



HHS Public Access

Author manuscript

JAMA Pediatr. Author manuscript; available in PMC 2021 March 23.

Published in final edited form as:

JAMA Pediatr. 2020 October 01; 174(10): 925–927. doi:10.1001/jamapediatrics.2020.3408.

Study Designs and Statistical Methods for Studies of Child and Adolescent Health Policies

Benjamin French, PhD,

Vanderbilt University Medical Center, Nashville, Tennessee; and Statistical Editor, *JAMA Pediatrics*.

Elizabeth A. Stuart, PhD

Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

Rigorous empirical studies of the effect of a policy intervention seek to consider (or estimate) what outcomes are (or would be) with the policy compared with what outcomes are (or would be) without the policy. For example, consider whether decriminalization of adult marijuana use (medical or recreational) is associated with adolescent marijuana use.¹ As detailed below, one can use data over time from states that did and did not decriminalize adult marijuana use and compare observed trends in adolescent marijuana use among states with the policy change to expected (or predicted) trends in marijuana use, had the policy change not occurred, to estimate the policy effect. Of note, the policy effect could also be estimated in settings in which there is not a comparison group, such as if marijuana were decriminalized nationwide. We focus on settings often referred to as *group panel data*, for which there are aggregate data available on groups of interest with outcomes measured over time both before and after the policy change and ideally with comparison groups that did not experience a policy change; individual-level data could also be available within the groups. The data in some cases correspond with full population data at each time point; in others, there might be repeated cross-sections of data, such as annual surveys of marijuana use among 10th graders. As long as the data can be thought of as representative of the unit under study, either data structure can be appropriate. We broadly consider the selection of data to examine (eg, the units to study, the time period to examine) as well as the statistical methods that can be used to estimate policy effects using that data.

Study Designs

In general, a randomized experiment is a particularly strong design for estimating causal effects. Therefore, researchers should look for opportunities to randomize study units to policy conditions. While randomized studies might not be feasible for national or state-level policies, there could be opportunities to randomize policies at a local level, such as randomizing different clinician groups to implement depression screening at their pediatric practice. A common design for such studies is a wait-list control or stepped-wedge design, which essentially randomizes the time at which each study unit receives the policy

Corresponding Author: Benjamin French, PhD, Department of Biostatistics, Vanderbilt University Medical Center, 2525 West End Ave, Ste 1100, Nashville, TN 37203 (b.french@vumc.org).

intervention.² A stepped-wedge design is often particularly effective when resources to roll out the intervention are constrained, such as having limited trainers who are able to travel to each pediatric practice and the selection of who to train first, second, and so on could be done randomly.

However, there are many situations in which randomization of policy conditions is not feasible or there is interest in effects beyond those study units that would agree to be randomized. In these contexts, a careful study design is still of utmost importance. Temporal ordering is crucial, with covariates measured before policy implementation and policy implementation measured before outcomes. Unfortunately, some policy studies simply use cross-sectional data to compare outcomes (eg, adolescent marijuana use) between study units (eg, states) with and without a particular policy (eg, decriminalization of adult marijuana use) but where everything is measured at the same point in time. Cross-sectional studies should be viewed as simple correlational analyses because this sort of association cannot disentangle policy effects from preexisting differences between the units that do and do not implement the policy. Likewise, researchers and readers should be wary of studies that perform simple pre-vs-post comparisons of outcomes before a policy was implemented with those after within the same study units. These pre-vs-post comparisons potentially conflate trends over time with the effect of the policy and can lead to particularly misleading policy effects if there are underlying temporal changes in the outcome in the absence of the policy change (eg, increasing marijuana use among adolescents).

The strongest policy evaluation designs take advantage of variation in time, before and after a policy is implemented, and space, with some units that implemented the policy and some that did not. These quasi-experimental methods, known as comparative interrupted time series or difference in differences, essentially estimate the policy effect by comparing trends over time between study units with and without the policy of interest.^{3,4} Note that the policy could have been implemented at different times in different study units; this should be accounted for in analyses (eg, by anchoring time at the timing of policy change, not necessarily at a particular calendar time). A body of statistical literature has established methods to do such modeling well in a way that allows accurate estimation and inference regarding policy effects. Briefly, the key assumption underlying these methods is that the treated and comparison units would have experienced the same trends in the outcomes in the absence of the policy change. This assumption leads to a need for careful selection of comparison units; various methods exist for this, including some based on propensity scores.^{5,6}

Statistical Models

Statistical modeling to estimate policy effects is also nuanced, with several important analysis issues to consider given the available data.⁷ Analyses typically proceed via a regression model appropriate for the distribution of the outcome under study, such as a Poisson regression model for count outcomes. As with any regression-based analysis, formulating a model for estimating the policy effect requires specifying the form for the exposure of interest (ie, the policy change), as well as for key adjustment covariates. Therefore, a primary consideration is modeling the form of the policy effect, such as

whether the policy is associated with an abrupt and sustained change in outcome values, with a gradual change in outcome values over time, or a combination of the 2. In terms of the statistical model, these analysis choices can be represented as including a policy indicator variable and/or its interaction with a variable (or variables for nonlinear terms) for time. Model specification requires close collaboration with substantive experts who know how the policy was implemented and how quickly it was likely to have an effect.

As noted earlier, rigorous policy change studies should leverage variation across time to inform estimation of the policy effect. Therefore, such variation must be accounted for in the statistical analysis. First, adjustment for calendar time is required to disentangle the policy effect from any background temporal trend. Second, collecting repeated measures on the same study units over time introduces within-unit correlation that must be acknowledged in the analysis. Failure to account for correlation can result in invalid association estimates and incorrect confidence intervals. Various statistical methods developed for the analysis of correlated data are directly applicable, such as generalized estimating equations and generalized linear mixed-effects models.^{8,9} Generalized estimating equations combine a population-averaged model for the mean outcome with an assumed structure for correlation, along with a robust variance estimator. Generalized linear mixed-effects models include unit-level random effects in a model for the unit-specific mean outcome to formulate a correlation structure; random effects can also be used to quantify heterogeneity in the policy effect across study units. These methods have benefits and drawbacks that must be weighed for the policy context under study, including their underlying assumptions, target of inference, and robustness to missing data.

The design of policy evaluations and the statistical analysis of the resultant data require close, ongoing collaboration among members of the research team.¹⁰ For example, it is crucial to understand what other factors might have influenced adolescent marijuana use around the same time as decriminalization occurred or whether other relevant policy changes were made at the same time. Investigators must collaborate with statisticians to efficiently design the study and properly interpret the results. In turn, statisticians must collaborate with investigators to ensure that the statistical analysis is congruent with the study's goals and scientific context. We hope that the considerations overviewed herein serve as a starting point for such collaboration.

Conflict of Interest Disclosures:

Dr Stuart was supported by the Bloomberg American Health Initiative and the National Institute of Drug Abuse (grant P50DA046351). No other disclosures were reported.

REFERENCES

1. Cerdá M, Wall M, Feng T, et al. Association of state recreational marijuana laws with adolescent marijuana use. *JAMA Pediatr.* 2017;171(2):142–149. doi:10.1001/jamapediatrics.2016.3624 [PubMed: 28027345]
2. Ellenberg SS. The stepped-wedge clinical trial: evaluation by rolling deployment. *JAMA.* 2018;319(6):607–608. doi:10.1001/jama.2017.21993 [PubMed: 29450512]

3. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. 2014;312(22):2401–2402. doi:10.1001/jama.2014.16153 [PubMed: 25490331]
4. Zeldow B, Hatfield L. Differences-in-differences. Health Policy Data Science Lab website. 1 27, 2018. Accessed June 30, 2020. <http://diff.healthpolicydatascience.org>
5. Basu S, Meghani A, Siddiqi A. Evaluating the health impact of large-scale public policy changes: classical and novel approaches. *Annu Rev Public Health*. 2017;38:351–370. doi:10.1146/annurevpublhealth-031816-044208 [PubMed: 28384086]
6. Stuart EA, Huskamp HA, Duckworth K, et al. Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Serv Outcomes Res Methodol*. 2014;14(4): 166–182. doi:10.1007/s10742-014-0123-z [PubMed: 25530705]
7. French B, Heagerty PJ. Analysis of longitudinal data to evaluate a policy change. *Stat Med*. 2008; 27(24):5005–5025. doi:10.1002/sim.3340 [PubMed: 18618416]
8. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat Med*. 2009;28(2):221–239. doi:10.1002/sim.3478 [PubMed: 19012297]
9. Detry MA, Ma Y. Analyzing repeated measurements using mixed models. *JAMA*. 2016; 315(4):407–408. doi:10.1001/jama.2015.19394 [PubMed: 26813213]
10. McGinty EE, Stuart EA, Caleb Alexander G, Barry CL, Bicket MC, Rutkow L. Protocol: mixed-methods study to evaluate implementation, enforcement, and outcomes of U.S. state laws intended to curb high-risk opioid prescribing. *Implement Sci*. 2018;13(1):37. doi:10.1186/s13012-018-0719-8 [PubMed: 29482599]