Can synthetic controls improve causal inference in interrupted time series evaluations of public health interventions?

Michelle Degli Esposti,[1]* Thees Spreckelsen,[1] Antonio Gasparrini,[2,3] Douglas J Wiebe,[4] Carl Bonander,[5] Alexa R Yakubovich,[6] David K Humphreys[1]

[1]Department of Social Policy and Intervention, University of Oxford, Oxford, UK; [2]Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London, UK; [3]Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, London, UK; [4]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA; [5]Health Economics and Policy, School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Gothenburg, Sweden; and [6]Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Canada.

*Corresponding author. Department of Social Policy and Intervention, University of Oxford, 32 Wellington Square, Oxford OX1 2ER, UK. E-mail: michelle.degliesposti@spi.ox.ac.uk

Word count: 3858 words (for paper plus abstract but not including tables or references)

Abstract

Interrupted time series designs are a valuable quasi-experimental approach for evaluating public health interventions. Interrupted time series extends a single group pre-post comparison by using multiple timepoints to control for underlying trends. But history bias – confounding by unexpected events occurring at the same time of the intervention – threatens the validity of this design and limits causal inference. Synthetic control methodology, a popular data-driven technique for deriving a control series from a pool of unexposed populations, is increasingly recommended. In this paper, we evaluate if and when synthetic controls can strengthen an interrupted time series design. First, we summarise the main observational study designs used in evaluative research, highlighting their respective uses, strengths, biases, and design extensions for addressing these biases. Second, we outline when the use of synthetic controls can strengthen interrupted time series studies and when their combined use may be problematic. Third, we provide recommendations for using synthetic controls in interrupted time series and, using a real-world example, we illustrate the potential pitfalls of using a data-driven approach to identify a suitable control series. Finally, we emphasize the importance of theoretical approaches for informing study design and argue that synthetic control methods are not always well-suited for generating a counterfactual that minimises critical threats to interrupted time series studies. Advances in synthetic control methods bring new opportunities to conduct rigorous research in evaluating public health interventions. However, incorporating synthetic controls in interrupted time series studies may not always nullify important threats to validity nor improve causal inference.

Key words: Interrupted time series, synthetic controls, quasi-experimental, causal inference, history bias

Key messages

- Interrupted time series design is a valuable quasi-experimental design for evaluating public health interventions.
- Synthetic control methodology is a data-driven technique for deriving a control series from a pool of unexposed populations.
- Incorporating synthetic controls in interrupted time series studies is increasingly recommended as a useful tool for improving causal inference.
- Theoretical considerations of plausible time-varying confounding are recommended in order to determine whether synthetic controls are able to minimise critical threats to validity in interrupted time series evaluations.

Introduction

Interrupted time series (ITS) is an increasingly popular quasi-experimental design for evaluating public health interventions when randomisation is not possible. ITS extends a single group pre-post (or before-after) comparison by using regularly collected data to model the underlying trend in the outcome, thereby controlling for gradual and cyclical patterns over time (e.g., secular trends, seasonality).[1,2] Such designs are especially well-suited for evaluating the effects of interventions that target population-level outcomes over a clearly defined time period,[1–4] and have been used to evaluate a range of public health interventions including new vaccines, traffic speed zones, and changes to drug packaging.[5–8] They have also been used to retrospectively evaluate the health impacts of unplanned events, such as the financial crisis,[9] and the unintended consequences of other events, such as the release of Netflix's *13 Reasons Why*.[10]

In using a single group to evaluate the impact of an intervention, ITS studies are susceptible to bias. In particular, a simple ITS is susceptible to time-varying confounding since it is unable to disentangle true effects of the intervention from some other simultaneously occurring event or co-intervention. This bias is known as *history bias* and threatens the validity of causal inference in ITS.[11] However, such threats to validity can be mitigated by adding additional design features, such as a control series, that can strengthen and extend a basic ITS to a controlled ITS (CITS) design.[12] A CITS design includes both a within- and between-group component, which can account for confounding from other co-occurring events by adding an external control series that has been exposed to such events but not to the intervention.

Synthetic control methodology, popularised by Abadie and colleagues, is a data-driven technique that derives a comparison unit from a weighted average of eligible comparison units (the "donor pool") that minimizes the differences between preintervention trends in the treated and synthetic control series.[13–15] The method is well-suited for deriving a counterfactual for between-group studies, such as case-control and controlled before-after studies where contemporaneous time series data are available for treated and untreated units, especially when there is no obvious single comparator.[16] The strengths and limitations of synthetic controls have been discussed in full elsewhere,[1,3,15–18] and has been recently described as "the most important innovation in the evaluation literature in the last fifteen years".[19] Synthetic controls are increasingly recommended in evaluations of public health interventions – including ITS studies – with the promise of improving causal inference.[13,14,16,20,21]

This paper aims to critically examine if and when synthetic controls can strengthen an ITS evaluation. Specifically, we use an ITS lens to critique the data-driven approach of using synthetic controls to identify a control series in a CITS study design. We use schematic diagrams and a worked example to describe the main threats to the validity of ITS studies and demonstrate when and how synthetic controls can be used in CITS and when their use may be problematic. Datasets and R code are provided in the supplementary material so that readers may reproduce the worked example.

Evaluative study designs

Study designs used to evaluate public health interventions in observational research are defined in Box 1. A counterfactual is defined as what would have happened had the intervention not taken place and, in the context of these evaluative study designs, a counterfactual represents the main comparator used to estimate the effect of the intervention. The different evaluative study designs fall under two main approaches for identifying a counterfactual: between-group and within-group designs. Both approaches, their biases, extensions, and applications are described in Table 1.

Box 1. Definitions for evaluative study designs in observational research

Case-control study*

Compares the outcome(s) of interest in two groups, one of which received the intervention ('case') versus another that did not receive the intervention ('control')

Before-after study

Compares the outcome(s) of interest in the same group before and after it receives an intervention

Controlled before-after study

Combines the two study designs above by comparing the before-after change in the outcome(s) of interest between the case and control group, typically using difference-in-difference estimation

Interrupted time series study

Extends the before-after design by comparing the outcome(s) of interest at multiple time points before and after an intervention in the same group, thus allowing for control of the underlying trend

Controlled interrupted time series study

Combines the interrupted time series and the controlled before-after designs by comparing the before-after change in the outcome(s) of interest between the case and control series, while controlling for the underlying trend

Adapted from Cochrane definitions for non-randomised controlled study designs.[22] Note the definitions provided here relate to the study designs referenced in this paper and is not a comprehensive list of all evaluative study designs in observational research.

*We are using the term case-control to refer to study designs with aggregated units as the unit of analysis (e.g., geographical areas which do and do not experience an intervention), rather than individuals as the unit of analyses, where cases refer to individuals with a disease of interest and controls to individuals who do not have the disease.

Table 1. Different approaches for identifying a counterfactual

Counterfactual model	**Within-group (+ control)** *Time period before the population received the intervention*	**Between-group (+ synthetic control)** *External population that did not receive the intervention*
Reason for use	Pre-intervention period is the most suitable counterfactual as there is no available external comparator that does not suffer from time-invariant confounding	Comparison group is the most suitable counterfactual as there is an appropriate external comparator and time-varying confounding poses the greatest threat to internal validity
Simple study design	Before-after	Case-control*
Main strength	Control of time-invariant factors	Control of time-varying factors
Main threats to internal validity	Maturation & history bias (time-varying confounding)	Selection bias† (time-invariant confounding)
Extended study design	Interrupted time series	Controlled before-after
Causal identification	Predicted counterfactual; modelling and forecasting using multiple timepoints (ie time series) for the treated population before and after the intervention	External counterfactual; difference-in-difference estimation using data for the treated and control population before and after the intervention
Main threats to internal validity	History	Non-parallel trends
Additional extensions to minimise bias	Controlled interrupted time series	Controlled before-after with synthetic control
Causal identification	Predicted + external counterfactual; difference-in-difference estimation using modelling and forecasting for treated and control series before and after the intervention	Matched + external counterfactual; difference-in-difference estimation using multiple timepoints (ie time series) for the treated and a matched control population before and after the intervention
Applied examples in public health evaluations	Ejlerskov *et al.*[22] Evaluating the effect of UK supermarket policies for reducing less-healthy food at checkouts	Abadie *et al.*[13] Evaluating the effect of California's Tobacco Control Program

*We are using the term case-control to refer to study designs with aggregated units as the unit of analysis (e.g., geographical areas which do and do not experience an intervention), rather than individuals as the unit of analyses, where cases refer to individuals with a disease of interest and controls to individuals who do not have the disease.
†We are defining selection bias as systematic differences between the treated and control groups that confound the true effects of the intervention with group differences.

Using interrupted time series to evaluate public health interventions

A simple ITS design uses a within-group approach to identify a counterfactual. ITS advances a basic before-after design by using multiple timepoints to model underlying trends and account for time-varying confounding from secular trends and seasonality (Figure 1a). These underlying trends are then used to predict the expected time series and serve as the counterfactual (methodology described in full elsewhere).[1–3,9] The main reason for using an ITS design is when the pre-intervention period is the most suitable counterfactual, often because the treated group is unique and so there is no suitable comparison group (Table 1). A common example is when a country implements a country-wide public health intervention, such as Australia enacting new cycle helmet legislation.[23] In such examples, it is difficult to identify a suitable comparison group as it is unlikely that any other country shares the same legislative context, number of bicycles, road structure, rates of road accidents and cycling-related injuries, or population size and density as Australia. As a result, Australia before the new cycle helmet legislation serves as the more suitable counterfactual (as compared to another country), especially when controlling for underlying trends.

Addressing history bias to improve causal inference in interrupted time series evaluations

However, interrupted time series is susceptible to bias. In particular, time-varying confounding from unexpected events or co-interventions that happen at the same time of the intervention (i.e., *history bias*).[11,12] These unexpected events are independent of underlying trends and cannot be predicted or forecasted, such as an economic recession or an environmental disaster. Thus, history bias poses the greatest threat to the validity of ITS studies and undermines causal inference. Figure 1b illustrates history bias and shows that, if another event simultaneously occurred, a simple ITS design is unable to estimate the intervention's effect on the outcome.

History bias can be potentially mitigated by adding additional design features to strengthen a basic ITS design and improve causal inference. The most common addition is selecting a well-chosen control series that *has not* been exposed to the intervention but has been exposed to the confounding other event, producing a CITS (Box 1 & Table 1).[12,24–26] This is schematically illustrated in Figure 1c, where the control series has been exposed to the confounding event but not the intervention. The benefit of including a control series depends on its ability to rule out history bias and, as a result, researchers are required to case-select an appropriate control. Specifically, an appropriate control should be selected based on having experienced the relevant time-varying confounders and being affected by these confounders to the same extent as the intervention series while not having experienced the intervention itself. If an evaluation of a control series, which experienced the confounding event, shows no impact on the outcome, then this additional design element serves as a useful sensitivity analysis or robustness check to the analysis of the intervention series and/or can be formally incorporated into the one model using difference-in-difference estimation. This accounts for history bias (at least by the known other event) and thus helps to strengthen causal inference.[12]

Using synthetic controls to improve causal inference in between-group evaluations

Synthetic control methodology is a data-driven technique that generates a control series by calculating a weighted average from a pool of potential controls (the so-called *donor pool*). The weighting algorithms use prediction errors, such as mean squared prediction errors (MSPE), to minimise differences between the treated and control series' pre-intervention trends.[13–15] The underlying assumption is that the donor pool can be used to accurately estimate the counterfactual based on pre-intervention trends in the outcome and (possibly) other covariates. The method is designed to improve causal inference by reducing selection bias when using a between-group approach to estimate the counterfactual.[14] Selection bias describes systematic differences in the treated and control groups that may confound effect estimates, and is pervasive in the between-group approach for identifying a counterfactual.[11] While adding a before-after element may reduce selection bias – extending a case-control design to a controlled before-after design – the benefits of this design extension rely on the parallel trends assumption (i.e., in the absence of the intervention, differences in the slopes of the treated and control series are constant over time). This assumption is the main threat to the validity of a controlled before-after design. Synthetic control methods are thus instrumental for improving causal inference for the between-group approach for identifying a counterfactual.[16,19]

Using synthetic controls to improve causal inference in interrupted time series evaluations

The use of synthetic controls is increasingly recommended for ITS.[13,14,16,20,21] As outlined above, a control series only strengthens an ITS design when it is able to address history bias. An appropriate control series should experience the time-varying confounder(s) but not the intervention (Figure 1b & 1c). In contrast, synthetic control methodology shifts the selection of a control series away from a theoretically and contextually informed case-selection approach to a data-driven method. As a result, synthetic control methods may derive a control series that has a parallel trend to the intervention series prior to implementation (as illustrated in Figure 1d). However, there is no guarantee that the controls in the donor pool have experienced the all-important confounding from other events. This key issue of interpolation bias, irrespective of synthetic control fit, is also acknowledged by Abadie *et al.*[13] (p. 495). In this instance, a synthetic control series would fail to improve causal inference in a CITS design as the vertical blue line representing other confounding events in Figure 1d would not be reflected in the synthetic control series.

Problems and benefits

The benefits of using synthetic controls in ITS depends on the nature of the history bias and the distribution of the confounding event among the donor pool. The donor pool is a group of potential controls that have not experienced the intervention, which are then weighted according to the matching algorithm.[28] If the majority of the potential controls in the donor pool have not experienced the confounding event, then it is likely that the derived synthetic control is unable to rule out this confounding and minimise history bias. On the other hand, if all the potential controls in the donor pool are hypothesized to have experienced the confounding event, then the use of synthetic controls in ITS may be beneficial.

An example using a real-world evaluation of an intervention is instructive here. Humphreys *et al.*[29] evaluated the impact of Florida's "stand your ground" (SYG) law on homicide.[29] The law, enacted October 1, 2005, extends the right to use lethal force in self-defense in public places when threat is perceived. The authors used an ITS design to examine whether extending the right to use self-defence escalated aggressive encounters and increased rates of homicide in Florida and checked for history bias by also analysing homicide rates in a selected subset of comparison states that did not have SYG laws. Recent guidance instead recommends using synthetic control approaches to identify a control series from a donor pool of all comparison states that do not have SYG laws.[20,21,31]

However, the appropriateness of this recommendation depends on the other events occurring at the same time that Florida's SYG law was enacted. The other events could range from state-specific legislative changes (e.g., enacting concealed carry laws) to country-wide events (e.g., an economic recession). If state-specific events pose the most likely threat to an ITS study's validity, then only the subset of states that have enacted the concealed carry law (i.e., the other confounding event) but have not enacted SYG laws (i.e., the intervention) should be selected (see Figure 2a). Under these conditions, a synthetic control method would be inappropriate because it would instead involve deriving a control series based on *all* states without SYG laws, without considering exposure to the confounding event that may bias the intervention effect estimate. This is because the matching algorithm constructs the synthetic control weights by minimising differences in pre-intervention trends.[13–15] Hence, the weighting procedure is unable to capture important time confounding events happening at the same time as the intervention, if such events cannot be predicted from pre-intervention trends alone (i.e., unexpected or abrupt events, also known as "shocks"). In this scenario the use of synthetic controls instead of case-selection could be problematic since it uses a CITS design but fails to address history bias.

If, however, country-wide events pose the most likely threat to an ITS study's validity, then the use of synthetic controls may strengthen an ITS design. This is because all comparison states in the donor pool have experienced the other confounding event. For instance, in the example of an SYG evaluation, a country-wide confounding event could have been the economic recession. If this is the primary hypothesised confounder, all donor states would have been equivalent in their ability to address history bias (assuming the recession affected all states equally; Figure 2b). Under these conditions, the benefits of using synthetic control methods hold – it reduces selection bias by minimising systematic differences between the treated and control series, satisfies the parallel trends assumption, and removes researcher bias in the selection of the control series.[16] It is important to emphasise that in an ITS design history bias is the most critical threat to validity and so these benefits are only of value if history bias is already accounted for.[1,26]

Recommendations for using synthetic controls in interrupted time series

Because synthetic control methods are not designed to address the critical threat to validity for ITS, we would argue that the benefits of incorporating them within ITS designs may be limited to certain situations. First, researchers should identify plausible confounding events or co-interventions that occurred at the same time as the intervention and affect the outcome of interest.[24] The identification of such confounding events should be guided by a researcher's knowledge of the wider literature and temporal context of the intervention. This includes an informed understanding of the theoretical mechanisms by which the intervention is proposed to impact the outcome of interest and how other coinciding events or co-interventions may also impact the same outcome. Insights from qualitative research and previous empirical studies, as well as graphical models such as causal directed acyclic graphs (DAGs), may facilitate the identification of key confounding events.[32–34] When all control units are hypothesized to have experienced these confounding events (Figure 2b), then synthetic control methods can be used to disentangle bias from other confounding events without restricting the donor pool. Alternatively, synthetic control methods can be used when the donor pool can be feasibly restricted so that all control units entered into the donor pool are hypothesized to have experienced the confounding events while still having a sufficient number of donors for synthetic control weighting. This strategy for restricting the donor pool combines theoretically informed case-selection with data-driven techniques to identify the most appropriate control series for mitigating bias in ITS and improving causal inference.

If concerns remain around whether the donor pool (restricted or otherwise) is able to capture important time-varying confounders that threaten the internal validity of the ITS design, then researchers should opt to carry out additional robustness checks and reconsider the appropriateness of synthetic control methodology to derive the control series for a CITS study.

The use of placebo tests would help to evaluate the robustness of the synthetic control itself.[15] “In-time” placebo tests artificially reassign the intervention to occur during the pre-intervention period, while “in-space” placebo tests artificially reassign the intervention to control units in the donor pool (full methodological details described elsewhere).[15,35] Although placebo tests are valuable extensions to generally assess a researcher’s confidence in the intervention effects estimated from the synthetic control, they are less able to speak to the specific concerns raised here about the appropriateness of synthetic controls in mitigating against time-varying confounding from abrupt events (or “shocks”) that occur at the same time as the intervention. In order to address this specific concern of history bias within an ITS design, we recommend selecting the most appropriate control series based on theoretical considerations.[12] For example, a carefully case-selected geographical area or negative control outcome, which are hypothesized to be affected by other confounding events but not the intervention, would be better equipped to assess whether the intervention effect estimate is impacted by history bias, as illustrated in Figure 1c.[9,36,37] Therefore, when there are concerns that the donor pool is unable to capture the important time-varying confounders, a case-selected control area or a negative control outcome instead of a synthetic control may be the preferable control series for a CITS study.

An illustrative example

Keeping with the same real-world evaluation of Florida's SYG law, the most plausible time-confounding events to affect homicide rates are US-wide (e.g., the 2008 economic recession).[29] We therefore extended the original study's CITS design from four comparison states (New York, New Jersey, Ohio, Virginia) to a synthetic control (donor pool: 15 comparison states). Monthly homicide rates from January 1999 to December 2014 were obtained at the state-level from the Centers for Disease Control and Prevention's (CDC) Wide-ranging Online Data for Epidemiologic Research (WONDER). The weights for the synthetic control were identified using underlying trends in homicide rates, as well as state-level characteristics (e.g., demographics; Supplementary Table 1). The donor pool consisted of 15 comparison states that did not pass SYG laws between January 1999 to December 2014, including by statue or case law (Supplementary Table 2).[38] Because Florida's homicide counts were outside the range of comparison states, we directly model homicides rates – which did fall within the range of comparisons states and thus within the convex hull – using a segmented log-linked Gaussian generalized linear model and exponentiated the coefficients to represent relative risk (RR). Synthetic control analyses were run in R (version 3.5.2) using the Synth package.[28] We also ran a sensitivity analysis specifying Newey-West standard errors to account for residual autocorrelation at lag 1.[20] This sensitivity analysis replicates our findings hence the results are not presented here but reproducible code is provided in the supplementary material. All data and code are available from the Open Science Framework (https://osf.io/dtyvq/?view_only=2b35d8f8e4af4b369246cfe4e218c001).

After adjusting for underlying trends, the original CITS found that the SYG law was associated with an increase in homicide rates in Florida (RR=1.25; 95%CI:1.18–1.32; p<.0001) but had no effect in the four comparison states (RR=1.05; 95%CI:0.98–1.13; p=0.19258)

(Figure 3a). The CITS estimated a 18.6% increase in homicide rates following the implementation of SYG relative to the comparison states (RR=1.19; 95%CI:1.08–1.30; p=0.00033). The synthetic control series showed good pre-intervention fit (Figure 3b). But, unlike the four comparison states, the synthetic control showed a small significant increase in homicide rates following the enactment of Florida's SYG law (RR=1.10; 95%CI:1.03–1.19; p=0.00714). Although this attenuated the estimated intervention effect from 18.6% to 12.9%, the increase remained significant (RR=1.13; 95%CI:1.03–1.24; p=0.01324).

These analyses identify an increase in monthly homicide rates in Florida following the SYG law. Because the confounding was from a US-wide social or economic event, the synthetic control series was an appropriate control. It was also more sensitive to identifying, thus mitigating, history bias than the original four comparison states used – strengthening causal inference in this ITS evaluation. However, if the hypothesized confounding was state-specific, localized to certain comparison states, then using a synthetic control series instead of case-selection could be inappropriate.

We demonstrate this limitation theoretically using a synthetic control in ITS by extending the worked example. First, we simulated an artificial time-confounding effect in October 2005 that increased homicides in only two comparison states: Maryland and Connecticut (cf. Figure 2a). Second, we performed the analyses on these artificial data for a CITS using these two comparison states (Maryland and Connecticut) and a CITS using a synthetic control (same donor pool: 15 comparison states). We found a 30.2% step increase in the control series (RR=1.30; 95%CI:1.18–1.44; p<.0001), which reflected the artificially simulated confounding event in the comparison states (Figure 4a). After removing this confounding effect through CITS estimation, there was no increase in homicides rates in Florida relative to comparison states (RR= 0.96; 95%CI:0.84–1.08; p=0.48535). In contrast, the synthetic control was insensitive to the confounding event, showing the same results with these artificial data to the original data. This meant that the synthetic control was unable to adequately remove the confounding effect and continued to estimate that Florida's SYG law resulted in a 12.9% increase in homicide rates (RR=1.13; 95%CI:1.03–1.24; p=0.01324). The reason that this example of a synthetic control is insensitive to the simulated confounding event is because the matching algorithm(s) in synthetic control methodology is unable to control for time-varying confounding that cannot be predicted from the preintervention period alone, such as the abrupt one-off event specified here. Consequently, the algorithm is blind to this type of confounding and the derived synthetic control is not designed to rule out this threat to internal validity.

Conclusion

There are many advantages to using synthetic control approaches. It helps to reduce selection bias, satisfies the parallel trends assumption, and removes researcher bias when selecting from a pool of eligible controls.[16] Such advantages speak directly to the main threats of validity for between-group approaches and can thus help to improve causal inference for these public health evaluations. However, the advantages of synthetic control methods are not necessarily aligned with the main critical threat to the validity of an ITS evaluation: history bias. We showed when and how synthetic control approaches may be beneficial and when and how they may be problematic for ITS. We thus express caution against viewing synthetic control methods as a panacea in CITS designs. Instead, we recommend that researchers base their selection of a control series on their main source of confounding that may bias their effect estimates in an ITS public health evaluation. If the donor pool can reliably reflect this confounding, then the use of synthetic controls can be useful. But if it cannot, then case-selecting an appropriate control area or using a negative control outcome may be the more reliable method for minimizing critical bias and improving causal inference. These observations underline the importance of prioritising a theoretical approach above a purely data-driven approach when aiming to strengthen an ITS study design.

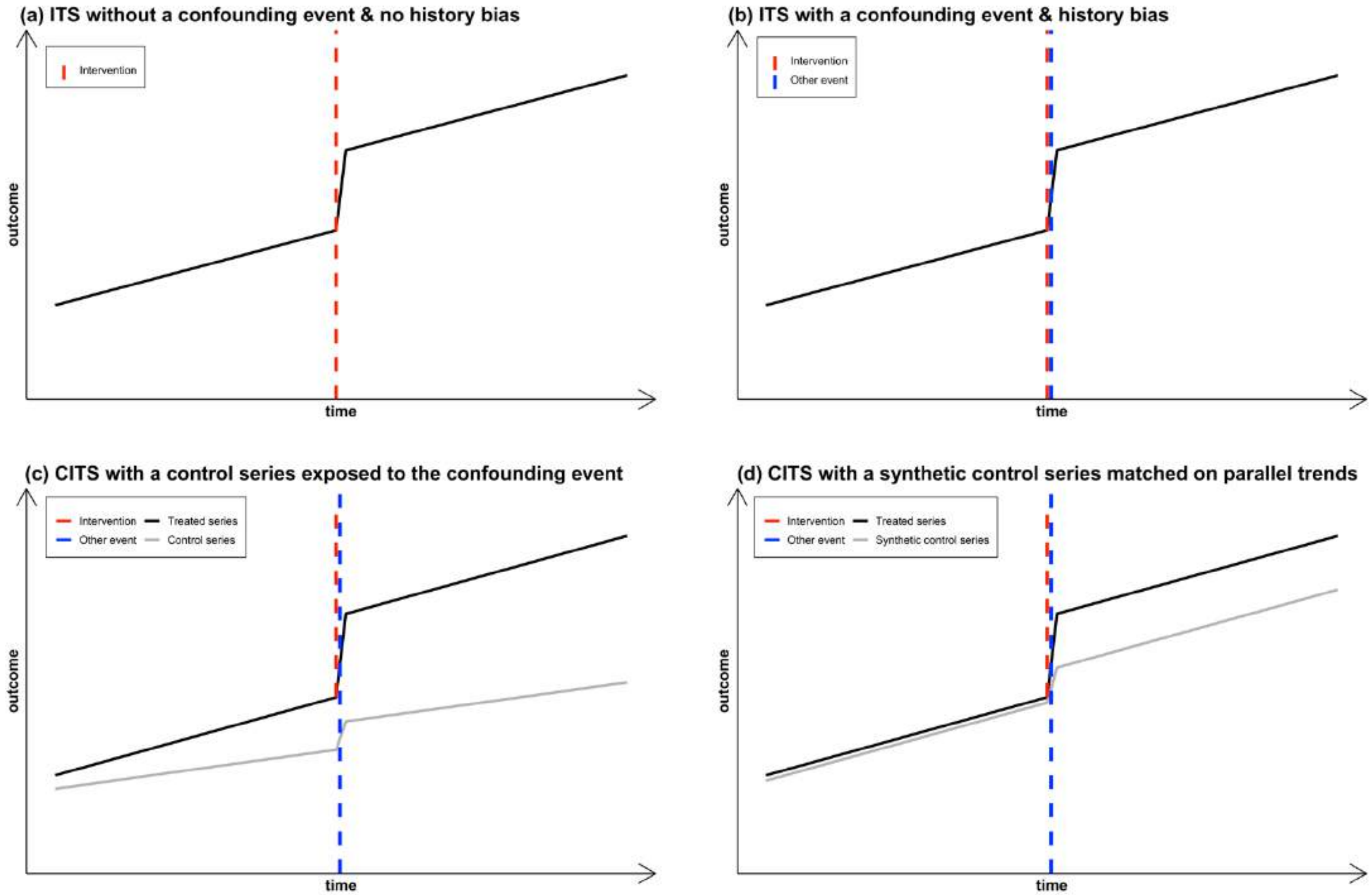

Figure 1. Schematic diagram showing history bias from a confounding other event in interrupted time series designs (simple and controlled). Adapted from Lopez-Bernal *et al.*[12]

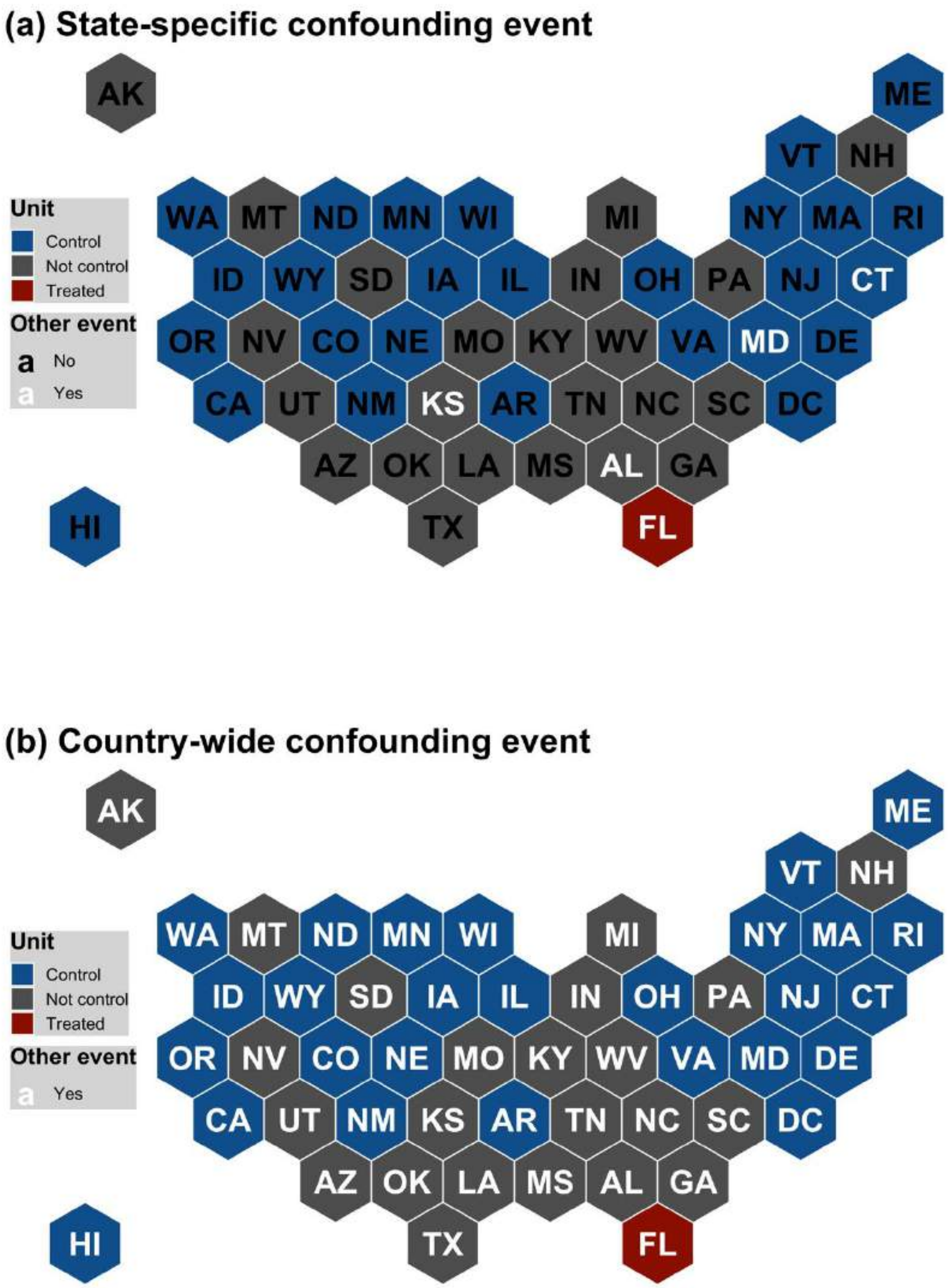

Figure 2. Illustrative example of when the use of synthetic controls in interrupted time series is problematic (state-specific confounding (a)) and beneficial (country-wide confounding (b)).

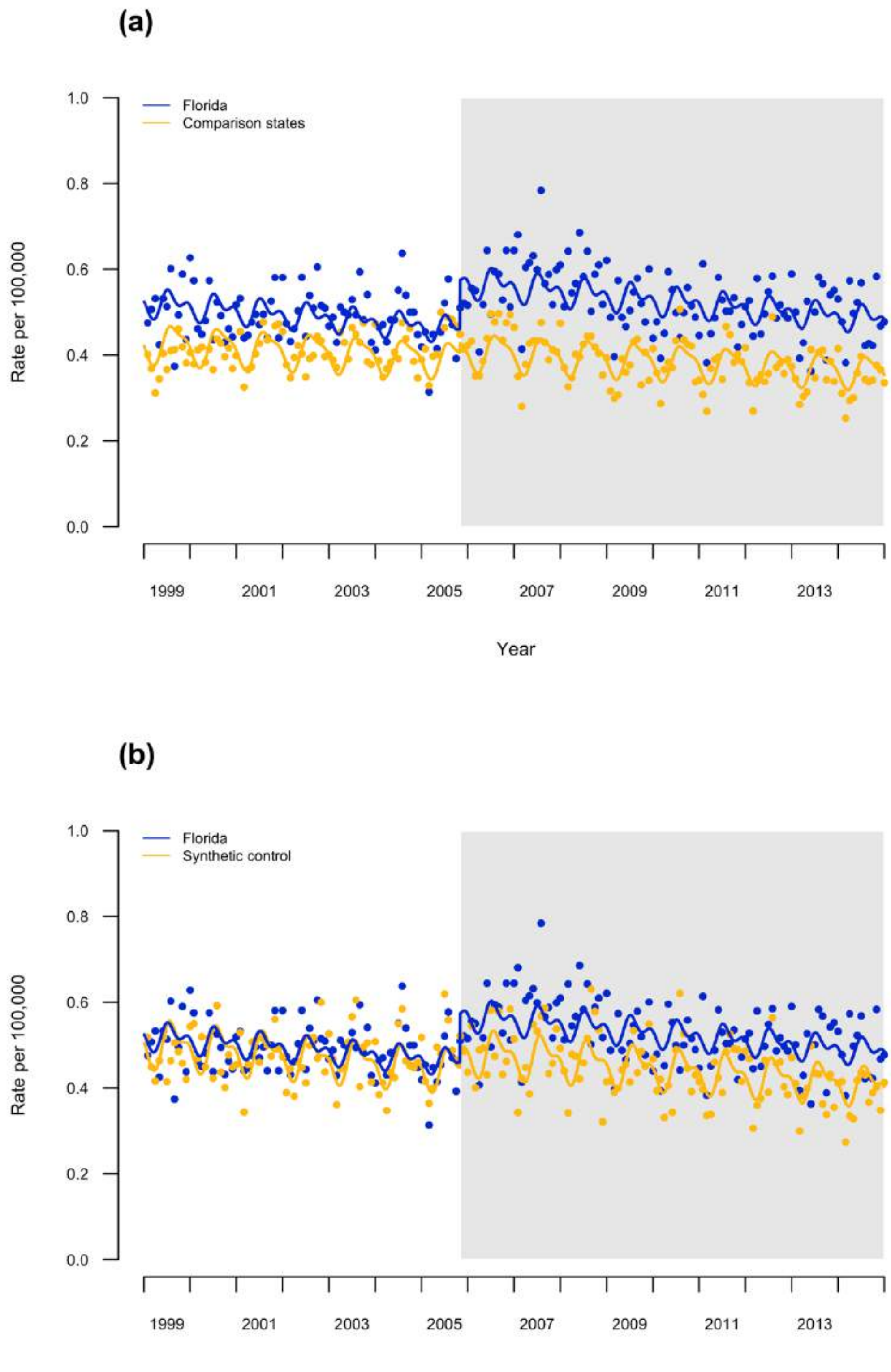

Figure 3. The impact of stand your ground (SYG) law on homicide rates in Florida, 1999-2014, using real data with a US-wide confounding event: (a) CITS with four comparison states (New York, New Jersey, Ohio, Virginia); (b) CITS with synthetic control (donor pool: 15 comparison states). Adapted from Humphreys *et al.* (2017).[30]

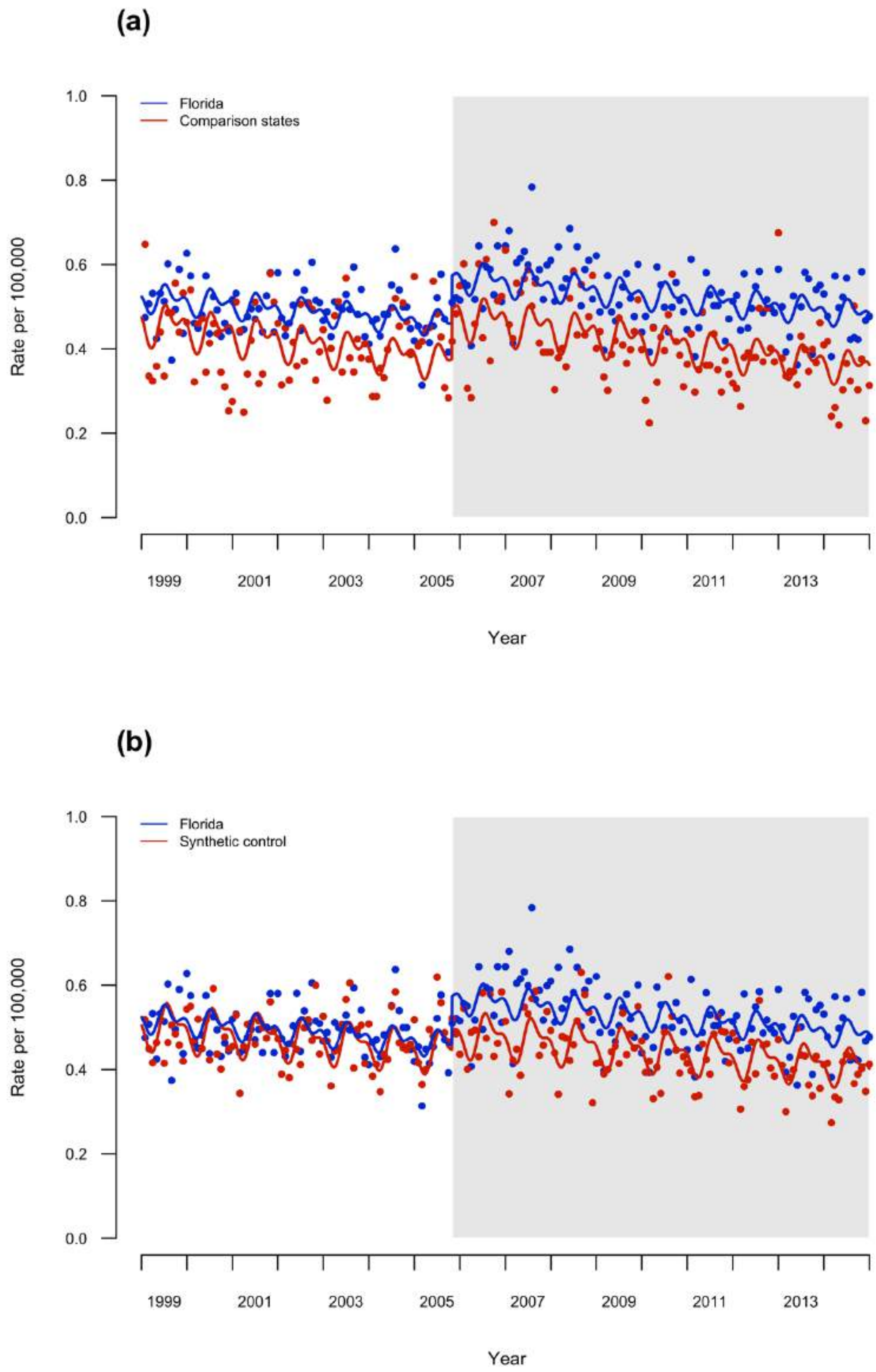

Figure 4. The impact of stand your ground (SYG) law on homicide rates in Florida, , 1999-2014, using artificial data with a state-specific confounding event: (a) CITS with two comparison states affected by the confounding event (Maryland, Connecticut); (b) CITS with synthetic control (donor pool: 15 comparison states). Adapted from Humphreys *et al.* (2017).[30]

Funding

This work was supported by The Joyce Foundation (US).

References

1. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348–55.

2. Wagner A, Soumerai S, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002;27(4):299–309.

3. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. Int J Epidemiol. 2013;42(4):1187–95.

4. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci. 2000;1(1):31–49.

5. Lau WCY, Murray M, El-Turki A, Saxena S, Ladhani S, Long P, et al. Impact of pneumococcal conjugate vaccines on childhood otitis media in the United Kingdom. Vaccine. 2015;33(39):5072–9.

6. Dennis J, Ramsay T, Turgeon AF, Zarychanski R. Helmet legislation and admissions to hospital for cycling related head injuries in Canadian provinces and territories: interrupted time series analysis. BMJ. 2013;346:f2674.

7. Hawton K, Bergen H, Simkin S, Dodd S, Pocock P, Bernal W, et al. Long term effect of reduced pack sizes of paracetamol on poisoning deaths and liver transplant activity in England and Wales: interrupted time series analyses. BMJ. 2013;346:f403.

8. Grundy C, Steinbach R, Edwards P, Green J, Armstrong B, Wilkinson P. Effect of 20 mph traffic speed zones on road injuries in London, 1986-2006: controlled interrupted time series analysis. BMJ. 2009;339:b4469.

9. Lopez Bernal JA, Gasparrini A, Artundo CM, McKee M. The effect of the late 2000s financial crisis on suicides in Spain: an interrupted time-series analysis. Eur J Public Health. 2013;23(5):732–6.

10. Bridge JA, Greenhouse JB, Ruch D, Stevens J, Ackerman J, Sheftall AH, et al. Association between the release of Netflix's 13 Reasons Why and suicide rates in the United States: an interrupted times series analysis. J Am Acad Child Adolesc Psychiatry. 2019;

11. Cook TD, Campbell DT, Shadish W. Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin Boston, MA; 2002.

12. Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. Int J Epidemiol. 2018;47(6):2082–93.

13. 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.

14. Abadie A, Gardeazabal J. The economic costs of conflict: a case study of the Basque Country. Am Econ Rev. 2003;93(1):113–32.

15. Abadie A, Diamond A, Hainmueller J. Comparative politics and the synthetic control method. Am J Pol Sci. 2015;59(2):495–510.

16. 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):673–8.

17. Powell D. Imperfect synthetic controls: did the Massachusetts Health Care Reform save lives? RAND Labor & Population; 2018 p. 44. Report No.: WR-1246.

18. Bonander C. Compared with what? Estimating the effects of injury prevention policies using the synthetic control method. Injury Prevention. 2018 Jun 1;24(Suppl 1):i60–6.

19. Athey S, Imbens GW. The state of applied econometrics: causality and policy evaluation. Econ Perspec. 2017;31(2):3–32.

20. Linden A. Combining synthetic controls and interrupted time series analysis to improve causal inference in program evaluation. J Eval Clin Pract. 2018;24(2):447–53.

21. Linden A. Challenges to validity in single-group interrupted time series analysis. J Eval Clin Pract. 2017;23(2):413–8.

22. Cochrane: Childhood Cancer. Non-randomised controlled study (NRS) designs [Internet]. Available from: https://childhoodcancer.cochrane.org/non-randomised-controlled-study-nrs-designs

23. Olivier J, Boufous S, Grzebieta R. The impact of bicycle helmet legislation on cycling fatalities in Australia. Int J Epidemiol. 2019 Aug 1;48(4):1197–203.

24. Lopez Bernal J, Cummins S, Gasparrini A. Difference in difference, controlled interrupted time series and synthetic controls. Int J Epidemiol. 2019;47(6):2082–93.

25. Linden A. Conducting interrupted time-series analysis for single- and multiple-group comparisons. The Stata Journal. 2015;15(2):480–500.

26. St.Clair T, Cook TD, Hallberg K. Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment. Am J Eval. 2014;35(3):311–27.

27. Campbell DT, Russo MJ. Social experimentation. Vol. 1. Sage Publications, Inc; 1999.

28. Abadie A, Diamond A, Hainmueller J. Synth: An R Package for synthetic control methods in comparative case studies. J Stat Softw. 2011;42(1):1–17.

29. Humphreys DK, Gasparrini A, Wiebe DJ. Evaluating the impact of Florida's "Stand Your Ground" self-defense law on homicide and suicide by firearm: an interrupted time series study. JAMA Intern Med. 2017 01;177(1):44–50.

30. Catalfamo C. Stand Your Ground: Florida's Castle Doctrine for the twenty-first century. Rutgers J L & Pub Pol'y. 2006 2007;4(3):504–45.

31. Benmarhnia T, Rudolph KE. A rose by any other name still needs to be identified (with plausible assumptions). Int J Epidemiol. 2019;1–2.

32. Steiner PM, Kim Y, Hall CE, Su D. Graphical models for quasi-experimental designs. Sociological Methods & Research. 2017 Mar;46(2):155–88.

33. Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. Respirology. 2014;19(3):303–11.

34. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Med Res Methodol. 2008 Oct 30;8(1):70.

35. McClelland R, Gault S. The synthetic control method as a tool to understand state policy. Washington, DC: Urban-Brookings Tax Policy Center. 2017;

36. Tchetgen Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. American journal of epidemiology. 2014;179(5):633–40.

37. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016 Dec 1;45(6):1866–86.

38. American Bar Association. National task force on Stand Your Ground laws: final report and recommendations [Internet]. 2015 p. 1–66. Available from: http://www.americanbar.org/content/dam/aba/images/diversity/SYG_Report_Book.pdf