

RESEARCH

Open Access



A comparison of four quasi-experimental methods: an analysis of the introduction of activity-based funding in Ireland

Gintare Valentelyte^{1,3*} , Conor Keegan²  and Jan Sorensen³ 

Abstract

Background Health services research often relies on quasi-experimental study designs in the estimation of treatment effects of a policy change or an intervention. The aim of this study is to compare some of the commonly used non-experimental methods in estimating intervention effects, and to highlight their relative strengths and weaknesses. We estimate the effects of Activity-Based Funding, a hospital financing reform of Irish public hospitals, introduced in 2016.

Methods We estimate and compare four analytical methods: Interrupted time series analysis, Difference-in-Differences, Propensity Score Matching Difference-in-Differences and the Synthetic Control method. Specifically, we focus on the comparison between the control-treatment methods and the non-control-treatment approach, interrupted time series analysis. Our empirical example evaluated the length of stay impact post hip replacement surgery, following the introduction of Activity-Based Funding in Ireland. We also contribute to the very limited research reporting the impacts of Activity-Based-Funding within the Irish context.

Results Interrupted time-series analysis produced statistically significant results different in interpretation, while the Difference-in-Differences, Propensity Score Matching Difference-in-Differences and Synthetic Control methods incorporating control groups, suggested no statistically significant intervention effect, on patient length of stay.

Conclusion Our analysis confirms that different analytical methods for estimating intervention effects provide different assessments of the intervention effects. It is crucial that researchers employ appropriate designs which incorporate a counterfactual framework. Such methods tend to be more robust and provide a stronger basis for evidence-based policy-making.

Keywords Interrupted time-series, Difference-in-differences, Propensity score matching, Synthetic control, Activity-based funding, Policy evaluation

*Correspondence:

Gintare Valentelyte
gintarevalentelyte@rcsi.com

¹Structured Population and Health services Research Education (SPHeRE) Programme, School of Population Health, RCSI University of Medicine and Health Sciences, Mercer Street Lower, Dublin, Ireland

²Economic and Social Research Institute (ESRI), Whitaker Square, Dublin, Ireland

³Healthcare Outcome Research Centre (HORC), School of Population Health, RCSI University of Medicine and Health Sciences, Dublin, Ireland



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

In health services research, quasi-experimental methods continue to be the main approaches used in the identification of impacts of policy interventions. These methods provide alternatives to randomised experiments e.g. Randomised Controlled Trials (RCTs), which are less prevalent in health policy research, particularly for larger scale interventions. Examples of previously conducted experiments include the RAND Health Insurance Experiment [1] and the Oregon Health Insurance Experiment [2] which have since led to the restructuring of health insurance plan policies across the United States. Although such large-scale experiments can generate robust evidence for informing health policy decisions, they are often too complex, expensive, unethical or infeasible to implement for larger scale policies and interventions [3, 4]. Quasi-experimental methods provide an alternative means to policy evaluation, using non-experimental data sources, where randomisation is infeasible or unethical when the intervention already occurred and its evaluation occurred later [3].

The evaluation of policy impacts, regardless of analytical approach, is aimed at identifying causal effects of a policy change. A concise guide highlights the approaches which are appropriate for evaluating the impact of health policies [3]. A recent review identified a number of methods appropriate for estimating intervention effects [5]. Additionally, several control-treatment approaches have recently been compared in terms of their relative performance [6, 7].

However, there is limited empirical evidence in the health services research field comparing control-treatment analytical approaches to non-control-treatment approaches, used for estimating health intervention or policy effects. We use an empirical example of Activity-Based Funding (ABF), a hospital financing intervention, to estimate the policy impact using four non-experimental methods: Interrupted Time-Series (ITS), Difference-in-Differences (DiD), Propensity Score Matching Difference-in-Differences (PSM DiD), and Synthetic Control (SC). A review of the application of these methods in the literature examining ABF impacts has recently been undertaken [5]. Out of 19 identified studies, six studies employed ITS, seven employed DiD and one study employed the SC approach [5]. The identified effects, as assessed by reporting on a set of hospital outcomes, varied based on the analytical method that was used. The studies which employed ITS all reported statistically significant effects post-ABF which have led to increased levels of hospital activity [8, 9], and reductions in patient length of stay (LOS) [10–13]. In contrast, the evidence is more mixed, among the remaining studies which employed control-treatment methods. For example, significant increases in hospital activity were

reported in three studies which used the DiD approach [14–16], while another study found no significant impacts in terms of activity [17]. Similarly, contrasting evidence in terms of changes in LOS [16, 18, 19] and mortality [18, 20] were also reported. Therefore, the overall evidence on the impacts of ABF on hospital outcomes can be considered mixed, and as highlighted by Palmer et al. (2014) [21] *'Inferences regarding the impact of ABF are limited both by inevitable study design constraints (randomized trials of ABF are unlikely to be feasible) and by avoidable weaknesses in methodology of many studies'* [21].

The aim of this study is to compare these analytical methods in their estimation of intervention effects, using an empirical case of ABF introduction in Ireland. Specifically, we focus on the comparison of control-treatment analytical approaches (DiD, PSM DiD, SC), to ITS, a commonly used non-control-treatment approach for evaluating policies and interventions. Additionally, we contribute to the very limited research evidence assessing the impacts of ABF within the Irish context.

ABF and the Irish health system

Activity-based funding (ABF) is a financing model that incentivises hospitals to deliver care more efficiently [22]. Under ABF, hospitals receive prospectively set payments based on the number and type of patients treated [22]. Services provided to patients are reflected by an efficient price of providing those services and adjustments incorporated for different patient populations served. Prices are determined prospectively e.g. in terms of Diagnosis Related Groups (DRGs), and reflect differences in hospital activity, based on types of diagnosis and procedures provided to patients [23]. DRGs provide transparent price differences, directly linking hospital services provision to hospital payments. In theory, this incentivises hospitals to deliver more efficient healthcare (e.g. shorten LOS) and to be more transparent in their allocation of resources and finances [22, 24].

The Irish healthcare system is predominantly a public health system, with the majority of health expenditure raised through general taxation (72%), and remainder through out-of-pocket payments (13%) and voluntary private health insurance (15%) [25]. In Ireland, most hospital care is delivered in public hospitals and this care is mostly government-financed, with approximately one-fifth of care delivered in public hospitals privately financed [25, 26]. Patients who receive private day or inpatient treatment in public hospitals are required to pay private accommodation and consultant charges. The majority of private patient activity in public hospitals is funded through private health insurance with the remainder through out-of-pocket payments. Public or private patient status relates to whether the hospital patient saw their consultant on a public or private basis

[27]. For non-consultant hospital staff, the same publicly funded staff are employed in delivering care to both publicly and privately financed patients [27].

Traditionally, all Irish public hospitals were funded on a budgetary block grant basis based on historical performance, making it difficult to measure and monitor activity and funding of public hospital care [28]. On the 1st January 2016, a major financing reform was introduced, and funding of public patients in most public hospitals moved to ABF [29]. ABF was introduced centrally by the Health Services Executive (HSE), responsible for delivery of public health services in Ireland. All public inpatient activity is funded under ABF, while all outpatient and Emergency Department (ED) activity continues to be funded using block budgets [30]. The ABF funding model is based on prospectively set average DRG prices, and additionally financially penalises hospitals for long patient LOS [30]. Additionally, the amount of activity that a hospital can carry out as well as the maximum funding it can receive, is capped, to preserve the overall health budget provided to a particular hospital [30]. Public patient reimbursement is based on the average price of DRGs, in contrast to private patients who are reimbursed at a per-diem basis [30].

Thus, this key difference in reimbursement between public and private patients treated in the same hospitals, lends itself to a naturally occurring control group for our analysis using the control-treatment approaches.

Methods

Estimation models

Interrupted time-series analysis

Interrupted Time Series (ITS) analysis identifies intervention effects by comparing the level and trend of outcomes pre and post intervention [31]. Often, ITS compares outcome changes for a single population and does not specify a control group against which intervention effects can be compared [32]. This can bias the estimated intervention effects, as a defined control group often eliminates any unmeasured group or time-invariant confounders from the intervention itself [33]. Therefore, ITS can overestimate the effects of an intervention producing misleading estimation results [4].

The ITS analysis model can be presented as [34, 35],

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t + \epsilon_t$$

Where Y_t is the outcome measured at time t , T is the time since the start of the study, X_t is a dummy variable representing the intervention (0=pre-intervention period, 1=post-intervention period), and TX is an interaction term; β_0 represents the intercept of the outcome (baseline level at $T=0$), β_1 is the change in outcome until

the introduction of the intervention (pre-intervention trend), β_2 is the change in the outcome following the intervention (the level change), β_3 represents the difference between pre-intervention and post-intervention slopes of the outcome (treatment effect over time).

Potential outcomes framework

Alternatively, analytical approaches such as Difference-in-Differences (DiD), Propensity Score Matching Difference-in-Differences (PSM DiD) and Synthetic Control (SC) overcome some of the shortcomings of ITS. These approaches are based on the counterfactual framework and the idea of potential outcomes which quantify the estimation of causal effects of a policy or an intervention¹. The potential outcomes framework defines a causal effect for an individual as the difference in outcomes that would have been observed for that individual with and without being exposed to an intervention [36, 37]. Since we can never observe both potential outcomes for any one individual (we cannot go back in time to expose them to the intervention), we cannot compute the individual treatment effect [36]. Researchers therefore focus on average causal effects across populations guided by this potential outcomes framework [3, 36, 37]. Therefore in practice, estimation is always related to the counterfactual outcome, which is represented by the control group [36, 38]². Consequently, it is for this reason all of these analytical approaches use a clearly defined control group in estimation, against which the outcomes for a group affected by the intervention are compared. The inclusion of a control group improves the robustness of the estimated intervention effects, by approximating experimental designs such as a RCT, the gold standard [38].

Difference-in-differences analysis

The DiD approach estimates causal effects by comparing the observed outcome changes pre intervention with the counterfactual outcomes post intervention, between a naturally occurring control group and a treatment group exposed to the intervention change [33]. The key advantage of the DiD approach is its use of the intervention itself as a naturally occurring experiment, allowing

¹ The treatment effect in terms of potential outcomes: where $Y^0(i, t)$ is the outcome that individual i would attain at time t in absence of treatment and $Y^1(i, t)$ is the outcome that individual i would attain at time t if exposed to treatment. The treatment effect on the outcome for individual i at time t is: $Y^1(i, t) - Y^0(i, t)$. The fundamental identification problem is that for any individual i and time t , both potential outcomes $Y^0(i, t)$ and $Y^1(i, t)$ are not observed and we cannot compute the individual treatment effect. We only observe the outcome $Y(i, t)$ expressed as: $Y(i, t) = Y^0(i, t)(1 - D(i, t)) + Y^1(i, t)D(i, t)$, [$D(i, t) = 0$ control and $D(i, t) = 1$ treatment]. Since treatment occurs after period $t=0$, we can denote $D(i) = D(i, 1)$, then we have $Y(i, 0) = Y^0(i, 0)$ and $Y(i, 1) = Y^0(i, 1)(1 - D(i)) + Y^1(i, 1)D(i)$ (Rubin (1974)).

² The change in outcomes from pre to post-intervention in the control group is a proxy for the counterfactual change in untreated potential outcomes in the treatment group.

to eliminate any exogenous effects from events occurring simultaneously to the intervention [33, 38].

The DiD approach estimates the average treatment effect on the treated (ATT) across individual units at a particular time point, represented by the general DiD model as [3, 6, 33, 38],

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 X_t + \beta_3 (D_i * X_t) + h_i + \lambda_t + \epsilon_{it}$$

Where Y_{it} is the value of the outcome observed for unit i at time t , D_i is an indicator of unit i being in a treatment group (vs. control group), X_t is a dummy variable representing the intervention period (0=pre-intervention period, 1=post-intervention period), and $D_i * X_t$ is the interaction term between the two; β_1 represents the estimated average difference in Y between the treatment and control groups, β_2 is the expected average change in Y from before to after the onset of the intervention, β_3 is the DiD estimator which captures the difference in outcomes before and after the intervention between the treatment and control groups i.e. the estimated average treatment effect on the treated (ATT), h_i is a vector of hospital fixed effects³ which capture unobserved time-invariant differences amongst hospitals (e.g. management), λ_t captures time fixed effects for each quarter t , and ϵ_{it} represents exogenous, unobserved idiosyncratic shocks.

However, DiD relies on the parallel trends assumption which states that, in the absence of treatment, the average outcomes for the treated and control groups would have followed parallel trends over time [33]. This parallel trends assumption can be represented as [33, 38],

$$\frac{E[Y^0(1) - Y^0(0) | D = 1]}{E[Y^0(1) - Y^0(0) | D = 0]}$$

Where $Y^0(0)$ is the outcome pre-intervention observed for all units in both the treatment ($D=1$) and control ($D=0$) groups; $Y^0(1)$ is the outcome post-intervention observed only for the control group and represents the unobserved counterfactual for units in the treatment group ($D=1$). This assumption cannot be statistically tested, as it applies to the unobserved counterfactual post-intervention [33, 38]. However, it is possible to examine the pre-treatment trends between both groups, by re-running the DiD model which includes an

interaction between time and the treatment dummy, in the pre-intervention period [39].

Propensity score matching difference-in-differences

PSM DiD is an extension to the standard DiD approach. Using this approach, outcomes between treatment and control groups are compared, after matching them with similar observable factors, followed by estimation by DiD [40–42]. Combining the PSM approach with DiD allows further elimination of any time-invariant differences between the treatment and control groups, and allows selection on observables and unobservables which are constant over time [40, 43]. Additionally, matching on the propensity score accounts for imbalances in the distribution of the covariates between the treatment and control groups [40]⁴. We present this model as follows [40],

$$Y = E(Y_{1i|D=1} - Y_{1i|D=0}) | P(x_{0i}) - E(Y_{0i|D=1} - Y_{0i|D=0}) | P(x_{0i})$$

Where Y_{1i} and Y_{0i} is the outcome in the post-intervention and pre-intervention period for individual patient episode i respectively, $D_i = 1$ indicates individual patient episode i is in the treatment group, $D_i = 0$ indicates individual patient episode i is in the control group, $P(x_{0i})$ represents the probability of treatment assignment conditional on observed characteristics in the pre-intervention period.

In our final PSM DiD estimation model we estimate the average treatment effect on the treated (ATT) using nearest neighbour matching propensity scores, by selecting the one comparison unit i.e. patient episode whose propensity score is nearest to the treated unit in question. We present our estimation model as follows:

$$Y^{(PSMDiD)} = \frac{1}{N_{D_1}} \sum_{i \in D_1 \cap S} \left[(Y_{i,t+1}^1 - Y_{i,t}^0) - \sum_{j \in D_0 \cap S} w_{ij} (Y_{j,t+1}^0 - Y_{j,t}^0) \right]$$

Where D_1 and D_0 represent the treatment and control groups respectively, w_{ij} the nearest neighbour matching weights, and S is the area of common covariate support⁵.

³ The unit used is at discharge level but we only have one observation per discharge by definition therefore we cannot apply discharge fixed effects and instead have to include hospital fixed effects.

⁴ Matching on the propensity score works because it imposes the same distribution of the covariates for both the control and treatment groups (Rosenbaum and Rubin (1983)).

⁵ The common support condition guarantees that only units with suitable control cases are considered by dropping treatment observations whose propensity score is higher than the maximum or less than the minimum propensity score of the controls.

Additionally, PSM makes the parallel trends assumption more plausible as the control groups are based on similar propensity scores in the PSM DiD approach. PSM forms statistical twin pairs before conducting DiD estimation, thus increasing the credibility of the identification of the treatment effect [40]. Instead, PSM relies on the conditional independence assumption (CIA). This assumption states that, in absence of the intervention, the expected outcomes for the treated and control groups would have been the same, conditional on their past outcomes and observed characteristics pre-intervention [40, 44]. However, it is also important to note, that even if covariate balance is achieved in PSM DiD, this does not necessarily mean that there will be balance across variables that were not used to build the propensity score [40, 44]. It is for this reason that the CIA assumption is still required.

Furthermore, recent developments of the DiD approach have highlighted that additional assumptions are necessary to ensure the estimated treatment effects are unbiased [45]. It is proposed that estimates will remain consistent after conditioning on a vector of pre-treatment covariates [45]. This was our motivation for employing the PSM DiD approach, as it accounts for pre-intervention characteristics, which allow to further minimise estimation bias. PSM DiD achieves this by properly applied propensity scores, based on matched pre-intervention characteristics, thus eliminating observations that are not similar between treatment and control groups [41]. Further developments have been made to account for multiple treatment groups, which receive treatment at various time periods i.e. differential timing DiD [46]. However, this does not affect our analysis, as the introduction of ABF in our empirical example took place at one time.

Synthetic control

The Synthetic Control (SC) method estimates the ATT by constructing a counterfactual treatment-free outcome for the treated unit using the weighted average of available control units pre-intervention [44, 47, 48]. The weights are chosen so that the outcomes and covariates for the treated unit and the synthetic control are similar in the pre-treatment period [44, 48]. This assumption may not hold in reality, particularly when estimating policy impacts, thus alternative analytical approaches which avoid the parallel trends assumption have been considered.

The SC approach becomes particularly useful in cases when a naturally occurring control group cannot be established, or in cases where the parallel trends assumption does not hold, and can often complement other analytical approaches [48]. Similarly to PSM, the SC method also relies on the CIA, and controls for pre-treatment

outcomes and covariates by re-weighting treated observations, using a semiparametric approach [44]. For a single treated unit the synthetic control is formed by finding the vector of weights W that minimises [44]:

$$(X_1 - X_0W)^\top V(X_1 - X_0W)$$

Where W represents the vector of weights that are positive and sum to 1, X_1 contains the pre-treatment outcomes and covariates for the treated unit, X_0 contains the pre-treatment outcomes and covariates for the control unit, and V is a positive matrix capturing the relative importance of the chosen variables as predictors of the outcome.

The choice of V is important as W^* depends on the choice of V . The synthetic control $W^*(V)$ is meant to reproduce the behaviour of the outcome variable for the treated unit in the absence of the treatment. Often a V that minimises the mean squared prediction error is chosen [44, 48]:

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

Where T_0 is the pre-intervention period, Y_{1t} is the outcome post-intervention at time t , Y_{jt} is the value of the covariates for unit j at time t , $W_j^*(V)$ is the synthetic control for unit j , W^* is a vector of optimally chosen weights.

Similarly, we limit biases in our estimated treatment effects [45] using the SC approach, which restricts the synthetic control weights to be positive and sum to one and such that the chosen weights minimise the mean squared prediction error with respect to the outcome [49].

Data and methods

In our empirical example analysis, we used national Hospital In-Patient Enquiry (HIPE) administrative activity data from 2013 to 2019 for 19 public acute hospitals providing orthopaedic services in Ireland. HIPE data used in our analysis record and classify all activity (public and private) in Irish public hospitals [27]. We divided our data into quarterly time periods ($n=27$) based on admission date. Data were available for 12 quarters pre-ABF introduction, and 15 quarters post-ABF introduction. We assessed the impact of ABF on patient average LOS, following elective hip replacement surgery, for a total of 19,565 hospital patient episodes.

For each analysis, we included hospital fixed effects and controlled for the same covariates: Age categories

(reference category 60–69 years), average number of diagnoses, average number of additional procedures (additional to hip replacement), Diagnosis-Related Group (DRG) complexity (split by minor and major complexity) and interaction variables: Age categories by average number of diagnoses, age categories by average number of additional procedures, age categories by DRG complexity.

We estimated the ITS model using ordinary least squares and included public patient episodes only. Following guidance from previous studies [32, 50], we accounted for seasonality by including indicator variables for elapsed time since ABF introduction. Additionally, we checked for presence of autocorrelation by plotting the residuals and the partial autocorrelation function [32, 50].

For the remaining models, we used treatment and control groups consisting of public and private patient episodes, respectively, and estimated the average treatment effects on the treated (ATT). We used the key differences in reimbursement between public (DRG payments) and private (per-diem payments) patient episodes, to differentiate our treatment group from the control group. The identification strategy exploits the fact that per-diem funding of private patient care remained unchanged over the study period. Any change in outcome between public and private patients before and after the introduction of ABF should be due to the policy introduction.

In our DiD analysis, we controlled for common aggregate shock changes by including dummy variables for each time period (time fixed effects). We additionally examined the parallel trends assumption by interacting the time and treatment indicators in the pre-ABF period (see Supplementary Tables 4, Additional File 6).

We estimated PSM DiD in a number of steps⁶: First we estimated propensity scores to treatment based on our list of covariates, using a probit regression. Second, we matched the observations in the treatment group (public patient episodes) with observations in the control group (private patient episodes) as per estimated propensity scores with the common support condition imposed. Finally, we compared the changes in the average LOS of the treated and matched controls by DiD estimation.

The SC estimation⁷ was conducted at the hospital level. It has been reported that the SC approach used in our analysis works best with aggregate-level data [44, 48, 52]. We incorporated the nested option in our estimation, a fully nested optimization procedure that searches among all (diagonal) positive semidefinite matrices and sets of

weights for the best fitting convex combination of the control units [44, 52]. The synthetic control group composition consisted of private patient episodes based on characteristics from 9 different public hospitals from the sample of 19 hospitals used in our analysis [see Supplementary Tables 1, Additional File 2].

To examine whether the estimated effects from all analyses still hold, we conducted sensitivity analysis and re-estimated each analytical model using trimmed LOS at 7 days (at the 90th percentile of the LOS distribution). As illustrated by the distribution of LOS in Supplementary Fig. 1, Additional File 1, this allowed for the exclusion of outlier LOS values. Additionally, to test the robustness of the estimated treatment effects, we tested the empirical strength of each model by inclusion and exclusion of certain covariates. We also examined the trends in the pre-ABF period across all DiD models, to check whether the trends were similar across the treatment and control groups.

Results

Table 1 summarises the key descriptive statistics of the data analysed. Over the study period, the overall average LOS for this sample of patient episodes was 5.2 days (5.3 and 5.0 days for public and private patients, respectively). The majority (31.7%) of patients were aged 60–69 years (30.9% of public and 33.8% private patients, respectively). The average number of additional diagnoses was 2.5 for public and 2.1 for private patients (overall average of 2.4), and average additional procedures were 3.3 for public and 2.8 for private patients. The DRG complexity indicates that most patients (95.7%) had undergone minor complexity hip replacement surgery.

We illustrate the estimated intervention effects for each of the models in Fig. 1. We observe a clear reduction in the average LOS from the ITS estimates (Fig. 1a). However, the DiD and PSM DiD estimates are very similar, and we do not observe a clear effect on the average LOS, with most coefficients distributed closely around zero (Fig. 1b and c). Similarly, the SC approach could not identify a clear effect (Fig. 1d). Additionally, both the SC (Fig. 1d & Supplementary Tables 1, Additional File 2) and PSM DiD (Supplementary Fig. 2, Additional File 3) approaches achieved good balance between the treated (public patient episodes) and control (private patient episodes) groups. Our examination of the pre-ABF trends did not identify any significant differences between treatment and control groups (see Supplementary Tables 4, Additional File 6).

Table 2 summarises the estimated treatment effects for each estimation model⁸. The ITS analysis suggested ABF

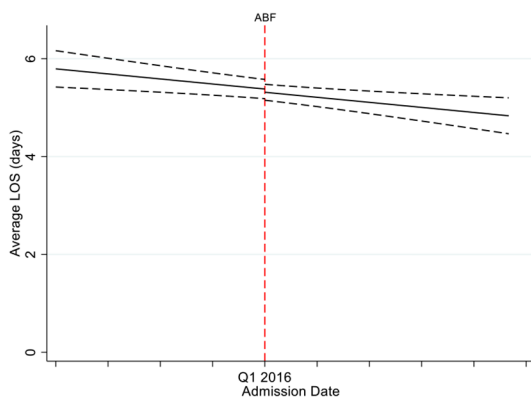
⁶ Using the *psmatch2* Stata command using nearest neighbour matching which showed the best balancing properties after comparing several algorithms [51].

⁷ Using the *synth* Stata command [44, 52].

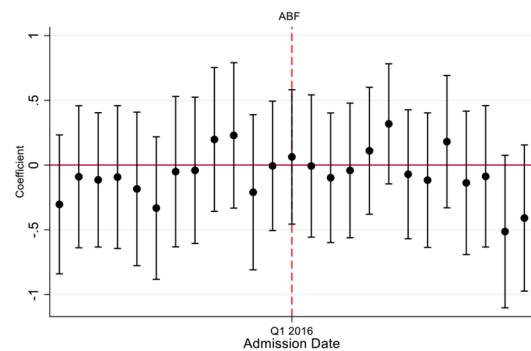
⁸ Reported p-values for ITS and DiD are for the hypothesis that ATT=0. For DiD PSM, reported p-values are conditional on the matched data. For SC,

Table 1 Descriptive Statistics of key covariates used in all models by treatment and control group

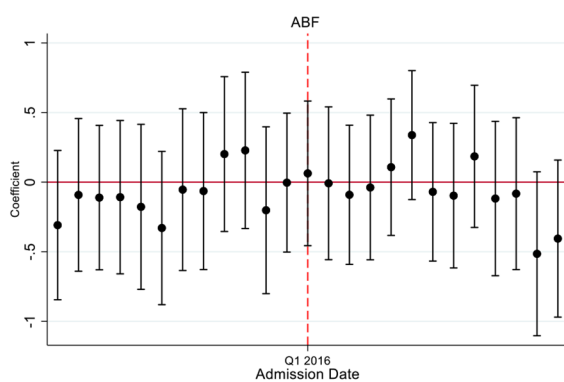
	Treatment (Public)	Control (Private)	Total
Average LOS	5.3 days	5.0 days	5.2 days
Age group (%)			
< 30	96 (0.7%)	18 (0.3%)	114 (0.6%)
30–39	329 (2.3%)	82 (1.5%)	411 (2.1%)
40–49	1026 (7.3%)	373 (6.8%)	1399 (7.2%)
50–59	2573 (18.3%)	982 (17.8%)	3555 (18.2%)
60–69	4338 (30.9%)	1868 (33.8%)	6206 (31.7%)
70–79	4151 (29.6%)	1633 (29.6%)	5784 (29.6%)
+ 80	1530 (10.9%)	566 (10.2%)	2096 (10.6%)
Average number of additional diagnoses	2.5	2.1	2.4
Average number of additional procedures	3.3	2.8	3.1
DRG complexity			
- major	13,389 (95.3%)	5322 (96.4%)	18,711 (95.7%)
- minor	654 (4.7%)	200 (3.6%)	854 (4.3%)
N	14,043	5522	19,565



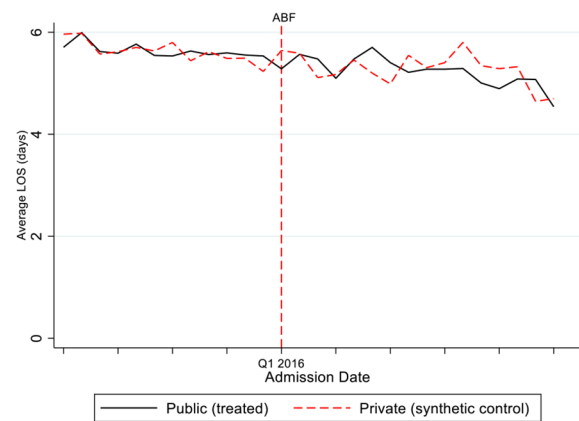
a. Interrupted Time Series analysis



b. Differences-in-Differences analysis



c. PSM Difference-in-Differences analysis



d. Synthetic Control analysis

Fig. 1 Model estimates

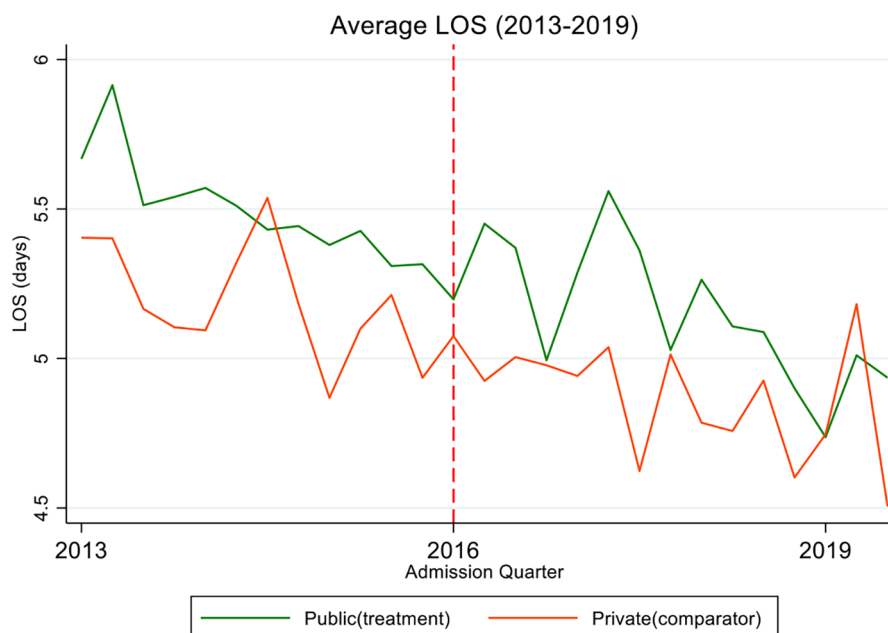
reported p-values were calculated using placebo-tests in a procedure akin to permutation tests (Abadie et al. 2010). This involved iteratively resampling from the control pool, and in each iteration re-assigning each control unit as a 'placebo treated unit', with a probability according to the proportion of

treated units in the original sample. The synthetic control method was then applied on these 'placebo data' and ATT calculated for the placebo treated versus control units. The p-value for the ATT was calculated according to the proportion of the replicates in which the absolute value of the placebo-

Table 2 Estimated Treatment Effects by estimation model

Estimation Model	Estimated Treatment Effect (Std. Error)	t	p-value	R ²	Observations	FE
ITS	-0.734 (0.162)	-4.54	0.001***	0.43	N = 14,043	Yes
DiD	-0.012 (0.280)	-0.04	0.966	0.42	N = 19,565	Yes
PSM DiD	-0.013 (0.279)	-0.05	0.961	0.42	N = 19,549 ^d	Yes
SC	-0.053 (0.136) ^b	-	0.400	-	N = 28 ^c	-

Note: The treatment effect parameters for ITS, DiD and PSM DiD all indicate a marginal change (reduction) in the average LOS. The DiD and PSM DiD estimates are almost identical, indicating that the matched propensity scores did not have a substantial impact on the overall DiD estimates. In contrast, the difference between ITS and DiD estimates is substantial. All models control for hospital Fixed Effects (FE) except for SC estimation at the hospital level; ^d16 observations in the treatment group were not matched; ^bThe SC method relies on minimising the RMSPE; ^c due to aggregated data at hospital level; Significance level: ***; p < 0.01; robust standard errors in parenthesis

**Fig. 2** Average LOS by quarter 2013–2019 for treatment and control groups

had the largest and statistically significant impact on the average LOS for public patients, a reduction of 0.7 days ($p < 0.01$). However, this effect could not be observed with the control-treatment approaches, although we also see a negative but smaller effect on the average LOS from the DiD, PSM DiD and SC estimates. The effect is not statistically significant for any of these models. As illustrated in Fig. 2 below, we observe a generally declining trend in the average LOS for both the public and private patients in our data. This explains the statistically significant effects of ITS, relative to the control-treatment methods,

which differentiate out the average LOS effects between both public and private patient episodes.

The results from our sensitivity analysis (Supplementary Tables 2, Additional File 4) revealed no material change for the ITS estimates, which remained statistically significant ($p < 0.001$). The estimated treatment effects from the control-treatment approaches remained small, and not statistically significant. Similarly, additional robustness testing of the estimated treatment effects by each model (and pre-ABF trend examination) remained consistent with the main results (Supplementary Tables 3, Additional File 5).

ATT exceeded the estimated ATT. It should be noted that the p-value based on placebo tests relate to falsification tests, while the p-values reported for the other methods relate to sampling uncertainty. Hence the p-values between each estimated model are not directly comparable.

Discussion

In this study we compared the key analytical methods that have been used in the evaluation of policy interventions and used the introduction of Activity-Based Funding (ABF) in Irish public hospitals as an illustrative policy case. Specifically, we compared several control-treatment methods (DiD, PSM DiD, SC), to a non-control-treatment approach, ITS. We contribute to the limited empirical evidence in the health services research field comparing control-treatment analytical approaches to non-control-treatment approaches, based on recent evidence highlighting the common use of these methods in estimation of health intervention or policy effects [5]. Additionally, we contribute to the very limited research evidence on the evaluation of the ABF policy within the Irish context. We were able to utilise an important dimension of the funding changes, by exploiting the fact that both publicly and privately financed patients are treated in public hospitals in Ireland and over the period of analysis, private patients were not subject to a change in their funding.

From our comparative methods analysis, ITS produced statistically significant estimates, indicating a reduction in LOS post ABF introduction, relative to control-treatment approaches, which did not indicate any significant effects. This is in line with the results from other studies, which have estimated ABF effects using ITS, and have reported significant reductions in LOS [10–13]. Caution should be taken when considering ITS, as the estimates may not truly capture the effects of the intervention of interest. This could lead to incorrect inferences, and potentially to misguided assessment of impacts from policy changes across the hospital sector. For instance, the estimated reduction in LOS for Irish public patients, may incorrectly indicate that the ABF reform has been successful. From a policy perspective, the importance of the resulting ABF effects, would be informed by the size of ITS estimates, providing potentially misleading evidence on the funding reform.

Further, caution should be taken, as ITS analysis does not include a control group, relative to the other methods we considered which incorporated a control and treatment groups. Therefore the conclusions drawn from the ITS analysis will differ to those drawn from the control-treatment approaches. Additionally, our findings from ITS analysis align with a recent study which tested the empirical strength of the ITS approach, by comparing the estimated ITS results to the results from a RCT [4]. Relative to a RCT, ITS produced misleading results, primarily driven by the lack of control group, and ITS model assumptions [4]. This would suggest, a comparison of the slope of outcomes before and after an intervention may lead to biased estimates when evaluating causal effects on outcomes affected over time, due to influences by

simultaneous and other unobservable factors at the time of the intervention.

However, over the study period, the average LOS for both public (treatment) and private (control) patient cases shows a reducing trend over time (Fig. 2). By limiting the analysis to the public patients only, the ITS approach ignores the system level effect for all patients treated (public and private), across public hospitals, and picks up a statistically significant and negative effect. In contrast, the control-treatment approaches account for the simultaneous downward trend in private (control) patient activity, thus approximating a natural experiment (e.g. a RCT) more closely, and producing more robust estimates, relative to ITS.

It is important to note that often no comparison group may be available, limiting the analysis to the ITS approach. This may be driven by various data limitations. For example, the data available over a period may only partially be available for a specific intervention. Therefore, conventional regression modelling may be the only feasible approach to account for pre-intervention differences, even though there is evidence that these methods may provide biased results, most notably in the presence of time-dependent confounders [4]. Additionally, certain intervention and policy evaluations may not be feasible under a control-treatment design, and for which the ITS approach is more suitable. This applies to studies which focus on a specific patient [53] or hospital group [10], or policies at a more aggregate or population level [54], for which it is difficult to identify a naturally occurring control group. Therefore, the inclusion of a control group in these instances would not be suitable, suggesting a before-after comparison in the level and trend of outcomes using ITS analysis as a more suitable approach. Additionally, ITS models may be more effective in the evaluation of policy and intervention effects when the control-treatment specific assumptions of parallel trends and the common independence assumptions do not hold [55].

Additionally, ITS has been highlighted as an effective approach to study short-term policy and intervention effects, as estimation of long-term effects can be biased due to the presence of simultaneous shocks to the outcome of interest [56]. In contrast, control-treatment approaches such as DiD and SC have been recognised as more appropriate and robust for estimation of long-term intervention effects [57], as these allow intervention effects to change over time [38, 49]. Despite recent improvements and developments of the ITS approach [34, 35], the benefits of adopting control-treatment approaches for health intervention and policy evaluation, have been previously highlighted [33].

It should be noted that all of the methods applied in this study are limited to the evaluation of a single policy.

Therefore, any other smaller scale simultaneous policies that are implemented during the period of analysis are difficult to differentiate out in many instances. However, the control-treatment methods account for any unmeasured group or time-invariant confounders from the main intervention itself by incorporating a control group [33]. For example, the introduction of ABF in our empirical example may have been accompanied by a hospital-wide discharge policy aimed at reducing LOS. In this instance, ITS may attribute the reduction in LOS as the impact of ABF entirely, although this is a hospital policy effect. Alternatively, the inclusion of a control group (e.g. patients targeted in the LOS policy, but not to ABF) would difference out the ABF effect from the LOS policy, and would capture effects specific to ABF introduction. In this case, ITS may overestimate the impacts of ABF relative to the other approaches and may further contribute to different evidence base for policy decisions.

This study has several limitations. First, we limited our ITS analysis to a single group (public patient episodes) despite recent developments to ITS for multiple group comparisons [34]. However, this was informed by a recent review, which identified that ITS was employed to estimate intervention effects for a single group [5]. Second, for each of the control-treatment methods, we assumed that any individual shocks following ABF introduction had the same expected effect on the average LOS for the treatment and control groups. Third, we assumed that all of the models were correctly specified in terms of their respective identification and functional form assumptions. However, if either the identification or the functional assumptions are violated, the estimates can be biased, particularly as highlighted in recent literature on DiD approaches [45]. Fourth, we limited our focus on two key assumptions applicable to the quasi-experimental approaches i.e. parallel trends and conditional independence, and did not focus on other assumptions e.g. common shock assumption. Fifth, recent research evidence has addressed the issues related to intervention ‘spillover effects’ i.e. the unintended consequences of health-related interventions beyond those initially intended [58]. It is possible that the differing estimated effects, based on the analytical method used, may have, or could lead to spillover effects as a result. However, given the nature of the data used in our analysis, and our focus on a single procedure in our empirical analysis, it is difficult to identify any potential spillover effects, which may have been linked to ABF. More exploration of such effects may be necessary in future research. Finally, caution should be taken in generalising the reported ABF effects in this study given that our empirical example focused on one procedural group in one particular country.

Conclusion

In health services research it is not always feasible to conduct experimental analysis and we therefore often rely on observational analysis to identify the impact of policy interventions. We demonstrated that ITS analysis produces results different in interpretation relative to control-treatment approaches such as DiD, PSM DiD and SC. Our comparative method analysis therefore suggests that choice of analytical method should be carefully considered and researchers should strive to employ more appropriate designs incorporating control and treatment groups. These methods are more robust and provide a stronger basis for evidence-based policy-making and evidence for informing future financing reform and policy.

List of abbreviations

RCT	Randomised Controlled Trial
ITS	Interrupted Time Series
DiD	Difference-in-Differences
PSM	Propensity Score Matching
PSM DiD	Propensity Score Matching Difference-in-Differences
SC	Synthetic Control
CIA	Conditional Independence Assumption
ABF	Activity-Based Funding
ATT	Average Treatment effect on the Treated
HSE	Health Service Executive
HIPE	Hospital In-Patient Enquiry
LOS	Length of Stay
DRG	Diagnosis-Related Group

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12913-022-08657-0>.

Supplementary Material 1

Acknowledgements

The authors wish to thank the Data Analytics team at the Healthcare Pricing Office (HPO) for granting access to the data used in this study. This study was conducted as part of the Health Research Board (HRB) SPHeRE Programme (Grant No. SPHeRE-2018-1). The Health Research Board (HRB) supports excellent research that improves people’s health, patient care and health service delivery. An earlier version of this work has been previously presented at the virtual International Health Economics Association (IHEA) Congress 2021.

Author contribution

JS and GV conceived the study. GV drafted and edited the manuscript and performed statistical analysis. CK and JS critically revised the manuscript. All authors approved the final draft.

Funding

This research was funded by the Health Research Board SPHeRE-2018-1.

Data Availability

The data that support the findings of this study were made available under a strict user agreement with the Healthcare Pricing Office. Access to the data may only be sought directly from the Healthcare Pricing Office.

Declarations

Ethics approval and consent to participate

Ethical approval for this study was granted by the Research Ethics Committee of the Royal College of Surgeons of Ireland (REC201910019). We confirm that all methods in this study were carried out in accordance with their specifications and other relevant guidelines and regulations. The ethics committee recognized that explicit consent to participate in the study was not required, as the data used in this study were retrospective, routinely collected, and anonymised. The data controller, the Healthcare Pricing Office, responsible for holding and managing the national Hospital In-Patient Enquiry (HIPE) database, granted access and permission to use the data in this study. The Healthcare Pricing Office ensured strict data user agreements were followed, and the data were anonymized by limiting certain combinations of data that could lead to patient identification. This was in line with the Healthcare Pricing Office adherence to The Data Protection Acts 1998 to 2018 and Regulation (EU) 2016/679 of the European Parliament and the council of 27 of 27 April 2016 also known as the General Data Protection Regulation or GDPR (HPO Data Protection Statement Version 1.2, May 2020, Healthcare Pricing Office. [available at: https://hpo.ie/data_protection/HPO_Data_Protection_Statement_Version_1.2_May2020_Covid_1.pdf]

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 13 December 2021 / Accepted: 16 September 2022

Published online: 03 November 2022

References

- Brook RH, Keeler EB, Lohr KN, Newhouse JP, Ware JE, Rogers WH, et al. The Health Insurance Experiment: A Classic RAND Study Speaks to the Current Health Care Reform Debate. Santa Monica: RAND Corporation; 2006.
- Finkelstein A, Taubman S, Wright B, Bernstein M, Gruber J, Newhouse JP, et al. The Oregon Health Insurance Experiment: Evidence from the first year. *Q J Econ*. 2012;127(3):1057–106.
- Jones AM, Rice N. *Econometric evaluation of health policies*. Oxford: Oxford University Press; 2011.
- Baicker KS, T., Testing the Validity of the Single Interrupted Time Series Design. CID Working Papers 364, Center for International Development at Harvard University. 2019.
- Valentelyte G, Keegan C, Sorensen J. Analytical methods to assess the impacts of activity-based funding (ABF): a scoping review. *Health Econ Rev*. 2021;11(1):17.
- O'Neill S, Kreif N, Grieve R, Sutton M, Sekhon JS. Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Serv Outcomes Res Methodol*. 2016;16:1–21.
- O'Neill S, Kreif N, Sutton M, Grieve R. A comparison of methods for health policy evaluation with controlled pre-post designs. *Health Serv Res*. 2020;55(2):328–38.
- Sutherland JM, Liu G, Crump RT, Law M. Paying for volume: British Columbia's experiment with funding hospitals based on activity. *Health Policy*. 2016;120(11):1322–8.
- Januleviciute J, Askildsen JE, Kaarboe O, Siciliani L, Sutton M. How do Hospitals Respond to Price Changes? Evidence from Norway. *Health Econ (United Kingdom)*. 2016;25(5):620–36.
- Shmueli A, Intrator O, Israeli A. The effects of introducing prospective payments to general hospitals on length of stay, quality of care, and hospitals' income: the early experience of Israel. *Soc Sci Med*. 2002;55(6):981–9.
- Perelman J, Closon MC. Hospital response to prospective financing of inpatient days: The Belgian case. *Health Policy*. 2007;84(2–3):200–9.
- Martinussen PE, Hagen TP. Reimbursement systems, organisational forms and patient selection: Evidence from day surgery in Norway. *Health Econ Policy Law*. 2009;4(2):139–58.
- Theurl E, Winner H. The impact of hospital financing on the length of stay: Evidence from Austria. *Health Policy*. 2007;82(3):375–89.
- Gaughan J, Gutacker N, Grašič K, Kreif N, Siciliani L, Street A. Paying for efficiency: Incentivising same-day discharges in the English NHS. *J Health Econ*. 2019;68:102226-.
- Allen T, Fichera E, Sutton M. Can Payers Use Prices to Improve Quality? Evidence from English Hospitals. *Health Econ*. 2016;25(1):56–70.
- Verzulli R, Fiorentini G, Lippi Bruni M, Ugolini C. Price Changes in Regulated Healthcare Markets: Do Public Hospitals Respond and How? *Health Econ*. 2017;26(11):1429–46.
- Krabbe-Alkemade YJFM, Groot TLM, Lindeboom M. Competition in the Dutch hospital sector: an analysis of health care volume and cost. *Eur J Health Econ*. 2017;18(2):139–53.
- Hamada H, Sekimoto M, Imanaka Y. Effects of the per diem prospective payment system with DRG-like grouping system (DPC/PDPS) on resource usage and healthcare quality in Japan. *Health Policy*. 2012;107(2):194–201.
- Farrar S, Yi D, Sutton M, Chalkley M, Sussex J, Scott A. Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. *BMJ (Online)*. 2009;339(7720):554–6.
- Cooper Z, Gibbons S, Jones S, McGuire A. Does Hospital Competition Save Lives? Evidence From The English NHS Patient Choice Reforms*. *Econ J*. 2011;121(554):F228-F60.
- Palmer KS, Agoritsas T, Martin D, Scott T, Mulla SM, Miller AP, et al. Activity-based funding of hospitals and its impact on mortality, readmission, discharge destination, severity of illness, and volume of care: a systematic review and meta-analysis. *PLoS ONE*. 2014;9(10):e109975.
- Street A, Vitikainen K, Bjorvatn A, Hvenegaard A. Introducing activity-based financing: a review of experience in Australia, Denmark, Norway and Sweden. Working Papers 030cherp, Centre for Health Economics, University of York. 2007.
- Street A, Maynard A. Activity based financing in England: the need for continual refinement of payment by results. *Health Econ Policy Law*. 2007;2(4):419–27.
- Shleifer A. A Theory of Yardstick Competition. *RAND J Econ*. 1985;16(3):319–27.
- Brick A, Nolan A, O'Reilly J, Smith S. Resource Allocation, Financing and Sustainability in Health Care. Evidence for the Expert Group on Resource Allocation and Financing in the Health Sector. Dublin: The Economic and Social Research Institute (ESRI); 2010. July 9, 2010.
- Keegan C, Connolly S, Wren MA. Measuring healthcare expenditure: different methods, different results. *Ir J Med Sci (1971 -)*. 2018;187(1):13–23.
- Healthcare Pricing Office. Activity in Acute Public Hospitals in Ireland. 2021.
- Department of Health. Future Health. A Strategic Framework for Reform of the Health Service 2012–2015. Dublin; 2012.
- Health Service Executive (HSE). Activity-Based Funding Programme Implementation Plan 2015–2017. Dublin; 2015.
- Healthcare Pricing Office. Introduction to the Price Setting Process for Admitted Patients V1.0 26May2015. 2015.
- Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ (Clinical research ed)*. 2015;350:h2750.
- Bernal JL, 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.
- Blundell R, Costa Dias M. Evaluation Methods for Non-Experimental Data. *Fisc Stud*. 2000;21(4):427–68.
- Linden A. Conducting Interrupted Time-series Analysis for Single- and Multiple-group Comparisons. *Stata J*. 2015;15(2):480–500.
- Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. *J Eval Clin Pract*. 2011;17(6):1231–8.
- Rubin DB. Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *J Am Stat Assoc*. 2005;100(469):322–31.
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educational Psychol*. 1974;66(5):688–701.
- Angrist JDP, Pischke J-K. *Parallel Worlds: Fixed Effects, Differences-in-differences, and Panel Data*. Mostly Harmless Econometrics: An Empiricist's Companion: Princeton University Press; 2009.
- Basu S, Meghani A, Siddiqi A. Evaluating the Health Impact of Large-Scale Public Policy Changes: Classical and Novel Approaches. *Annu Rev Public Health*. 2017;38:351–70.
- Heckman JJ, Ichimura H, Todd PE. Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Rev Econ Stud*. 1997;64(4):605–54.

41. Heckman J, Ichimura H, Smith J, Todd PE. Characterizing Selection Bias Using Experimental Data. *Econometrica*. 1998;66(5):1017–98.
42. Song Y, Sun W. Health Consequences of Rural-to-Urban Migration: Evidence from Panel Data in China. *Health Econ*. 2016;25(10):1252–67.
43. Glazerman S, Levy DM, Myers D. Nonexperimental Replications of Social Experiments: A Systematic Review 2003.
44. 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.
45. Sant'Anna PHC, Zhao J. Doubly robust difference-in-differences estimators. *J Econ*. 2020;219(1):101–22.
46. Callaway B, Sant'Anna PHC. Difference-in-Differences with multiple time periods. *Journal of Econometrics*. 2020.
47. Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. *Health Econ*. 2016;25(12):1514–28.
48. Bouttell J, Craig P, Lewsey J, Robinson M, Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Commun Health*. 2018;72(8):673.
49. Abadie A. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *J Econ Lit*. 2021;59(2):391–425.
50. Cruz M, Bender M, Ombao H. A robust interrupted time series model for analyzing complex health care intervention data. *Stat Med*. 2017;36(29):4660–76.
51. Leuven E, Sianesi B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Boston College Department of Economics; 2003.
52. Abadie A, Diamond AJ, Hainmueller J. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science* 2014, Forthcoming, Formerly MIT Political Science Department Research Paper No 2011-25. 2014.
53. Epstein RA, Feix J, Arbogast PG, Beckjord SH, Bobo WV. Changes to the financial responsibility for juvenile court ordered psychiatric evaluations reduce inpatient services utilization: an interrupted time series study. *BMC Health Serv Res*. 2012;12(1):136.
54. Pincus D, Widdifield J, Palmer KS, Paterson JM, Li A, Huang A, et al. Effects of hospital funding reform on wait times for hip fracture surgery: a population-based interrupted time-series analysis. *BMC Health Serv Res*. 2021;21(1):576.
55. Hudson J, Fielding S, Ramsay CR. Methodology and reporting characteristics of studies using interrupted time series design in healthcare. *BMC Med Res Methodol*. 2019;19(1):137.
56. Ewusie JE, Soobiah C, Blondal E, Beyene J, Thabane L, Hamid JS. Methods, Applications and Challenges in the Analysis of Interrupted Time Series Data: A Scoping Review. *J Multidiscip Healthc*. 2020;13:411–23.
57. Aragón MJ, Chalkley M, Kreif N. The long-run effects of diagnosis related group payment on hospital lengths of stay in a publicly funded health care system: Evidence from 15 years of micro data. *Health Economics*. 2022;n/a(n/a).
58. Francetic I, Meacock R, Elliott J, Kristensen SR, Britteon P, Lugo-Palacios DG, et al. Framework for identification and measurement of spillover effects in policy implementation: intended non-intended targeted non-targeted spillovers (INTENTS). *Implement Sci Commun*. 2022;3(1):30.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.