American Journal of Epidemiology **Submitted Manuscript**

Special Collection: Pharmacoepidemiology: From Real World Data to Real World Evidence

Title: Quasi-experimental methods for pharmacoepidemiology: difference-in-differences and synthetic control methods with case studies for vaccine evaluation

Authors: Lee Kennedy-Shaffer, PhD

ORCiD ID: 0000-0001-7604-3638

Correspondence Address: Box 0226, 124 Raymond Ave, Poughkeepsie, NY, 12604-0226 (email: lkennedyshaffer@vassar.edu)

Joint Authorship: N/A

Affiliations: Department of Mathematics and Statistics, Vassar College

Key words: Difference-in-differences; natural experiments; quasi-experiments; staggered treatment adoption; synthetic control method; vaccine effectiveness

Acknowledgments[1]**:** N/A

Funding: N/A

Conflict of Interest: The author has no conflicts to report.

Disclaimer: The views expressed in this article are those of the author alone.

Data Availability Statement: N/A

ABSTRACT

Difference-in-differences and synthetic control methods have become common study designs for evaluating the effects of policy changes, including health policies. They also have potential for providing real-world effectiveness and safety evidence in pharmacoepidemiology. To effectively add to the toolkit of the field, however, designs—including both their benefits and drawbacks—must be well understood. Quasi-experimental designs provide an opportunity to estimate the average treatment effect on the treated without requiring the measurement of all possible confounding factors, and to assess population-level effects. This requires, however, other key assumptions, including the parallel trends or stable weighting assumptions, a lack of other concurrent events that could alter time trends, and an absence of contamination between exposed and unexposed units. The targeted estimands are also highly specific to the settings of the study, and combining across units or time periods can be challenging. Case studies are presented for three vaccine evaluation studies, showcasing some of these challenges and opportunities in a specific field of pharmacoepidemiology. These methods provide feasible and

[1] Study investigators, conference presentations, preprint publication information, thanks.

valuable sources of evidence in various pharmacoepidemiologic settings and can be improved through research to identify and weigh the advantages and disadvantages in those settings.

Abbreviations: COVID-19, coronavirus disease 2019; DiD, difference-in-differences; SCM, synthetic control method; ATT, average treatment effect on the treated; TWFE, two-way fixed effects; BCG, bacille Calmette-Guérin

INTRODUCTION

Determining the safety and efficacy of drugs and biologics is crucial, yet it faces many statistical, epidemiological, regulatory, and logistical challenges. A related yet distinct challenge—determining the effects of health and social policies—has incorporated quasi-experimental study designs into its set of methods. Quasi-experimental methods can provide a convenient means for assessing the real-world effectiveness and safety of pharmacological products—especially those with population-level effects—in well-defined populations and specific circumstances that mitigate the risks of bias.

Quasi-experiments (in some cases also known as "natural experiments") here refer to observational studies that identify a causal effect by taking advantage of circumstances that create variation in exposure status in ways that avoid the usual sources of confounding for that exposure and the outcome of interest.[1,2] This often comes through changes in the exposure that are not otherwise causally connected to the outcome, often referred to as exogeneity.[2,3]

A suite of such approaches developed or formalized in the quantitative social sciences — including instrumental variables, regression discontinuity designs, interrupted time series, difference-in-differences, and synthetic control methods, among others—have become major empirical approaches in those fields[3,4] and have become popular in epidemiology as well.[5] For example, during the coronavirus disease 2019 (COVID-19) pandemic, they have been used to assess the effectiveness of universal masking in schools[6], vaccine lotteries[7], and other interventions.[8,9] Along with this rise in the use of the methods have come articles describing the uses and limitations of these methods in health policy evaluation.[1,2,10–16]

This article will focus on the difference-in-differences (DiD) and synthetic control method (SCM), quasi-experimental designs that are both controlled (i.e., include exposed and

unexposed units) and longitudinal (i.e., exploit a change in exposure status over time to adjust for unmeasured confounders).[11,17]

Despite their expanding use in health policy research, these methods have not been extensively studied for use in evaluating drugs and biologics. This article provides a brief overview of these methods (including recent developments from the econometrics literature), advice on settings where they are beneficial for pharmacoepidemiology (exemplified through three case studies of vaccine effectiveness evaluation), and discussion of the tradeoffs and limitations they face. Finally, the article describes future research that can improve their use in pharmacoepidemiology.

OVERVIEW OF METHODS

DiD and SCM analyses estimate the average treatment effect on the treated (ATT) by constructing the potential outcome for the exposed units in the (counterfactual) absence of the exposure using both cross-unit and within-unit comparisons. Below, I give a brief overview of these methods and their estimators, along with recent extensions. For a more comprehensive treatment, see the relevant econometrics[3,4] or public health review literature.[1,10,14,15,17]

Difference-in-Differences

Used (in function if not name) by John Snow in his investigation of the 1854 London cholera outbreak,[3,18,19] the difference-in-differences method grew in popularity after Card and Krueger's 1994 study on minimum wage increases.[20] DiD methods rely on the parallel trends assumption: in the absence of exposure, the exposed and unexposed units would have equivalent trends over time in their expected outcomes. For the canonical "two-by-two" DiD analysis, with one pre-exposure and one post-exposure period for the exposed unit, and the same two time periods for the unexposed units, the ATT estimate is given by the difference between the pre-

post difference in the exposed unit and the pre-post difference in the unexposed unit (see Figure 1A). For multiple units and time points, this is commonly extended into the following two-way fixed effects (TWFE) regression model:[21]

$$E[Y_{it}|i,t] = \alpha_i + \gamma_t + \beta D_{it},$$

where Y_{it} is the outcome of interest for unit *i* in time period *t*, α_i are the unit-specific fixed effects, γ_t are the period-specific fixed effects, and D_{it} is an indicator of whether unit *i* was exposed in period *t*. The ATT estimate is then given by $\hat{\beta}$. The fixed effects control for within-unit confounding over time and within-period confounding between exposure groups.

Extensions of the DiD Model

Several extensions to the DiD model aim to improve the validity of the parallel trends assumption. For example, covariates can be added when parallel trends holds only conditionally, or the so-called difference-in-difference-in-differences (or triple-differences) model can be used to remove non-parallel trends.[4] Both of these have analogous approaches in the epidemiology literature: adjusting for measured covariates in cohort studies for the former, and using negative controls to adjust for residual bias for the latter.[22,23]

The functional form of the parallel trends assumption can also be changed by transforming the outcome variable. For example, if the counterfactual outcomes are judged more likely to have equivalent multiplicative trends than linear trends, a logarithmic transformation can be used on the outcome. Importantly, only one functional form can truly have parallel trends, so the modeling and estimand choices determine the validity of the causal assumptions made.[24]

Recently, literature has extensively investigated DiD in cases where there are units with different exposure onset times (e.g., "staggered treatment adoption"; see Figure 1B).[25] The TWFE model is subject to biased estimation of the ATT under this setting as the unit and time

fixed effects do not properly capture treatment effect heterogeneity.[21,25] Various approaches have been proposed to ameliorate this problem in the DiD and related stepped-wedge trials literature;[26–30] these include different estimators and explicit weighting of time- and unit-specific estimators.[3,26,27]

Synthetic Control Method

The SCM, developed for comparative case studies, has also become a common method for policy evaluation.[12,31] The core of the design is modeling the counterfactual outcome of the exposed unit using a weighted average of the unexposed units (see Figure 1C). For a setting with one exposed unit and *n* unexposed units (with outcomes in period *t* denoted Y_{1t} and $Y_{01t}, \dots, Y_{0nt}$, respectively), the estimator for the ATT in period *t* is:

$$\hat{\theta}_t = Y_{1t} - \sum_{i=1}^{n} w_i Y_{0it},$$

where $\{w_1, w_2, \dots, w_n\}$ are nonnegative weights summing to 1. These weights are chosen by minimizing the pre-exposure difference between the outcomes of the not-yet-exposed unit and the weighted average of the outcomes of the unexposed units; minimizing covariate differences can also be incorporated.[3,31,32]

The major assumption of SCM is that these weights are stable, in the sense that a weighting scheme that matches the pre-exposure outcome trends well will also give an unbiased estimate of the potential outcome after the exposure time point.[31] A long pre-exposure time series on which to optimize the weights provides the best justification for this assumption under many data-generating processes.[31] For multiple exposed units and/or multiple post-exposure periods, the individual estimators can be combined to target the desired estimand.[29,33]

Extensions of the Synthetic Control Method

One reason for the growth of SCM in comparative case studies is its interpretability: all unexposed units have a nonnegative weight showing their contribution to the estimated counterfactual, allowing discussion of the reasonableness of the counterfactual. However, this also imposes restrictions on the model; SCM does not work when the outcomes of the exposed unit fall outside the convex hull (loosely, the range) of the unexposed units' outcomes.[31] Thinking of past outcomes as the "adjustment variables", this requirement is roughly analogous to positivity requirements in cohort studies: adjustment fails if there are combinations of covariates for which there are only exposed units.

Several extensions of SCM—including the generalized SCM[34], augmented SCM (see Figure 1D)[35], and Bayesian structural time series model[36]—trade off some interpretability by allowing outcome modeling and extrapolation outside of the convex hull. This allows the use of more control series, including those on different scales than the outcome itself.[37] This extrapolation, however, can lead to bias, especially when there are few pre-treatment periods or unexposed units.[34]

Quantifying Uncertainty and Negative Controls

Pharmacoepidemiologic studies require not just point estimates but also quantification of uncertainty, often through p-values, confidence intervals, or Bayesian alternatives.[38] For quasi-experiments, this can be more challenging, as they often involve one or a limited number of exposed observations, and rarely form a sample of an identifiable study population. Because of this, model-based estimates of variability—even in cases when they can be computed, such as regression-based DiD and generalized SCM—may not have the desired interpretation or may require additional design-based assumptions.[27]

Beyond statistical inference, it is desirable to quantify the causal or model-based uncertainty. Negative controls, sensitivity analyses, and robustness checks provide a means to assess the required assumptions to some extent. These include assessing the validity of the assumptions for the particular setting through models and graphical inspection of pre-exposure time trends, as well as assessing the model performance in settings with known effects.[17]

Repeated negative controls—often called placebo tests or dummy analyses—conducted for different time points (where no change in exposure occurs), different units (i.e., only using untreated units), or different outcomes (where the exposure should have no effect) generate null distributions using cases that should identify effects of zero.[3,31] Comparing the observed estimate to these distributions allows hypothesis testing. Bayesian frameworks are also feasible, but may be sensitive to the prior distributions used.[36] Properly reporting and interpreting these measures of uncertainty remains a challenge for these quasi-experimental methods, especially as they are applied to new fields.

CASE STUDIES: QUASI-EXPERIMENTAL VACCINE EVALUATION

Case studies of the use of DiD and SCM in vaccine evaluation can illuminate the benefits and challenges of these designs. Vaccine evaluation studies demonstrate the use of large-scale routinely collected data sets in the context of exogenous variation to identify key public health effects as a supplement to randomized trial evidence or when it is not available. They also show potential pitfalls: the risk of concurrent events and misspecification leading to bias and the limited generalizability of targeted estimands.

Case 1: SCM Analysis of Meningococcal Vaccines (Prunas et al. 2022)

Prunas et al. analyzed the impact of early childhood meningococcal vaccination programs in Brazil and England in the 2010s (see Table 1).[39] Real-world estimates of vaccine effectiveness

for hard-to-predict and highly variable diseases like invasive meningococcal disease are challenging, making quasi-experiments particularly useful. Using SCM on the logarithmic scale with Bayesian time series modeling, the authors compared meningococcal disease incidence after the program roll-out to a synthetic control constructed from time series of other (non-targeted) diseases and the targeted disease in older age groups who were ineligible for vaccination.

To demonstrate the value of the method, the authors conducted placebo tests using other quasi-experimental designs and found that the SCM better captured seasonality and non-linear time trends; notably it avoids the need to specify a parametric time trend. It thus likely suffers from less bias in the desired analyses, although there could still be extrapolation bias due to the time series modeling.[32,36]

Within-region controls (here, time series of non-targeted populations and outcomes) are particularly valuable for quasi-experimental evaluation of vaccines and drugs. By avoiding comparisons across geographic areas, they mitigate the risk of bias due to concurrent events or changes. However, they come with a potential increased risk of spillover or contamination, especially in infectious disease settings.[40] For the non-targeted age groups, bias could be caused by indirect protection of the vaccine (i.e., a younger child receiving the vaccine is less likely to infect an older child in the household). For the non-targeted diseases, there is perhaps less risk of bias, but it could still occur due to off-target effects of vaccination or changes in health-seeking behavior and diagnosis in the wake of the vaccination policy. The authors used placebo tests and sensitivity analyses that dropped the older age groups as candidate control series, finding robust results that suggest these biases were minimal in this setting.[39]

The results demonstrated effectiveness of the vaccination program against the targeted infections. Within each country of analysis, results on different early childhood age groups

showed remarkable similarity, indicating that there may be some generalizability of these effects. However, as ATT estimands, these may still not be appropriate to transport to other countries with different infection risks. On the other hand, these estimates assess the overall effect of the program, rather than the individual-level direct effect, so can provide an additional piece of evidence that is rarely achieved in pre-authorization trials.[41] This is useful for public health policy-makers as real-world evidence of both the vaccine itself and the vaccination program.[42]

DiD Analysis of an Off-Target Vaccine Against COVID-19

Early in the COVID-19 pandemic, researchers sought to use existing drugs and vaccines to prevent infection, illness, and severe outcomes until COVID-19-specific drugs and vaccines could be developed and authorized. The bacille Calmette-Guérin (BCG) vaccine against tuberculosis was one such candidate because of its hypothesized off-target effects[43], which were supported by cross-sectional analyses showing better COVID-19 outcomes in countries with high BCG vaccination coverage.[44] While randomized controlled trials would eventually occur in specific populations[43], the emergency situation called for interim evidence to confirm or repute this hypothesis. A preprint by Matsuura et al. in 2020 re-examined this hypothesis with a DiD analysis using existing country-level data (see Table 2).[44]

The authors identified countries that changed BCG vaccination recommendations at some point in the past and determined for various age groups in each country whether they were in a national vaccination cohort or not. They then conducted a TWFE analysis of the relationship between being in a BCG vaccination cohort and the log of confirmed COVID-19 cases for that age group and country, finding no protective effect of the vaccine.[44]

In this study, incorporating within-country controls helped mitigate the bias of comparing across countries cross-sectionally. However, the use of age cohorts as the distinguishing feature

risks contamination—as individuals of different ages interact and may provide indirect protection—and concurrent events, since many non-pharmaceutical interventions and recommendations in the COVID-19 pandemic were targeted to specific age groups. Identification for this design requires that the difference in log-cases between age cohorts would be constant (parallel multiplicative trends) across countries absent differential BCG vaccination policies. This would be threatened by differential age-targeted policies.

As in the previous study, the generalizability of the ATT is limited, since it is specific to the age cohorts and countries studied, as well as the time since BCG vaccination inherent in those age groups. However, the staggered adoption that occurs in this setting (different countries changed vaccination rules for different age cohorts) risks a larger problem of bias or incorrect interpretation of the ATT. The estimate given by the TWFE model is a weighted average of individual effects in each country-age cohort combination, with some given potentially negative weights.[21] Careful specification of the estimand and the analysis method in quasi-experimental analyses, especially with staggered adoption, is thus crucial. In this case, individual two-by-two DiD analyses may be more useful, providing a range of estimated treatment effects in the different age group-country combinations.[21,45]

The challenges of estimand interpretation and generalizability limit the internal and external validity of this approach. Nonetheless, it provided useful exploratory evidence that controlling to some extent for country-level factors could account for the correlation observed in a cross-sectional analysis. As this negative finding was borne out by a randomized controlled trial[43], this study provided immediate useful interim evidence before that trial was concluded.

DiD and SCM Analyses of the Indirect Protection of a COVID-19 Vaccine

After the authorization of COVID-19 vaccines to prevent symptomatic and severe illness in adults, questions remained about the effectiveness of the vaccines in providing indirect protection by preventing infection and transmission.[41] Winner et al. sought to assess the indirect protection of children via adult vaccination by comparing infection rates among Austrian children in the district of Schwaz—which had high adult vaccination uptake due to a campaign during a localized outbreak— to other districts with much lower rates as of March 2021. They used both a SCM analysis with other Austrian districts as controls and DiD analyses comparing Schwaz municipalities to neighboring municipalities outside the district (see Table 3).[46] All analyses reported strong indirect protection of children.[46]

With a clear time point for the mass vaccination campaign, quasi-experimental methods are a natural design for this setting, and the neighboring municipalities and other districts in the country form a large pool of potential controls. Selecting unexposed units within the same country increases the plausibility of similar non-pharmaceutical interventions and other policies, supporting the assumptions of stable weights and parallel trends. The risk of cross-border contamination through indirect protection remains but would, if anything, yield conservative estimates. Negative controls included an in-space placebo test for SCM (i.e., using the same method on non-targeted districts to create a null distribution) and assessment of pre-exposure parallel trends for DiD.[46]

While policies spurred by a particular need create variation, there is risk in using them as quasi-experiments. In this case, a large outbreak in Schwaz spurred the vaccination campaign. This creates concurrent events and anticipation: the past outbreak and the vaccination both influence infection dynamics going forward and may violate the assumptions needed for these methods.[8] The authors sought to use districts with similar prior infection trends to mitigate this

bias. Parallel trends could also be threatened by the use of linear functional forms for a non-linear outcome like infections, although the authors used both a linear TWFE model and a logarithmic two-by-two DiD model here.[8,17,47]

Once again, the results are highly contingent to the campaign and the setting in which it was enacted. The differential outcomes result from both the vaccination itself and any changes in behavior that may have resulted from a mass vaccination campaign; this point is well-described in the quasi-experimental literature for policy evaluation.[1,19] For pharmacoepidemiologic purposes, then, this makes it difficult to ascribe the full effect to the pharmaceutical itself in such cases. In addition, indirect protection in infectious disease outbreaks is specific to the prior evolution of the outbreak and the population at risk, limiting the generalizability and interpretability of estimates.[23,40,48,49] Nonetheless, the finding of a large indirect effect could inform public health practice and motivate the study of this effect in other settings, even if precise estimates are not transportable.[41,42]

DISCUSSION

Quasi-experimental methods provide a promising set of designs for pharmacoepidemiology. The designs discussed here, among others, provide a class of observational study designs that can target useful public health-relevant estimands and provide meaningful real-world evidence in advance of or as a supplement to randomized controlled trials.[14,50,51] In the case of vaccines, for example, these designs can identify potential off-target effects of pharmaceutical products, encouraging further research into those effects,[44,51] and identify rare safety signals through the use of large, routinely-collected data.[52] In addition, these studies can provide important post-marketing evaluation of the effects of products on their target

outcomes in actual use, including indirect effects and changing effectiveness over time as conditions change.[53,54]

By avoiding the need to measure and model all confounders between exposure and outcome, quasi-experimental methods can achieve high internal validity with routine data sources.[3,4,32,52] This, however, requires other counterfactual assumptions, such as parallel trends or stable weights. Justifying these assumptions, especially the appropriate functional form, requires an understanding of the setting and likely effect.[8,17,24] As seen in the case studies, both linear and logarithmic scales are used in vaccine evaluation studies, leading to different assumptions and different estimands.

Moreover, spillover and contamination of control units by the exposure and anticipation or lagged effects of the intervention can cause bias, often towards the null.[8,27,32] For pharmacoepidemiology, this could occur through off-label prescribing or, as in the vaccine cases, indirect effects within networks and communities. Attributing the observed causal effect to the exposure also requires a lack of concurrent events in the study units, which may be threatened by changes in behavior that co-occur with pharmaceutical interventions, like behavior changes that follow vaccination.[5,8,9,17]

As in all medical studies, understanding the estimand targeted by these methods is crucial to properly using the evidence.[55] When aggregate population data are used, these methods target aggregate or population-level effects, and thus cannot be transported to individuals.[19] But this also provides useful real-world evidence, such as the overall effect measure in vaccine studies, as seen in the case studies.[23] This is rarely targeted by randomized trials and provides useful additional regulatory and public health evidence.[41,42,53]

ATT estimands from quasi-experimental studies, however, are specific to the setting, exposure, outcomes, and exposed unit(s), as well as the contrast identified and other statistical analysis choices. Since the exposed units are rarely randomly selected (i.e., the mass vaccination campaign in Schwaz occurred because of a prior outbreak[46]), the ATT differs from the average treatment effect that is usually targeted by randomized trials and used in regulatory decisions.[56] It may be of specific interest in safety studies, as it represents the effect of removing the exposure from all exposed units (e.g., compare to[6]). In general, however, it may be less transportable to populations yet to be exposed or require further assumptions.[5,16]

Compared to other observational study designs, quasi-experimental designs tend to trade away some external validity for internal validity, although these aspects should ideally be considered in concert.[3,57] Researchers need to be careful about generalizing results of these methods to other settings and populations, especially for outcomes that have marked spatiotemporal patterns (e.g., infectious diseases) or for populations with specific health needs and vulnerabilities (e.g., specific age groups or people with co-morbidities).[19]

Achieving the potential benefits of these designs will require further research and careful reporting and interpretation. The general methodology of these designs in the field of pharmacoepidemiology—as well as how to report results and place evidence in the context of other studies routinely conducted in the field—will be important for ensuring validity and appropriate interpretation.[9,11] Connecting the literature across various disciplines will improve uptake of the most appropriate methods as well. Moreover, research on the most appropriate designs for specific settings, considering the populations, interventions, and outcomes, is crucial.[50] This can include simulation studies in those settings to observe the relationship between estimates and true estimands.[15]

For example, quasi-experimental vaccine evaluation studies will need to determine appropriate lag times, functional forms, and time frames for the necessary DiD and SCM assumptions to hold.[14,52] This can build on the existing health policy and stepped-wedge trials literature.[8,47,48] Understanding and communicating the tradeoffs of identifying population-level effects but potentially losing generalizability will be key to appropriately contextualizing this evidence.

Negative controls, placebo tests, and sensitivity analyses also have an important role to play. Assessing methods for sensitivity to the selection of control units and time periods can inform both internal and external validity. Quantitative placebo tests can also allow the reporting of causal uncertainty alongside statistical uncertainty.[3] Investigators can improve generalizability by using multiple populations (as in the meningococcal vaccine case study[39]), different methodological approaches or models (as in the COVID-19 vaccine case study[46]), or multiple units and exposure points (as in the BCG vaccine case study[44]). The latter, however, targets an estimand that averages potentially heterogeneous treatment effects, which may or may not have greater generalizability—depending on the setting—and can be statistically inefficient.[21,30,48] The proposed alternatives vary in their targeted estimands, required assumptions, and statistical properties.[25–27,45,58] Investigation of the assumptions and relative performance of these methods in specific pharmacoepidemiologic settings (see, e.g.[8,9,17,47]), alongside the appropriate reporting of these challenges, is needed.

Quasi-experimental methods can alleviate some of the challenges of pharmacoepidemiology, while introducing others. Understanding and using them provides another tool for regulators, physicians, and public health policymakers to understand the benefits and risks of drugs and biologics.

REFERENCES

1. 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(1):39-56. doi:10.1146/annurev-publhealth-031816-044327

2. Matthay EC, Hagan E, Gottlieb LM, et al. Alternative causal inference methods in population health research: Evaluating tradeoffs and triangulating evidence. *SSM - Popul Health*. 2020;10:100526. doi:10.1016/j.ssmph.2019.100526

3. Cunningham S. *Causal Inference: The Mixtape*. New Haven, CT: Yale University Press; 2021.

4. Huntington-Klein N. *The Effect: An Introduction to Research Design and Causality*. Boca Raton, FL: CRC Press; 2022.

5. Matthay EC, Glymour MM. Causal inference challenges and new directions for epidemiologic research on the health effects of social policies. *Curr Epidemiol Rep*. 2022;9(1):22-37. doi:10.1007/s40471-022-00288-7

6. Cowger TL, Murray EJ, Clarke J, et al. Lifting universal masking in schools — Covid-19 incidence among students and staff. *N Engl J Med*. 2022;387(21):1935-1946. doi:10.1056/NEJMoa2211029

7. Brehm ME, Brehm PA, Saavedra M. The Ohio vaccine lottery and starting vaccination rates. *Am J Health Econ*. 2022;8(3):387-411. doi:10.1086/718512

8. Goodman-Bacon A, Marcus J. Using difference-in-differences to identify causal effects of COVID-19 policies. *Surv Res Methods*. 2020;14(2):153-158. doi:10.18148/srm/2020.v14i2.7723

9. Haber NA, Clarke-Deelder E, Feller A, et al. Problems with evidence assessment in COVID-19 health policy impact evaluation: a systematic review of study design and evidence strength. *BMJ Open*. 2022;12(1):e053820. doi:10.1136/bmjopen-2021-053820

10. 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(1):351-370. doi:10.1146/annurev-publhealth-031816-044208

11. Waddington H, Aloe AM, Becker BJ, et al. Quasi-experimental study designs series—paper 6: risk of bias assessment. *J Clin Epidemiol*. 2017;89:43-52. doi:10.1016/j.jclinepi.2017.02.015

12. Bouttell J, Craig P, Lewsey J, Robinson M, Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Community Health*. 2018;72(8):673-678. doi:10.1136/jech-2017-210106

13. Wing C, Simon K, Bello-Gomez RA. Designing difference in difference studies: Best practices for public health policy research. *Annu Rev Public Health*. 2018;39(1):453-469. doi:10.1146/annurev-publhealth-040617-013507

14. Lopez Bernal JA, Andrews N, Amirthalingam G. The use of quasi-experimental designs for vaccine evaluation. *Clin Infect Dis*. 2019;68(10):1769-1776. doi:10.1093/cid/ciy906

15. Nianogo RA, Benmarhnia T, O'Neill S. A comparison of quasi-experimental methods with data before and after an intervention: an introduction for epidemiologists and a simulation study. *Int J Epidemiol*. 2023;52(5):1522–1533. doi:10.1093/ije/dyad032.

16. Rothbard S, Etheridge JC, Murray EJ. A tutorial on applying the difference-in-differences method to health data [published online ahead of print September 7, 2023]. *Curr Epidemiol Rep*. (doi:10.1007/s40471-023-00327-x).

17. Haber NA, Clarke-Deelder E, Salomon JA, Feller A, Stuart EA. Impact evaluation of coronavirus disease 2019 policy: A guide to common design issues. *Am J Epidemiol*. 2021;190(11):2474-2486. doi:10.1093/aje/kwab185

18. Coleman T. Causality in the time of cholera: John Snow as a prototype for causal inference [preprint]. *SSRN*. 2019. (doi:10.2139/ssrn.3262234). Accessed May 23, 2023.

19. Caniglia EC, Murray EJ. Difference-in-difference in the time of cholera: A gentle introduction for epidemiologists. *Curr Epidemiol Rep*. 2020;7(4):203-211. doi:10.1007/s40471-020-00245-2

20. Card D, Krueger A. Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *Am Econ Rev*. 1994;84(4):772-793.

21. Goodman-Bacon A. Difference-in-differences with variation in treatment timing. *J Econom*. 2021;225(2):254-277. doi:10.1016/j.jeconom.2021.03.014

22. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388. doi:10.1097/EDE.0b013e3181d61eeb

23. Halloran ME, Hudgens MG. Estimating population effects of vaccination using large, routinely collected data. *Stat Med*. 2018;37(2):294-301. doi:10.1002/sim.7392

24. Kahn-Lang A, Lang K. The promise and pitfalls of differences-in-differences: Reflections on *16 and Pregnant* and other applications. *J Bus Econ Stat*. 2020;38(3):613-620. doi:10.1080/07350015.2018.1546591

25. Sun L, Abraham S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J Econom*. 2021;225(2):175-199. doi:10.1016/j.jeconom.2020.09.006

26. De Chaisemartin C, D'Haultfœuille X. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *Econom J*. 2023;26(3):C1–C30. doi:10.1093/ectj/utac017.

27. Roth J, Sant'Anna PHC, Bilinski A, Poe J. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *J Econom*. 2023;235(2):2218-2244. doi:10.1016/j.jeconom.2023.03.008

28. Thompson JA, Davey C, Fielding K, Hargreaves JR, Hayes RJ. Robust analysis of stepped wedge trials using cluster-level summaries within periods. *Stat Med*. 2018;37(16):2487-2500. doi:10.1002/sim.7668

29. Kennedy-Shaffer L, Gruttola V, Lipsitch M. Novel methods for the analysis of stepped wedge cluster randomized trials. *Stat Med*. 2020;39(7):815-844. doi:10.1002/sim.8451

30. Lindner S, McConnell KJ. Heterogeneous treatment effects and bias in the analysis of the stepped wedge design. *Health Serv Outcomes Res Methodol*. 2021;21(4):419-438. doi:10.1007/s10742-021-00244-w

31. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J Am Stat Assoc*. 2010;105(490):493-505. doi:10.1198/jasa.2009.ap08746

32. Abadie A. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *J Econ Lit*. 2021;59(2):391-425. doi:10.1257/jel.20191450

33. Ben-Michael E, Feller A, Rothstein J. Synthetic controls with staggered adoption. *J R Stat Soc Ser B Stat Methodol*. 2022;84(2):351-381. doi:10.1111/rssb.12448

34. Xu Y. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Polit Anal*. 2017;25(1):57-76. doi:10.1017/pan.2016.2

35. Ben-Michael E, Feller A, Rothstein J. The augmented synthetic control method. *J Am Stat Assoc*. 2021;116(536):1789-1803. doi:10.1080/01621459.2021.1929245

36. Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL. Inferring causal impact using Bayesian structural time-series models. *Ann Appl Stat*. 2015;9(1). doi:10.1214/14-AOAS788

37. Bruhn CAW, Hetterich S, Schuck-Paim C, et al. Estimating the population-level impact of vaccines using synthetic controls. *Proc Natl Acad Sci*. 2017;114(7):1524-1529. doi:10.1073/pnas.1612833114

38. Kennedy-Shaffer L. When the alpha is the omega: P-values, "substantial evidence," and the 0.05 standard at FDA. *Food Drug Law J*. 2017;72(4):595-635.

39. Prunas O, Weinberger DM, Medini D, Tizzoni M, Argante L. Evaluating the impact of meningococcal vaccines with synthetic controls. *Am J Epidemiol*. 2022;191(4):724-734. doi:10.1093/aje/kwab266

40. Halloran ME, Longini IM, Struchiner CJ. *Design and Analysis of Vaccine Studies*. New York: Springer; 2010.

41. Kennedy-Shaffer L. Public health impacts of vaccines for COVID-19 and beyond: opportunities to overcome technical and regulatory barriers for randomized trials. *Am J Public Health*. 2023;113(7):778-785. doi:10.2105/AJPH.2023.307302

42. Jones M, Jetelina KK. Vaccination has more to offer than direct clinical benefit: FDA's vaccine licensure process ignores population health and social determinants of disease. *Am J Epidemiol*. 2024;193(1):1–5. doi:10.1093/aje/kwad161.

43. Pittet LF, Messina NL, Orsini F, et al. Randomized trial of BCG vaccine to protect against Covid-19 in health care workers. *N Engl J Med*. 2023;388(17):1582-1596. doi:10.1056/NEJMoa2212616

44. Matsuura H, Fukui M, Kawaguchi K. Does TB vaccination reduce COVID-19 infection?: No evidence from a regression discontinuity and difference-in-differences analysis [preprint]. *Research Square*. 2020. (doi:10.21203/rs.3.rs-56123/v1). Accessed May 15, 2023.

45. De Chaisemartin C, D'Haultfœuille X. Two-way fixed effects estimators with heterogeneous treatment effects. *Am Econ Rev*. 2020;110(9):2964-2996. doi:10.1257/aer.20181169

46. Winner H, Kimpel J, Krammer F, Von Laer D, Paetzold J. Can high COVID-19 vaccination rates in adults help protect unvaccinated children? Evidence from a unique mass vaccination campaign, Schwaz/Austria, March 2021. *Eurosurveillance*. 2022;27(39). doi:10.2807/1560-7917.ES.2022.27.39.2101027

47. Callaway B, Li T. Policy evaluation during a pandemic. *J Econom*. 2023;236(1):105454. doi:10.1016/j.jeconom.2023.03.009

48. Kennedy-Shaffer L, Lipsitch M. Statistical properties of stepped wedge cluster-randomized trials in infectious disease outbreaks. *Am J Epidemiol*. 2020;189(11):1324-1332. doi:10.1093/aje/kwaa141

49. Kahn R, Schrag SJ, Verani JR, Lipsitch M. Identifying and alleviating bias due to differential depletion of susceptible people in postmarketing evaluations of COVID-19 vaccines. *Am J Epidemiol*. 2022;191(5):800-811. doi:10.1093/aje/kwac015

50. Basta NE, Halloran ME. Evaluating the effectiveness of vaccines using a regression discontinuity design. *Am J Epidemiol*. 2019;188(6):987-990. doi:10.1093/aje/kwz043

51. Eyting M, Xie M, Hess S, Geldsetzer P. Causal evidence that herpes zoster vaccination prevents a proportion of dementia cases [preprint]. *medRxiv*. 2023. (doi:10.1101/2023.05.23.23290253). Accessed May 26, 2023.

52. Pescarini JM, Teixeira CSS, Cruz EP, et al. Métodos para avaliação da efetividade de vacinas para COVID-19 com ênfase em abordagens quase-experimentais. *Ciênc Saúde Coletiva*. 2021;26(11):5599-5614. doi:10.1590/1413-812320212611.18622021

53. Clemens J. Evaluating new vaccines for developing countries. Efficacy or effectiveness? *JAMA*. 1996;275(5):390-397. doi:10.1001/jama.275.5.390

54. Saadatian-Elahi M, Horstick O, Breiman RF, et al. Beyond efficacy: The full public health impact of vaccines. *Vaccine*. 2016;34(9):1139-1147. doi:10.1016/j.vaccine.2016.01.021

55. Kahan BC, Cro S, Li F, Harhay MO. Eliminating ambiguous treatment effects using estimands. *Am J Epidemiol*. 2023;192(6):987-994. doi:10.1093/aje/kwad036

56. Permutt T. Do covariates change the estimand? *Stat Biopharm Res*. 2020;12(1):45-53. doi:10.1080/19466315.2019.1647874

57. Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. *Am J Epidemiol*. 2019;188(2):438-443. doi:10.1093/aje/kwy228

58. Callaway B, Sant'Anna PHC. Difference-in-differences with multiple time periods. *J Econom*. 2021;225(2):200-230. doi:10.1016/j.jeconom.2020.12.001

TABLES

Table 1. Summary of Meningococcal Vaccine Analyses in Prunas et al. 2022.

	Brazil Analysis	**England Analysis**
Research Question	Did meningococcal vaccination programs reduce early childhood invasive meningococcal disease incidence?	
Exposure	Within the target age group (<1-year and 1–4 years) of the MenC vaccination program	Within the target age group (18–51 weeks and 1-year-old) of the MenB vaccination program
Outcome	Total MenC cases (monthly)	Total MenB cases (quarterly)
Setting	Brazil, 2007–2013	England, 2011–2019
Control Series	Cases for various off-target infectious diseases in target age group; outcome of interest in non-targeted age groups	
Method	SCM with Bayesian variable selection	
Scale	Logarithmic	
Identification Assumption	Weighted average of control series is the expected outcome absent vaccination program	
Results	Vaccine effectiveness: • 69% (95% CrI: 51, 80) in <1-year group • 64% (95% CrI: 55, 70) in 1–4 year group	Vaccine effectiveness: • 75% (95% CrI: 69, 80) in 18–51-week group • 72% (95% CrI: 65, 79) in 1-year-old group
Negative Controls/ Placebo Tests	• Models tested on non-targeted age groups and compared to other methods • Sensitivity analyses excluding non-targeted age groups as control series	

Abbreviations: CrI, credible interval; MenB, meningococcal serogroup B; MenC, meningococcal serogroup C; SCM, synthetic control method

Table 2. Summary of BCG Vaccine Off-Target Effects vs. COVID-19 Analysis in Matsuura et al. 2020.

Research Question	Did prior BCG vaccination recommendation reduce COVID-19 incidence?
Exposure	Within age-group cohort covered by national BCG vaccination recommendation
Outcome	Confirmed COVID-19 cases per 1,000
Setting	Various countries, unknown dates of analysis for outcome
Control Series	Outcome of interest in age group cohorts not included in national BCG vaccination recommendations
Method	DiD TWFE model, with country and age-group fixed effects
Scale	Logarithmic
Identification Assumption	Expected ratio of infection rates across age-groups and countries are equal absent BCG vaccination recommendation
Results	Do not support hypothesis of protective effect of BCG vaccine
Negative Controls/ Placebo Tests	None reported
Additional Details	• Includes BCG strain as a covariate to address hypotheses generated from cross-sectional analysis of countries • Exposure onset is by age cohort, not calendar time

Abbreviations: BCG, bacilli Calmette-Guérin; COVID-19, coronavirus disease 2019; DiD, difference-in-differences; TWFE, two-way fixed effects

Table 3. Summary of COVID-19 Vaccine Indirect Protection Analyses in Winner et al. 2022.

	SCM Analysis	**DiD Analysis**
Research Question	Did a mass vaccination campaign in adults reduce COVID-19 incidence in children?	
Exposure	Within ineligible age cohort (under 16 years old) in the district of Schwaz	
Outcome	Cumulative daily SARS-CoV-2 infections per 100,000	
Setting	Schwaz District of Austria, January–May 2021	
Control Series	Outcome of interest in other districts	Outcome of interest in bordering municipalities
Method	SCM using infection history, population size, geographic area, and number of municipalities as covariates	DiD, two models: • TWFE (municipality and week), with effect size varying by week • Two-by-two, averaging across weeks
Scale	Linear	• Linear (TWFE) • Logarithmic (two-by-two)
Identification Assumption	Weighted average of control series is the expected outcome absent vaccination campaign	Additive (TWFE) or multiplicative (two-by-two) change in expected outcome pre- to post-campaign is the same across municipalities
Results	Vaccine effectiveness: 675.3 avoided infections per 100,000 children in Schwaz (95% CI: 146.9–1,232.6)	Vaccine effectiveness: • Significant decrease after second dose in campaign (TWFE) • 64.5% (95% CI: 30.2, 82.0) (two-by-two)
Negative Controls/ Placebo Tests	In-space placebo test	Pre-exposure trends
Additional Details	All analyses reported results also for adults aged 16–50 (targeted by vaccination program and thus expected to exhibit a larger effect)	

Abbreviations: CI, confidence interval; COVID-19, coronavirus disease 2019; DiD, difference-in-differences; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SCM, synthetic control method; TWFE, two-way fixed effects

FIGURES

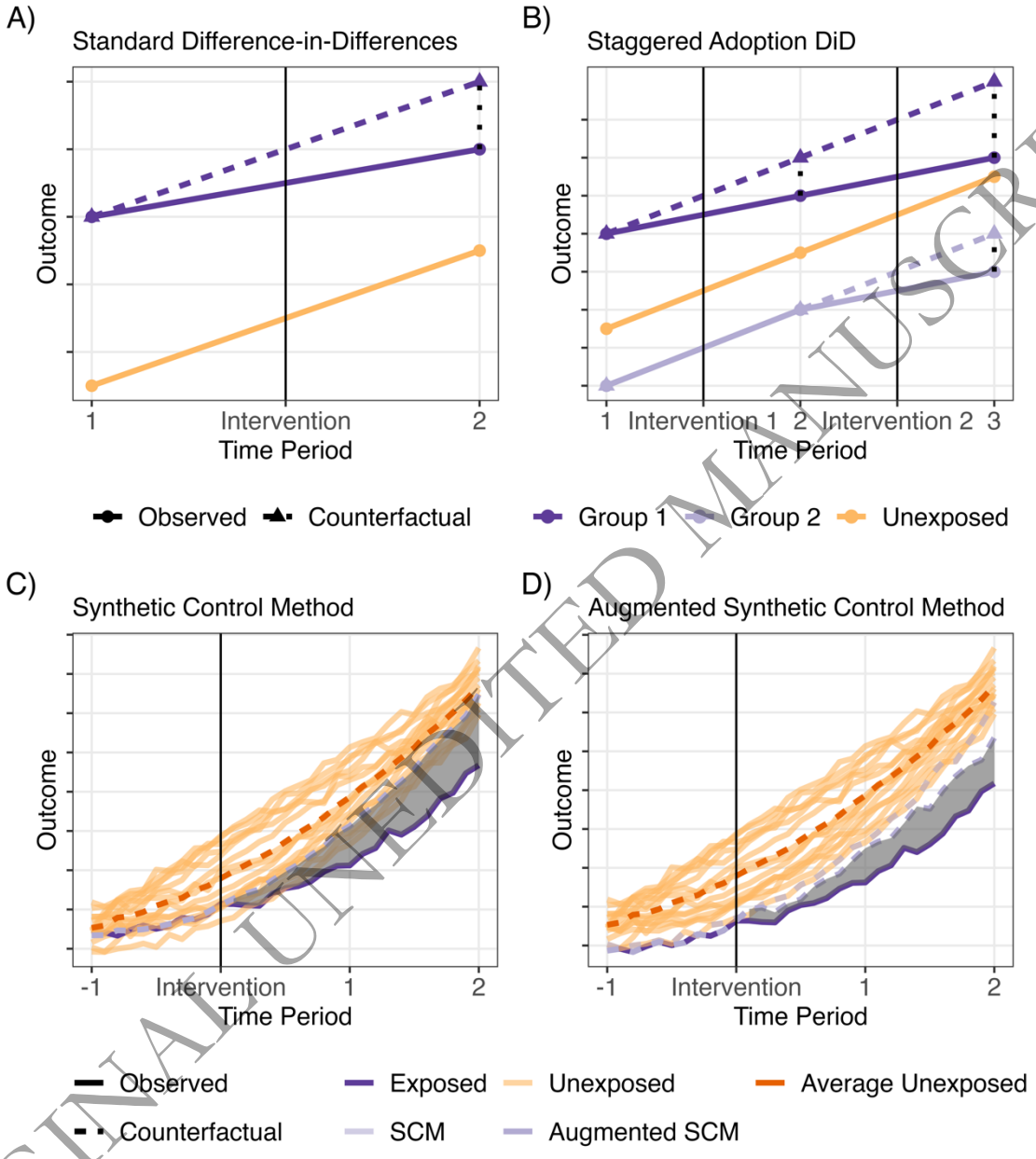

Figure 1. Schematics showing simulated observed and counterfactual data for four quasi-experimental methods: DiD (A), DiD with staggered adoption (B), SCM (C), and augmented SCM (D). Vertical lines represent the beginning of exposure in exposed units/groups. Black dotted lines (A,B) and gray shaded regions (C,D) represent estimated effects. Note that different counterfactuals are possible with additional assumptions (B) and that augmented SCM is shown here when the exposed unit lies outside of the convex hull of the unexposed units (D). Abbreviations: DiD, difference-in-differences; SCM, synthetic control method