Compared with what? Estimating the effects of injury prevention policies using the synthetic control method

Carl Bonander

Correspondence to
Dr Carl Bonander, Centre for Public Safety, Faculty of Health, Science and Technology, Karlstad University, Karlstad SE-651 88, Sweden; carl.bonander@kau.se

ABSTRACT

Introduction This paper discusses the application of the synthetic control method to injury-related interventions using aggregate data from public information systems. The method selects and determines the optimal control unit in the data by minimising the difference between the pre-intervention outcomes in one treated unit (eg, a state) and a weighted combination of potential control units.

Method I demonstrate the synthetic control method by an application to Florida's post-2010 policy and law enforcement initiatives aimed at bringing down opioid overdose deaths. Using opioid-related mortality data for a panel of 46 states observed from 1999 to 2015, the analysis suggests that a weighted combination of Maine (46.1%), Pennsylvania (34.4%), Nevada (5.4%), Washington (5.3%), West Virginia (4.3%) and Oklahoma (3.4%) best predicts the preintervention trajectory of opioid-related deaths in Florida between 1999 and 2009. Model specification and placebo tests, as well as an iterative leave-*k*-out sensitivity analysis are used as falsification tests.

Results The results indicate that the policies have decreased the incidence of opioid-related deaths in Florida by roughly 40% (or −6.19 deaths per 100.000 person-years) by 2015 compared with the evolution projected by the synthetic control unit. Sensitivity analyses yield an average estimate of −4.55 deaths per 100.000 person-years (2.5th percentile: −1.24, 97.5th percentile: −7.92). The estimated cumulative effect in terms of deaths prevented in the postperiod is 3705 (2.5th percentile: 1302, 97.5th percentile: 6412).

Discussion Recommendations for practice, future research and potential pitfalls, especially concerning low-count data, are discussed. Replication codes for Stata are provided.

INTRODUCTION

With increasing requirements on decision makers to apply and implement evidence-based policies comes a greater need for stronger evidence related to the population-level impact of interventions.[1] The focus of this supplement to *Injury Prevention* is to highlight effective strategies for achieving population-level impact. To enhance this discussion, this paper focuses on recent advances in quantitative evaluation methods stemming from a seminal paper by Abadie and Gardeazabal,[2] and later formalised by Abadie *et al*.[3] The quasiexperimental method, which is especially useful for estimating effects in case studies of unique interventions in small samples

panel data, is called the synthetic control method. Under ideal conditions, it outperforms both the interrupted time series and the difference-in-differences methods in terms of internal validity.[4] It is now widely considered one of the most credible quasiexperimental approaches for policy evaluation in political science and economics, especially when coupled with supplementary analyses in the form of sensitivity and placebo tests.[5] Yet, to my knowledge, its usage remains sparse in the fields of public health and injury prevention, with a few exceptions. For instance, Crifasi *et al*[6] employ the method to study the effects of changes in permit-to-purchase handgun laws in Connecticut and Missouri on suicide rates, and DeAngelo and Hansen[7] study the causal effects of police enforcement on traffic fatalities exploiting variations in roadway troopers caused by mass layoffs in Oregon. Sampaio[8] also uses the synthetic control method to estimate the effect of New York's ban on the use of handheld cell phones while driving. In the hopes of introducing the method to a broader audience, the purpose of this paper is to discuss its application to injury-related policy changes. To this end, I apply the method to study the effects of Florida's post-2010 regulations and enforcement of the prescription of opioids aimed at combatting the opioid epidemic in the USA. I also discuss current research and developments of the synthetic control method and note some potential limitations and pitfalls relating specifically to the type of data commonly found in injury surveillance datasets.

THE SYNTHETIC CONTROL METHOD

To motivate the use of the synthetic control method for quasiexperimental studies of injury control interventions, we can begin by considering its theoretical underpinnings using the potential outcomes framework.[9]

First, let us define the data requirements involved. To apply the method, we must have dataset containing repeated observations for N units (eg, regions, countries, groups or individuals) over T time periods (eg, months and years), commonly referred to as panel or time series cross-sectional data. Typically, the method is applied to aggregate data, which are readily available to injury epidemiologists through public health information systems (eg, CDC's Wide-ranging Online Data for Epidemiologic Research (WONDER) database, which contains panel data for states and counties in the USA). The dataset must be balanced with respect
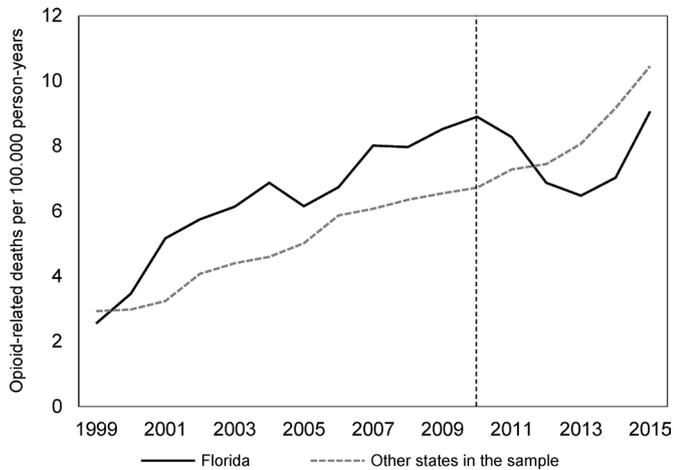
Figure 1 Trends in opioid-related deaths per 100.000 person-years in Florida and the unweighted donor pool (the vertical line indicates the start of the intervention period).

Figure 3 Dynamic effect estimates for the Florida post-2010 opioid overdose interventions on the incidence of opioid-related deaths per 100.000 person-years.

to the temporal dimension, which means that an outcome variable Y_{jt} is observed for all units j over the same time periods (see appended data for an example).

Let D_{jt} represent some intervention or natural experiment of interest coded as 1 for observations in the postintervention period in the treated unit and 0 otherwise. For a binary intervention, such as D_{jt}, we can define two potential outcomes for each unit j and time period . Using the synthetic control method, our goal is to estimate the causal effect of an intervention on a single intervention unit, $j = 1$. The effect is defined as $\alpha_{1t} = Y^t_{1t} - Y^c_{1t}$, where Y^t_{1t} is the observed potential outcome in the postintervention period Y^c_{1t} and is a an unobservable counterfactual state in which the unit was not exposed to the intervention. To estimate the effect, α_{1t}, we must therefore attempt to quantify Y^c_{1t} using a set of concurrent control units from a so-called donor pool of untreated units $j = 2, ..., j + 1$. The question then becomes: how do we select our control units in such a way that they are most likely to represent Y^c_{1t}? And what if there is no such control in the data, or what if we are uncertain that we have found the optimal control unit?

Abadie and Gardeazabal[2] provide a potential answer to these questions that exploit preintervention data and matching methods to avoid many of the pitfalls of classic quasiexperimental
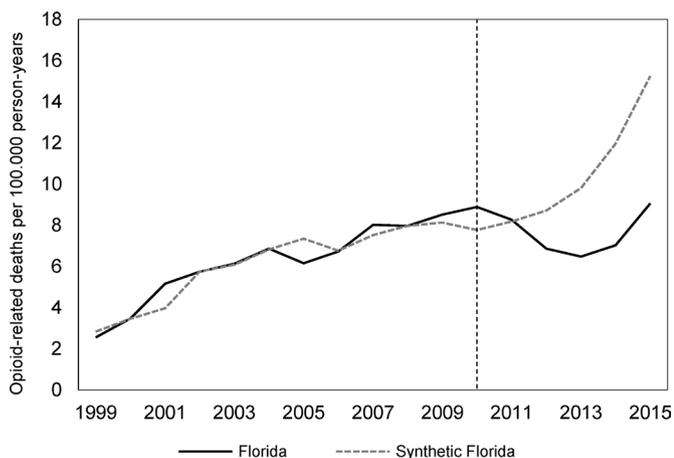
evaluation methods. The synthetic control method, which Abadie *et al* later chose to call it,[3] is an extension of the difference-in-differences method (or controlled before–after study) that enables the analyst to automatically select and generate a control unit from a donor pool of potential controls. This is achieved through convex optimisation and weighting such that the synthetic control unit closely resembles the intervention unit in the preintervention period on the outcome variable and a set of user-specified covariates, by simultaneously minimising the difference in preintervention covariate composition and root mean squared prediction error (RMSPE) in the same period.

As such, the method is more likely to produce internally valid estimates as compared with a manual selection of controls (which is sensitive to user specification error or biased selection of control units). This is because it optimises the control selection by directly addressing the causal assumptions behind the difference-in-differences approach: the common trends assumption (ie, that both the treated and control groups must follow parallel paths on the outcome over time). Thus, the selection is mainly data driven and tailored towards causal inference theory. A major strength of the method is also that the control unit is a weighted average of all potential controls in the data. By also considering all possible weighted combinations of control units, we do not limit the search for the best control group to the states as they appear originally in the data (e.g. by simple omission or inclusion of certain states as control units), which should increase the chance of finding a valid counterfactual.

Specifically, with the goal of minimising the preintervention RMSPE, the synthetic control method obtains a set of weights W_j for each $j = 2, ..., j + 1$ units in the donor pool, stored in a vector $W^* = (W_2, ..., W_{J+1})$ that is constrained to be non-negative and sum to 1, the effect estimator

$$\alpha_{1t} = Y_{1t} - \sum_{J=2}^{J+1} W^*_j Y_{jt}$$

yields an unbiased estimate of the effect at time t under the assumption that the synthetic control sufficiently captures all unobserved time-varying confounding factors throughout the study period using the preintervention period as training data. As Abadie *et al* show in their mathematical proofs,[3] the method will capture these factors if the synthetic control unit is accurately able to predict the preintervention evolution of the outcome in

Figure 2 Trends in opioid-related deaths per 100.000 person-years in Florida and the synthetic control unit (the vertical line indicates the start of the intervention period).
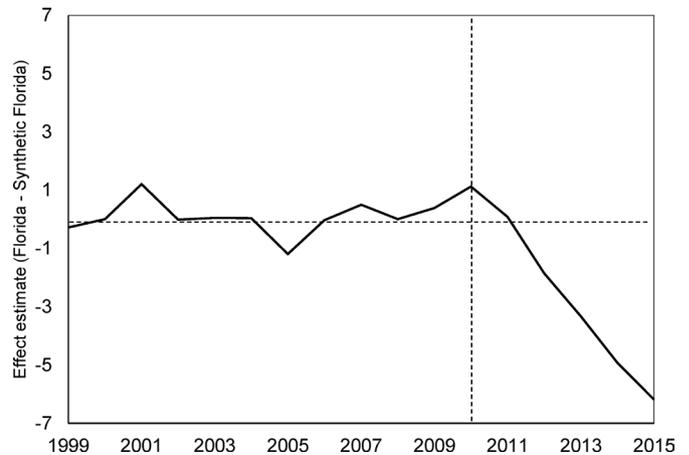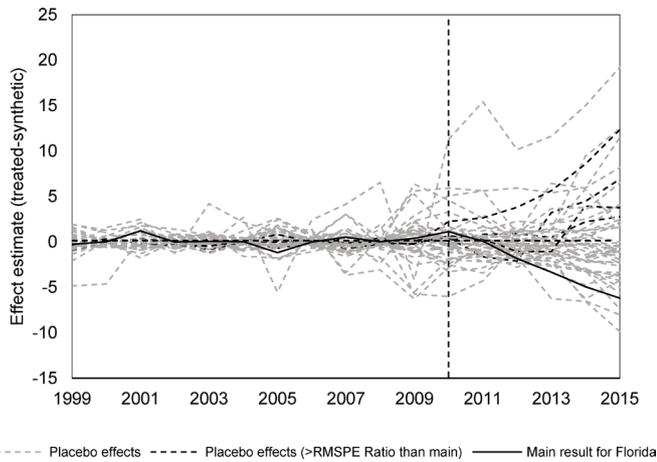
Figure 4 Dynamic effects for Florida compared to placebo studies on the 45 other states in the donor pool.

$$RMSPE = \left(\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} W_j^* Y_{jt} \right)^2 \right)^{\frac{1}{2}}$$

where T_0 is the total number of time periods in the preintervention period (note that the RMSPE can be analogously defined for the postintervention period).[10] We will use this measure to perform permutation tests below.

APPLICATION

To demonstrate the process, let us apply the method to a real case using publicly available injury data (the dataset and replication codes for Stata are available in the online appendix).

The case

Since 2010, Florida has adopted a multifaceted approach to combat the opioid epidemic. The state has regulated pain clinics, stopped healthcare providers from dispensing prescription opioids from the offices and established a prescription drug monitoring programme. Using time series analysis without concurrent controls, Johnson et al[11] find evidence that these interventions have reduced opioid-related deaths in Florida. Delcher et al[12] also arrive at similar conclusions using autoregressive integrated moving average (ARIMA) intervention models, and Kennedy-Hendricks et al[13] also find evidence of an effect using a difference-in-differences approach with North Carolina as a control state. The CDC also note on their website that this might be the first substantial reduction in drug overdose mortality in any state during the last decade.[14] However, to obtain a higher level of confidence in this result, although intuitive and already quite convincingly estimated in the studies above, we might consider estimating the effects using the synthetic control method. Using this approach, the best control is selected based on the data, and we do not need to make out-of-sample predictions based on preintervention trends (as in ARIMA intervention or interrupted time series analyses) to estimate the shape of the effect over time. Poisoning deaths are now the leading cause of injury death in the USA, and the problem continues to rise in most states.[15] It is therefore paramount that effective interventions are well documented and their effects estimated using credible estimation methods.

the treated unit, at least as the number of time points in the preperiod approach infinity. As a result, the risk of bias increases if the preintervention fit is poor or if the preperiod is too short (especially if the variance in the outcome variable is large). I will return to these points later in the discussion. For now, it is sufficient to note that we must also rely on the assumptions that there are no concurrent events specific to the intervention group in the postperiod that can explain the results and that there are no spillover effects into one or many of the groups in the donor pool, especially if these receive large weights. Thus, the method is not immune to bias, and it is up to the analyst to determine the sensitivity to such errors and, for example, remove invalid controls from the donor pool. Nonetheless, assuming the risk of bias from these violations is determined to be acceptable, we can proceed with the analysis.

As noted above, the optimal weights are estimated by minimising the RMSPE for the preintervention period, which measures the fit of the synthetic control unit. The RMSPE is given by

Data

I extract state-years data from the multiple cause of death records contained in CDC WONDER for all states and years from 1999 to 2015, using International Statistical Classification of Diseases, 10th revision (ICD-10), external cause codes X40–X44, X60–X64 and Y10–Y14, in combination with the opioid-related poisoning codes T40.0–T40.4 and T40.6 to identify opioid-related deaths of all intents, including deaths from prescribed opioids, opium and heroin. The rationale for using a composite measure such as this, even though the intervention targets prescription opioids, is that it will also capture potential substitution to illicit drugs. (Note, however, that similar results are similar even if illicit drugs are excluded (available on request).)

To illustrate the use of preintervention covariates in the matching procedure, I also collect the following economic variables: per capita real gross domestic product (GDP) (in dollars) and per capita personal income (in dollars), from the Bureau of Economic Analysis, as well as unemployment rate in per cent (of midyear population aged 16+ years) from the Bureau of Labor Statistics. In an ideal case, this covariate list could probably
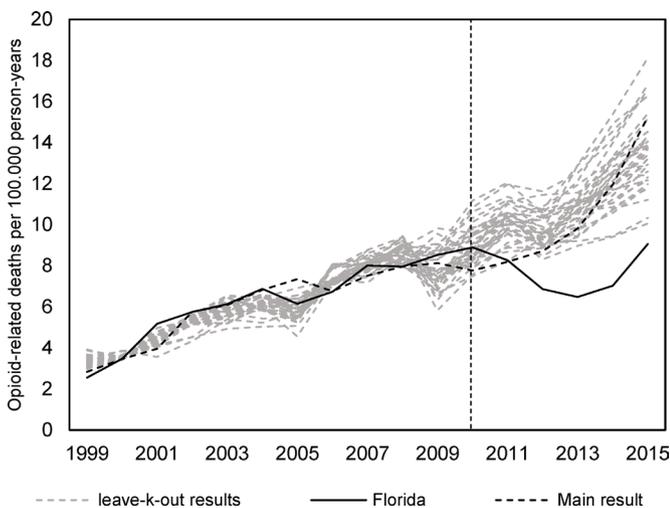

Figure 5 Results from a leave-k-out analysis that iteratively reduces the donor pool by excluding the most influential state from the synthetic control unit until the pre-intervention prediction errors are twice as large as in the main analysis (n = 35 iterations).

be expanded to include even more predictors of opioid-related deaths and other demographic variables. However, as is discussed below, covariates tend to play a relatively minor role in the analysis.

Note that as opposed to conventional matching methods (which do not place any specific importance on covariates in the matching procedure), the synthetic control method assigns variable weights to the included variables based on the predictive power of the covariates on the preintervention outcomes, meaning that poor predictors will automatically be considered less important in the matching process. An important feature that Abadie *et al* also suggest is to match on a linear combination of preintervention outcomes in order to capture changes in unobservable variables as well, rendering it less important to include covariates in the model. In fact, it is unlikely that we are ever aware of, or able to perfectly measure, all relevant predictors. The strongest predictor, which also requires the least amount of assumptions about the data generation process, is therefore the observed preintervention outcomes themselves.[16] In fact, Botusaru and Ferman[17] have shown that the inclusion of covariates is likely unnecessary in synthetic control analyses as long as a perfect match on preintervention outcomes can be obtained, and covariates are usually assigned small variable weights in favour of the preintervention outcomes themselves, which tend to receive the highest importance weights. However, an argument for including covariates is that it may result in a counterfactual that is structurally more similar to the treatment unit and as a method to reduce the risk of overfitting on random variability in large samples (see Discussion for details).[3]

Note that the *Synth* algorithm for Stata requires complete outcome data without gaps in the time series for all units in the data but can accommodate incomplete data on preintervention covariates by matching on, for example, averages or data from specific years (however, the covariate data are complete for all periods in this case). States with missing values on the outcome variable must be excluded if the missing values are not imputed, which would require further testing and adjustment for imputation errors, and is therefore not done here for the sake of parsimony. Due to the CDC's suppression policy of data points with less than 10 deaths,[18] this mainly affects states with low counts of opioid-related deaths in the current data. The states with gaps that are omitted from the analysis are Alaska, Nebraska, North Dakota, South Dakota and Wyoming.

Results

We begin by examining the raw data. Figure 1 compares the incidence of opioid-related deaths per 100.000 person-years in Florida to the remaining states in the donor pool (after removing states with gaps on the outcome). As we can see there, the overall sample of untreated states provides a poor counterfactual. We can clearly see an indication of an effect at the time of the intervention, but the information contained in figure 1 is not enough to quantify its shape or size.

Let us move to the creation of the synthetic control unit. While Abadie *et al* argue that the major strength of the method is that it circumvents researcher specification error in the choice of the optimal control unit,[3] we are still faced with the dilemma of specifying a vector of preintervention outcomes to match on. Ferman *et al* list a set of common choices that follow logical rules: (1) matching on the average of all preintervention outcomes, (2) matching on the entire range of preintervention outcomes (point-by-point), (3) matching on the first half, (4) matching on the first three-fourths, (5) matching on values from

Table 1 Unit weights for the synthetic control unit for Florida

State	Weight	State (cont.)	Weight (cont.)
Alabama	0	Montana	0
Alaska	–	Nebraska	–
Arizona	0	Nevada	0.054
Arkansas	0	New Hampshire	0
California	0	New Jersey	0
Colorado	0	New Mexico	0
Connecticut	0.011	New York	0
Delaware	0	North Carolina	0
District of Columbia	0	North Dakota	–
Florida	*	Ohio	0
Georgia	0	Oklahoma	0.034
Hawaii	0	Oregon	0
Idaho	0	Pennsylvania	0.344
Illinois	0	Rhode Island	0
Indiana	0	South Carolina	0
Iowa	0	South Dakota	–
Kansas	0	Tennessee	0
Kentucky	0	Texas	0
Louisiana	0	Utah	0
Maine	0.461	Vermont	0
Maryland	0	Virginia	0
Massachusetts	0	Washington	0.053
Michigan	0	West Virginia	0.043
Minnesota	0	Wisconsin	0
Mississippi	0	Wyoming	–
Missouri	0		

Notes: '–'=removed from donor pool due to missing data.
*Treated unit. The weights are restricted to be non-negative and must sum to one.

every odd year or (6) matching on even years.[19] Since this choice might affect the results and lead to cherry picking, they suggest running all these analyses and choosing the model that minimises the preintervention RMSPE. One problem with this procedure is that (2) will almost invariantly minimise the RMSPE since the matching takes place on all years, and Kaul *et al*[20] warn against doing this if other covariates are included in the model since their variable weights will then automatically be set to zero (see below for an explanation of these weights, which are different from the unit weights discussed above). Since a set of covariates are included in the analysis, I select the model that minimises the RMSPE of the other options, which in this case is (6).

The resulting 'synthetic Florida' is a weighted average of the outcomes in Maine (46.1%), Pennsylvania (34.4%), Nevada (5.4%), Washington (5.3%), West Virginia (4.3%) and Oklahoma (3.4%). The entire unit weight matrix is presented in table 1. Figure 2 compares the synthetic control unit to the observed outcomes in Florida. As we can see there, the model fits the data well in the preperiod and indicates an effect in the expected direction in the postperiod. Results for the other specifications are displayed in online appendix figure A1 , where we can see that they also show similar effect estimates (note that the main result happens to be the most conservative of the options).

Table 2 compares the actual Florida to the synthetic control, as well as the unweighted donor pool, to probe their resemblance in the preintervention period. This analysis also confirms what figure 1 indicates, that is, that the unweighted donor pool provides a poor counterfactual, at least in terms of preintervention incidence rates and per capita GDP. However, synthetic Florida provides a better predictor balance on all predictors

Table 2 Preintervention predictor balance check between Florida, synthetic Florida and the unweighted donor pool

	Florida	Synthetic Florida	Unweighted donor pool	V-weights
Incidence rate (100.000 population) in year:				
2000	3.47	3.46	2.98	0.156
2002	5.75	5.76	4.08	0.160
2004	6.87	6.83	4.59	0.189
2006	6.73	6.76	5.87	0.246
2008	7.97	7.97	6.35	0.192
Covariates (average 1999–2009)				
Per capita gross domestic product (in dollars)	40 898.64	41 471.36	46 981.3	0.048
Per capita personal income (in dollars)	34 324.45	33 367.03	34 042.4	0.005
Unemployment rate (in per cent)	5.08	5.07	5.25	0.004

except for per capita personal income. Note that this does not necessarily mean that synthetic Florida is a biased counterfactual. Recall that the optimisation routine assigns variable weights based on the predictive power of each covariate (that, similar to the unit weights, also sum to one) and assigns lower weights to poor predictors so that they are given less importance in the matching process. These weights are displayed in the V-weights column in table 2, where we can see that the vector of preintervention outcomes are given considerably higher weight than the covariates. Of the covariates, per capita GDP is assigned the highest weight (0.048), which indicates that this variable has some predictive power, while the other two are assigned much smaller weights (0.005 and 0.004).

The time-varying effect estimate, which is calculated by taking the difference between the outcomes in Florida and the synthetic control unit, is displayed in figure 3. There we can see that the effect appears to be delayed by 2 years and has been gradually increasing in magnitude over time (in the short run). However, more postintervention data are needed to determine the long-run shape and magnitude of the effect. In 2015, the estimated effect was −6.19 opioid-related deaths per 100.000 person-years, which corresponds to a relative effect of −40.60%. Using population size estimates of Florida from the US Census Bureau to derive the number of opioid-related deaths prevented from 2010 to 2015, the estimates in figure 3 suggest that the cumulative effect amounts to 3013 opioid overdose deaths prevented over the course of the postperiod.

Inference procedures

While attempts have been made to derive sampling-based inference statistics for the synthetic control method, these methods are still under development.[21] Arguing that sampling-based inference does not say anything about the validity of the results, Abadie et al[3] suggest instead using permutation tests in the form of placebo studies on untreated units. The logic behind these tests are that we can pretend that the intervention took place in another unit by running the same synthetic control procedure on other states to check if similar results manifest elsewhere as well, which would weaken the case for causality. It should be

noted that by random error, or by the occurrence of other events that we are unaware of, we are likely to find some effects in other states as well. However, if the estimate for Florida lies on the tail of the distribution of estimated effects, we obtain greater confidence in the results.

To avoid assigning an implausibly large weight to placebo studies with a poor fitting synthetic control unit, it is also important to account for the preintervention fit in these permutation tests, since we worry mainly about placebo studies that outperform the main analysis in terms of preintervention fit and effect size. To this end, Abadie et al suggest calculating the ratio between the postintervention and preintervention RMSPE for all placebo studies as well as the main analysis.[3] From this, we can derive a pseudo P value by calculating the probability of finding a post/pre-RMSPE ratio greater than or equal to the ratio in Florida. It is worth noting that the use of the term P value is slightly misleading since it is based on permutation tests and therefore not comparable with a real P value (eg, with a sample size of 10 states, the minimum possible 'P value' will be 1/10=0.10). We must therefore be willing to rely on more arbitrary decision rules when running these tests.

The inference procedures suggest a pseudo P value of 5/46=0.11, meaning that four placebo runs outperforms or equals the effect estimate for Florida when preintervention fit (RMSPE) is accounted for. These states are Connecticut, Indiana, Ohio and South Carolina, and the black dashed lines in figure 4 indicate their respective placebo results. The solid black line shows the result for Florida, and the grey dashed lines indicate placebo studies with poorer fit in the preperiod. As can be seen there, the placebo studies with good fit that outperformed Florida in terms of post/pre-RMSPE ratio all indicate effects in the opposite direction, which would yield a one-sided pseudo P value of 1/46=0.02. My assessment of this, in absence of a clear decision rule, is that the evidence for a causal effect of the Florida initiatives is strong.

Sensitivity analysis

As a sensitivity analysis to ensure that the results are not produced as a consequence of single influential state in the synthetic control unit, I perform a leave-k-out analysis in which highly influential states are iteratively removed from the donor pool. Specifically, this is performed iteratively so that each iteration reduces the donor pool by one (k = 1 in the first iteration, k = 2 in the second and so on) and refits the synthetic control model using the restricted donor pool. For the first iteration, this means that we drop Maine from the donor pool based on the unit weights in table 1. After refitting the model, we find that Hawaii receives the highest unit weight and drop Hawaii (along with Maine) for the next iteration. This process can be repeated until only one state remains, but since we mainly care about synthetic controls with low prediction errors, I perform the iterations until the preintervention RMSPE is more than twice that of the main analysis (which yields 36 different synthetic controls including the main result). The results from this exercise are presented in figure 5 , where we can see that the synthetic control results are robust to the exclusion of influential states, which lends more credibility to the analysis. Notice also that the delay in the effect that we observed in the main analysis is not present in all iterations, which means that we are less certain about the shape and size of the effect during the first 2 years. The mean, median, 2.5th and 97.5th percentiles of these iterations are presented in online appendix table A1. The average estimate in 2015 was −4.55 deaths per 100.000 person-years (2.5th percentile:

−1.24, 97.5th percentile: −7.92). The iterations suggest that the cumulative effect in terms of deaths prevented in the postperiod is 3705 (2.5th percentile: 1302, 97.5th percentile: 6412).

CONCLUDING DISCUSSION

I have demonstrated the potential value of the synthetic control method for measuring the population-level impact of injury prevention policies. It is especially useful for unique interventions and policy changes for which standard panel data approaches are inefficient. For the most part, it also provides a transparent and clear analysis that focuses on the causal assumptions of the difference-in-differences model, in ways that make it easy to assess the quality of the analysis by studying how well the synthetic control matches the pretreatment outcomes and covariates of the intervention unit. If the analysis is coupled with a transparent reporting of sensitivity analyses and placebo checks, we might stand to gain a greater level of confidence in estimates of the population-level impact of injury-related interventions compared with other quasiexperimental methods.

Yet, there are some caveats with the method that prospective analysts should be aware of. One is the lack of standardised decision rules for statistical inference (see, eg, Ferman and Pinto for a discussion of this issue).[21] Even though Abadie et al argue that the synthetic control method forces the analyst to think about the causal assumptions of the model rather than sampling error,[3] the optimisation routine still generates weights based on the observed outcomes in the preperiod. Hence, the bias of the synthetic control method will partly be a function of the random year-to-year variability of the data and is thus not free of this type of uncertainty. In a sense, the method based on the assumption that the random variability in large sample aggregate data is small to none,[3] which limits its usage to panel data with small amounts of white noise. This may hold true for stable macroeconomic time series (such as GDP per capita), for which the method was developed, but in aggregate epidemiological datasets containing injury or disease incidence rates, the variance is also a function of the amount of events that occur.[22] In practice, this means that it will likely be less useful for smaller regions, or uncommon types of injury events, since the low counts and high variance in such series will make it harder to find an optimal synthetic control unit due to matching on noise rather than the actual signal of the trend. In these cases, it may be better to use alternative methods.

An additional limitation that should be noted is that constraints are placed on the donor weights so that they are non-negative and must sum to one. According to Abadie et al,[3] this choice was made to minimise the risk of extrapolation bias and overfitting,[23] since allowing for any positive or negative weight to be associated with the donors will produce a perfect match in almost any case, but will likely result in poor out of sample predictions.[24] Still, there are machine learning regularisation methods (lasso, ridge or elastic net regressions) that can be used to avoid this type of bias without imposing an explicit restriction on the weights, as discussed and tested by Doudchenko and Imbens,[16] but these are not currently implemented in the *Synth* routines for Stata or R. In my experience, this constraint mainly becomes an issue if the intervention unit has the highest or lowest injury rate of all units before the intervention, meaning that there is no convex combination of donor units that can be interpolated to generate a synthetic control unit. If this is the case, the machine learning methods may be more appropriate if other quasiexperimental methods (eg, interrupted time series or difference-in-differences) appear infeasible.

Another limitation, which is more specific to my presentation of the method rather than its current state, is its limitation to a single intervention unit. This was at least the case with the original synthetic control method proposed by Abadie et al and used in their algorithms for Stata and R.[3] [10] However, Cavello et al and Kreif et al have examined extensions of the method to panel data with multiple intervention units and interventions with time-varying implementation dates (which can be useful to study, eg, the effects of US state laws with varying enactment years).[25] [26] Robbins et al[27] also consider extensions to multiple outcome measures in high-dimensional data settings, and Klößner and Pfieffer[28] develop multivariate synthetic control models that allow for time series predictors. Xu[29] also merges the synthetic control method with the interactive fixed effects framework in a recent paper, and Sills et al[30] discuss and implement a bootstrapping method to produce CIs for the counterfactual. As we can see, there are many developments not covered here, and prospective analysts will likely benefit from studying these recent advances as well.

What is already known on the subject

► Valid policy evaluation is important for effective injury prevention and control.
► Randomized trials can seldom be used to estimate population-level intervention effects.
► The synthetic control method is gaining traction in the quantitative social sciences, but remains underused in injury research.

What this study adds

► This paper serves as a basic introduction to the synthetic control method, complete with replication files and data.
► We apply the synthetic control method to study the impact of policy changes in Florida on opioid-related deaths.
► The method is discussed in the light of issues related to injury outcome data.

REFERENCES
1 Cartwright N, Hardie J. *Evidence-based policy: a practical guide todoing it better*. Oxford: Oxford University Press, 2012.
2 Abadie A, Gardeazabal J. The economic costs of conflict: a case study of the basque country. *Am Econ Rev* 2003;93:113–32.
3 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;105:493–505.
4 Gardeazabal J, Vega-Bayo A. An empirical comparison between the syntheticcontrol method and hsiao et al.s panel data approach to programevaluation. *J Appl Econ* 2017;32:983–1002.

5 Athey S, Imbens GW. The state of applied econometrics: causality and policy evaluation. *J Econ Perspect* 2017;31:3–32.

6 Crifasi CK, Meyers JS, Vernick JS, *et al*. Effects of changes in permit-to-purchase handgun laws in Connecticut and Missouri on suicide rates. *Prev Med* 2015;79:43–9.

7 DeAngelo G, Hansen B. Life and death in the fast lane: police enforcement and traffic fatalities. *Am Econ J Econ Policy* 2014;6:231–57.

8 Sampaio B. Identifying peer states for transportation policy analysis with an application to New York's handheld cell phone ban. *Transp Transp Sci* 2014;10:1–14.

9 Rubin DB. Causal inference using potential outcomes. *J Am Stat Assoc* 2005;100:322–31.

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

11 Johnson H, Paulozzi L, Porucznik C, *et al*. Decline in drug overdose deaths after state policy changes - Florida, 2010-2012. *MMWR Morb Mortal Wkly Rep* 2014;63:569–74.

12 Delcher C, Wagenaar AC, Goldberger BA, *et al*. Abrupt decline in oxycodone-caused mortality after implementation of Florida's prescription drug monitoring program. *Drug Alcohol Depend* 2015;150:63–8.

13 Kennedy-Hendricks A, Richey M, McGinty EE, *et al*. Opioid overdose deaths and Florida's crackdown on pill mills. *Am J Public Health* 2016;106:291–7.

14 Centers for Disease Control and Prevention. State successes | drug overdose | cdc injurycenter. https://www.cdc.gov/drugoverdose/policy/successes.html (accessed 22 Mar 2017).

15 Shiels MS, Chernyavskiy P, Anderson WF, *et al*. Trends in premature mortality in the USA by sex, race, and ethnicity from 1999 to 2014: an analysis of death certificate data. *Lancet* 2017;389:1043–54.

16 Doudchenko N, Imbens GW. 2016. Balancing, regression, difference-in-differences and synthetic control methods: a synthesis: National Bureau of Economic Research. NBER Report No.22791.

17 Botosaru I, Ferman B. On the role of covariates in the synthetic controlmethod working paper. 2017 https://mpra.ub.uni-muenchen.de/80796/ (accessed 30 Aug 2017).

18 Centers for Disease Control and Prevention. Frequently asked questions. https://wonder.cdc.gov/wonder/help/faq.html#Privacy (accessed 30 Aug 2017).

19 Ferman B, de P, Possebom VA. Cherry picking with synthetic controls EESP Working Paper, 2016. http://bibliotecadigital.fgv.br/dspace/handle/10438/16583 (accessed 22 Mar 2017).

20 Kaul A, Klößner S, Pfeifer G, *et al*. Synthetic control methods: never use all pre-intervention outcomes as economic predictors.Working paper. 2016 http://www.oekonometrie.uni-saarland.de/papers/SCM_Predictors.pdf (accessed 31 May 2017).

21 Ferman B, Pinto C. Revisiting the synthetic control estimator FGV/EESP Working Paper. 421, 2016. No. 421 2016. https://mpra.ub.uni-muenchen.de/75128/

22 Boyle P, Parkin DM. Cancer registration: principles and methods. Statistical methods for registries. *IARC Sci Publ* 1991;95:126–58.

23 King G, Zeng L. The dangers of extreme counterfactuals. *Polit Anal* 2006;14:131–59.

24 Hastie T, Tibshirani R. Generalized additive models. *Statistical Science* 1986;1:297–310.

25 Cavallo E, Galiani S, Noy I, *et al*. Catastrophic natural disasters and economic growth. *Rev Econ Stat* 2013;95:1549–61.

26 Kreif N, Grieve R, Hangartner D, *et al*. Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Econ* 2016;25:1514–28.

27 Robbins MW, Saunders J, Kilmer B. A framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention. *J Am Stat Assoc* 2017;112:109–26.

28 Klößner S, Pfeifer G. Synthesizing cash for clunkers: stabilizing the car market, hurting the environment. German economic association: annual conference 2015 (muenster): economic development - theory and policy no. 113207, 2015. https://ideas.repec.org/p/zbw/vfsc15/113207.html

29 Xu Y. Generalized synthetic control method: causal inference with interactive fixed effects models. *Political Analysis* 2017;25:57–76.

30 Sills EO, Herrera D, Kirkpatrick AJ, *et al*. Estimating the Impacts of Local Policy Innovation: The Synthetic Control Method Applied to Tropical Deforestation. *PLoS One* 2015;10:e0132590.