ELSEVIER

Journal of Clinical Epidemiology 67 (2014) 247–253

Journal of Clinical Epidemiology

In randomization we trust? There are overlooked problems in experimenting with people in behavioral intervention trials☆

Jim McCambridge[a,*], Kypros Kypri[b], Diana Elbourne[c]

[a]Department of Social & Environmental Health, Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK

[b]Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, Newcastle, Australia

[c]Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London, UK

Accepted 13 September 2013; Published online 4 December 2013

Abstract

Objectives: Behavioral intervention trials may be susceptible to poorly understood forms of bias stemming from research participation. This article considers how assessment and other prerandomization research activities may introduce bias that is not fully prevented by randomization.

Study Design and Setting: This is a hypothesis-generating discussion article.

Results: An additivity assumption underlying conventional thinking in trial design and analysis is problematic in behavioral intervention trials. Postrandomization sources of bias are somewhat better known within the clinical epidemiological and trials literatures. Neglect of attention to possible research participation effects means that unintended participant behavior change stemming from artifacts of the research process has unknown potential to bias estimates of behavioral intervention effects.

Conclusion: Studies are needed to evaluate how research participation effects are introduced, and we make suggestions for how research in this area may be taken forward, including how these issues may be addressed in the design and conduct of trials. It is proposed that attention to possible research participation effects can improve the design of trials evaluating behavioral and other interventions and inform the interpretation of existing evidence.

Keywords: Behavior; Trials; Bias; Research participation; Intervention; Hawthorne effect

1. Introduction

Randomized controlled trials (RCTs) are widely accepted as the most rigorous research designs for the evaluation of the effects of interventions. Behavioral intervention trials are studies in which the primary purpose is to evaluate attempts to influence behavior or the consequences of any resultant behavior change. They are important to public health as lifestyle behavioral risk factors contribute strongly to a wide range of health problems [1]. Data from our best behavioral intervention trials may not, however, be as robust as we currently believe, and it has been suggested that research participation may account for more observed change than evaluated interventions [2]. It has long been known that participants may react in unintended ways to being studied and this may lead to change [3]. It is suggested that this entails largely overlooked potential for bias in behavioral intervention trials. Valid inferences about the true effects of behavioral interventions are hampered by our inability to identify and rule out alternative explanations for behavior change. These concerns have much wider relevance as almost all trials and other types of human research depend on the cooperation of their participants, which may be unwittingly influenced by the way studies are conducted.

2. Assessment and other aspects of research participation may change behavior

Taking part in trials typically involves both recruitment and baseline assessment activities before randomization, and subsequently exposure to study conditions and assessment at follow-up. Any or all of these research activities

Funding: The work on this article was supported by a Wellcome Trust Research Career Development Fellowship in Basic Biomedical Science to the first author (WT086516MA). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

* Corresponding author. Tel.: +44-20-7927-2945.

E-mail address: Jim.McCambridge@lshtm.ac.uk (J. McCambridge).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2013.09.004

What is new?

- An additivity assumption underlying conventional trial design and analysis is problematic in behavioral intervention trials.
- Pre and postrandomization research participation effects may interact with evaluated interventions.
- Randomization does not fully prevent the introduction of bias via these mechanisms.
- New conceptual and empirical work is needed to better understand these problems.
- Research artifacts in other types of trials should also be amenable to control.

may influence participant cognitions, emotions, and behavior. Formally signing a consent form, for example, may lead to or strengthen commitment to behavior change. Questions answered for research assessment purposes may stimulate new thinking about the behavior, which also may be a prelude to action [4,5].

It is difficult to point to any well-established coherent body of literature investigating these issues. There exist, however, somewhat disparate strands of relevant research, and thinking about research, which relate to different parts of the research process being investigated, or have their origins in specific disciplines or research contexts, or are concerned with specific methodological problems in research. For example, assessment reactivity effects in trials of brief alcohol interventions jeopardize the safety of inferences made because although reactivity effects may be small, the effects of the interventions being evaluated are also small [6]. In this field, because assessment is an integral component of the brief interventions being evaluated, research assessments produce contamination in the form of unwitting exposure of the control group to intervention content [7].

There is a plethora of labels and constructs that have been developed to describe and study similar phenomena. For example, within health psychology, assessment reactivity is conceptualized as ‘‘mere measurement,’’ ‘‘question-behavior,’’ or ‘‘self-generated validity’’ effects [4,5,8]. Synthesizing this type of literature is challenging as many findings have been generated incidentally to the main purposes of the research being undertaken. The idea that being assessed itself influences behavior has, however, been established in the literature for approximately one 100 years [3]. The Hawthorne effect, usually taken to mean that monitoring of a behavior for research purposes changes performance of that behavior, is approximately 60 years old [9]. This is probably the most recognizable term used to describe the effects of being assessed across disciplines [10–12].

Around the same time, an alteration to basic experimental design, the Solomon four-group design, was developed to allow quantification of the size of baseline assessment effects and to control for them [3]. Campbell [13] subsequently proposed that assessments may interact with interventions to either strengthen or weaken observed effects, thus producing biased estimates of effects. The construct of ‘‘demand characteristics’’ [14,15] was subsequently introduced in psychology, referring to the ways in which study participants adjust their responses according to their perceptions of the implicit preferences or expectations of researchers, to be ‘‘good subjects’’ [16].

Four recent systematic reviews summarize and evaluate empirical data on assessment reactivity in brief alcohol intervention trials [7], the Hawthorne effect [17], applications of Solomon four-group designs [18], and demand characteristic studies in nonlaboratory settings [19]. Collectively, these reviews demonstrate that being assessed can impact on behaviors, with small effects usually having been identified, albeit inconsistently, on both self-reported and objectively ascertained outcomes. These are due to being interviewed, completing questionnaires, or being observed. These four reviews do not, however, provide strong evidence of assessment effects as there were substantial weaknesses in the primary studies. Strong and consistent themes to emerge from these studies are the need for a new generation of primary studies dedicated to estimate the size of assessment and other possible research participation effects, and the mechanisms of their production, and the circumstances in which they occur.

3. Overlooked prerandomization sources of bias in behavioral intervention trials

The example provided in Box 1 suggests that in such cases, reliable effect estimation has been precluded and thus that randomization has not protected against some form of bias. The reason for this is the violation of a key assumption in conventional trial design and analysis on which the capacity of randomization to prevent bias depends. This is the additivity assumption [20] that the effects of the intervention being evaluated are independent of any possible prerandomization effects of research participation. In simple terms, this implies that it does not matter whether assessment changes behavior or participants react to some other aspect of being researched before randomization because with sufficiently large numbers, randomization guarantees between-group equivalence and ensures that randomized groups differ only in outcomes as a function of the intervention being studied.

Attention has previously been drawn to this additivity assumption in pharmacological trials in mental health [20], although its implications are rarely considered more widely. This assumption is untenable in behavioral intervention trials, most obviously where the research and intervention