Research in Developmental Disabilities xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

The impact of response-guided baseline phase extensions on treatment effect estimates

Seang-Hwane Joo[a,*], John M. Ferron[a], S. Natasha Beretvas[b], Mariola Moeyaert[c], Wim Van den Noortgate[d]

[a] Department of Educational and Psychological Studies, University of South Florida, Tampa, FL, USA
[b] Department of Educational Psychology, University of Texas, Austin, TX, USA
[c] Department of Educational Psychology and Methodology, State University of New York, Albany, NY, USA
[d] Faculty of Psychological and Educational Sciences, Katholieke Universiteit Leuven, Belgium

ARTICLE INFO

Number of reviews completed is 2
The paper is part of a special issue on Single case experimental studies

Keywords:
Single-case research
Response-guided
Multilevel modeling
Monte Carlo study

ABSTRACT

Background: When developmental disabilities researchers use multiple-baseline designs they are encouraged to delay the start of an intervention until the baseline stabilizes or until preceding cases have responded to intervention. Using ongoing visual analyses to guide the timing of the start of the intervention can help to resolve potential ambiguities in the graphical display; however, these forms of response-guided experimentation have been criticized as a potential source of bias in treatment effect estimation and inference.

Aims and methods: Monte Carlo simulations were used to examine the bias and precision of average treatment effect estimates obtained from multilevel models of four-case multiple-baseline studies with series lengths that varied from 19 to 49 observations per case. We varied the size of the average treatment effect, the factors used to guide intervention decisions (baseline stability, response to intervention, both, or neither), and whether the ongoing analysis was masked or not.

Results: None of the methods of responding to the data led to appreciable bias in the treatment effect estimates. Furthermore, as timing-of-intervention decisions became responsive to more factors, baselines became longer and treatment effect estimates became more precise.

Conclusions: Although the study was conducted under limited conditions, the response-guided practices did not lead to substantial bias. By extending baseline phases they reduced estimation error and thus improved the treatment effect estimates obtained from multilevel models.

What this paper adds

This Monte Carlo study contributes to the single-case design literature by addressing the concern with response-guided experimentation. The study examined the bias and precision of treatment effect estimates obtained from multilevel models under conditions with response-guided experimentation. It was found that under the simulated conditions, response-guided experimentation did not result in substantial bias of the treatment effect estimates using multilevel models.

* Corresponding author.
E-mail address: sjoo@mail.usf.edu (S.-H. Joo).

https://doi.org/10.1016/j.ridd.2017.12.018
Received 26 July 2017; Received in revised form 1 December 2017; Accepted 18 December 2017

1. Introduction

Single-case researchers frequently adopt a form of response-guided experimentation where decisions about the design of the study are made based on an ongoing visual analysis (e.g. Gast, 2009; Kazdin, 2010). For example, multiple-baseline researchers may delay the start of intervention until the data document a stable baseline pattern so that baseline trends can be reliably extended, or the researchers may wait for a case in a multiple-baseline design to respond to intervention prior to intervening with the next case. Using ongoing visual analyses to guide the timing of interventions can help to resolve what would be ambiguities in the graphical display and thus increase the analyst's sensitivity to detecting effects in the graphical display.

Although well intentioned, these response-guided experimental strategies may bias the intervention effect estimates or inferences. These response-guided strategies have been compared to the strategy determining sample size in a group comparison study by repeatedly testing for differences as the sample is gathered, an approach that is known to increase Type I error rates (Allison, Franklin, & Heshka, 1992). A simulation study showed an increase in the number of false detections of effects by randomization tests when the single-case data were gathered in a response-guided manner (Ferron, Foster-Johnson, & Kromrey, 2003). Furthermore, a study of visual analysis of graphs of randomly generated observations showed that when the line separating the hypothetical baseline and treatment phases was placed after a stable set of observations, as opposed to placed randomly, that visual analysts were more likely to incorrectly conclude that there were effects (Todman & Dugard, 1999). These concerns led to the development of a method of masking graphs in an ongoing visual analysis (Ferron & Jones, 2006).

1.1. Masked visual analysis

Masking graphs in a visual analysis, or masked visual analysis (MVA), was first proposed in single-case research by Mawhinney and Austin (1999). In MVA, the transition points between different phases (i.e., from baseline to treatment) are purposely concealed from the visual analysts, and the visual analysts are tasked with determining when the treatment was initiated. If a single-case research design involves multiple participants such as multiple-baseline design, the visual analysts are tasked with deciding the intervention initiation points for each case. Later, response-guided or ongoing MVA was developed with more detailed procedures to ensure control over Type I error rates (Ferron & Jones, 2006; Ferron & Levin, 2014).

For the response-guided MVA method (Ferron & Jones, 2006), a single-case research team is divided into two separate groups; an intervention team and an analysis team. The intervention team is responsible for interactions with the cases and data collection, whereas the analysis team is responsible for using visual analysis of a masked graph to make decisions about baseline stability and response to intervention. The graphs are 'masked' because information about the independent variable (i.e., which case will enter intervention first and whether the observation is part of a baseline or treatment phase) is not marked on the graph or made explicit to the analysis team. The analysis team analyses the stability of the masked data one session at a time and the data collection continues until all cases show a stable pattern for the baseline observations. Once stability is obtained, the analysis team directs the intervention team to randomly select a case to begin the intervention phase. The intervention team does so and the information about which case is in the intervention phase is not given to the analysis team. The collected data are still masked and sent to the analysis team, and the analysis team continues analyzing the data until there is sufficient data to demonstrate that one case has initiated the intervention. Then the second case is randomly selected to begin the intervention. This process continues until all cases participate in the intervention. By computing the probability of selecting the intervention order for all cases when there is no true effect this method theoretically controls the Type I error rate (Ferron & Jones, 2006). In addition, a recent Monte Carlo study has shown that the response-guided MVA controls Type I errors to the nominal level (Ferron, Joo & Levin, 2017).

If single-case researchers were interested in integrating several studies, including a meta-analysis of single-case studies (Shadish, 2014), they could do so by combining probabilities (Rosenthal, 1978; Solmi & Onghena, 2014). Although probability estimates of response-guided experimentation using MVA are accurate, MVA was developed for and is limited to the estimation of probabilities. Researchers who want to estimate and synthesize effect sizes, rather than probabilities, must turn to other analyses that may or may not be negatively impacted by response-guided experimentation. One widely used approach for obtaining effect size estimates of single-case data that is of particular interest in the current study involves the use of multilevel models (Van den Noortgate & Onghena, 2003a, 2003b).

1.2. Multilevel models

Multilevel models have been utilized for analyzing single-case data because they take variability within- and between-cases into account when estimating the treatment effect (e.g., Van den Noortgate & Onghena, 2003a,b). In principle, multilevel models are specifically developed for analyzing hierarchically structured data, where lower-level units are nested in higher-level units. Hierarchical structured data are often found in behavioral and social science studies. For example, in educational settings, students are nested in classes and classes are nested in schools. Similarly, multiple-baseline data can be considered as hierarchical because the repeated observations are nested within cases.

Multilevel models for multiple-baseline studies are also advantageous over multiple single-level models. For example, multilevel models provide not only individual cases' treatment effect estimates, but also the average treatment effect estimate across cases. Although multilevel models were developed based on the assumption of a relatively large number of second-level units, a number of simulation studies have shown that multilevel models with small sample size adjustments produce unbiased treatment effect estimates and reliable statistical inferences with as few as four cases (e.g., Ferron, Bell, Hess, Rendina-Gobioff, & Hibbard, 2009; Ferron,