Regression Discontinuity Designs in Health
A Systematic Review

Michele Hilton Boon, Peter Craig, Hilary Thomson, Mhairi Campbell, and Laurence Moore

Background: Regression discontinuity designs are non-randomized study designs that permit strong causal inference with relatively weak assumptions. Interest in these designs is growing but there is limited knowledge of the extent of their application in health. We aimed to conduct a comprehensive systematic review of the use of regression discontinuity designs in health research.

Methods: We included studies that used regression discontinuity designs to investigate the physical or mental health outcomes of any interventions or exposures in any populations. We searched 32 health, social science, and gray literature databases (1 January 1960 to 1 January 2019). We critically appraised studies using eight criteria adapted from the What Works Clearinghouse Standards for regression discontinuity designs. We conducted a narrative synthesis, analyzing the forcing variables and threshold rules used in each study.

Results: The literature search retrieved 7658 records, producing 325 studies that met the inclusion criteria. A broad range of health topics was represented. The forcing variables used to implement the design were age, socioeconomic measures, date or time of exposure or implementation, environmental measures such as air quality, geographic location, and clinical measures that act as a threshold for treatment. Twelve percent of the studies fully met the eight quality appraisal criteria. Fifteen percent of studies reported a prespecified primary outcome or study protocol.

Conclusions: This systematic review demonstrates that regression discontinuity designs have been widely applied in health research and could be used more widely still. Shortcomings in study quality and reporting suggest that the potential benefits of this method have not yet been fully realized.

Keywords: Evaluation; Intervention; Natural experiment; Policy; Regression discontinuity; Systematic review

Regression discontinuity designs are of increasing interest in epidemiology as a method of analyzing natural experiments, evaluating interventions, and supporting causal inference in the absence of randomized trials.[1] Thistlethwaite and Campbell (1960) first proposed the regression discontinuity design based on the insight that, given an eligibility rule based on a cutoff value for a continuous variable whose measurement cannot be precisely manipulated by participants or administrators, treatment assignment for participants close to the threshold value will be effectively random; therefore, the causal effect of the treatment can be estimated by comparing outcomes for groups just above and just below the cutoff, without any bias due to unobserved confounding.[2,3] Regression discontinuity designs are attractive because they allow the evaluation of causal effects of interventions or exposures using real-world data; furthermore, the method requires relatively weak assumptions that can be empirically tested.[4] The ability to assign participants to an intervention based on risk, severity, or need is a potential advantage over a randomized controlled trial design in terms of acceptability to stakeholders and ethical requirements.[5] The main limitation of the design noted in the literature is the need for larger sample sizes than in randomized experiments.[6,7]

Following its initial presentation in the 1960s, uptake of the design was limited, partly due to a belief that few situations existed in which it could be applied.[8] Two reviews of the design in health have identified a small number of applications. Moscoe et al[9] identified 32 studies from medicine, epidemiology, or public health that used regression discontinuity designs. Accordingly, they argued that the design is likely under-used in these fields. A limitation of the review was that the search was restricted to a single database (PubMed). Venkataramani et al[10] presented 13 studies as examples of the regression discontinuity design in healthcare and as support for their assertion that the design could be applied usefully and widely in clinical medicine and health policy; this review was nonsystematic.

We aimed to conduct a comprehensive systematic review to determine how regression discontinuity designs have been used to analyze the health effects of interventions or natural experiments. Our objectives were: (1) to map the use of regression discontinuity designs in settings and policy areas relevant to public health; (2) to identify what forcing variables have been used and what interventions or exposures have been investigated with these designs; and (3) to assess the quality of reporting of health-related regression discontinuity design studies.

METHODS

We published the review protocol in the PROSPERO international prospective register of systematic reviews (reference number CRD42015025117).

Inclusion Criteria

We included primary, empirical studies from any field of research that (1) reported an analysis which the authors described as regression discontinuity and (2) had an outcome that measured any aspect of physical or mental health or wellbeing.

Search Strategy

We searched 32 health and social science databases for records containing the phrase "regression discontinuity" or "regression-discontinuity" in title, abstract, keyword, or full text (see eTable 1; http://links.lww.com/EDE/B738 in the supplemental digital content for the list of databases and search strategies). The databases were selected to ensure coverage of disciplines relevant to social determinants of health and to public policy, particularly those such as education and economics in which regression discontinuity designs are more commonly used. The date range covered was 1 January 1960 (year of first publication describing the design) to 1 January 2019 (last update search conducted in week 1 of March 2019). No language restrictions were applied. We examined reference lists of review articles and included studies to identify additional studies.

Study Selection

Retrieved references were compiled in an EndNote X7 library and duplicates were manually removed. A random 10% sample (random number sequence generated in Stata version 13) was screened independently for eligibility by two reviewers (M.H.B. and H.T.) based on the record title and abstract. We resolved disagreements by discussion. We recorded reasons for exclusion of studies in EndNote.

Data Extraction

A coding framework was designed to record information about the publication, research topic, study design, outcomes, country of first author's institution, and country of data source. We described the implementation of regression discontinuity in each study in terms of the forcing variable used, the intervention or exposure under investigation, the health-related outcome(s) measured, whether a primary outcome was specified, and whether a study protocol was reported. For journal articles, we additionally described the academic discipline of the journal in a method derived from Stuckler et al[11] and modified to create seven categories. One reviewer (M.H.B.) coded all studies, with 30% of the studies coded by a second reviewer (M.C., H.T., or P.C.) as a check for quality and consistency.

Quality Assessment

One reviewer (M.H.B.) appraised all studies with 30% of the studies appraised by a second reviewer (M.C., H.T., or P.C.) as a check for quality and consistency. To date, the only formally developed and validated quality assessment tool specific to regression discontinuity designs has been produced by What Works Clearinghouse (WWC), an online resource center funded by the United States Department of Education.[12] The 2010 version (1.0) of the WWC standards for regression discontinuity designs comprises three screening questions ("qualifying criteria") and four standards that each involve multiple criteria. According to the WWC Standards, studies that do not meet the three qualifying criteria are not considered valid regression discontinuity designs. However, in this review we applied all quality assessment criteria to each study, irrespective of whether it met the qualifying criteria, given that failure to meet the third criterion in particular may have reflected inadequate reporting rather than a genuinely confounded forcing variable. We further adapted the WWC tool by only judging whether individual criteria were satisfied and not whether the overall standards were met, rephrasing multi-component criteria as single-component yes or no questions, and by omitting optional criteria that are only intended to apply under specific study conditions. We did not apply the WWC standard for attrition, which requires evaluations of educational interventions using regression discontinuity designs to report study attrition in the same manner as randomized controlled trials, because nearly all regression discontinuity designs in health were retrospective analyses of existing datasets rather than prospective studies. Answers of yes, no, or unclear were recorded for the eight resulting criteria:

(1) Was a forcing variable used for treatment assignment?
(2) Was the forcing variable ordinal with at least four unique values on either side of the cutoff?
(3) Was the forcing variable cutoff used only for assignment to the treatment under investigation, and not to other treatments simultaneously?
(4) Was there a description of how the forcing variable was scored and how treatment assignment occurred?
(5) Was smoothness of the forcing variable around the cutoff investigated?
(6) Did the study compare baseline measurements of key covariates for treated and non-treated groups?
(7) Were falsification tests conducted?

(8) Were robustness checks of the statistical model conducted, such as sensitivity analyses relating to bandwidth selection and functional form?

Criteria 1, 2, and 3 relate to conditions for valid implementation of regression discontinuity designs with sufficient data points to model the relationship between the forcing variable and the outcome. If the cutoff is used to assign participants to multiple treatments simultaneously, for example when family income is used to assess eligibility for several benefits, the forcing variable is said to be confounded and regression discontinuity designs cannot be used to evaluate the effect of any one of those treatments alone (although they could be used to evaluate the effect of the totality of the treatments).

Criteria 4 and 5 relate to investigation of how the forcing variable was implemented in the context under investigation. The study should have a description of the scoring and assignment process that supports the integrity of the forcing variable, i.e. a narrative account showing that treatment assignment was not open to manipulation. This narrative should be supported by a histogram or density test that demonstrates smoothness of the forcing variable around the cutoff (bunching on either side could suggest manipulation).

Criterion 6 relates to the potential for selection bias in the study. Criterion 7 relates to falsification tests, which may look for unexplained discontinuities at values other than the cutoff or check for discontinuities at the cutoff in outcomes that ought not to have been affected by the intervention.

Criterion 8 assesses the quality of the statistical analysis. As the results of regression discontinuity design analyses are sensitive to model specification, robustness checks of the model should be conducted, for example through sensitivity analyses.

Data Synthesis

We conducted a narrative synthesis that aimed to describe patterns and commonalities across studies. The main method of narrative synthesis used was thematic summary, in which a descriptive coding framework is developed to allow the grouping of studies in order to compare their characteristics.[13] Count data were collected in relation to numbers of publications by year, by topic, and by academic discipline to enable identification of trends in the use of regression discontinuity designs.

RESULTS

Quantity of Studies Using Regression Discontinuity Designs in Health

eFigure 1; http://links.lww.com/EDE/B738 in the supplemental digital content shows the study selection process as a flowchart. The searches retrieved 7658 records, of which 5049 were duplicates. Of the unique records, we excluded 2305 (1784 abstracts and 521 full-text articles screened) because they did not meet the inclusion criteria. We identified 21

additional studies through examination of full text: 19 from the reference lists of full-text articles, and two doctoral theses that each contained two separate RD studies. In total, 325 studies were included.

The number of studies published by year (Figure 1) shows that the use of regression discontinuity designs in health-related applications is increasing over time, with the greatest increase in output taking place in the past decade.

Topics and Academic Disciplines of Publications

eTable 2; http://links.lww.com/EDE/B738 in the supplemental digital content provides an overview of the synthesis, with studies organized by subject area and by the type of intervention or exposure under investigation. Detailed characteristics of the included studies, including context, forcing variable used, intervention or exposure, and outcomes, along with references for all 325 studies, appear in eTables 3–12; http://links.lww.com/EDE/B738 in the supplemental digital content.

One-third (106/325; 33%) of included studies investigated public health policy-related questions (other than health insurance programs), namely alcohol policies (n = 25), the organization of health services (n = 26), disease prevention and screening (n = 13), early years interventions (n = 11), nutritional exposures or interventions (n = 11), tobacco (n = 8), air quality (n = 7), and road safety (n = 5). The public health policy issue investigated by the largest number of studies was minimum legal drinking age legislation (n = 24). One-third of the studies evaluated either health insurance programmes (n = 54) or the health effects of education (n = 54). The remaining studies considered clinical treatments in physical (n = 23) or mental health (n = 8), the health effects of social programs in low- and middle-income (n = 35) or high-income (n = 23) countries, and health effects of (non-interventional) exposure to disasters, social conditions, or family conditions (n = 22).

More than two-thirds (230/325; 71%) of the publications identified appeared in peer-reviewed journals, with the remainder identified from gray literature sources (including 52 working articles, 37 theses, two reports, two book chapters, and two conference articles). Of the peer-reviewed articles, over half (127/230; 55%) appeared in journals indexed in Web of Science as economics or health economics journals. Just over one-third (113/325; 35%) of the included studies were indexed in Medline at the time of the search.

The geographic coverage of regression discontinuity design output is global, with authors and data from six continents represented in the included studies, but dominated by the United States: over half of the publications (186/325; 57%) had a first author affiliated with a US institution and nearly half (134/325; 41%) used US data for analysis.

Thematic Analysis of Forcing Variables and Threshold Rules

Analysis of the study designs focused on the forcing variable as it is a fundamental requirement of regression
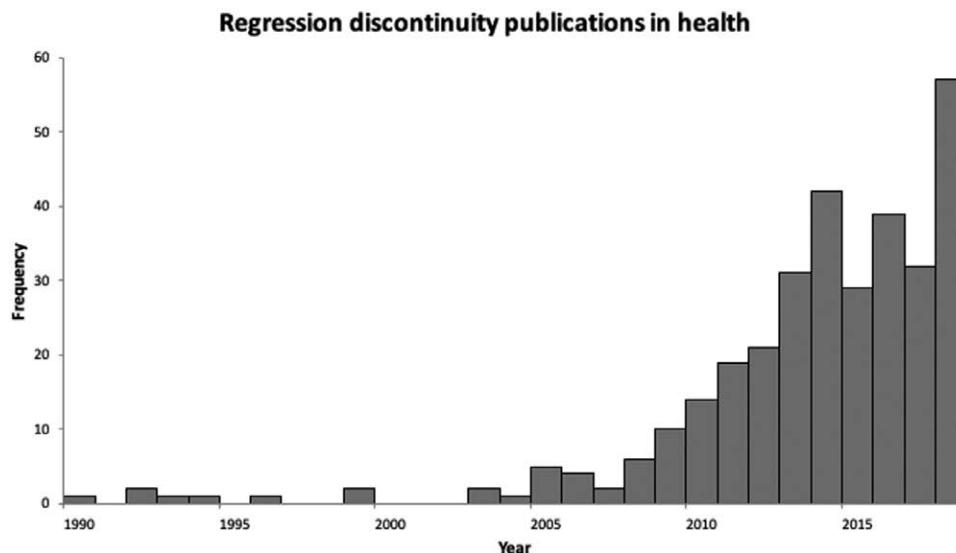
FIGURE 1. Histogram of regression discontinuity studies of health outcomes published by year between January 1980 and January 2019.

discontinuity designs. Thematic analysis of the forcing variables demonstrated that six types of variable have been used to apply these designs in the study of health outcomes:

(1) age;
(2) socioeconomic measures such as poverty indices, literacy rates, or income;
(3) clinical measures that act as a threshold for an intervention;
(4) environmental measures;
(5) geographic location; and
(6) date or time.

For a regression discontinuity design to be implemented, a threshold rule must exist according to which the value of the forcing variable is used to assign people or study units into treated and untreated groups. Thematic analysis demonstrated that four sources of threshold rules are common to the included studies:

(1) program eligibility rules for social programs and other complex interventions;
(2) clinical decision-making rules or guidelines;
(3) thresholds imposed by legislation to enable or restrict activities that affect health; and
(4) date or time of the implementation or occurrence of changes or events (such as policy changes or natural disasters) that produce a change in exposure status.

The Table provides examples of each type of forcing variable and of the threshold rules used.

Quality Assessment

The eight quality assessment criteria were fully met by 12% (40/325) of the studies (Figure 2). Fifteen percent

(49/325) of studies reported a pre-specified primary outcome or study protocol.

Almost all studies (323/325; 99%) clearly reported the forcing variable used (criterion 1) and most (295/325; 91%) reported the use of at least four discrete values of the forcing variable on either side of the cutoff value (criterion 2). In the included studies, 172/325 (53%) provided enough information to support a conclusion that the forcing variable was not confounded, 13/325 (4%) used a cutoff that was clearly used to assign people to additional treatments other than the one under investigation, and 140/325 (43%) used a forcing variable that could conceivably be confounded without reporting clear evidence to the contrary (criterion 3).

Of the included studies, 291/325 (89%) provided some account of scoring and treatment assignment demonstrating the integrity of the forcing variable (criterion 4), and 158/325 (49%) reported a density test or histogram of the forcing variable (criterion 5).

Just over two-thirds of studies (224/325; 69%) examined whether treatment and control groups showed baseline equivalence on any covariates (criterion 6), but less than half (152/325; 47%) conducted falsification tests (criterion 7).

Finally, over three-quarters (256/325; 79%) of studies reported robustness checks (criterion 8).

DISCUSSION

Key Findings

Through a comprehensive search of 32 databases, this systematic review has identified 325 studies that apply regression discontinuity designs to investigate health-related research questions. The findings confirm that the designs are suitable for evaluation of health interventions and health policy, while also

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc.

TABLE. Thematic Analysis of Forcing Variables and Threshold Rules Used in Regression Discontinuity Studies of Health Outcomes

Type of Forcing Variable	Number of Studies	Measurement Used	Threshold Rule
Age	110	Age in days, months, weeks, or years	Age threshold for: • Starting school • Leaving school • Legal drinking age • Treatment or benefit eligibility • Insurance eligibility • Retirement
Date/time	107	Calendar date, month, or year Time in minutes, hours, or days	Date or time of: • Implementation of policy/ legislation • Repeal of policy/legislation • Disaster or major incident • Change in situation or conditions
Socioeconomic measure	57	Company payroll total Dropout risk score Family income Household acreage Investment cost Poverty or literacy rate Poverty or welfare index Predicted probability of borrowing microcredit Program quality score Vote share or margin	Benefit or program eligibility Election outcome Legislated threshold
Clinical measure	31	Addiction severity measure Birthweight Blood lead levels Body mass index Cardiovascular risk CD4 count Down syndrome risk Exeter Alcohol Scale Hospital Safety Score Parity Positive Symptoms Scale Posttraumatic Stress Disorder Reaction Index Staffing numbers Systolic blood pressure Time of birth Visual acuity Weeks of gestation	Risk threshold for intervention Guideline threshold for intervention Legislated threshold for intervention
Environmental measure	6	Ozone forecasts Air pollution levels	Policy threshold for action Legislated threshold for action
Geographical location	9	Political boundary Distance from boundary Latitude and longitude	Program eligibility
Other	6	Class size Number of schools School test score Village population Draft lottery number	Policy threshold for intervention/ exposure Program eligibility

Although 325 studies are included in the review, the total number of analyses is 326 because one study conducted two RD analyses using two different forcing variables (age and date).

showing that they have been more widely applied in health research than previously appreciated. The synthesis identified six categories of forcing variable (age, date/time, socioeconomic measures, clinical measures, environmental measures, and geographic boundaries) and four types of threshold rules (program eligibility, treatment thresholds, legislated thresholds, and dates
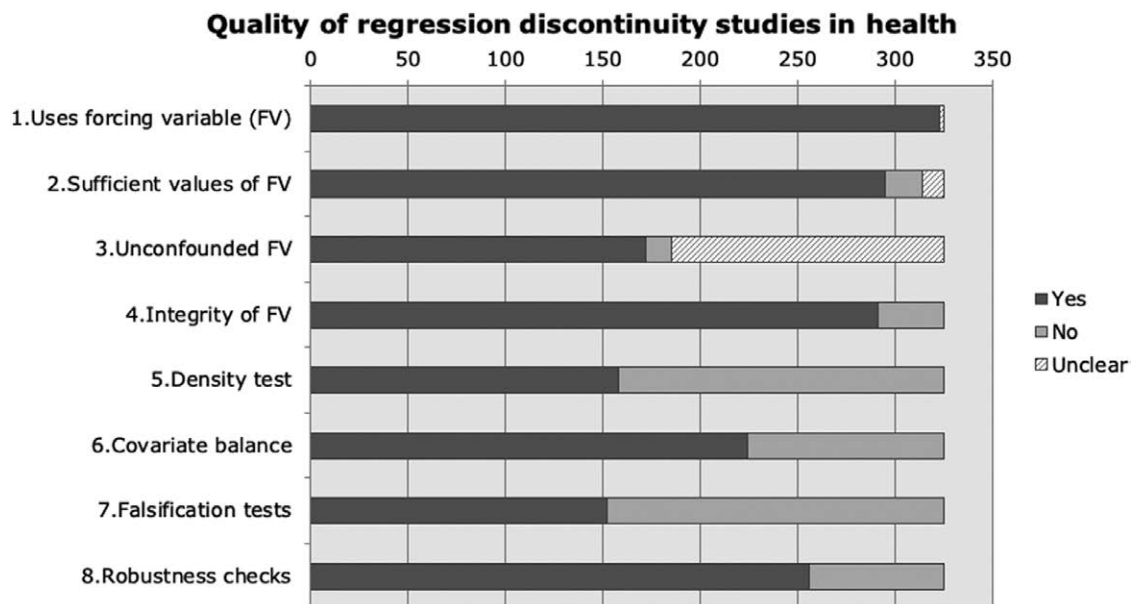
FIGURE 2. Summary of quality assessments of regression discontinuity studies that report health outcomes. The bar chart shows the number of studies (total = 325) judged as yes, no, or unclear as to whether they meet eight criteria derived from the What Works Clearinghouse Standards for RD Version 1.0 (Schochet et al., 2010). FV, forcing variable.

of implementation) that have been used to implement regression discontinuity designs in health studies.

In assessing the quality of these studies against eight criteria specific to regression discontinuity designs, this review demonstrates the need for improvement in the design and reporting of studies that use regression discontinuity. The most frequently encountered issues in study quality related to unclear reporting of how treatment assignment occurred, failing to show clearly that the forcing variable was not manipulated or confounded, not reporting a density test or histogram of the forcing variable, and not reporting any falsification tests (i.e., looking for discontinuities at non-cutoff values of the forcing variable or in outcomes that ought not to be affected by the treatment). These are not trivial issues, as they assess whether the key identifying assumptions of the design have been met and therefore, whether any effect can plausibly be attributed to the intervention or exposure under investigation.

Contribution to Knowledge About Regression Discontinuity Designs and Natural Experiments

In addition to the varied performance against the quality criteria, some limitations of regression discontinuity designs are apparent which may need to be considered in future evidence synthesis work and further development of standards of reporting. Previously the chief limitation of the design, apart from the perceived difficulty of finding situations in which it could be used, was thought to be the need for large sample sizes to achieve adequate statistical power. Many of the studies examined in this review used very large datasets and thus sample size was less of a concern. However, by exploring functional form and conducting robustness checks

in the absence of a study protocol or prospectively chosen primary outcome, many studies inadvertently created problems in terms of transparency, interpretation, and synthesis. Studies using regression discontinuity designs frequently present the results of multiple analyses, including different stratification of data (e.g., by gender or age), different choices of bandwidth, and different model specifications. While robustness checks are recommended, they may inadvertently create a risk of data-dredging in the primary study, and of bias in the selection of results for inclusion in subsequent evidence synthesis.

By identifying a large number of studies that evaluate clinical interventions, national policy changes, and effects of social determinants of health, this review offers substantial evidence that regression discontinuity designs have greater applicability in health research or policy evaluation than has previously been appreciated.[8,9] This review joins a small number of other systematic reviews that have investigated the application of innovative non-randomized study designs and natural experiment methods to medicine, epidemiology, and public health. This review identified more examples of regression discontinuity designs than another systematic review identified of instrumental variable studies in epidemiology and medicine, suggesting that, although good instruments may be hard to find, good forcing variables may be less so.[14] These findings also support the conclusion of Moscoe et al[9] that regression discontinuity designs are probably under-used in health research: although numerous relevant applications of the design can be identified, few have been replicated or extended to other contexts, and the results suggest that the

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc.

potential to do so exists. Also, regression discontinuity designs are not yet as commonly applied as, for example, propensity score matching has been in medicine; a systematic review on that topic identified 296 studies published in a 6-month period in PubMed alone.[15]

Strengths and Limitations

The strengths of this review include systematic searching of 32 databases and critical appraisal of included studies against eight design-specific criteria, leading to a comprehensive overview of regression discontinuity designs in health. The review was conducted according to a protocol registered in PROSPERO with no significant deviations from the protocol. The unexpectedly large number of studies meeting the inclusion criteria led to the study's main limitation, which was the lack of resource necessary to involve two reviewers in all steps of study selection and critical appraisal. However, this limitation was addressed by double-sifting a random sample of the search results, piloting the critical appraisal method with two reviewers, and having two reviewers appraise and extract data from 30% of studies as a check for quality and consistency.

Implications

This review has two main implications for researchers interested in identifying natural experiments or implementing regression discontinuity designs. First, we offer a comprehensive overview of forcing variables and threshold rules used in regression discontinuity studies with health outcomes, which may help researchers to identify situations in which these designs may be used to evaluate interventions or policies. Second, we show the strengths and weaknesses of the existing literature in terms of study quality, which point to several considerations that researchers should take into account in order to produce high-quality evidence using regression discontinuity designs. Researchers should provide a full account of the choice of forcing variable and how it was implemented in the context of the study setting; provide evidence that treatment assignment was free from manipulation; state whether the same value of the forcing variable was used simultaneously to assign the participants to other treatments that could affect the outcome; explore the sensitivity of results to bandwidth choice and model specification; and investigate and rule out rival hypotheses.[16] A report of a potentially valid design may undermine its own conclusions if it fails to demonstrate that the assumptions of regression discontinuity designs, particularly the unconfoundedness of the forcing variable, are met. Our results point to the need for reporting standards for regression discontinuity designs to improve quality, as STROBE and CONSORT have achieved with other study designs.

CONCLUSIONS

This systematic review provides a detailed illustration of how regression discontinuity designs can be implemented to investigate the effects of a wide range of interventions and exposures on health outcomes. We have conducted an exhaustive search of 32 databases to provide the most comprehensive review to date of the use of these designs in public health and related policy areas. We have identified substantially greater use of regression discontinuity designs in health than has been previously recognized, demonstrating the relevance to health research and wide potential scope for application of the method, while highlighting substantial shortcomings in study reporting.

REFERENCES

1. Craig P, Cooper C, Gunnell D, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *J Epidemiol Community Health.* 2012;66:1182–1186.
2. Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *J Educ Psychol.* 1960;51:309–317.
3. Imbens GW, Lemieux T. Regression discontinuity designs: a guide to practice. *J Econom.* 2008;142:615–635.
4. Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annu Rev Public Health.* 2017;38:39–56.
5. Labrecque JA, Kaufman JS. Commentary: can a quasi-experimental design be a better idea than an experimental one? *Epidemiology.* 2016;27:500–502.
6. Lee DS, Lemieux T. Regression discontinuity designs in economics. *J Econ Lit.* 2010;48:281–355.
7. van Leeuwen N, Lingsma HF, de Craen AJ, et al. Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes. *Epidemiology.* 2016;27:503–511.
8. Cook TD. "Waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. *J Econom.* 2008;142:636–654.
9. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *J Clin Epidemiol.* 2015;68:122–133.
10. Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ.* 2016;352:i1216.
11. Stuckler D, Reeves A, Karanikolos M, McKee M. The health effects of the global financial crisis: can we reconcile the differing views? A network analysis of literature across disciplines. *Health Econ Policy Law.* 2015;10:83–99. Available at: http://journals.cambridge.org/article_S1744133114000255. Accessed 12 October 2020.
12. Schochet P, Cook TD, Deke J, Imbens GW, Lockwood JR, Porter J, et al. *Standards for Regression Discontinuity Designs.* 2010. Available at: https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_rd.pdf. Accessed 12 October 2020.
13. Gough D, Thomas J, Oliver S. Clarifying differences between review designs and methods. *Syst Rev.* 2012;1:28.
14. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology.* 2013;24:363–369.
15. Ali MS, Groenwold RH, Belitser SV, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *J Clin Epidemiol.* 2015;68:112–121.
16. Ludwig J, Miller DL. Does head start improve children's life chances? Evidence from a regression discontinuity design. *Q J Econ.* 2007;122:159–208.