Journal of Contextual Behavioral Science 3 (2014) 51–64

Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

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

Randomization tests for single-case experiments: State of the art, state of the science, and state of the application

Mieke Heyvaert [a,b,*], Patrick Onghena [a]

[a] Faculty of Psychology and Educational Sciences, KU Leuven, Belgium
[b] Postdoctoral Fellow of the Research Foundation - Flanders, Belgium

ARTICLE INFO

Article history:
Received 11 January 2013
Received in revised form
23 August 2013
Accepted 24 October 2013

Keywords:
Single-subject experimental design
Randomization test
Phase design
Alternation design
Statistical software

ABSTRACT

A single-case experimental design is a research design that can be used to evaluate the effect of an intervention for a single entity. There are two important schedules to include randomization into the design of single-case experiments: phase designs and alternation designs. We present these two schedules and provide a detailed example for each schedule. For both examples, we illustrate the use of a free software package that assists researchers in designing and analyzing single-case experiments using randomization tests. Furthermore, we discuss several additions (simultaneous and sequential replication designs; meta-analysis of single-case experimental studies) and alternatives (statistical and visual analysis methods).

1. State of the art

1.1. Single-case experiments

A single-case experiment (SCE) is an experiment that can be used to evaluate the effect of an independent variable for a single entity, for example a single patient, a single therapist–patient dyad, or a single family. An experimental approach is used: the independent variable is manipulated by the experimenter. The dependent variable is measured repeatedly for this entity under different levels of the independent variable. For example, in research on interventions for reducing challenging behavior among persons with autism, a simple SCE involves one person with autism showing certain challenging behavior (e.g., aggressive episodes) that is treated by an intervention (e.g., a behavioral intervention). The number of aggressive episodes of that person is repeatedly measured before (=baseline phase), during (=intervention phase), and after (=withdrawal phase) the behavioral intervention, during a certain period of time (e.g., 8 weeks). Based on the difference between the number of aggressive episodes per day (i.e., the dependent variable) in the baseline phase versus in the intervention phase (i.e., the independent variable), the effect of the behavioral intervention on the challenging behavior can be determined for this person.

SCEs have a long history in the behavioral sciences, with pioneers like Ebbinghaus, Fechner, Stratton, and Wundt in the 19th century, and Pavlov, Sidman, and Skinner in the 20th century (Barlow, Nock, & Hersen, 2009; Blampied, 1999; Kratochwill & Mace, 1984). We refer the reader interested in the historical and philosophical foundations of SCEs to the chapter of Ittenbach and Lawhead (1996). SCE research is often applied in several subdisciplines of the behavioral sciences, such as clinical psychology, counseling psychology, neuropsychology, school psychology, psychopharmacology, social work, education, and special education (e.g., Barlow et al., 2009; Dugard, File, & Todman, 2012; Heyvaert, Maes, Van den Noortgate, Kuppens, & Onghena, 2012; Horner et al. 2005; Kratochwill & Levin, 1992, 2010; Maggin, O'Keeffe, & Johnson, 2011; Perdices & Tate, 2009; Rapoff & Stark, 2008; Zhan & Ottenbacher, 2001).

There are several reasons to explain the growing interest in and popularity of SCE research in the behavioral sciences. We list three basic reasons and five pragmatic reasons. A first basic reason is that the focus of SCE research on the individual case parallels the care for the individual patient in applied clinical settings (cf. Hayes, 1981). SCEs render results that are easily understood by clinicians who work at the level of individual patients (Rapoff & Stark, 2008). A second basic reason is that sometimes one SCE is sufficient to refute a hypothesis, or to confirm the presence of a phenomenon (Edelson, 1985; Onghena, 2005). A third basic reason concerns the growing importance of evidence-based practice, accountability, and of evaluating interventions at the level of the

* Corresponding author at: Methodology of Educational Sciences Research Group, Andreas Vesaliusstraat 2, Box 3762, B-3000 Leuven, Belgium. Tel.: +32 1632 6265; fax: +32 1632 6200.
E-mail address: mieke.heyvaert@ppw.kuleuven.be (M. Heyvaert).

2212-1447/$ - see front matter © 2013 Association for Contextual Behavioral Science. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcbs.2013.10.002

individual participant (Barlow, Hayes, & Nelson, 1985; Horner et al., 2005; Zhan & Ottenbacher, 2001).

A first pragmatic reason is that SCE research is one of the only eligible design options if rare or unique conditions are involved (e.g., a patient with a rare psychological disorder). A second pragmatic reason is that it is sometimes sufficient to use SCE research (or a few replications) when the between-case variability is very small, for instance for very homogeneous groups. In that case, it can be far more interesting to study the within-case variability than the between-case variability. A third pragmatic reason for the growing interest in SCE research in the behavioral sciences is its feasibility and flexibility (Hacker, 1980; McReynolds & Thompson, 1986). A fourth pragmatic reason for the growing impact of SCE research is the present-day availability of several methods for the design and analysis of SCE research (cf. Barlow et al., 2009; Dugard et al., 2012; Edgington & Onghena, 2007, pp. 225–259; Franklin, Allison, & Gorman, 1996; Kazdin, 2011; Parker, Vannest, & Davis, 2011), as well as software for the design and analysis of SCE research (e.g., Bulté & Onghena, 2008, 2009, 2012; Koehler & Levin, 2000; Van den Noortgate & Onghena, 2003a, 2007). A fifth pragmatic reason for the popularity of SCE research is its small-scale design: a small-scale design is less harmful and less costly than a large-scale design.

1.2. Randomization

In group-comparison studies, randomization concerns the random assignment of participants to control and treatment groups. In SCEs, randomization concerns the random assignment of measurement times to baseline and treatment conditions. Applying randomization increases the methodological quality of a study, whether it is a group-comparison study or an SCE. For group-comparison studies, the 'randomization' element of the randomized controlled trial (RCT) design led to its status of the gold standard for evaluating the efficacy of treatments. Accordingly, reporting tools for RCTs and tools for evaluating the methodological quality of RCTs include items on randomization (e.g., Altman et al., 2001; Maher, Sherrington, Herbert, Moseley, & Elkins, 2003). Concerning SCEs as well, the advantages of including randomization in the design of the experiment are described in several books and articles (e.g., Barlow et al., 2009; Dugard et al., 2012; Edgington & Onghena, 2007; Kazdin, 2011; Kratochwill & Levin, 1992, 2010). Accordingly, we notice that some recently developed SCE reporting tools and tools for evaluating the methodological quality of SCEs include items on randomization too (e.g., Romeiser-Logan, Hickman, Harris, & Heriza, 2008; Task Force on Evidence-Based Interventions in School Psychology, 2003).

We argue that randomization is as important for SCEs as it is for group-comparison studies. A researcher conducting a *nonrandomized* SCE has to be very careful when attributing changes in the outcomes to changes in the treatment conditions, because it is possible that the observed response trend of the participant might have been there *without* any treatment manipulation. However, in a *randomized* SCE the random assignment of measurement times to baseline and treatment conditions provides control for sources of bias (Edgington, 1987, 1996). For SCEs the randomization process can render control over—both known and unknown—confounding variables that are time-related, such as maturation effects (Onghena, 2005). As such, randomization can increase the methodological quality of an SCE by strengthening the internal validity of SCEs (Edgington & Onghena, 2007; Kratochwill & Levin, 2010; Onghena & Edgington, 2005). However, there are limits on the control established, depending on the SCE design and the applied randomization procedure. When measurement times are randomly assigned to baseline and treatment conditions (cf. Section 2.3), intervention assignment is unrelated to time. When only the intervention start points are randomly determined (cf. Section 2.1), there is still some relationship between treatment assignment and time.

1.3. Randomization tests

Randomization tests (RTs) are statistical significance tests based on the random assignment of experimental units to treatments (Edgington & Onghena, 2007). They are used to test hypotheses about treatment effects. Using RTs increases the methodological quality of an SCE by improving the statistical conclusion validity of SCEs (Edgington & Onghena, 2007; Kratochwill & Levin, 2010; Onghena & Edgington, 2005). Let us look in more detail at the steps involved in RT (cf. Edgington & Onghena, 2007; Ferron & Ware, 1995). As a prerequisite, the randomization method requires a researcher to design his experimental study so that it involves random assignment. A priori, all possible random assignments are recorded. Randomly one of these assignments is chosen: our actual experiment will follow this assignment. Then, the researcher chooses an appropriate test statistic, runs the experiment, collects the data, and calculates the test statistic based on the collected data. We then look at all possible random assignments that were recorded at the beginning of our study: each of these assignments involved a different way of dividing the data. We calculate the chosen test statistic for each of these assignments. Based on this, we can determine the statistical significance of our test statistic: we look where our obtained test statistic falls within the distribution of all possible test statistic values. The p value of the RT is calculated as the proportion of possible test statistic values that is as extreme, or even more extreme, than the value of the test statistic based on the collected data.

2. Randomization tests for single-case experiments

There are two important schedules to include randomization into the design of SCEs: phase designs and alternation designs. In the former, the moment of phase change is randomly determined. In the latter, the treatment alternation is randomly determined. We now consecutively present these two schedules (Sections 2.1 and 2.3), and describe a detailed example for each schedule (Sections 2.2 and 2.4).

2.1. Randomization tests for single-case phase designs

When using phase designs, all measurement times are divided into phases and several consecutive measurements are taken in each phase (Edgington, 1975, 1980; Onghena, 1992). The AB design is the most basic phase design: the A phase is the baseline or control phase and the B phase is the intervention phase. There exist several extensions and variations of the AB phase design: ABA designs (i.e., reversal and withdrawal designs), BAB designs, ABAB designs, ABABAB designs, and so on (Barlow et al., 2009, pp. 135–161; Kazdin, 2011, pp. 121–143). In Fig. 1, three examples of single-case phase designs are presented. It is also possible that more than one treatment is evaluated by means of a single-case design. With the B phase representing the consecutive measurements taken under the first treatment and the C phase representing the consecutive measurements taken under the second treatment, several other extensions and variations are possible, such as ABACA designs, ABCB designs, and so on (Barlow et al., 2009, pp. 162–175).

In all these phase designs, the sequence of the phases is fixed before the start of the actual experiment. The incorporation of randomization in phase designs concerns the moment of phase change. For instance, in an AB design the randomization concerns the moment when the intervention (B phase) starts. In an ABAB design the randomization concerns the moment when the first