predictors to control for pre-existing differences between therapists; this can improve power for detecting a non-zero treatment effect by reducing between-therapist variation (see simulation results in Raudenbush, 1997). Equation (3) also allows for an interaction of therapist experience × treatment. The $\gamma_{001}^{(t)}$ and $\gamma_{001}^{(c)}$ are simple slopes of therapist experience in each study arm—obtained automatically from the model specification without needing to do post-hoc probing. In our third example, from Table III, both simple slopes are significantly different from 0: $\hat{\gamma}_{001}^{(t)} = 1.259$ (.231), $p < .001$ and $\hat{\gamma}_{001}^{(c)} = 0.394$ (.182), $p = .031$. In order to instead include a main effect of therapist experience, we would constrain $\gamma_{001}^{(t)} = \gamma_{001}^{(c)}$ in Equation (3). In order to *test* whether treatment moderates the effect of therapist experience (i.e., whether these two simple slopes are different from *each other*), we could use a LRT comparing models with and without the constraint $\gamma_{001}^{(t)} = \gamma_{001}^{(c)}$. In our third example, this LRT indicates that there is an interaction between therapist experience and treatment (LRT = $-2(-436.92 - -433.472) = 6.896$ (d$f = 1$), $p < .05$). Therapist experience has a stronger effect on depression outcomes in the group therapy arm than the individual therapy arm.

Equation (3) also includes a therapy-group level predictor in the treatment arm (therapy-group-average level of cognitive functioning, $\bar{x}_{.jk}$), whose effect is $\gamma_{010}^{(t)}$. Researchers may want to include therapy-group level predictors (e.g., therapy group size, proportion of the group that is female, proportion of the group with a conduct disorder diagnosis, therapy-group-average age, or therapy-group-average level of cognitive functioning) to help

explain therapy group variation in the treatment arm only. For instance, lower therapy-group-average cognitive functioning may be associated with worse therapy-group-average depression scores if it implies limited sophistication in processing and insights. Therapy group size may also have a negative effect on average depression outcomes in the therapy group, if it implies less intimacy and disclosure. Bauer et al. (2008) and Sterba et al. (2014) discuss a data management step to prevent listwise deletion of individuals in the unclustered arm when therapy-group-level predictors are missing by design for these individuals.

Equation (3) also includes an individual-level (level 1) predictor, individual cognitive functioning, in each arm. In the individual therapy arm of the Equation (3) MA-PN model, $\gamma_{100}^{(c)}$ is the effect of individual cognitive functioning (x_{ijk}). In the group therapy arm, $\gamma_{100}^{(t)}$ is the effect of therapy-group-mean centered individual cognitive functioning ($x_{ijk} - \bar{x}_{.jk}$). Specifically, $\gamma_{100}^{(t)}$ is the effect of the *deviation* of person *i*'s cognitive functioning from their therapy-group-average cognitive functioning on the *deviation* of person *i*'s depression score from their therapy-group-average depression score. In order to test whether the effect of cognitive functioning is *moderated by treatment*, we can use a LRT to compare a model imposing the constraint $\gamma_{100}^{(t)} = \gamma_{010}^{(t)}$ with a model imposing the constraint $\gamma_{100}^{(t)} = \gamma_{010}^{(t)} = \gamma_{100}^{(c)}$. We do not need to construct or include product terms. In our third example, from Table III, we do not find evidence of an interaction between cognitive functioning and treatment (LRT = −2(−435.820 − −435.065) = 1.510 ($df = 1$), $p > .05$).

Assessing Mediators of Treatment Effects in Partially Nested Designs

In treatment outcome studies interest often focuses on what variables, such as coping skills, mediate the effect of treatment on outcomes, such as depression (e.g., Christensen et al., 2015; Liu et al., 2015). This can be investigated in partially nested designs as well (e.g., Chu, 2013; Lachowicz et al., 2015). To illustrate, we expand the MA-PN model in Equation (1) to include a coping skills mediator, denoted m_{ij}, in Equation (4). Superscripts *y* and *m* in Equation (4) pertain to coefficients in the y_{ij} or m_{ij} equations. In the group therapy arm, the effect of the mediator can be split into the effect $\gamma_{01}^{(y)}$ of therapy-group average coping skills, $\bar{m}_{.j}$, and the effect $\gamma_{10}^{(yt)}$ of an individual's deviation from the therapy group average coping skills ($m_{ij} - \bar{m}_{.j}$) (e.g., Zhang, Zyphur, & Preacher, 2009), as shown in Equation (4). Receiving treatment may improve depression scores via improving average coping skills. Note that because treatment does not vary within cluster it can affect m_{ij} and $\bar{m}_{.j}$ but not ($m_{ij} - \bar{m}_{.j}$).

$$
\begin{aligned}
&\text{Group therapy arm:}\\
&\text{Level 1:}\quad y_{ij} = \beta_{0j}^{(yt)} + \gamma_{10}^{(yt)}(m_{ij} - \bar{m}_{.j}) + r_{ij}^{(yt)}\\
&\qquad\quad m_{ij} = \beta_{0j}^{(mt)} + r_{ij}^{(mt)}\\
&\text{Level 2:}\quad \beta_{0j}^{(yt)} = \gamma_{00}^{(yt)} + \gamma_{01}^{(y)}\bar{m}_{.j} + u_{0j}^{(yt)}\\
&\qquad\quad \beta_{0j}^{(mt)} = \gamma_{00}^{(mt)} + u_{0j}^{(mt)}\\
&\text{Where: } \begin{bmatrix} r_{ij}^{(yt)} \\ r_{ij}^{(mt)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^{2(yt)} & \\ 0 & \sigma^{2(mt)} \end{bmatrix}\right)\\
&\qquad \begin{bmatrix} u_{0j}^{(yt)} \\ u_{0j}^{(mt)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^{(yt)} & \\ 0 & \tau^{(mt)} \end{bmatrix}\right)
\end{aligned}
\qquad (4)
$$

$$
\begin{aligned}
&\text{Control (wait − list) arm:}\\
&y_{ij} = \gamma_{00}^{(yc)} + \gamma_{01}^{(y)} m_{ij} + r_{ij}^{(yc)}\\
&m_{ij} = \gamma_{00}^{(mc)} + r_{ij}^{(mc)}\\
&\begin{bmatrix} r_{ij}^{(yc)} \\ r_{ij}^{(mc)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^{2(yc)} & \\ 0 & \sigma^{2(mc)} \end{bmatrix}\right)
\end{aligned}
$$

Table IV provides estimates of the Equation (4) parameters for our fourth simulated example (the associated dataset is available in the Online Supplement [equation4.dat] and the associated syntax is on p. 4 of the Online Appendix). This Table IV example uses the same sample sizes and design as in the Table I example.

The *indirect effect* (mediation effect) is computed as $(\gamma_{00}^{(mt)} - \gamma_{00}^{(mc)}) \times \gamma_{01}^{(y)}$. Following Sterba et al. (2014) and Lachowicz et al. (2015) a confidence interval for this indirect effect can be obtained using the Monte Carlo procedure described in Preacher and Selig (2012) and implemented in an online utility (Selig & Preacher, 2008). In our fourth example, from Table IV, this indirect effect is (1.675 − .849) × .696 = .575, and a Monte Carlo CI_{95} = (.045, .775) indicates a significant indirect effect of treatment on depression outcomes, through coping skills. Note that we could also consider variables other than coping skills as a mediator (see Lachowicz et al., 2015) and/or we could allow the effect of coping skills on depression to vary randomly across therapy groups, in the treatment arm (see Chu, 2013).

Finally, to investigate one kind of *moderated mediation*, we could compare the Equation (4) model to a model in which the across-arms equality constraint on $\gamma_{01}^{(y)}$ is relaxed, using a LRT. For our fourth example, from Table IV, this LRT statistic is −2(−391.450 − −391.449) = .002 ($df = 1$), $p > .05$.

Hence, we lack evidence that the indirect effect differs across treatment vs. control arms.

Methodological Issues to Consider When Analyzing Data from Partially Nested Trials

When to Consider Using the MA-PN Model

Previous authors have debated whether the MA-PN model should be a default analysis approach for partially nested designs. One possibility that has been raised is to consider a statistically significant ICC to be a prerequisite for fitting MA-PN models. However, this possibility has statistical drawbacks. This use of the test of ICC = 0 reverses the reject-support logic of hypothesis testing and can be underpowered for small ICCs that are still capable of meaningfully inflating Type I error rates (Baldwin et al., 2011; Kenny, Mannetti, Pierro, Livi, & Kashy, 2002; Lee & Thompson, 2005b; Roberts & Roberts, 2005; Walwyn & Roberts, 2010; relatedly see Crits-Christoph & Mintz, 1991). Even when ICC = 0, simulations have found little decrement in power associated with using MA-PN models so long as the number of clusters was not small (such as $\leq$5 clusters in the treatment arm; Baldwin et al., 2011). Also, when ICC = 0, simulations have found little bias associated with using MA-PN models so long as boundary constraints were not placed on variances by the estimation method (e.g., Tessler, 2014). When the number of clusters in the treatment arm is small, treating therapists as fixed effects (see Baldwin et al., 2011 for review) remains an option, but limits generalizations to clusters in the study (see Serlin et al., 2003). To ensure sufficient clusters in the treatment arm to investigate and explain between-cluster variation, a priori power analyses are essential. For secondary data analyses of existing partially nested trials, on balance the strategy of fitting a MA-PN model as a default where estimable—without requiring a statistically significant ICC—seems to balance risks and benefits. This strategy is consistent with recommendations from methodologists concerning fully nested randomized controlled trials (e.g., Kenny et al., 2002; Murray, Hannan, & Baker, 1996; Roberts & Roberts, 2005; Rohde, 2008).

Power Analysis for MA-PN Models

Many individually-randomized psychotherapy trials use one or two therapy groups and one or two therapists—severely limiting researchers' ability to decompose variability into between-cluster and within-cluster components (for reviews, see Crits-Christoph & Mintz, 1991; Rohde, 2008). Several studies have investigated power for partially nested designs, either deriving formulas for optimal allocation of subjects for particular designs or providing simulation results showing how power changes as design factors are manipulated (Baldwin et al., 2011; Candel & van Breukelen, 2009, 2010; Lohr et al., 2014; Moerbeek & Wong, 2008; Roberts & Roberts, 2005; Sanders, 2011; Tessler, 2014; Walwyn & Roberts, 2010). Overall findings indicate that, for a fixed total sample size, power is more a function of the number of clusters in the treatment arm than either the cluster size in the treatment arm (necessarily small for therapy groups) or number of singletons in the control arm. Moreover, power decreases with increasing ICC and decreases slightly with unbalanced cluster sizes. Monte Carlo power analysis procedures (e.g., Muthén & Muthén, 2002) allow researchers to flexibly assess power under any generating MA-PN model; see also Roberts (2008).

Estimation Methods and Software for Fitting MA-PN Models

MA-PN models are most commonly fit with maximum likelihood (or restricted maximum likelihood) estimation under missing at random assumptions for outcomes. Researchers are encouraged to include, as covariates, any predictors that may be associated with differential patterns of missing data across individuals, therapists, or therapy-groups (e.g., therapy group cohesion as a level 2 predictor potentially related to attrition in the treatment arm). Basic MA-PN models can be fit using conventional multilevel software packages (e.g., SAS Proc Mixed, R, HLM, MLwiN; see syntax in Baldwin et al., 2011; Bauer et al., 2008; Lohr et al., 2014; Sanders, 2011). However, more general software (e.g., M*plus*) allows certain extensions of the MA-PN model that would be cumbersome (e.g., the partially nested multivariate multilevel model in Equation (4)) or impossible (e.g., partially nested multilevel structural equation models) to specify in conventional multilevel software (see syntax in Lachowicz et al., 2015; Sterba et al., 2014).

Note that these software packages differ in the reference distribution used for testing fixed effects. Test statistics for fixed effects (e.g., the treatment effect) often are computed using the z reference distribution (e.g., Stata, M*plus*). Particularly when the number of clusters in the clustered arm is < 8, there are advantages to testing fixed effects using t-tests with approximated degrees of freedom (e.g., see the Kenward-Roger method available only in SAS) when fitting applicable MA-PN models (Baldwin et al., 2011).

Bayesian methods have also been employed to fit MA-PN models (e.g., Baldwin & Fellingham, 2013; Chu, 2013; Lee & Thompson, 2005b). In simulations comparing likelihood and Bayesian methods in this context, Baldwin and Fellingham (2013) found that Bayesian methods can yield more bias but greater efficiency for estimating variance components, under carefully chosen priors. Under diffuse priors, Baldwin and Fellingham (2013) found Bayesian methods to be less efficient for estimating these parameters. These differences across methods were more pronounced at smaller sample sizes.

Conclusions

This review integrated cutting-edge developments in modeling partially nested data and highlighted those of particular relevance to psychotherapy researchers. We began by differentiating variants of individually-randomized fully nested designs vs. partially nested designs. We then presented several model specifications that allow psychotherapy researchers to explain variability at the individual level, therapy-group level (where present), and therapist-level; assess treatment effects; investigate moderation by treatment; and investigate mediators of the treatment effect. For illustration, the Online Appendix contains annotated software syntax for fitting the models in Equations (1)–(4), and the Online Supplement contains four simulated datasets that were generated with the respective nesting structures analyzed in Equations (1)–(4) in order to produce the empirical model fitting results in Tables I–IV. Finally, we reviewed recent work on effect size, power, and estimation methods in the context of partially nested data.

This review gives psychotherapy researchers scaffolding for evaluating and employing new and different kinds of models in the context of partial nesting. Recent examples not discussed here specifically include partially nested cross-classification models (Lai & Kwok, 2014; Luo et al., 2015) and partially nested multiple membership models (Roberts & Walwyn, 2013). In future research, partial nesting could also be accommodated in models for rolling group therapy trials (Andridge, Shoben, Muller, & Murray, 2014; Bauer, Gottfredson, Dean, & Zucker, 2012; Tasca, Balfour, Ritchie, & Bissada, 2007), by allowing for time-varying group effects only in the treatment arm. Psychotherapy researchers are encouraged to consider the sources of full and partial nesting that arise in individually-randomized trials. Therapist-level and therapy-group-level variability need not be viewed as a methodological nuisance; they can be investigated and predicted in order to yield substantive insights about the treatment process in individually-randomized trials.

Supplemental data

Supplemental data for this article can be accessed at doi:10.1080/10503307.2015.1114688.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes

[1] For instance, Lohr et al. (2014) considered partially nested *cluster randomized trials*. Partially nested designs can arise in cluster randomized trials (e.g., Cornfield, 1978), such as when pre-existing clusters (e.g., schools) are randomly assigned to treatment or control arms and then tutoring/study groups are formed within schools in the treatment arm. Since partially nested cluster randomized trials are more common in education and public health settings, rather than clinical settings, they are not discussed here.

[2] For the present discussion it is not relevant whether full nesting of patients within therapist arises from an arrangement where therapists are "nested" within study arm or "crossed" with study arm (e.g. Kazdin, 1986). Either arrangement could exist in a "fully nested" design and each can be accommodated by slightly different specifications of a conventional two-level multilevel model (see later discussion).

[3] Note that the term multiple-group multilevel model is more common for related models (e.g., Asparouhov & Muthén, 2012) but we use the term "multiple-arm" because in this paper "group" refers to therapy group.

[4] If very few therapists delivered treatment, estimating outcome variability at the therapist level in Equation (2) could be unreliable and estimation problems could be encountered (Tessler, 2014). See later section on power analysis.

[5] Note that in Hedges and Citkowicz (2015) residual variances were assumed homoscedastic across study arm, which is not assumed here in this section.

[6] Slopes of individual-level predictors could in theory be allowed to randomly vary across level 2 or level 3 units (therapy-groups or therapists) in the group therapy arm, and could vary across level 3 units (therapists) in the individual therapy arm. Slopes of therapy-group level predictors could in theory be allowed to randomly vary across level 3 units (therapists) in the group therapy arm. For simplicity we have expanded the Equation (2) MA-PN model to include only fixed slopes of predictors (denoted with γ's) in Equation (3). The decision regarding whether to model slopes as fixed or random may be based on theoretical grounds (i.e., does substantive theory predict that this slope should differ across therapy groups?) and/or empirical grounds (i.e., does a LRT comparing a model with a fixed slope versus random slope in the treatment arm support the necessity of including the random slope?).

References

Andridge, R., Shoben, A., Muller, K., & Murray, D. (2014). Analytic methods for individually randomized group treatment trials and group-randomized trials when subjects belong to multiple groups. *Statistics in Medicine*, *33*, 2178–2190.

Asparouhov, T., & Muthén, B. (2012). Multiple group multilevel analysis: M*plus* Web Note 16. *Technical appendix*. Retrieved from statmodel.com.

Baldwin, S., Bauer, D., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods*, *16*, 149–165.

Baldwin, S. & Fellingham, G. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*, 151–164.

Baldwin, S., Murray, D., & Shadish, W. (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.

Bauer, D. J., Gottfredson, N. C., Dean, D., & Zucker, R. A. (2012). Analyzing repeated measures data on individuals nested within groups: Accounting for dynamic group effects. *Psychological Methods*, *18*, 1–14.

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research*, *43*, 210–236.

Candel, M., & van Breukelen, G. (2009). Varying cluster sizes in trials with clusters in one treatment arm: Sample size adjustments when testing treatment effects with linear mixed models. *Statistics in Medicine*, *28*, 2307–2324.

Candel, M., & van Breukelen, G. (2010). *D*-optimality of unequal versus equal cluster sizes for mixed effects linear regression analysis of randomized trials with clusters in one treatment arm. *Computational Statistics and Data Analysis*, *54*, 1906–1920.

Christensen, S., Frostholm, L., Ornbol, E., & Schroder, A. (2015). Changes in illness perceptions mediated the effect of cognitive behavioural therapy in severe functional somatic syndromes. *Journal of Psychosomatic Research*, *78*, 363–370.

Chu, Y. (2013). *Bayesian mediation analysis for partially clustered designs* (Master of Science in Statistics). University of Texas at Austin. Retrieved from http://repositories.lib.utexas.edu/handle/2152/22571

Compas, B. E., Forehand, R., Keller, G., Champion, A., Rakow, A., Reeslund, K. L., … Cole, D. A. (2009). Randomized clinical trial of a family cognitive-behavioral preventive intervention for children of depressed parents. *Journal of Consulting and Clinical Psychology*, *77*, 1007–1020.

Compas, B. E., Forehand, R., Thigpen, J., Hardcastle, E., Garai, E., McKee, L., … Sterba, S. (2015). Efficacy and moderators of a family group cognitive behavioral preventive intervention for children of depressed parents. *Journal of Consulting and Clinical Psychology*, *83*, 541–553.

Conrad, I., Kleinstäuber, M., Jasper, K., Hiller, W., Andersson, G. & Weise, C. (2015). The changeability and predictive value of dysfunctional cognitions in cognitive behavior therapy for chronic tinnitus. *International Journal of Behavioral Medicine*, *22*, 239–250.

Cornfield, J. (1978). Randomization by group: A formal analysis. *American Journal of Epidemiology*, *108*, 100–102.

Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting Clinical Psychology*, *59*, 20–26.

Dishion, T., Poulin, F., & Burraston, B. (2001). Peer group dynamics associated with iatrogenic effects in group interventions with high-risk young adolescents. *New Directions for Child and Adolescent Development*, *91*, 79–92.

Emsley, R., Dunn, G., & White, I. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research*, *19*, 237–270.

Fuchs, L. S., Sterba, S. K., Fuchs, D., & Malone, A. (under review). Does evidence-based fraction intervention address the needs of very low performing students? *Journal of Research on Educational Effectiveness*.

Hedges, L., & Citkowicz, M. (2015). Estimating effect size when there is clustering in one treatment group. *Behavior Research Methods*. doi:10.3758/s13428-014-0538-z

Hedman, E., Alaoui, S., Lindefors, N., Andersson, E., Rück, C., & Ghaderi, A., … Liotsson, B. (2014). Clinical effectiveness and cost-effectiveness of Internet- vs. group-based cognitive behavior therapy for social anxiety disorder: 4-Year follow-up of a randomized trial. *Behaviour Research and Therapy*, *59*, 20–29.

Jewell, J. Malone, M., Rose, P., Sturgeon, D., & Owens, S. (2015). A multiyear follow-up study examining the effectiveness of a cognitive behavioral group therapy program on the recidivism of juveniles on probation. *International Journal of Offender Therapy and Comparative Criminology*, *59*, 259–272.

Kazdin, A. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. *Journal of Consulting and Clinical Psychology*, *54*, 95–105.

Kenny, D., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, *83*, 126–137.

Kim, D., Wampold, B., & Bolt, D. (2006). Therapist effects in psychotherapy: A random effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research*, *16*, 161–172.

Korendijk, E., Maas, C., Hox, J., & Moerbeek, M. (2012). The robustness of the parameter and standard error estimates in trials with partially nested data: A simulation study. In E. Korendijk (Ed.). *Robustness and optimal design issues for cluster randomized trials* (Dissertation, pp. 59–94). Utrecht University. Retrieved from http://dspace.library.uu.nl/bitstream/handle/1874/240965/korendijk.pdf?sequence=1

Lachowicz, M. J., Sterba, S. K., & Preacher, K. J. (2015). Investigating multilevel mediation with fully or partially nested data. *Group Processes & Intergroup Relations*, *18*, 274–289.

Lai, M., & Kwok, O.-M. (2014). Standardized mean differences in two-level cross-classified random effects models. *Journal of Educational and Behavioral Statistics*, *39*, 282–302.

Lee, K., & Thompson, S. (2005a). Clustering by health professional in individually randomised trials. *British Medical Journal*, *330*, 142–144.

Lee, K. J., & Thompson, S. G. (2005b). The use of random effects models to allow for clustering in individually randomized trials. *Clinical Trials*, *2*, 163–173.

Liu, Q.-X., Fang, X.-Y., Yan, N., Zhou, Z.-K., Yuan, X.-J., Lan, J., Liu, C.-Y. (2015). Multi-family group therapy for adolescent internet addiction: Exploring the underlying mechanisms. *Addictive Behaviors*, *42*, 1–8.

Lohr, S., Schochet, P. Z., & Sanders, E. (2014). *Partially nested randomized controlled trials in education research: A guide to design and analysis* (NCER 2014–2000) Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Luo, W., Cappaert, K., & Ning, L. (2015). Modelling partially cross-classified multilevel data. *British Journal of Mathematical and Statistical Psychology*. doi:10.1111/bmsp.12050

Marshall, A., Spredbury, J., Cheston, R., Coleman, P., Ballinger, C., Mullee, M., … Bartlett, E. (2015). A pilot randomized controlled trial to compare changes in quality of life for participants with early diagnosis dementia who attend a 'Living Well with

Dementia' group compared to waiting list control. *Aging & Mental Health*, *19*, 526–535.

Martindale, C. (1978). The therapist-as-fixed-effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *46*, 1525–1530.

Mehta, P. (2015, May). *Multilevel structural equation modeling using xxm*. Paper presented at the 5th annual modern modeling methods conference, Storrs, CT.

Moerbeek, M., & Wong, W. (2008). Sample size formulae for trials comparing group and individual treatments in a multilevel model. *Statistics in Medicine*, *27*, 2850–2864.

Murray, D. Hannan, P., & Baker, W. (1996). A Monte Carlo study of alternative responses to intraclass correlation in community trials: Is it ever possible to avoid Cornfield's penalties? *Evaluation Review*, *20*, 313–337.

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus users' guide*. (7th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599–620.

Pals, S., Murray, D., Alfano, C., Shadish, W., Hannan, P., & Baker, W. (2008). Individually randomized group treatment trials: A critical appraisal of frequently used design and analytic approaches. *American Journal of Public Health*, *98*, 1418–1424.

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*, 77–98.

Raudenbush, S. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*, 173–185.

Roberts, C. (1999). The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Statistics in Medicine*, *18*, 2605–2615.

Roberts, C. (2008). *cluspower* [*Computer software and manual*]. Retrieved from http://personalpages.manchester.ac.uk/staff/chris.roberts/

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials*, *2*, 152–162.

Roberts, C., & Walwyn, R. (2013). Design and analysis of non-pharmacological treatment trials with multiple therapists per patient. *Statistics in Medicine*, *32*, 81–98.

Rohde, P. (2008). Statistical analysis of group-administered intervention data: Re-analysis of two randomized trials. *Psychotherapy Research*, *18*, 365–376.

Sanders, E. (2011). *Multilevel analysis methods for partially nested cluster randomized trials* (Unpublished dissertation). University of Washington. Retrieved from http://gradworks.umi.com/34/52/3452760.html

Selig, J. P., & Preacher, K. J. (2008, June). *Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects* [Computer software]. Retrieved from http://quantpsy.org/

Serlin, R., Wampold, B., & Levin, J. (2003). Should providers of treatment be regarded as a random or fixed factor? If it ain't broke, don't "fix" it: A comment on Siemer and Joormann. *Psychological Methods*, *8*, 524–534.

Sterba, S. K., Preacher, K. J., Hardcastle, E. J., Forehand, R., Cole, D. A., & Compas, B. E. (2014). Structural equation modeling approaches for analyzing partially nested data. *Multivariate Behavioral Research*, *49*, 93–118.

Sundquist, J., Lilja A., Palmer, L. Memon, A., Wang, X., Johansson, L. & Sundquist, K. (2015). Mindfulness group therapy in primary care patients with depression, anxiety, stress, and adjustment disorders: Randomised controlled trial. *British Journal of Psychiatry*, *206*, 128–135.

Talley, A. (2013). *The impact of nonnormal and heteroscedastic level one residuals in partially clustered data* (Masters of Arts). University of Texas at Austin. Retrieved from http://repositories.lib.utexas.edu/handle/2152/22630

Tasca, G., Balfour, L., Ritchie, K., & Bissada, H. (2007). The relationship between attachment scales and group therapy alliance growth differs by treatment type for women with binge-eating disorder. *Group Dynamics: Theory, Research, and Practice*, *11*, 1–14.

Teismann, T., von Brachel, R., Hanning, S., Grillenberger, M., Hebermehl, L., Hornstein, I., & Willutzki, U. (2014). A randomized controlled trial on the effectiveness of a rumination-focused group treatment for residual depression. *Psychotherapy Research*, *24*, 80–90.

Tessler, J. (2014). *Three-level models for partially nested data structures* (Unpublished dissertation). University of California at Los Angeles. Retrieved from https://escholarship.org/uc/item/21q2m89j

Walwyn, R., & Roberts, C. (2010). Therapist variation within randomised trials of psychotherapy: Implications for precision, internal, and external validity. *Statistical Methods in Medical Research*, *19*, 291–315.

Wampold, B., & Brown, G. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, *73*, 914–923.

Wesner, A. C., Gomes, J. B., Detzel, T., Blaya, C., Manfro, G., & Heldt, E. (2014). Effect of cognitive-behavioral group therapy for panic disorder in changing coping strategies. *Comprehensive Psychiatry*, *55*, 87–92.

Wolff, N., Huening, J., Shi, J., Frueh, C., Hoover, D., & McHugo, G. (2015). Implementation and effectiveness of integrated trauma and addiction treatment for incarcerated men. *Journal of Anxiety Disorders*, *30*, 66–80.

Zhang, W., Yan, T., Du, Y., & Liu, X. (2014). Brief report: Effects of solution-focused brief therapy group-work on promoting post-traumatic growth of mothers who have a child with ASD. *Journal of Autism & Developmental Disorders*, *44*, 2052–2056.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*, 695–719.