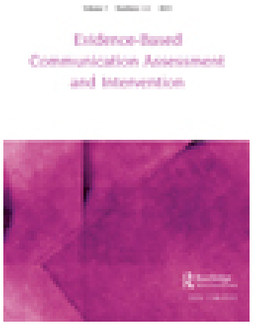
Taylor & Francis
Taylor & Francis Group



Evidence-Based Communication Assessment and Intervention

ISSN: 1748-9539 (Print) 1748-9547 (Online) Journal homepage: https://www.tandfonline.com/loi/tebc20

Using Bayesian methods to test mediators of intervention outcomes in single-case experimental designs

Milica Miočević, Fayette Klaassen, Gemma Geuke, Mariola Moeyaert & Marija Maric

To cite this article: Milica Miočević, Fayette Klaassen, Gemma Geuke, Mariola Moeyaert & Marija Maric (2020): Using Bayesian methods to test mediators of intervention outcomes in single-case experimental designs, Evidence-Based Communication Assessment and Intervention, DOI: 10.1080/17489539.2020.1732029

To link to this article: https://doi.org/10.1080/17489539.2020.1732029

Evidence-Based Communication Assessment and Intervention, 2020
https://doi.org/10.1080/17489539.2020.1732029

Using Bayesian methods to test mediators of intervention outcomes in single-case experimental designs

Milica Miočević[1], Fayette Klaassen[2], Gemma Geuke[3], Mariola Moeyaert[4] & Marija Maric[5]
[1]Department of Psychology, McGill University, Montreal, QC, Canada; [2]Department of Methods & Statistics, Utrecht University, Utrecht, The Netherlands; [3]Department of Sociology, Utrecht University, Utrecht, The Netherlands; [4]Department of Educational Psychology and Methodology, University at Albany - CUNY, Albany, NY, USA; [5]Department of Developmental Psychology, University of Amsterdam, Amsterdam, The Netherlands

Abstract
Single-Case Experimental Designs (SCEDs) have lately been recognized as a valuable alternative to large-group studies. SCEDs form a great tool for the evaluation of treatment effectiveness in heterogeneous and low-incidence conditions, which are common in the field of communication disorders. Mediation analysis is indispensable in treatment research because it informs researchers about the mechanism through which the intervention leads to changes (e.g., communication skills) in the outcome of interest (e.g., developmental outcomes). Despite the increasing popularity of both SCEDs and mediation analysis, there are currently no methods for estimating mediated effects for a single individual. This paper describes how Bayesian piecewise regression analysis can be used for mediation analysis in SCEDs. A Playskin Lift™ dataset from one infant born preterm who is at risk for cognitive developmental delays is used to illustrate two approaches to mediation analysis in SCEDs: Bayesian computation of the mediated effect and Bayesian informative hypothesis testing. Annotated R code is provided so researchers can easily fit the proposed models to their own SCED data set. Advantages and limitations of the method are discussed.

Keywords: *Bayesian statistics; single-case; single-subject; mediator analysis; hypothesis testing.*

The methodology of single-case experimental designs (SCEDs) is a rigorous scientific research approach that can be used to evaluate the effectiveness of an intervention (Horner et al., 2005; Kazdin, 2011). SCEDs have shown to be a prime alternative for large-group studies either as an initial study leading to specific hypothesis to be tested in a group study, or as a stand-alone research study. This second option is especially important in heterogeneous populations or populations with rare incidence rates which may not be uncommon in communication disorders research. Because SCEDs can also easily be incorporated in clinical practice, they have the potential to enhance evidence-based practice and stimulate collaboration between research and practice, unifying research questions that emerge from clinical practice on one hand, and, on the other hand, research methodology to test these questions on a single-client level.

The ultimate goal of SCED research methodology is to evaluate whether there is a functional relationship between the intervention and change in the outcome measure of interest (Kratochwill et al., 2010). For this purpose, a case is measured repeatedly over time during a baseline condition that is "interrupted" by an intervention (also referred to as "treatment" in the remainder of the paper). By using SCED methodology, a case serves as its own control, detailed information related to changes across time can be obtained, and case-specific intervention effects can be estimated

For correspondence: Milica Miočević, Department of Psychology, McGill University, Canada. E-mail: milica.miocevic@mcgill.ca

(Barlow, Nock, & Hersen, 2009). Because of these advantages, the methodology has become increasingly popular over time and has been the method of choice for over a thousand studies to date (Wiessenekker, 2019). SCEDs are used across a variety of different research fields ranging from rehabilitation and clinical psychology to special education and communication disorders, and are known under several different names such as interrupted time series, single-subject experimental design, intrasubject designs, among others (Smith, 2012).

Together with the increasing interest in using SCEDs to establish an evidence base for the effectiveness of treatments, there is a need for methods to quantify the size of the intervention effect. During the last decade, there have been efforts to develop and empirically validate indices and effect sizes to report the strength and statistical significance of effects. However, there is no best index and some indices might be better in some conditions compared to others (Manolov & Moeyaert, 2017; Vannest, Peltier, & Haas, 2018). Nonparametric nonoverlap indices quantify the degree of non-overlap between the baseline and the treatment data clouds, such as Nonoverlap of All Pairs (NAP; Parker & Vannest, 2009), Tau-U (Parker, Vannest, Davis, & Sauber, 2011), Tau-C (Tarlow, 2017), Improvement Rate Difference (IRD; Parker, Vannest, & Brown, 2009) and the Percent of Data Exceeding the Phase A Median Trend (PEM-T; Wolery, Busick, Reichow, & Barton, 2010) just to name a few. Parametric approaches on the other hand allow for a quantification of the size of a treatment effect together with an estimate of the standard error. Some popular parametric approaches are regression-based effect sizes (i.e. Center, Skiba, & Casey, 1985; van den Noortgate & Onghena, 2003a, 2003b), multilevel modeling (Shadish, Rindskopf, & Hedges, 2008), hierarchical linear modeling (Parker et al., 2009), standardized mean differences (e.g. Cohen's *d*, Hedge's *g*; Shadish, Hedges, & Pustejovsky, 2014) and the between-case standardized difference (Hedges, Pustejovsky, & Shadish, 2012, 2013). All of these approaches can be used to test the effectiveness of a treatment; that is, they provide an answer to the question: "Does the treatment work for this individual client?". However, none of the above methods allow researchers to evaluate *how* the treatment worked for a particular client, i.e. what was the mechanism of change.

Mechanisms through which treatments achieve effects: Mediation analysis

When studying effects on a group level, scientists implicitly assume that interventions work the same for all group members, and neglect the fact that the same intervention might achieve its effects through different mechanisms for different clients. Identification of individual mechanisms could lead to identification of the most potent treatment techniques, that is, techniques that are affecting these mechanisms (Maric, Wiers, & Prins, 2012). For example, finding out that negative cognition for client 1 diagnosed with depression was reduced through Cognitive Restructuring and not through Behavioral Activation allows us to tailor the treatment to client 1 by making sure it includes a treatment phase that targets Cognitive Restructuring. However, without examining effects at the individual level, we cannot evaluate the mechanism through which a treatment works (or does not work) for a given person. Generalizing relationships from the group-level to the individual level is not recommended (Cattell, 1952).

Mediation analysis is used to evaluate intermediate variables (mediators; M) that transmit the effect of an independent variable (X) on a dependent variable (Y) (MacKinnon, 2008). It provides an answer to a question: "How does the treatment work, through which mechanisms?" For example, Maric, Heyne, MacKinnon, Van Widenfelt, and Westenberg

(2013) found that self-efficacy mediated the relationship between cognitive-behavioral therapy (CBT) and school-related fear in adolescents. Thus, the theory tested by mediation analysis in clinical settings is that a certain intervention will produce changes in the mediator and that these changes will, in turn, affect intervention outcomes (MacKinnon, 2008). So far, these intervention theories have, unfortunately, only been tested in large-group studies. In the remainder of this section, we describe a single mediation model (see Figure 1) and the most frequent data-analytic approaches to testing for mediation.

The effects of interest in the single mediator model (Figure 1) can be computed using three equations:

$$Y = i_1 + cX + e_1 \quad (1)$$

$$M = i_2 + aX + e_2 \quad (2)$$

$$Y = i_3 + c'X + bM + e_3 \quad (3)$$

where X is the independent variable, M is the mediator, and Y is the dependent variable. Intercepts are i_1, i_2, and i_3, c is the total effect of the independent variable on the dependent variable, a is the coefficient relating the independent variable to the mediator, b is the coefficient relating the mediator to the dependent variable in the model containing the independent variable, c' is the coefficient relating the independent variable to the dependent variable (also called the direct effect), and e_1, e_2, and e_3 are error terms assumed to follow a normal distribution with a mean of 0 and variances of σ^2_{e1}, σ^2_{e2} and σ^2_{e3} (respectively).

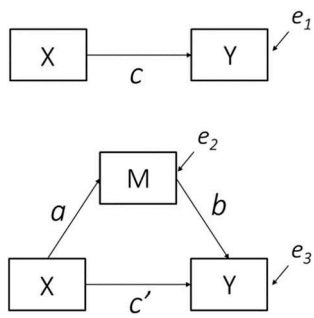


Figure 1. Top panel: total effect of the independent variable on the outcome. Bottom panel: Single mediator model. The intercepts are included in the two models, but not in the figure

One of the first approaches to testing for mediation was described in papers by Judd and Kenny (1981) and Baron and Kenny (1986), and it consists of four steps: (1) establishing that the independent variable affects the dependent variable (i.e. significant coefficient c in Equation (1)); (2) establishing that the independent variable affects the mediator (i.e. significant coefficient a in Equation (2)); (3) establishing that the effect of the mediator on the outcome, controlling for the independent variable, is nonzero (i.e. significant coefficient b in Equation (3)); (4) establishing that the effect of the independent variable on the dependent variable is weaker when we control for the effect of the mediator than when we do not control for the effect of the mediator (i.e. coefficient c' in Equation (3) should be smaller than coefficient c in Equation (1)). This approach falls under the category of causal steps approaches to mediation analysis, and one of the less stringent and more powerful causal steps methods is called the joint significance test, which only requires steps 2 and 3. However, none of the causal steps approaches provide a numerical estimate of the value of the indirect (mediated) effect, and they have less power to detect the mediated effect relative to methods that compute and test the significance of the mediated effect directly (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002).

The mediated (indirect) effect is most often computed as the product of coefficients ab, and in linear models with no missing values, we obtain the same value of the mediated effect if we compute it as the difference of

coefficients $c - c'$ (MacKinnon, Warsi, & Dwyer, 1995). Modern approaches to mediation analysis test the significance of the mediated effect by computing confidence intervals for the mediated effect and evaluate whether 0 is in the interval. Modern methods that have the most power either model the distribution of the mediated effect appropriately (i.e. using the distribution of the product of two normal variates; Craig, 1936; Lomnicki, 1967; MacKinnon et al., 2002; MacKinnon, Lockwood, & Williams, 2004) or do not make any assumptions about the distribution of the mediated effect (e.g. bootstrap and Bayesian methods; MacKinnon et al., 2004; Yuan & MacKinnon, 2009).

Bayesian mediation analysis

The mediated effect can be computed and evaluated in the frequentist (classical) framework using methods such as ordinary least squares regression (OLS) or structural equation models fit using Maximum Likelihood estimation. It is also possible, and sometimes more advantageous, to do mediation analysis in the Bayesian framework (Miočević, MacKinnon, & Levy, 2017; Yuan & MacKinnon, 2009). In the Bayesian framework, the analysis starts by specifying prior distributions for all freely estimated parameters in the model. In the case of the single mediator model, the parameters that are assigned priors are those from Equations (2) and (3): the intercepts i_2 and i_3, regression paths a, b, and c', and residual variances σ^2_{e2} and σ^2_{e3}. The next step of a Bayesian analysis requires updating the prior distributions with the observed data using Bayes' theorem, in order to obtain the posterior distribution of the model parameters: $p(\theta|data) \propto p(data|\theta)\,p(\theta)$, where p (θ| data) denotes the posterior distribution of the parameters, p(data| θ) denotes the likelihood function based on the observed data, and p(θ) denotes the prior distribution for the set of freely estimated parameters. The inferences about the parameters of interest are based on the posterior distributions that can be summarized to obtain a point summary (e.g. mean or median) or an interval summary. The distribution of the mediated effect is approximated using values from the posterior distributions for coefficients a and b. These distributions can be obtained using Markov Chain Monte Carlo (MCMC), implemented in various software (for a tutorial on using MCMC, see Sinharay, 2004). The MCMC draws can be used to approximate the posteriors, but also for hypothesis testing. Bayesian statistics have a unique take on hypothesis testing which allows for quantifying the *probability* that a parameter (e.g. the mediated effect) is greater than a clinically significant value (thus providing a measure of the degree to which a clinical hypothesis is supported) and for quantifying relative evidence for different hypotheses using a Bayes factor (Kass & Raftery, 1995). Bayesian hypothesis testing is very flexible in terms of hypotheses that can be compared. Expectations about the directions of the effect (e.g. the sign of a regression coefficient) can be formulated as so-called *informative* hypothesis (Klugkist, Laudy, & Hoijtink, 2005). There are at least three advantages of Bayesian over frequentist hypothesis testing:

(1) Rather than evaluating simple null and alternative hypotheses, Bayesian hypothesis testing allows for formulating hypotheses that express expectations about a combination of parameters (e.g. a and b paths in mediation analysis), and a combination of (in)equalities for these parameters. This would not be possible using frequentist hypothesis testing. Moreover, these hypotheses reflect direct expectations we have from the

theory with regards to the model parameters.
(2) Rather than comparing one hypothesis to another, Bayesian hypothesis tests allow us to devise a set of competing hypotheses and find which hypothesis is most supported.
(3) The conclusion of a Bayesian hypothesis test is much more intuitive – and in line with Bayesian statistics: it provides the probability that a hypothesis is the best hypothesis. Not "true", but the best, from the set considered.

For the sake of space, we cannot provide a more extensive description of Bayesian methods for mediation analysis and informative hypothesis testing, and we refer the interested reader to chapters by Miočević and Van de Schoot (2019), the paper by Yuan and MacKinnon (2009), the book by Hoijtink (2012) and the paper by Béland, Klugkist, Raîche, and Magis (2012).

The above methods are frequently used for group-level mediation analyses. There have been at least two proposed methods for mediation analysis in context of SCEDs (Gaynor & Harris, 2008; Geuke, Maric, Miočević, Wolters, & de Haan, 2019). However, the proposed methods do not yield a numerical estimate of the mediated effect, nor do they allow the researcher to quantify the support of the mediation hypothesis from the data. Knowledge about individual participants' mediators of treatment outcomes could inform treatment-decision making and lead to a more evidence-based practice (Maric, Prins, & Ollendick, 2015). Furthermore, knowing the mediator(s) that transmit the effect of an intervention on the outcome(s) of interest can help in tailoring the treatment to each client.

This study: SCEDs mediation analysis. In this paper, we describe two methods for evaluating whether there is a mediated effect: a method that can compute the value of the mediated effect using repeated measures of a hypothesized mediator and an outcome of interest collected from a single participant, and a method that tests whether this mediated effect is different from 0 (or any other user-specified clinically relevant value). The methods developed and described in this paper will use Bayesian estimation for the parameters in the mediation model, and this is the first paper (to our knowledge) that includes both parameter estimation and informative hypothesis testing for mediation models.

We will focus on the regression-based effect size originally introduced by Center et al. (1985) because of its flexibility. In order to estimate the regression-based effect size, a piecewise regression can be run which results in the estimate of the outcome score at the start of the SCED, the time trend during the baseline, the immediate intervention effect (i.e. change in outcome score at the start of the intervention phase) and the difference in time trend between the baseline phase and the intervention phase. This results in two regression-based effect sizes of interest, namely an immediate intervention effect and an intervention effect on the time trend.

The following sections describe the data for the empirical example and how Bayesian piecewise regression analysis can be used to test for mediation in a SCED.

METHOD

Empirical example

The dataset for the empirical example comes from a study of the effectiveness of wearing the Playskin Lift™ exoskeletal garment on object exploration and cognitive outcomes in infants that were born preterm and/or had brain injuries (Babik, Cunha, Moeyaert, Hall, & Lobo, 2019). The exoskeletal garment was designed to assist antigravitational movement

of the infant and improve function and strength of their arms, which was hypothesized to aid object grasping and exploration, also at moments when the garment was no longer worn (Lobo et al., 2016). For a more detailed and comprehensive description of the dataset and measurement procedure of this study, the reader is referred to the article by Babik et al. (2019). We simplified the data set from the original study for the purposes of illustrating the proposed methods, and the results of the empirical example should not be used to make generalizations about the utility of the exoskeleton.

The dataset is a multiple baseline $A_1B_1A_2$-design, which means that it consists of three phases: the first phase is a baseline phase (A_1), which was designed to assess the baseline level of the infant's scores on various variables of object exploration and reaching. The amount of measurement occasions in this baseline phase was alternated across participants, ranging from 3 to 5 occasions. During this phase, the exoskeletal garment was not worn, except for during a subset of assessments. The second phase (B_1) is the treatment phase, in which parents were asked to perform a structured set of daily exercises of 40 min with the infants using the exoskeletal garment. The third phase (A_2) was a follow-up phase, which was designed to assess whether there were remaining effects of using the exoskeletal garment after the treatment was stopped, and was similar to the baseline phase. As mentioned before, because the effect of the intervention on the outcome score is replicated across multiple participants, the SCED study is more externally valid (i.e. more generalized conclusions about the intervention effectiveness can be obtained).

At each measurement occasion, six types of assessments were conducted. Each assessment consisted of a toy presentation to the infant, after which the reaction of the infant was measured in a structured manner. This assessment was conducted in 2 × 3 conditions, both with the exoskeletal garment off and on, and with the toy presented at hip, chest, or eye level. All assessments were recorded on video. For each of these assessments, several variables were recorded, such as grasping ability and the percentage of time the infant looked at the toy.

For the purposes of the current example, a subset of the variables of one participant will be used to illustrate the suggested analysis methods. The mediation hypothesis was that daily exercise with the exoskeletal garment (X; treatment) leads to better grasping ability without wearing the garment (M; mediator), which leads the infant to be more interested in toys and to spend more time looking at the toy (Y; outcome). Grasping was measured as the percentage of the total assessment time in which the infant had any type of contact with the toy, that is, the sum of bimanual and unimanual contact. Looking was measured as the percentage of the total assessment time in which the infant directed their eyes at the toy. Data for the empirical example are plotted in Figure 2 using the raw data obtained from Babik et al. (2019). One condition of measurements was selected for the illustrative analysis here: with the exoskeletal garment off and the toy presented at the chest level, as one of the aims of the treatment in the study by Babik et al. (2019) was to improve the independent grasping abilities of the infants, that is, without wearing the exoskeletal garment. Note that, for a more complete analysis of this data, the proposed analysis can be repeated for all six conditions and that the methods we illustrate use only the baseline phase (i.e. A_1) and the intervention phase (B_1), but could be extended to include additional phases (e.g. A_2, which presents the maintenance phase in the present data set). Also note that using data of only one participant of a multiple baseline study does not allow the analysis to make generalizations (i.e. external validity) about the intervention and the mediation effect, that the mediation

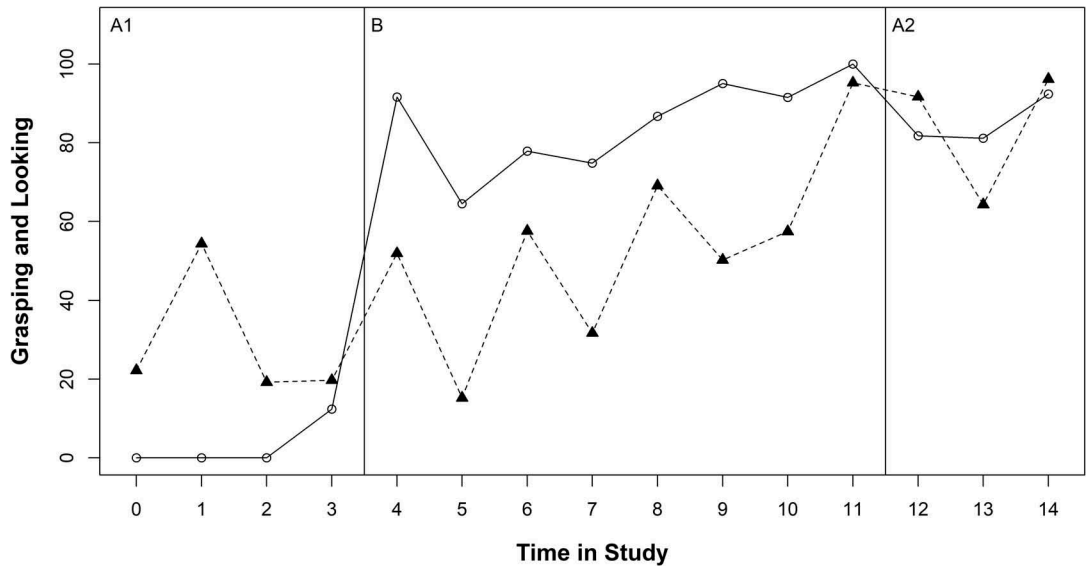


Figure 2. Graphical display of the scores of Grasping (dashed lines and triangles) and Looking (solid lines and points) of participant 201 of the study by Babik et al. (2019). Phases are denoted in the upper left corner of each phase. Reprinted with permission

model in this article may not be theoretically valid, and the data are used solely to illustrate the proposed methods.

For readers interested in using the example code provided as Supplemental Material, it is important to organize the data in a specific format for the code to work. The data set needs to contain the following variables: (1) Phase, which denotes whether a given observation belongs to the baseline phase (Phase = 0) or the treatment phase (Phase = 1); (2) Time1, which is equal to the value of the measurement occasion – 1 (and ranges from 0 to 11 in the present data set which uses a total of 12 measurement occasions in the analysis); (3) phase_time2, which denotes the time spent in the treatment phase, and has a value of 0 during the baseline phase and at the first occasion in the treatment phase, and values of 1, 2, 3, etc., for subsequent observations in the treatment phase; (4) ScoreM, which are scores on the mediator on occasions 1–12; (5) ScoreY, which denotes the score on the outcome at a given measurement occasion (in the present data set, there are 12 values of ScoreY); and (6) Tmed, which represents scores on the mediator with a missing value in the first row and scores on occasions 1–11 as values in the subsequent rows. The current formatting of the data set will yield a data set with the number of rows equal to the number of observations; also, the variables Tmed will be missing a value in the first row. This data format is necessary for executing the analyses for the proposed methods.

Data analysis

Most data analytic methods for SCEDs were developed with the goal of evaluating the effect of a change in phase on a single variable. In the single mediator model for SCEDs, both the hypothetical mediator and outcome are measured repeatedly over at least two phases (i.e. baseline

phase and intervention phase). Given that our goal is to compute the numerical value of the indirect effect, we automatically excluded methods that quantify percentage of nonoverlapping data (e.g. Schlosser, Lee, & Wendt, 2008; Scruggs, Mastropieri, & Casto, 1987). We opted for piecewise regression analysis because it allows for quantifying the change in the mediator due to the change in phase (*a* path in Figure 1) and change in outcome due to the change in the mediator (*b* path in Figure 1) controlling for the effect of phase. For the purposes of the current analyses, the equations for piecewise regression analyses of the mediator and outcome are as follows:

$$\begin{aligned} M = {} & b_{0M} + b_{1M}time1 + b_{2M}phase \\ & + b_{3M}phase\_time2 + e_M \end{aligned} \quad (4)$$

and

$$\begin{aligned} Y = {} & b_{0Y} + b_{1Y}time1 + b_{2Y}phase \\ & + b_{3Y}phase\_time2 + b_{4Y}M_{t-1} + e_Y. \end{aligned} \quad (5)$$

Due to the specific coding of the predictors, regression coefficients from the piecewise regression analysis provide estimates of the level of the first time point of phase A (b_{0M} for the mediator and b_{0Y} for the outcome), of the trend in phase A (b_{1M} for the mediator and b_{1Y} for the outcome), of the change in level at the start of phase B (b_{2M} for the mediator and b_{2Y} for the outcome) and of the change in trend between the two phases (b_{3M} for the mediator and b_{3Y} for the outcome; Manolov & Moeyaert, 2017). The additional term in the equation for the outcome represents the lagged effect of the mediator (b_{4Y}).

There are two reasonable definitions for the effect of the treatment on the mediator (*a* path in Figure 1) in this context: the effect of phase change can either be measured as the change in level (b_{2M}), or as the change in trend between the two phases (b_{3M}). Defining the *a* path as the change in *level* between phases allows for computing the indirect effect of the phase change on the outcome through changes in the level of the mediator. Defining the *a* path as the change in *trend* between two phases leads to an indirect effect that quantifies the effect of change in phase on the outcome through change in the trend of the mediator. The effect of the mediator on the outcome (*b* path in Figure 1) is represented by the b_{4Y} coefficient from Equation (5) and the direct effect (*c′* path in Figure 1) of phase on the outcome controlling for the effect of the mediator is represented either by coefficient b_{2Y} (if the direct effect is defined as a change in level) or using the coefficient b_{3Y} (if the direct effect is defined as the change in trend).

There are two ways to conceptualize the mediated effect in the present example: 1) as the product of coefficients $b_{2M}b_{4Y}$ which represents the change in the value of the outcome due to the change in the *level* of the mediator following a change in phase, and 2) as the product of coefficients $b_{3M}b_{4Y}$ which represents the change in the value of the outcome due to the change in the *trend* (slope) of the mediator following a change in phase. The procedures for evaluating whether these indirect effects are different from 0 require approximating the distributions of $b_{2M}b_{4Y}$ and $b_{3M}b_{4Y}$, and covariances between b_{2M} and b_{4Y} and between b_{3M} and b_{4Y}, which was more straightforward to obtain in the Bayesian framework. The mediated effect is evaluated using two approaches: parameter estimation and hypothesis testing. Both analyses were performed in R (R Core Team, 2013) using the packages rjags (Plummer, 2018) and the software JAGS (Plummer, 2003) for the Bayesian piecewise regression, the R package coda for the computation of intervals for the mediated effects (Plummer et al., 2018), and the R package bain for hypothesis testing (Gu, Hoijtink, Mulder, & van Lissa, 2019). The annotated R syntax for the analysis is available in the Supplemental

Material. The analysis consisted of the following five steps. Step 1–3 are preparation for step 4 (parameter estimation) and step 5 (hypothesis testing):

Step 1. Obtain frequentist estimates of the parameters in Equations (4) and (5) using the lm() function. The estimates and standard errors are shown in Table 1.

Step 2. Formulate priors for the parameters in the Bayesian estimation of the parameters in Equations (4) and (5). These priors have data-dependent mean hyperparameters and variance hyperparameters that are diffuse for the scale of the variables (as shown in the last column of Table 1). In other words, the priors for each intercept and regression coefficient encode the assumption that the best guess for these parameters is equal to the OLS estimate of that parameter, and the prior variances indicate limited confidence in these best guesses. Data-dependent priors are somewhat controversial because they lead to an underestimation of the uncertainty of the parameter estimate/posterior summary (Darnieder, 2011). However, in this situation, fitting the model with normal priors centered at 0 for each intercept and regression coefficient leads to posterior means and medians that are noticeably lower in absolute value relative to the frequentist estimates of the corresponding parameters (probably due to the small sample size). Using data dependent priors alleviates this issue (see, e.g. McNeish, 2016), as can be seen from the comparison of numerical values of posterior point summaries of all model parameters obtained using priors centered at 0 and priors centered at the corresponding OLS estimate (Appendix A).

Step 3. Fit a Bayesian model for Equations (4) and (5) and obtain Markov Chain Monte Carlo (MCMC) draws for all parameters. Preliminary analyses using the Potential Scale Reduction Factor (PSRF; Brooks & Gelman, 1998) and trace plots indicated that the chains converge to the posterior by 10,000 iterations. We discarded the first 10,000 iterations, and ran additional 10,000 iterations to approximate the posterior distribution. For the sake of brevity, we do not explain convergence diagnostics in detail, and for readers new to MCMC we recommend the paper by Sinharay (2004).

Table 1. Ordinary least squares estimates of parameters in Equations (4) and (5) for grasping (M) and looking (Y) and priors for the Bayesian analysis based on these results

Parameter	Estimate	Standard Error	p-Value	Prior
b_{0M} (Intercept)	35.272	16.417	0.064	N(35.272, 1000)
b_{1M} (Time1)	−4.266	8.775	0.640	N(−4.266, 1000)
b_{2M} (Phase)	13.260	27.166	0.639	N(13.260, 1000)
b_{3M} (phase_time2)	10.575	9.283	0.288	N(10.575, 1000)
b_{0Y} (Intercept)	−0.983	15.534	0.952	N(−0.983, 1000)
b_{1Y} (Time1)	5.868	6.547	0.405	N(5.868, 1000)
b_{2Y} (Phase)	57.837	15.342	0.009	N(57.837, 1000)
b_{3Y} (phase_time2)	−1.838	6.791	0.796	N(−1.838, 1000)
b_{4Y} (Tmed)	−0.208	0.185	0.304	N(−0.208, 1000)

Note: The coefficients in the table correspond to the coefficients in Equations (4) and (5), and the variable names in parentheses correspond to the labels in R output. The symbol N denotes a normal prior distribution where the first parameter represents the mean and the second parameter represents the variance. The analyses were run in rjags so the sample code contains the precision parametrization meaning that the second parameter in the normal priors is the precision and the residual precisions are assigned Gamma (G) priors with both hyperparameters equal to .5.

Step 4. Approximate and summarize the posterior distributions of the mediated effects. The first approach to evaluating the size of the mediated effects requires approximating the posterior distributions of these parameters by computing the products $b_{2M}b_{4Y}$ and $b_{3M}b_{4Y}$ using the 10,000 retained draws for these parameters. In order to make inferences about the values of the indirect effects, the posterior distributions need to be summarized using point and interval summaries. Here we use the posterior median instead of the posterior mean because the distribution of the product of two regression coefficients is often asymmetric (Craig, 1936; Lomnicki, 1967). The two options for interval summaries of the posterior are the equal-tail credibility intervals obtained using the $\alpha/2$ and $1\text{-}\alpha/2$ percentiles of the posterior distribution ($\alpha = 0.05$ for 95% credibility intervals), and the Highest Posterior Density (HPD) intervals which have the property that no value outside of the interval is more probable than values within the interval. Given the potential asymmetry of the posteriors for the indirect effects, we use 95% HPD intervals. The last summary of the posterior is the probability that the mediated effect is of the hypothesized sign (here, positive) computed as the proportion of posterior draws of the mediated effect that are either 0 or positive (as illustrated in Miočević et al., 2017). Instead of computing the probability that the mediated effect is positive (or negative), researchers can select a critical value other than 0 that is meaningful for the scale of the outcome and the research question in their study. The accompanying R code can be used to compute the probability that the mediated effect is greater than (or lower than) a user-specified critical value (denoted *crit* in the R syntax). Note that this type of probabilistic interpretation is only available in the Bayesian framework.

Step 5. Test hypotheses that the mediated effects are nonzero. The second approach to evaluating whether the indirect effects are different from 0 requires the specification of hypotheses that evaluate the presence of a mediated effect (akin to the joint significance test in the frequentist framework where the presence of a mediated effect is established if the *a*-path and *b*-path in the single mediator model are both significantly different from zero; for more on the logic and statistical properties of the joint significance test, see MacKinnon et al., 2002). A set of four hypotheses of interest, presented in Table 2, was defined for the Playskin Lift™ dataset presented in this paper. These hypotheses were formulated based on theoretical expectations for the current dataset. For other research questions, the expected signs of the *a* and *b*-paths may be different. Because the *a*-path can be conceptualized in two ways, this set of hypotheses was evaluated using both b_{2M} and b_{3M} as the *a*-path, while the *b*-path was conceptualized as b_{4Y}, as shown in the third and fourth columns of Table 2.

This set of hypotheses can be used to test the presence of a positive mediated effect. The first hypothesis specifies our main theoretical expectation, namely that both the *a*-path and the *b*-path are positive and different from zero. We can compare this hypothesis to its complement, *H1c*, that says that either the *a*-path, or the *b*-path, or both are not positive. This is a generic "catch-all" alternative hypothesis. By comparing *H1* to *H1c* we can evaluate whether there is a hypothesized positive mediated effect or not. Additionally, *H2* and *H3* are more precise falsifications of the hypothesized mediated effect under *H1*. *H2* specifies that the *a*-path is negative (as opposed to positive under *H1*), without placing any constraints on the *b*-path. *H3* specifies that the *b*-path is negative (as opposed to positive in *H1*), without placing any constraints on the *a*-path.

Bayes factors and/or posterior probabilities can be used to compare each pair of these hypotheses to each other and

Table 2. Mediation hypotheses for the Playskin Lift™ dataset

Hypothesis	In words	*a*-path as change in level	*a*-path as change in trend
H1: *a*-path > 0 & *b*-path > 0	Both the *a*-path and the *b*-path are positive	*H1*: $b_{2M} > 0$ & $b_{4Y} > 0$	*H1*: $b_{3M} > 0$ & $b_{4Y} > 0$
H1c: not H1	Either the *a*-path or the *b*-path or both are not positive	*H1c*: not *H1*	*H1c*: not *H1*
H2: *a*-path < 0	The *a*-path is in opposite direction (negative)	*H2*: $b_{2M} < 0$	*H2*: $b_{3M} < 0$
H3: *b*-path < 0	The *b*-path is in opposite direction (negative)	*H3*: $b_{4Y} < 0$	*H3*: $b_{4Y} < 0$

quantify the relative evidence for each hypothesis. The R package bain (Gu et al., 2019) was used to evaluate the above hypotheses. To obtain the Bayes factors, bain requires the sample size and the estimated covariance matrix for the parameters in the hypotheses, which we obtained from the MCMC output in Step 3. The interested reader is referred to as the bain manual (Hoijtink, Mulder, van Lissa, & Gu, 2019).

A Bayes factor quantifies the evidence for one hypothesis relative to another. For example, if *BF12* = 3, this means that the data are three times more likely to occur if *H1* is true compared to when *H2* is true. If all pairwise Bayes factors for a set of hypotheses are known, these can be used to update the prior probabilities of the hypotheses to obtain the posterior probabilities. Each hypothesis has a prior probability, that is, the probability that a hypothesis is true before observing the data. Using the posterior probabilities for a set of hypotheses, we can select the best hypothesis from a given set.

RESULTS

Across all 10 participants, the original study by Babik et al. (2019) found significant improvement of the mean of Grasping and Looking between the baseline and intervention phase. Looking and only unimanual grasping at the object had a significant immediate change at the beginning of the intervention phase. Compared to the time trend in the baseline phase, Grasping had a larger time trend (i.e. rate of improvement) in the intervention phase, but Looking did not have a significantly larger time trend in the intervention phase. Thus, there is some evidence for an effect of the independent variable on the dependent variable (path *c* in the top panel of Figure 1), and for an effect of the independent variable on the mediator (path *a* in the bottom panel of Figure 1). The mediation analysis presented below provides additional insights about whether the effect of the intervention on Looking is mediated by improvement in Grasping for one of the participants.

First method: Parameter estimation

The results from Step 4 require evaluating the posterior distribution of the mediated effects $b_{2M}b_{4Y}$ and $b_{3M}b_{4Y}$. The posterior summaries of the mediated effects are presented in Table 3 and shown in Figure 3. Note that the posterior medians for both mediated effects were negative. The Highest Posterior Density (HPD) intervals for the indirect effect through changes in the level of the mediator, $b_{2M}b_{4Y}$, ranged

Table 3. Posterior summaries of $b_{2M}b_{4Y}$ and $b_{3M}b_{4Y}$

	$b_{2M}b_{4Y}$	$b_{3M}b_{4Y}$
Posterior median	−0.205	−0.275
Posterior standard deviation	3.634	1.866
95% HPD interval	[−9.486, 5.310]	[−5.371, 1.898]
$p(ab \geq 0)$	38%	30%

from −9.486 to 5.310, thus indicating that 0 is among the most probable values for this effect. Furthermore, 38% of the posterior draws were positive, thus indicating that there is 38% probability that the indirect effect through changes in the level of the mediator is positive. The HPD intervals for the indirect effect through changes in the trend of the mediator, $b_{3M}b_{4Y}$, ranged from −5.371 to 1.898, thus indicating that 0 is, once again, among the most probable values for this effect. Furthermore, 30% of the posterior draws were positive, thus indicating that there is 30% probability that the indirect effect through changes in the trend of the mediator is positive. Overall, the posterior summaries suggest that there was no indirect effect of phase change on Looking through changes in level or trend of Grasping. Thus, in this case, no evidence of mediated effect was found. In situations where the indirect effect is nonzero, researchers can report the median and interpret it in units of the dependent variable. The magnitude and importance of indirect effects computed this way will depend on the scale of the outcome variable and the research setting.

Second method: Hypothesis testing

The results from the Bayesian hypothesis comparison for both representations of the *a*-path are presented in Table 4. *H3* has the highest posterior probability of the set of hypotheses for both conceptualizations of the *a*-path, indicating that the existence of a negative *b*-path receives the most evidence out of the considered set of hypotheses. The differences in results between the results for the two conceptualizations of the *a*-path are minimal, and for the sake of brevity, we will only discuss the results for the change in level. We find that *H3* (negative *b*-path)is .384/.208 ≈ 1.85 times more supported by the data than *H1* (both *a*-path and *b*-path are positive) and .384/

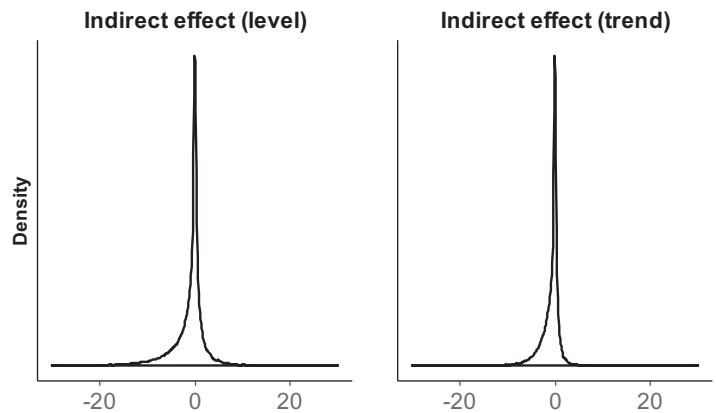


Figure 3. Plot of posteriors for the mediated effects through the changes in level ($b_{2M}b_{4Y}$) and trend ($b_{3M}b_{4Y}$)

Table 4. Posterior probabilities

	***a*-path as change in level**	***a*-path as change in trend**
H1	.208	.266
H1c	.279	.280
H2	.129	.047
H3	**.384**	**.407**

Note. Probabilities in **boldface** indicate the hypothesis with the highest probability. These probabilities were obtained with equal prior probabilities.

.279 ≈ 1,38 times more supported than *H1c* (the *a*-path and *b*-path are not both positive). There appears to be the least evidence for a negative *a*-path (*H2*), since each of the other hypotheses receives more support. There is a slight preference for H3 relative to H1 and H1c.

Note that the posterior probabilities in Table 4 were obtained using equal prior probabilities. That is, all hypotheses received the same prior weight in order to make a fair comparison. The findings do not match our expectations, as we expected the *b*-path to be positive. However, had there been prior research that supported our expectations and had we encoded our prior beliefs in subjective prior probabilities that favor *H1* and updated those with the evidence from the data, the posterior probabilities for *H1* would be higher.

DISCUSSION

Identifying mechanisms through which a certain intervention achieves its effects is extremely important for the identification of the most potent intervention components and therefore for the conduct of the more evidence-based personalized mental health care (Ng & Weisz, 2016). In the original SCED study that investigated effectiveness of a Playskin Lift™ intervention (Babik et al., 2019) two outcome variables were investigated: Looking at and Grasping for objects. Over the whole group of single-case participants, significant improvement of the mean of Grasping and Looking between the baseline and intervention phase was found. However, the theoretical hypothesis underlying Playskin Lift™ intervention points to the following: daily exercise with the exoskeletal garment would lead to better grasping ability, and this would, in turn, lead to infant looking more at toys. The testing of this mediating hypothesis was illustrated in the current study using data from one preterm born infant who underwent Playskin Lift™ intervention. The methods described in this paper allowed for the computation of the numerical value of the indirect or mediated effect and for testing whether this effect is of the hypothesized sign in SCEDs with two phases (i.e. a baseline phase, A_1, and a treatment phase, B_1) in a single-participant. Bayesian parameter estimation and informative hypothesis testing are two ways of approaching the same question; however, the results of each approach are interpreted differently and the two approaches may require different numbers of repeated measures of the same participant for optimal performance. We suggest using both approaches in tandem because together they provide more information about the mediated effect(s). In the case of our single participant, no mediated effect of Grasping was found on the Looking efforts of the participant.

We might conclude that for this infant, Playskin Lift™, does not affect looking behavior through changes in grasping behavior, but through some other mechanism, such as increases in parental guidance. In this way, individual mechanisms of change could be identified and the most potent treatment techniques that affect changes in these mechanisms. The fields of rehabilitation and communication disorders could profit from single-case methods in a substantial way because of phenomena such as (i) a great amount of interventions to treat diverse impairments and client needs; (ii) few

interventions are seen as evidence-based, as informed by information limited to group studies; and (iii) large heterogeneity in client populations.

Limitations and future directions

Note that the default coding of the predictors in piecewise regression in the syntax in the Supplemental Material assumes that the phase effect takes place in the first measurement of the second phase. However, change might not be immediate for all therapies, and the syntax needs to be modified to accommodate a different expectation about the timing of the effect. The same is true for the assumed timing of the effect of the mediator: there is a lag of 1 between the mediator and outcome, and this may not be suitable for all processes. Researchers can modify the code we provide to increase the time to the effect; however, in many situations it is very difficult to formulate a prior hypothesis about the appropriate amount of time necessary for changes in the hypothesized mediator to produce changes in the outcome. If a researcher is for instance interested in estimating the effect of the intervention at the third observation point in the intervention phase, then the time can be centered around that observation point. For more information of the influence of centering time on the estimated intervention effect using piecewise regression, see Moeyaert, Ugille, Ferron, Beretvas, and Van den Noortgate (2014).

The Playskin Lift™ dataset was limited to only 12 repeated measurements over time. A larger number of observation points is preferred to obtain more certainty in the results. A simulation study could provide more insight in how much the current sample size affects the results. Our results showed wide credibility intervals for the parameters and relatively comparable posterior probabilities and we do not know whether that is because there are indeed no indirect effects in the data and only a weak preference for one hypothesis over another, or whether we did not have a sufficient number of observations to obtain stronger evidence.

Finally, while Bayesian methods allow for more intuitive interpretations of indirect effects, they do not provide any more evidence than classical methods that the causal order of effects is correctly specified. Like classical methods for mediation analysis, Bayesian mediation analysis also requires the assumptions of no unmeasured confounders of the relationship between the mediator and outcome in order to make causal claims about the indirect effect (Miočević, Gonzalez, Valente, & MacKinnon, 2018).

The methods described in the paper have yet to be tested in simulation studies to evaluate the required number of observations per phase for adequate power to detect the mediated effect. Furthermore, future research should develop guidelines and sensitivity analyses for evaluating the timing of the effect of the treatment on the mediator and the effect of the mediator on the outcome. Future research is also needed to identify optimal ways of incorporating autoregressive effects in context of Bayesian mediation analysis of SCEDs.

Data of a single participant presented in this study were selected from a larger SCED data set, but the same (or different) mediation hypothesis can be tested for the other participants. This data set also used a multiple baseline SCED design (different SCEDs were randomized to different lengths of the baseline A phase). As a consequence, when we replicate our mediation analysis across the other participants, internal and external validity increases (Kratochwill et al., 2010). Because frequentist estimates of the regression-based effect sizes have a known sampling distribution, their inverse squared standard error can be used as a weight in meta-analyses. By synthesizing effect sizes across cases and studies, more generalized decisions can be made

related to the effectiveness of an intervention, which is a significant contribution to evidence-based practices and policy decisions (Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2013a, 2013b; Moeyaert et al., 2014). However, when combining effect sizes across studies, standardization of the outcome score is needed as it is unlikely that the same scale is used across different studies. Future research should extend the methods described in this paper to include standardization, as described by van den Noortgate and Onghena (2007) for frequentist regression-based effect sizes.

CONCLUSION

This paper illustrated two Bayesian methods for mediation analysis using repeated measures of the potential mediator and outcome of interest from a single participant. The two methods were illustrated using data of a single participant from the Playskin Lift™ intervention, and the syntax is provided so researchers can apply the new methods to their data. The new methods have yet to be examined in simulation studies to find out the optimal number of repeated measures required for adequate power to detect the indirect effect in SCEDs. Testing mediators of intervention effects in SCEDs conducted in the fields of rehabilitation and communication disorders can add valuable information about the mechanisms through which interventions achieve (or do not achieve) the desired effects for a given client.

Acknowledgements

This research was supported by a grant from The Netherlands Organization for Scientific Research (NWO): NWO 406-12-001 (Fayette Klaassen), a grant from the European Commission Horizon 2020 research and innovation program under grant agreement No. 792119 (Milica Miočević), and a grant from the Institute of Education Sciences under grant agreement No. R305D190022 (Mariola Moeyaert)

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

Funding

This work was supported by The Netherlands Organization for Scientific Research (NWO) [NWO 406-12-001]; European Commission Horizon 2020 research and innovation program [792119]; Institute of Education Sciences [R305D190022].

REFERENCES

Babik, I., Cunha, A., Moeyaert, M., Hall, M., & Lobo, M. (2019). Feasibility and effectiveness of intervention with the Playskin Lift™ exoskeletal garment for infants at risk. *Physical Therapy*, *99*(6), 666–676.

Barlow, D. H., Nock, M., & Hersen, M. (2009). *Single case experimental designs : Strategies for studying behavior for change*. New York, NY: Pearson/Allyn and Bacon.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Béland, S., Klugkist, I., Raîche, G., & Magis, D. (2012). A short introduction into Bayesian evaluation of informative hypotheses as an alternative to exploratory comparisons of multiple group means. *Tutorials in Quantitative Methods for Psychology*, *8*(2), 122–126.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Cattell, R. B. (1952). The three basic factor-analytic research designs—their interrelations and derivatives. *Psychological Bulletin*, *49*(5), 499–520.

Center, B. A., Skiba, R. J., & Casey, A. (1985). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education*, *19*(4), 387–400.

Core Team, R. (2013). *R: A language and environment for statistical computing* [Computer software manual].

Vienna, Austria. Retrieved from http://www.R-project.org/

Craig, C. C. (1936). On the frequency function of xy. *Annals of Mathematical Statistics*, *7*, 1–15.

Darnieder, W. F. (2011). *Bayesian methods for data-dependent priors* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Gaynor, S. T., & Harris, A. (2008). Single-participant assessment of treatment mediators: Strategy description and examples from a behavioral activation intervention for depressed adolescents. *Behavior Modification*, *32*, 372–402.

Geuke, G., Maric, M., Miočević, M., Wolters, L. H., & de Haan, E. (2019). Testing mediators of youth intervention outcomes using single-case experimental designs (SCEDs). *New Directions for Child and Adolescent Development*, *167*, 39–64.

Gu, X., Hoijtink, H., Mulder, J., & van Lissa, C. (2019). *bain: Bayes Factors for Informative Hypotheses* (R package version 0.2.0).

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3* (3), 224–239.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*(4), 324–341.

Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton: Chapman & Hall/CRC.

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*, 539–556. Advance online publication.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165–179.

Judd, C. M., & Kenny, D. A. (1981). Process analysis. *Evaluation Review*, *5*(5), 602–619.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kazdin, A. E. (2011). *Single-case research designs* (2nd ed.). New York, NY: Oxford University Press.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, *10*(4), 477–493.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What Works Clearinghouse*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf

Lobo, M. A., Koshy, J., Hall, M. L., Erol, O., Cao, H., Buckley, J. M., … Higginson, J. (2016). Playskin Lift: Development and initial testing of an exoskeletal garment to assist upper extremity mobility and function. *Physical Therapy*, *96*(3), 390–399.

Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society*, *29*, 513–524.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Routledge.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1), 83–104.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*(1), 41–62.

Manolov, R., & Moeyaert, M. (2017). How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. *Behavior Modification*, *41*(2), 179–228.

Maric, M., Heyne, D. A., MacKinnon, D. P., Van Widenfelt, B. M., & Westenberg, P. M. (2013). Cognitive mediation of cognitive-behavioural therapy outcomes for anxiety-based school refusal. *Behavioural and Cognitive Psychotherapy*, *41*(5), 549–564.

Maric, M., Prins, P. J. M., & Ollendick, T. H. (Eds.). (2015). *Moderators and mediators of youth treatment outcomes*. New York: Oxford University Press.

Maric, M., Wiers, R. W., & Prins, P. J. M. (2012). Ten ways to improve the use of statistical mediation analysis in the practice of child and adolescent treatment research. *Clinical Child and Family Psychology Review*, *15*, 177–191.

McNeish, D. M. (2016). Using data-dependent priors to mitigate small sample bias in latent growth models: A discussion and illustration using Mplus. *Journal of Educational and Behavioral Statistics*, *41*(1), 27–56.

Miočević, M., Gonzalez, O., Valente, M. J., & MacKinnon, D. P. (2018). A tutorial in Bayesian potential outcomes mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(1), 121–136.

Miočević, M., MacKinnon, D. P., & Levy, R. (2017). Power in Bayesian mediation analysis for small sample research. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(5), 666–683.

Miočević, M., & Van de Schoot, R. (2019). Gentle introduction to Bayesian statistics. In J. Edlund & A. L. Nichols (Eds.), *Advanced research methods and statistics for the behavioral and social sciences* (pp. 289–308). Cambridge, UK: Cambridge University Press.

Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2014). The influence of the design matrix on treatment effect estimates in the quantitative analyses of single-case experimental design research. *Behavior Modification*, *38*(5), 665–704.

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2013a). The three-level synthesis of standardized single-subject experimental data: A monte carlo simulation study. *Multivariate Behavioral Research*, *48*(5), 719–748.

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2013b). Three-level analysis of single-case experimental data: Empirical validation. *The Journal of Experimental Education*, *82* (1), 1–21.

Ng, M. Y., & Weisz, J. R. (2016). Annual research review: Building a science of personalized intervention for youth mental health. *Journal of Child Psychology and Psychiatry*, *57*, 216–236.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*(4), 357–367.

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single case research. *Exceptional Children*, *75*(2), 135–150.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-u. *Behavior Therapy*, *42*(2), 284–299.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling* (version 4.3.0). Retrieved from http://mcmc-jags.sourceforge.net/

Plummer, M. (2018). *rjags: Bayesian Graphical Models using MCMC* (R package version 4-8).

Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., … Magnusson, A. (2018). *coda: Output analysis and diagnostics for MCMC* (R package version 0.19-2).

Schlosser, R. W., Lee, D. L., & Wendt, O. (2008). Application of the Percentage of Non-overlapping Data in systematic reviews and meta-analyses: A systematic review of reporting characteristics. *Evidence-Based Communication Assessment and Intervention*, *2*, 163–187.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, *8*(2), 24–33.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*, *52*(2), 123–147.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta- analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 188–196.

Sinharay, S. (2004). Experiences with Markov chain Monte Carlo convergence assessment in two psychometric examples. *Journal of Educational and Behavioral Statistics*, *29*(4), 461–488.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*(4), 510–550.

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected tau. *Behavior Modification*, *41*(4), 427–467.

van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*(3), 325–346.

van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 1–10.

van den Noortgate, W., & Onghena, P. (2007). The aggregations of single-case research using hierarchical linear models. *The Behavior Analyst Today*, *8*(2), 196–209.

Vannest, K. J., Peltier, C., & Haas, A. (2018). Results reporting in single case experiments and single case meta-analysis. *Research in Developmental Disabilities*, *79*, 10–18.

Wiessenekker, M. (2019). *Efficacy of child and adolescent therapies: A meta-analysis of single-case studies* (Master thesis). University of Amsterdam, Amsterdam, The Netherlands.

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, *44*(1), 18–28.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301–322.