JAMA Guide to Statistics and Methods

Regression Discontinuity Approaches to Evaluate Threshold-Guided Interventions in Observational Studies

Matias D. Cattaneo, PhD; Rocío Titiunik, PhD

When using observational data to learn about the effect of an intervention, unmeasured differences between patients who receive different interventions can confound estimates of treatment effects.[1] Regression discontinuity (RD) approaches can address this problem when treatment (eg, intervention, program, or policy) depends on whether a patient measurement—such as a laboratory test result—crosses a prespecified threshold.[2] This measurement is often called the *score* or *running variable* in RD analyses. Threshold-based medical guidelines therefore create opportunities to use RD approaches in clinical research.[3]

Editor's Note

CME at jamacmelookup.com

Goulden and colleagues used an RD analysis to study the effect of contrast-enhanced computed tomographic pulmonary angiography (CTPA) on kidney function among patients undergoing D-dimer testing for suspected pulmonary embolism.[4] Typical clinical guidance at the time recommended CTPA when the D-dimer level was above 500 ng/mL, with some hospitals using a 460- or 470-ng/mL cutoff. The analysis compared patients with D-dimer levels just above and below the cutoff, using estimated glomerular filtration rate (eGFR) as the measure of kidney function.

Why Are RD Approaches Used for Identifying Causal Effects in Observational Studies?

When treatment assignments are not randomized, patient groups receiving different treatments often differ in terms of both measured and unmeasured characteristics that influence the outcome. Adjusting only for measured characteristics may not control for confounding by unmeasured characteristics. When treatment eligibility or treatment recommendation changes at a cutoff, patients with scores just above and below the cutoff are often likely to be similar in both their measured and unmeasured baseline characteristics.[5] An RD analysis compares those 2 groups to obtain a valid estimate of the treatment effect.

How Does an RD Approach Work and How Is It Evaluated?

The key assumption that justifies an RD analysis is the comparability of patients with measurements just below and above the cutoff. This assumption can be formalized in 2 ways: a continuity framework or a local randomization framework. Under the continuity framework, the average outcomes patients would experience under each intervention are assumed not to change abruptly at the cutoff, and therefore a jump in observed average outcomes at the cutoff can be attributed to the change in treatment.[6] Under the local randomization framework, whether a patient's score falls above or below the cutoff is assumed to be randomly assigned within a small region around the cutoff, permitting the use of analytic methods similar to those for randomized trials.[7]

An RD approach is called *sharp* when every patient receives the intervention indicated by the threshold rule and *fuzzy* when adherence to the rule is imperfect. In a sharp RD approach, the causal effect of interest is the average treatment effect for patients at or near the cutoff because there is no difference between the rule-defined treatment and the treatment received.[6]

In contrast, a fuzzy RD approach yields 3 related causal effects. The first is the effect of the threshold rule on the probability of receiving the intervention, called the *first-stage effect* or *take-up effect*. This effect is a measure of the concordance—or lack thereof—between actual care and care dictated by the rule. The second causal effect is the effect of the threshold rule on the outcome, analogous to the intention-to-treat effect of randomized treatment assignment on the outcome in a randomized clinical trial.[8] The third causal effect is the effect of receiving the intervention among the complier subgroup, namely those patients whose intervention was changed because their score fell above rather than below the cutoff. Estimating this effect, known as the *complier average treatment effect*, requires making additional assumptions for causal interpretation.[7]

An RD analysis requires data on the score and outcome, knowledge of the cutoff, and, for a fuzzy RD analysis, data on the intervention actually received. Baseline characteristics are also useful for evaluating the key RD assumption of comparability near the cutoff. The analysis focuses on patients within a small window around the cutoff. Under the continuity framework, separate local linear regressions of the effect of the score on the outcome are fitted above and below the cutoff, with the score centered at the cutoff. Each intercept estimates the average outcome at the cutoff on its side, and the difference between the intercepts is the estimated average effect of treatment assignment on the outcome. Under the local randomization framework, the effect of the threshold rule is commonly estimated by comparing average outcomes just above and below the cutoff within the chosen window. For a fuzzy RD analysis, under either framework, the complier average treatment effect is estimated by dividing the effect of the threshold rule on the outcome by its effect on the probability of receiving the intervention.[6,7]

To evaluate an RD analysis, investigators can use falsification checks that can reveal patterns inconsistent with the key RD assumption of comparability near the cutoff. The main check, analogous to evaluating covariate balance in a randomized clinical trial, is whether important baseline characteristics show a discontinuity (or jump) at the cutoff because, in a valid RD analysis, they should not. Other checks examine whether the distribution of scores jumps at the cutoff, conclusions are similar across reasonable window choices around the cutoff, apparent effects occur at artificial cutoffs where the treatment rule does not change, and results change after excluding observations closest to the cutoff. These checks

© 2026 American Medical Association. All rights reserved, including those for text and data mining, AI training, and similar technologies.

Downloaded from jamanetwork.com by Princeton University user on 11/02/2026

cannot prove that an RD analysis is valid, but they can reveal evidence against its assumptions.[6]

What Are the Limitations of RD Analysis?

An RD analysis requires a known threshold rule that changes eligibility for, recommendation of, or receipt of an intervention. Although RD analyses can be extended to more complex settings, including those involving multiple cutoffs,[9] many clinical questions do not arise from threshold-based decisions. Randomized clinical trials or other observational approaches must be used in those settings.

Even when an RD approach is applicable, causal interpretation may fail if patients or clinicians can influence the score, its timing, or how it is measured or recorded to obtain a preferred intervention because patients just above and below the cutoff may then differ systematically. Interpretation is also difficult when another aspect of care or policy changes at the same cutoff because any jump in the outcome cannot be uniquely attributed to the intervention of interest.

RD estimates rely primarily on observations close to the cutoff. Consequently, even a large study may yield imprecise estimates if few patients lie near the cutoff or, in a fuzzy RD analysis, if crossing the threshold only slightly changes the probability of receiving the intervention. Moreover, the estimated effects are local: they apply to patients at or near the cutoff and may not generalize to patients with substantially different scores. Extrapolation beyond the cutoff requires making additional strong assumptions.

How Did Goulden and Colleagues Apply RD?

Goulden and colleagues studied 156 028 emergency department patients who underwent D-dimer testing for suspected pulmonary embolism. Patients with D-dimer levels above the applicable cutoff were much more likely to receive contrast-enhanced CTPA than those below it, but adherence to this cutoff was imperfect, so the authors used a fuzzy RD analysis.[4]

The study illustrates how comparisons near the cutoff can reduce confounding. Across the full sample, patients whose D-dimer levels were above the cutoff were, on average, approximately aged 60 years, compared with 50 years among those whose levels were below the cutoff, a difference that could confound comparisons of kidney function. Near the cutoff, the age difference was much smaller, and other measured baseline characteristics were also similar.[4] Manipulation of the score was unlikely because patients could not control their measured D-dimer levels.

Using local linear regressions on either side of the cutoff, the investigators estimated (1) the effect of crossing the D-dimer cutoff on long-term eGFR and (2) the effect of receiving contrast-enhanced CTPA among patients whose receipt of CTPA was induced by the threshold rule—the complier average treatment effect.

How Should the Goulden Study Be Interpreted?

The estimated effect of crossing the D-dimer cutoff on long-term eGFR was −0.1 mL/min/1.73 m^2 (95% CI, −1.2 to 1.1). The estimated complier average treatment effect of receiving contrast-enhanced CTPA was −0.4 mL/min/1.73 m^2 (95% CI, −4.9 to 4.0).[4] Both estimates were close to 0 and both 95% CIs include no effect. Assuming the RD and fuzzy RD assumptions hold, these results provide no evidence that contrast-enhanced CTPA harms long-term kidney function among patients with D-dimer levels near the cutoff whose receipt of CTPA was affected by the threshold rule. They do not establish the absence of harm for all patients: the 95% CIs for the complier effect allow for both decreases and increases in eGFR, and the estimate may not generalize to patients with substantially different D-dimer levels or those whose receipt of CTPA would not change at the cutoff.

ARTICLE INFORMATION

Author Affiliations: Department of Economics, Princeton University, Princeton, New Jersey (Cattaneo); Department of Politics, Princeton University, Princeton, New Jersey (Titiunik).

Corresponding Author: Rocío Titiunik, PhD, Department of Politics, Princeton University, Fisher Hall No. 001, Princeton, NJ 08544 (titiunik@princeton.edu).

Published Online: November 2, 2026. doi:10.1001/jama.2026.22316

Conflict of Interest Disclosures: None reported.

Funding/Support: The authors gratefully acknowledge financial support from the National Science Foundation (SES-2019432, SES-2241575).

Role of the Funder/Sponsor: The National Science Foundation had no role in the preparation, review, or approval of the manuscript or decision to submit the manuscript for publication.

REFERENCES

1. Rosenbaum PR. *Design of Observational Studies*. Springer; 2010. doi:10.1007/978-1-4419-1213-8
2. Maciejewski ML, Basu A. Regression discontinuity design. *JAMA*. 2020;324(4):381-382. doi:10.1001/jama.2020.3822
3. Cattaneo MD, Keele L, Titiunik R. A guide to regression discontinuity designs in medical applications. *Stat Med*. 2023;42(24):4484-4513. doi:10.1002/sim.9861
4. Goulden R, Rowe BH, Abrahamowicz M, Strumpf E, Tamblyn R. Association of intravenous radiocontrast with kidney function: a regression discontinuity analysis. *JAMA Intern Med*. 2021;181(6):767-774. doi:10.1001/jamainternmed.2021.0916
5. Lee DS, Lemieux T. Regression discontinuity designs in economics. *J Econ Lit*. 2010;48(2):281-355. doi:10.1257/jel.48.2.281
6. Cattaneo MD, Idrobo N, Titiunik R. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press; 2020. doi:10.1017/9781108684606
7. Cattaneo MD, Idrobo N, Titiunik R. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge University Press; 2024. doi:10.1017/9781009441896
8. Detry MA, Lewis RJ. The intention-to-treat principle: how to assess the true effect of choosing a medical treatment. *JAMA*. 2014;312(1):85-86. doi:10.1001/jama.2014.7523
9. Cattaneo MD, Titiunik R. Regression discontinuity designs. *Annu Rev Econ*. 2022;14:821-851. doi:10.1146/annurev-economics-051520-021409

© 2026 American Medical Association. All rights reserved, including those for text and data mining, AI training, and similar technologies.

Downloaded from jamanetwork.com by Princeton University user on 11/02/2026