JAMA Guide to Statistics and Methods

Methods for Evaluating Changes in Health Care Policy
The Difference-in-Differences Approach

Justin B. Dimick, MD, MPH; Andrew M. Ryan, PhD

Observational studies are commonly used to evaluate the changes in outcomes associated with health care policy implementation. An important limitation in using observational studies in this context is the need to control for background changes in outcomes that occur with time (eg, secular trends affecting outcomes). The difference-in-differences approach is increasingly applied to address this problem.[1]

Related articles pages 2364 and 2374

In this issue of *JAMA*, studies by Rajaram and colleagues[2] and Patel and colleagues[3] used the difference-in-differences approach to evaluate the changes that occurred following the 2011 Accreditation Council for Graduate Medical Education (ACGME) duty hour reforms. The 2 studies were conducted with different data sources and study populations but used similar methods.

Use of the Method

Why Was the Difference-in-Differences Method Used?

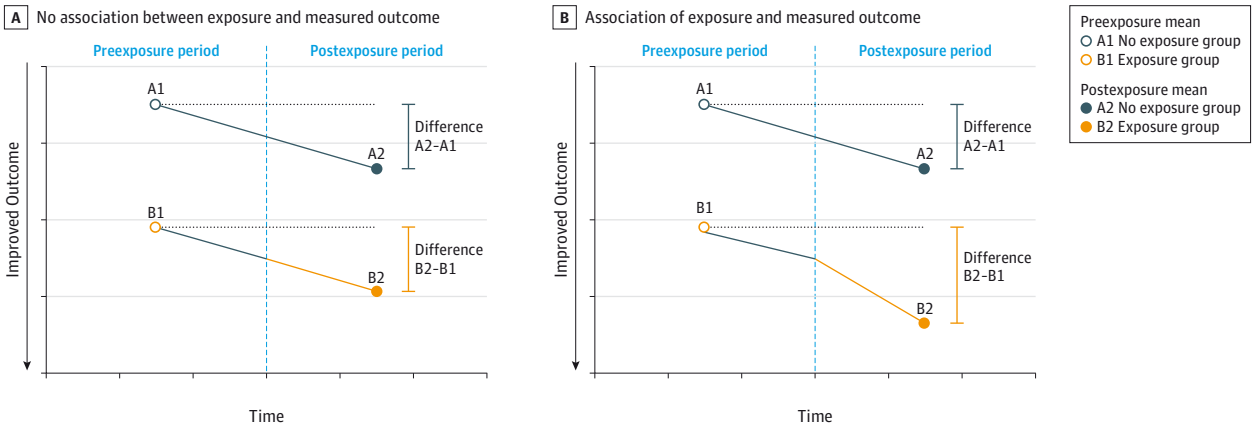
The association between policy changes and subsequent outcomes is often evaluated by pre-post assessments. Outcomes after implementation are compared with those before. This design is valid only if there are no underlying time-dependent trends in outcomes unrelated to the policy change. If clinical outcomes were already improving before the policy, then using a pre-post study would lead to the erroneous conclusion that the policy was associated with better outcomes.

The difference-in-differences study design addresses this problem by using a comparison group that is experiencing the same trends but is not exposed to the policy change.[4] Outcomes after and before the policy are compared between the study group and the comparison group without the exposure (group A) and the study group with the exposure (group B), which allows the investigator to subtract out the background changes in outcomes. Two differences in outcomes are important: the difference after vs before the policy change in the group exposed to the policy (B2 −B1, **Figure**) and the difference after vs before the date of the policy change in the unexposed group (A2 −A1). The change in outcomes that are related to implementation of the policy beyond background trends can then be estimated from the difference-in-differences analysis as follows: (B2 −B1) −(A2 −A1). If there is no relationship between policy implementation and subsequent outcomes, then the difference-in-differences estimate is equal to 0 (Figure, A). In contrast, if the policy is associated with beneficial changes, then the outcomes following implementation will improve to a greater extent in the exposed group. This will be shown by the difference-in-differences estimate (Figure, B).

These estimates are derived from regression models rather than simple subtraction. Using regression modeling allows the estimates to be adjusted for other factors (eg, patient or hospital characteristics) that may differ between the groups.[4] Regression models also offer a way to estimate the statistical significance of the association between policy change and outcomes, by including a variable that indicates if the observation is in the pre or post period and another variable that divides the groups into those exposed and unexposed to the policy.

Statistically, the association between policy implementation and outcomes is estimated by examining the interaction between the pre-post and exposed-unexposed variables. If the association exists,

Figure. Conceptual Illustration of a Difference-in-Differences Analysis for 2 Scenarios



this interaction term will be significantly different from zero. Other design and statistical issues should be considered when performing difference-in-differences analysis and are considered in detail elsewhere.[1,5]

What Are the Limitations of the Difference-in-Differences Method?

The 2 main assumptions of difference-in-differences analysis are parallel trends and common shocks.[4] The *parallel trends assumption* states that the trends in outcomes between the treated and comparison groups are the same prior to the intervention (Figure). If true, it is reasonable to assume that these parallel trends would continue for both groups even if the program was not implemented. This is tested empirically by examining the trends in both groups before the policy was implemented. In a regression model, this is evaluated by assessing the significance of the interaction term between time and policy exposure in the preintervention period. If the trends are significantly different prior to the intervention, a difference-in-differences analysis would be biased and a different comparison group should be sought.

In economics, a *shock* is an unexpected or unpredictable event (unrelated to the policy) that affects a system. The *common shocks assumptions* state that any events occurring during or after the time the policy changed will equally affect the treatment and comparison groups. A key limitation to implementing difference-in-differences design is finding a control group for which these assumptions are met. Ideally, the only difference between the comparison group and the study group would be exposure to the policy. In practice, such a group may be difficult to find.

Why Did the Authors Use the Difference-in-Differences Method?

The studies by Rajaram et al[2] and Patel et al[3] both used the difference-in-differences method to control for background trends in patient outcomes. The study by Rajaram et al, conducted using a large clinical registry for surgical patients (American College of Surgeons National Surgical Quality Improvement Program), evaluated several clinical outcomes (mortality, serious morbidity, readmission, failure to rescue) and American Board of Surgery pass rates after vs before the 2011 ACGME duty hour reforms.[2] The authors chose to use nonteaching hospitals as a control group, which makes the assumption that teaching and nonteaching hospitals have similar trends for improved outcomes prior to the ACGME policy changes. Similarly, the study by Patel et al, conducted using Medicare claims data, evaluated mortality and readmissions after vs before the ACGME duty hour reforms, also using a comparison group of nonteaching hospitals.[3]

How Should the Findings Be Interpreted?

Both studies found no association of the 2011 ACGME duty hour reform with clinical outcomes. After accounting for the slight background trend for improved outcomes among these populations using the difference-in-differences method, there was no additional improvement (or worsening) in outcomes associated with the ACGME policy. Both studies had strong comparison groups and neither appeared to violate the key assumptions of this approach. The rigorous approach and the consistency of the finding across outcomes make a compelling case that there was no association between implementation of the policy and the measured outcomes.

What Caveats Should the Reader Consider?

Difference-in-difference analyses must also account for spillover effects. Spillovers occur when some aspect of the policy spills over and influences clinical care in the hospitals unexposed to the policy (eg, nonteaching hospitals improved quality in some way in reaction to the ACGME duty hour reforms). Spillover can be evaluated by examining whether there is a measurable change in outcomes in the comparison group of hospitals at the time of the policy implementation. In the studies in this issue of *JAMA*, the lack of a change in outcomes among nonteaching hospitals at the time of the duty hour reforms suggests there were no associated spillover effects.

ARTICLE INFORMATION

Author Affiliations: The Center for Healthcare Outcomes and Policy, University of Michigan, Ann Arbor (Dimick, Ryan); Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor (Dimick, Ryan); Department of Surgery, School of Medicine, University of Michigan, Ann Arbor (Dimick); Department of Health Policy & Management, School of Public Health, University of Michigan, Ann Arbor (Dimick, Ryan).

Corresponding Author: Justin B. Dimick, MD, MPH, University of Michigan Medical School, 2800 Plymouth Rd, Bldg 16, Office 136E, Ann Arbor, MI 48109 (jdimick@umich.edu).

Section Editors: Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, *JAMA*.

Conflict of Interest Disclosures: Both authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Dimick reported that he receives grant funding from the National Institutes of Health, the Agency for Healthcare Research and Quality, and BlueCross BlueShield of Michigan Foundation; and is a cofounder of ArborMetrix Inc, a company that makes software for profiling hospital quality and efficiency. Dr Ryan reported that he receives grant funding from Agency for Healthcare Research and Quality.

REFERENCES

1. Ryan AM, Burgess J, Dimick, JB. Why we shouldn't be indifferent to specification in difference-in-differences analysis [published online December 9, 2014]. *Health Serv Res*. doi:10.1111/1475-6773.12270.

2. Rajaram R, Chung JW, Jones AT, et al. Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance. *JAMA*. doi:10.1001/jama.2014.15277.

3. Patel MS, Volpp KG, Small DS, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and readmissions among hospitalized Medicare patients. *JAMA*. doi:10.1001/jama.2014.15273.

4. Angrist JD, Pischke JS. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press; 2008.

5. Bertrand M, Duflo E, Mullainathan S. How much should we trust differences-in-differences estimates? *Q J Econ*. 2004;119:249-275.