Journal of Clinical Epidemiology 82 (2017) 94–102

The regression discontinuity design showed to be a valid alternative to a randomized controlled trial for estimating treatment effects

Iris L. Maas[a], Sandra Nolte[a,b], Otto B. Walter[a], Thomas Berger[c], Martin Hautzinger[d], Fritz Hohagen[e], Wolfgang Lutz[f], Björn Meyer[g,h], Johanna Schröder[i], Christina Späth[e], Jan Philipp Klein[e], Steffen Moritz[i], Matthias Rose[a,j,*]

[a]*Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany*
[b]*Population Health Strategic Research Centre, School of Health and Social Development, Deakin University, 221 Burwood Highway, Burwood, Victoria 3125, Australia*
[c]*Department of Clinical Psychology and Psychotherapy, University of Bern, Fabrikstrasse 8, Bern 3012, Switzerland*
[d]*Department of Psychology, Clinical Psychology and Psychotherapy, Eberhard Karls University Tübingen, Schleichstrasse 4, Tübingen 72074, Germany*
[e]*Department of Psychiatry and Psychotherapy, Lübeck University, Ratzeburger Allee 160, Lübeck 23538, Germany*
[f]*Department of Psychology, University of Trier, Trier 54286, Germany*
[g]*GAIA AG, Gertigstrasse 12-14, Hamburg 22303, Germany*
[h]*Department of Psychology, City University London, Northampton Square, London EC1V 0HB, UK*
[i]*Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany*
[j]*Outcomes Measurement Science, Department of Quantitative Health Sciences, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA*

Accepted 9 November 2016; Published online 16 November 2016

Abstract

Objectives: To compare treatment effect estimates obtained from a regression discontinuity (RD) design with results from an actual randomized controlled trial (RCT).

Study Design and Setting: Data from an RCT (EVIDENT), which studied the effect of an Internet intervention on depressive symptoms measured with the Patient Health Questionnaire (PHQ-9), were used to perform an RD analysis, in which treatment allocation was determined by a cutoff value at baseline (PHQ-9 = 10). A linear regression model was fitted to the data, selecting participants above the cutoff who had received the intervention (n = 317) and control participants below the cutoff (n = 187). Outcome was PHQ-9 sum score 12 weeks after baseline. Robustness of the effect estimate was studied; the estimate was compared with the RCT treatment effect.

Results: The final regression model showed a regression coefficient of −2.29 [95% confidence interval (CI): −3.72 to −.85] compared with a treatment effect found in the RCT of −1.57 (95% CI: −2.07 to −1.07).

Conclusion: Although the estimates obtained from two designs are not equal, their confidence intervals overlap, suggesting that an RD design can be a valid alternative for RCTs. This finding is particularly important for situations where an RCT may not be feasible or ethical as is often the case in clinical research settings.

Keywords: Regression discontinuity design; Randomized controlled trials; Causal inference; Nonrandom group assignment; Depression; PHQ-9

1. Introduction

Regression discontinuity (RD) is a quasi-experimental design used to make causal inference and estimate treatment effects. It has been mostly used in the fields of economics and educational or social sciences [1], but it is rarely applied in medical research. The design was introduced by Thistlethwaite and Campbell in 1960 [2] and aims to solve the problem of estimating causal effects when subjects are nonrandomly assigned to treatment groups. This applies particularly to situations where randomization is not ethical or feasible, which is often

Funding: This study was supported by the German Federal Ministry of Health, II A 5 - 2512 FSB 052. The funding body had no role in the design of the study, data collection, analysis, or interpretation of the data. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflict of interest: J.P.K. received payments for presentations, workshops, and books on psychotherapy for chronic depression and on psychiatric emergencies. B.M. is employed as research director at GAIA AG, the company that developed, owns, and operates the Internet intervention investigated in this trial. All the other authors report no relationships with commercial interests.

The EVIDENT trial is registered at ClinicalTrials.gov under identifier NCT01636752.

* Corresponding author. Tel.: +49 30 450 553 002; fax: +49 30 450 553 900.

E-mail address: matthias.rose@charite.de (M. Rose).

http://dx.doi.org/10.1016/j.jclinepi.2016.11.008

What is new?

Key findings

- Estimates from the regression discontinuity (RD) were comparable to treatment effect estimates from the randomized controlled trial (RCT).

What this adds to what was known?

- Estimates from an RD design are rarely compared to real-life data.

What is the implication and what should change now?

- RD is a technique suited for many cases in medicine and solves the problem of randomization when an RCT is not ethical or feasible. Assignment to treatment group is based on some cutoff score at baseline.
- The RD design appears suitable for use in clinical research.

the case in clinical settings. The idea of an RD design is that it estimates unbiased causal effects of a treatment while eliminating potential confounding effects [3]. In detail, within an RD design, subjects are not randomly assigned to treatment or control groups, but group membership depends on a threshold of a certain baseline variable. For example, subjects above the threshold are assigned to the intervention group, whereas people below the threshold are assigned to the control group. This leads to a design where groups are not exchangeable, with the exception of those subjects closest on either side to the cutoff. Nevertheless, causal inference can still be made under certain assumptions by estimating the treatment effect close to the chosen threshold. To assure random assignment closely around this prespecified threshold, the approach is based on the assumption of random noise in the measurement of the assignment variable [4].

Consider the example given by Thistlethwaite where scholars are selected into an honors program based on their grades [2]. As there is always some random measurement error (random noise) when grades are awarded, scholars with similar but not identical grades will not differ much from each other. Because of this random noise, scholars with similar grades—while falling just on opposite sides of the threshold—will be exchangeable. The closer to the threshold, the more likely is the counterfactual (CF) that they could also have ended up on the other side of the threshold. Hence, the assignment of groups just around the threshold can be considered random, given that random noise is present. Consequently, it can be assumed that no unobserved confounding occurs at that point, leading to the idea that regression estimates at that particular point can be considered valid treatment effects estimates. Strictly speaking, the RD design is not a statistical method but rather an alternative approach to designing a study and interpreting obtained estimates. It does not necessarily require complicated statistical tools, but instead, it relies heavily on the knowledge and description of the assignment rule.

The RD design has already been critically evaluated with use of simulated data [5], and estimates obtained by randomized controlled trials (RCTs) have been compared before with other designs (e.g., [6]), but there are not many recent studies that compare RCTs to RD in clinical research [7–9]. To date, the method has not yet been applied widely in the medical literature, despite plenty of possibilities and unexploited data sets to which this design could be applied [10]. However, over the last few years, RD has gained increasing attention in the medical field, particularly in health policy [11]. Although there have been recent articles that discussed RD, those were generally explaining the method itself or how to report on it [3,12], while few actually applied the method [7,13]. As pointed out before, the RD may be a useful design [10], but its application may need some more attention to show its usefulness. Therefore, this renewed interest in RD should be supported by comparing obtained estimates to real-life data. It remains to be shown that an RD design can be a valid and feasible alternative to randomization in clinical research. A solid proof of whether this approach works is to use data from an RCT design, in which not only CFs can be shown, but estimates can further be compared to real data.

This paper is aimed at demonstrating the usefulness of the RD design in a clinical setting. For this, we evaluate the applicability and performance of this design by checking its obtained estimates in comparison with real-life data from a recent RCT using self-report depression data from the EVIDENT trial [14,15]. Using data from the EVIDENT trial offers the unique opportunity to not only test the RD design by using a well-established cutoff, but in addition, the value of RD designs in clinical settings can be demonstrated as actual data on the CFs are available.

2. Methods

2.1. Description of RCT data

In this study, real-life data from the EVIDENT trial were used [15]. The aim of the trial was to determine the effectiveness of an Internet intervention (Deprexis) on participants with mild to moderate depressive symptoms, measured by the Patient Health Questionnaire (PHQ-9). The PHQ-9 is a widely used clinical tool for measuring depressive symptoms, with higher scores reflecting more severe symptoms. It has been validated extensively, showing good psychometric properties [16,17]. In the EVIDENT trial, participants were eligible when they