

Huseyin Naci

What to do (or not to do) when randomization is not possible

**Article (Accepted version)
(Refereed)**

Original citation: Naci, Huseyin (2017) *What to do (or not to do) when randomization is not possible*. [The Journal of Heart and Lung Transplantation](#). ISSN 1053-2498
DOI: [10.1016/j.healun.2017.07.008](https://doi.org/10.1016/j.healun.2017.07.008)

© 2017 [International Society for the Heart and Lung Transplantation](#)

This version available at: <http://eprints.lse.ac.uk/83660/>
Available in LSE Research Online: August 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

What to do (or not to do) when randomization is not possible

Huseyin Naci, PhD

Assistant Professor of Health Policy, LSE Health, Department of Health Policy, London School of Economics and Political Science, London WC2A 2AE, United Kingdom, h.naci@lse.ac.uk

Nonrandomized studies have a bad reputation in health care research. This is not entirely surprising: a key predictor of reversing established health care standards is the original adoption of an intervention on nonrandomized evidence alone.¹ An earlier systematic evaluation found that several influential nonrandomized studies were refuted or found to have exaggerated effects when later tested in randomized controlled trials.²

What should be done when randomization is not possible? In the linked article, Peng and colleagues evaluate the difference in wait times for receiving a donor heart between patients listed for a transplant before and after their 18th birthday and whether any difference in wait time is associated with lower likelihood of transplant and/or higher risk of mortality.³ It is not conceivable how a randomized controlled trial could be designed to address this important question.

Fortunately, not all nonrandomized studies are created equal, and strong quasi-experimental designs, when carefully used, can mimic randomization and produce valid findings.⁴ Peng and colleagues use one such quasi-experimental design – regression discontinuity design – and find that patients assigned after their 18th birthday to an adult allocation system had longer wait times and were less likely to receive a transplant than patients assigned before their 18th birthday to a paediatric allocation system, despite no observed differences in waitlist-associated mortality between the groups.

When interpreting these findings, four questions are important to consider. First, why is randomization often considered essential to establish causality in health care research? Second, in its absence, what adjustment methods are available to “mimic” randomization and what are their drawbacks? Third, what are the assumptions of using quasi-experimental designs as substitutes to randomized controlled trials? Fourth, what specific issues arise in Peng and colleagues’ use of the regression discontinuity design?

Why is randomization essential in health care research?

Randomized controlled trials use a random process (e.g., a coin toss) to assign study subjects to different intervention groups. These so-called “gold-standard” designs ensure that the findings can be attributed to the intervention and nothing else. Random allocation introduces unpredictability to intervention assignment and reduces the likelihood that systematic differences in groups may influence outcomes.

In nonrandomized designs, intervention assignment is deliberately influenced by the patient or the provider rather than randomly assigned by the researcher. This often results in differences in the baseline characteristics of patient groups receiving different interventions. In routine clinical practice, for instance, prescribing decisions may be guided by the prognosis of the patient: the worse the prognosis, the more intense the treatment.⁵ Therefore, in nonrandomized study designs, it is impossible to know if unknown or unmeasured factors that affect the outcomes of interest (e.g., indications for treatment, severity of illness) are evenly distributed across the intervention groups. Such differences are a severe threat to the validity of nonrandomized studies and explain why

providers, researchers, and policy makers are often reluctant to use evidence from such designs to reach conclusions about the effect of interventions, programs, or policies.

What analytic approaches are available to minimize differences between groups when randomization is not possible?

When randomization is not possible to assign individuals to different groups, a range of alternative analytic strategies can be employed.⁶ **Table 1** provides an overview of different adjustment methods and their advantages and disadvantages.

Researchers often use standard regression methods to adjust for differences in baseline patient characteristics included in different groups. A key limitation of standard regression methods in nonrandomized designs is that they are strictly limited to known and measured variables in the data at hand.

In recent years, a more advanced suite of analytic approaches has emerged. One such approach is propensity score matching. A propensity score expresses the probability of receiving a treatment based on a list of patient characteristics at the time of treatment initiation. Treatment effectiveness can then be measured among patients who have a similar (matched) propensity score, thus potentially controlling for confounding. While this approach can theoretically control for a large set of confounders,⁷ propensity score adjustment – like standard regression methods – cannot account for unknown or unmeasured factors and rarely addresses concerns in nonrandomized studies.^{8,9}

Another approach is instrumental variable analysis. Instrumental variable methods identify and exploit naturally occurring quasi-random events (referred to as “instruments”) that affect treatment assignment but does not otherwise affect outcomes, mimicking the randomization of patients into treatment.¹⁰ The validity of the instrumental variables approach depends on the researchers’ ability to assess whether the instrument chosen does not influence the outcome by any mechanism that is not mediated through the treatment – an assumption that is typically untestable using the data at hand.¹¹ Despite this threat to the validity of nonrandomized studies using instrumental variables, the use of instrumental variables in health care research has skyrocketed in recent years.

Table 1. Different analytic strategies

Analytic strategy	Advantages	Disadvantages
Standard regression analysis	Introduces balance between groups on known and measured factors	Unmeasured and unknown factors cannot be taken into account
Propensity score adjustment	Controls for larger numbers of known and measured factors than standard regression analysis	Unmeasured and unknown factors cannot be taken into account
Instrumental variable analysis	Theoretically adjusts for known <i>and</i> unknown factors	Assumptions are untestable with data at hand

What are the assumptions of different nonrandomized designs as substitutes to randomized trials?

In practice, testing the assumptions of different adjustment factors remain difficult. Previous empirical evaluations have warned against overreliance on such methods.¹² Instead, researchers can adopt different nonrandomized, quasi-experimental, designs. Study designs that address issues of internal validity in the absence of

randomization are referred to as “quasi-experimental” designs and include the controlled before-and-after design, interrupted time series analysis, and regression discontinuity design. **Table 2** provides an overview of different quasi-experimental designs and their advantages and disadvantages. The rest of this commentary focuses on the regression discontinuity design used in the study by Peng and colleagues.

Regression discontinuity designs are strong quasi-experimental designs that are increasingly used to evaluate the effect of clinical or policy decision rules based on specific eligibility criteria.¹³ Such decision rules are commonplace in clinical practice and health policy. Examples include LDL cholesterol thresholds for initiating statin therapy, income level for insurance coverage, and children’s date of birth for vaccination eligibility.

These designs are particularly useful when people are differentially assigned to an intervention, program or policy if they fall below or above an arbitrary cut-off or threshold.¹⁴ In the study by Peng and colleagues, age serves as one such decision rule: patients awaiting heart transplants before or after their 18th birthdays are eligible for inclusion in different allocation systems and thus potentially different wait lists. The regression discontinuity design exploits this decision rule and assumes that patients whose age is “just above” (e.g., a few months older) or “just below” (e.g., a few months younger) this cut-off have similar characteristics. This is akin to considering the arbitrary cut-off as a quasi-random assignment to different groups. People who are just above the arbitrary cut-off are “randomly” assigned to one group while people who are just below the cut-off are assigned to another. The validity of findings obtained from studies employing regression discontinuity designs relies on the assumption that people around the cut-off are similar. In such designs, it is essential to perform robustness checks by differing the size of the comparison groups (by varying the distance from the cut-off) and reporting the sensitivity of the findings to different specifications. In theory, regression discontinuity designs have fewer assumptions than other quasi-experimental designs

Table 2. Different quasi-experimental designs

Quasi-experimental design	Description	Assumptions	Advantages	Disadvantages
Controlled before-and-after design	Outcomes are observed before and after an event of interest (e.g., intervention, program, policy) in two groups – only one group is exposed to the event	Groups are similar on all known and unknown factors except for exposure to the event	Feasible alternative when randomization is not possible	Other factors could explain the observed changes in the outcome
Interrupted time series analysis ¹⁵	When a key outcome is observed over a long period of time, slope and level of an outcome is compared before	Observations after an event (e.g., intervention, program, policy) have a different slope or level from those before the event	Pre-event time period serves as a strong internal control and addresses most threats to internal validity	If there is an external event co-occurring with the event under investigation, it may not be possible to attribute the

	and after an event of interest			observed change in the outcome to either event
Regression discontinuity design ¹³	Takes advantage of decision rules in which people are differentially assigned to an intervention if they fall below or above an arbitrary cut-off	Groups with values “just above” and “just below” an arbitrary cut-off value are identical except for group allocation	Mimics randomization around the arbitrary cut-off	If group allocation around the “arbitrary” cut-off is not entirely arbitrary, groups may be systematically different from each other on factors that affect the outcome, leading to bias

What specific issues arise in Peng and colleagues’ use of this design?

Peng and colleagues admit that using regression discontinuity design was not their first choice and that they adopted this approach “[b]ecause of instability of the results using Cox regression when age was analysed as a dichotomous exposure variable using different age windows surrounding the 18th birthday.” Yet, the strength of the regression discontinuity design goes far beyond this issue: when used correctly, it can approximate random allocation of study subjects around an arbitrary cut-off.

Still, if the key assumption of similarity around the decision cut-off is violated, findings of regression discontinuity designs may share the limitations of weaker nonrandomized alternatives. In the study by Peng and colleagues, if the decision to assign patients in late adolescence to adult vs. paediatric allocation systems is influenced by either patient or clinician factors, this could bias the results. If clinicians have *a priori* expectation that patients would have to wait longer on the adult allocation system, this might influence the listing time for patients approaching their 18th birthday. Such manipulation by clinicians would introduce systematic differences between patients just below and just above the cut-off, violating the key assumption of the regression discontinuity design. Study authors do not address this issue directly and in fact imply that this assumption may not hold in their analysis. There is also an apparent lack of sensitivity analyses to check the robustness of findings.

Conclusion

Regression discontinuity designs, despite their strengths, are underutilized in health care research.¹⁶ As Peng and colleagues highlight, their study is among the first applications of this method in cardiovascular medicine. Introducing an established quasi-experimental method to an unfamiliar audience is no easy task. How do Peng and colleagues take up the challenge? Admittedly, they could do more to demystify the method and convincingly communicate and test for its assumptions. Substantially more robustness checks could go a long way to convince the reader that their findings stand to scrutiny.

Acknowledgements, funding sources and relevant disclosures

None.

References

- ¹ Prasad V, Gall V, Cifu A. The frequency of medical reversal. *Arch Intern Med*. 2011;171(18):1675-1676.
- ² Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294(2):218-228.
- ³ Peng DM, et al. Impact of the eighteenth birthday on waitlist outcomes among young adults listed for heart transplant: A regression discontinuity analysis. *J Heart Lung Transplant*. (in press in this issue)
- ⁴ Soumerai SB, Ceccarelli R, Koppel R. False Dichotomies and Health Policy Research Designs: Randomized Trials Are Not Always the Answer. *J Gen Intern Med*. 2017;32(2):204-209.
- ⁵ Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc*. 1999;47:749-754.
- ⁶ Agoritsas T, Merglen A, Shah ND, et al. Adjusted analyses in studies addressing therapy and harm: User's guides to the medical literature. *JAMA*. 2017;317:748-759.
- ⁷ Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512.
- ⁸ Freemantle N, Marston L, Walters K, et al. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *BMJ*. 2013-11-11 11:33:33 2013;347.
- ⁹ Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*. Feb 20 2007;26(4):734-753.
- ¹⁰ Stukel TA, Fisher ES, Wennberg DE, et al. Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods. *JAMA*. 2007;297(3):278-285.
- ¹¹ Garabedian LF, Zaslavsky AM, Soumerai SB. Instrumental variable analyses for observational comparative effectiveness research: the paired availability design. *Ann Intern Med*. 2014;161(11):841.
- ¹² Garabedian LF, Chu P, Toh S, et al. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Ann Intern Med*. 2014;161:131-138.
- ¹³ O'Keeffe AG, Geneletti S, Baio G, et al. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. *BMJ* 2014;349:g5293
- ¹⁴ Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ* 2016;352:i1216
- ¹⁵ Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46:348-355.
- ¹⁶ 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(2):122-33.