International Journal of Epidemiology, 2024, **53(6)**, dyae137
https://doi.org/10.1093/ije/dyae137
Education Corner

IEA International Epidemiological Association
OXFORD

Education Corner

Staggered interventions with no control groups

Brice Batomen [1,*] **and Tarik Benmarhnia** [2,3]

[1]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada, [2]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA and [3]Irset Institut de Recherche en Santé, Environnement et Travail, Inserm, University of Rennes, EHESP, Rennes, France

*Corresponding author. Department of Epidemiology, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Room 688, Toronto, ON M5T 3M7, Canada. E-mail: brice.kuimi@utoronto.ca

Abstract

The limitations of the two-way fixed effects for the impact evaluation of interventions that occur at different times for each group have meant that 'staggered interventions' have been highlighted in recent years in the econometric literature and, more recently, in epidemiology. Although many alternative strategies (such as staggered difference-in-differences) have been proposed, the focus has predominantly been on scenarios in which one or more control groups are available. However, control groups are often unavailable, due to limitations in the available data or because all units eventually receive the intervention. In this context, interrupted time series (ITS) designs can constitute an appropriate alternative. The extent to which common model specifications for ITS analyses are limited in the case of staggered interventions remains an underexplored area in the methodological literature. In this work, we aim to demonstrate that standard ITS model specifications typically yield biased results for staggered interventions and we propose alternative model specifications that were inspired by recent developments in the difference-in-differences literature to propose adapted analytical strategies.

Keywords: Staggered interventions, time series, quasi-experimental designs, longitudinal analyses, injury.

Key Messages

- In staggered intervention scenarios without control groups, common model specification for impact evaluations may yield biased estimates due to ill-defined post-intervention periods.
- Alternative model specifications that are drawn from the difference-in-differences literature for staggered interventions can be easily adopted when no control group is available.
- The adoption of these alternative models improves the validity of impact evaluations, especially if heterogeneity is expected across treated groups and across post-intervention time periods.

Introduction

Quasi-experimental designs that capitalize on the timing of a natural experiment are ubiquitous in the impact evaluation of non-randomized interventions.[1] These designs generally consist of pre-post (PP), interrupted time series (ITS) for the evaluation of treated groups without controls and difference-in-differences (DiD) approaches (including extensions such as synthetic control) when control groups are incorporated.[2] Recent years have seen a new body of methodological research that highlights the limitations of DiD when multiple groups have received interventions at different times—'staggered intervention'.[3] Briefly, because the post-intervention periods for control groups are not well defined,[3,4] the standard model specification of two-way fixed effects (TWFE) is biased, unless the intervention effects are non-dynamic, i.e. they are consistent across groups and throughout post-intervention periods.[3,5] Several mitigating strategies have been proposed and the topic remains an active area for methodological development.[4,6,7]

In practice, it often happens that control groups fail to meet the assumptions necessary for DiD analysis or that there may be a lack of available data for potential controls. Whereas designs such as synthetic control methods can mitigate the former,[8–10] practitioners often resort to ITS designs in the absence of suitable controls.[11–15] Although the standard model specifications for ITS in the context of staggered interventions bear similarities to TWFE, their limitations have not been explicitly highlighted in the literature.

The first objective of this paper is to describe the common model specifications for PP, ITS and DiD designs, and to elucidate potential biases within a staggered intervention framework. The second aim is to introduce alternative model specifications that were inspired by the rapidly growing staggered DiD methodology literature.

Methods

Brief review of common model specifications

Let us start with a single group who are receiving an intervention. The model specification for a PP analysis will be $E(Y) = \alpha + \delta Post$ (1), with Y representing the outcome being evaluated, α the average outcome value at the baseline,

Received: 3 April 2024. **Editorial Decision:** 15 August 2024. **Accepted:** 24 September 2024

Post an indicator for the post-intervention period, and its coefficient δ the estimate of interest, namely the average treatment effect among the treated (ATT). δ is an unbiased estimate only if certain assumptions are satisfied: (i) there are no unmeasured confounding or concurrent events that may influence the outcome; (ii) there are no underlying secular trends, (iii) there is the common shock assumption, i.e. exogenous forces affect the pre and post periods equally;[16] and (iv) there are no selection bias and measurement errors.[17] When several pre- and/or post-intervention measures are available, ITS can be applied and the following model specification is generally used for ITS: $E(Y) = \alpha + \beta_1 Time + \delta Post$ (2). *Time* can be coded from 1 to N (the number of total measurements) and its coefficient β_1 captures the underlying trend of the outcome. *Time* can be modelled by using various functions, ranging from simple linear to more complex nonlinear forms such as splines, or as fixed effects through the use of dummy variables for each period.[18,19] δ is still estimating the ATT and will be valid under a less stringent version of assumption (ii): the effect of time-varying factors on the outcome does not itself vary over time. In addition, compared with Model 1, Model 2 is less vulnerable to the regression to the mean.[20] However, the other assumptions are still necessary. No other intervention that is taking place on a larger scale may influence the outcome of interest (which is one of the reasons why the use of control groups when available is valuable). If there is an interest in examining changes in trends following the intervention, a product term that combines the intervention indicator and time can be incorporated.[9,21] For longer time-series data, it is essential to evaluate stationarity (constant mean and variance over time, meaning that the process that is -For longer time rt in generating the series does not change over time) and account for autocorrelations.[17,22]

With two time periods (one pre and one post) and two groups (one treated and one control), the ATT can be estimated by using the following specification: $E(Y) = \alpha + \beta_1 Treated + \beta_2 Post + \delta Treated * Post$ (3). Model 3 is the canonical DiD model (which can also be referred to as a TWFE model), in which δ provides a valid estimate of the ATT under relatively less stringent assumptions (i) and (iii) (concomitant event taking place at the same time as the intervention exerts similar effects on both the treated and control groups). However, in addition to the assumptions that have already been discussed, a new assumption is necessary: parallel trends in pre-intervention outcomes.[9] When multiple measurements are available for both pre and post intervention, the following specification is commonly used: $E(Y) = \alpha + \beta_1 Treated + \sum_{t=1}^{T} \beta_t Time + \sum_{g=1}^{N} \beta_g Group + \delta Post$ (4), with dummy variables for groups and time periods (which is also a TWFE model).[4]

Issues with staggered interventions

Model 4 is equivalent to Model 3 when there are only two time periods. However, with multiple time periods, Model 4 will be biased because of the impossibility to simultaneously and non-parametrically adjust for group-specific and time-specific unobserved confounders, unless we are willing to make an additional assumption of linear additive effects.[5] The inability of Model 4 to recover an unbiased ATT is exacerbated in the context of staggered interventions. Under such conditions, Model 4 remains valid only if the intervention effects are homogeneous across all groups and subsequent post-intervention periods. In other words, we are not expecting a dynamic treatment effect, which is a strong assumption.[3,23] The δ coefficient from Model 4 emerges as a weighted average of all two-group/two-period estimates within the data set, with the weights being a function of the variability in treatment, with groups that underwent the intervention mid-series receiving greater weights.[3,4]

In a staggered intervention setting with no control groups, the following model is often applied in a one-stage ITS: $E(Y) = \alpha + \sum_{g=1}^{N} \beta_g Group + \beta_1 Time + \delta Post$ (5). Similarly to the single ITS specification model (Model 2), *Time* can be modelled by using a simple linear function, nonlinear functions or fixed effects (equivalent to a TWFE). However, δ in Model 5 will also be biased unless the intervention effects are homogeneous across groups and time points. Redefinition of the *Time* variable to align with the intervention timing for each group (setting *Time* to 0 for the period immediately preceding the intervention and assigning subsequent positive values 1, 2, 3 … for periods thereafter, and negative values –1, –2, –3 … for periods before) will not address the bias.

Alternative model specification

Strategies proposed in the DiD literature to mitigate the aforementioned pitfalls can be adopted even in the absence of a control group. The following model specification can be used: $E(Y) = \alpha + \sum_{g=1}^{N} \beta_g Group + \beta_1 Time + \sum_{ct} \delta_{ct} Post_{ct}$ (6). $Post_{ct}$ are dummies for the post-intervention periods of each cohort, with all groups that undergo the intervention at the same time being classified into identical cohorts. Again, *Time* can be modelled by using a simple linear function, nonlinear functions or fixed effects (offer more flexibility but might not be statistically efficient). Upon the estimation of all $Post_{ct}$ values, several schemes can be employed to summarize them into a singular ATT.[4,7] Briefly, we can compute a weighted average of the ATTs across all cohort–post-time combinations, with weights being proportional to the number of groups in each cohort that are subjected to the intervention at specific time points. We can also average the ATTs across either the cohorts or the time periods. However, if the ATTs are heterogenous across time and/or cohorts, then provision of a single summary measure might be questionable. A summary of the six models can be found in Table 1.

Simulation study

Let us assume that six cities within a state have adhered to international recommendations regarding posted speed limits on urban roads.[24] However, the policy implementation timing varied, as shown in Table 2. Through the examination of three scenarios, we will compare Models 5 and 6 to assess the impact of reduced speed limits on road traffic collisions. The scenarios include: (i) a homogeneous and constant effect, in which the policy induces the same reduction in collision numbers across cities, with this reduction remaining consistent over the entire post period; (ii) a heterogeneous and variable effect; and finally (iii) a heterogeneous and variable effect with a secular trend in collisions.

The effect of the intervention in the first scenario is a diminution of 10 collisions (refer to the seventh column of Table 1). For the subsequent scenarios, the effect of the intervention varied according to the city and post-intervention periods (as indicated in the tenth and thirtieth columns of Table 1). Additionally, we have three cohorts: Cohort 2, encompassing a single city in which the intervention was

Table 1. Common model specifications for quasi-experimental design***

Design	Model specification	Name in the text	Assumptions
Pre-post	$E(Y) = \alpha + \delta Post$	Model 1	(i) no unmeasured confounding; (ii) no secular trends; (iii) common shock assumption, i.e. exogenous forces affect the pre and post periods equally; and (iv) no selection bias and measurement errors
Interrupted time series	$E(Y) = \alpha + \beta_1 Time + \delta Post$[a]	Model 2	Same assumptions, although a less stringent assumption (ii); the effect of time-varying factors on the outcome does not itself vary over time
Difference-in-differences	$E(Y) = \alpha + \beta_1 Treated + \beta_2 Post + \delta Treated * Post$	Model 3 or two-way fixed effect (TWFE)	(v) Parallel trends in pre-intervention outcomes; less stringent assumption (i) and (iii) (unmeasured events exert similar effects on both the treated and control groups)
Difference-in-differences with multiple time periods	$E(Y) = \alpha + \beta_1 Treated + \sum_{t=1}^{T} \beta_t Time + \sum_{g=1}^{N} \beta_g Group + \delta Post$	Model 4, also a TWFE	Same assumptions as Model 3 plus the assumption of linear additive effects
Staggered intervention with no control groups	$E(Y) = \alpha + \sum_{g=1}^{N} \beta_g Group + \beta_1 Time + \delta Post$	Model 5, also a TWFE	Valid only if the intervention has (vi) non-dynamic effects
	$E(Y) = \alpha + \sum_{g=1}^{N} \beta_g Group + \beta_1 Time + \sum_{ct} \delta_{ct} Post_{ct}$	Model 6	Valid if all assumptions (i) to (v) are satisfied

[a] For simplicity, we do not distinguish between changes in intercept and slope, but δ captures the average difference over the post period compared with the counterfactual.
Y is the outcome being evaluated. α is the intercept or the average outcome value when all the other covariates equal zero. Time can be coded from 1 until N (the number of total measurements), and its coefficient β_1 captures the underlying trend of the outcome. Post is an indicator for the post-intervention period. Treated is a dummy variable to distinguish treated and control groups. Group represents each unit. δ is the estimate of interest, namely the average treatment effect among the treated (ATT). δ_{ct} is the ATT across all cohort–post-time combinations, with all groups that undergo the intervention at the same time being classified into identical cohorts.

Table 2. Simulated scenario with six cities implementing the policy at different times

ID	Time	Post	Cohort	Y_{CE}	Y^0_{CE}	δ_{CE}	Y_{HE}	Y^0_{HE}	δ_{HE}	Y_{HEs}	Y^0_{HEs}	δ_{HEs}
1	1	0	2	50	50	–	50	50	–	50	50	–
1	2	1	2	40	50	–10	40	50	–10	35	45	–10
1	3	1	2	40	50	–10	40	50	–10	30	40	–10
1	4	1	2	40	50	–10	40	50	–10	25	35	–10
2	1	0	3	35	35	–	35	35	–	35	35	–
2	2	0	3	35	35	–	35	35	–	30	30	–
2	3	1	3	25	35	–10	28	35	–7	18	25	–7
2	4	1	3	25	35	–10	30	35	–5	15	20	–5
3	1	0	3	40	40	–	40	40	–	40	40	–
3	2	0	3	40	40	–	40	40	–	35	35	–
3	3	1	3	30	40	–10	40	40	0	30	30	0
3	4	1	3	30	40	–10	32	40	–8	17	25	–8
4	1	0	4	65	65	–	65	65	–	65	65	–
4	2	0	4	65	65	–	65	65	–	60	60	–
4	3	0	4	65	65	–	65	65	–	55	55	–
4	4	1	4	55	65	–10	53	65	–12	38	50	–12
5	1	0	4	55	55	–	55	55	–	55	55	–
5	2	0	4	55	55	–	55	55	–	50	50	–
5	3	0	4	55	55	–	55	55	–	45	45	–
5	4	1	4	45	55	–10	45	55	–10	25	35	–10
6	1	0	4	20	20	–	20	20	–	20	20	–
6	2	0	4	20	20	–	20	20	–	15	15	–
6	3	0	4	20	20	–	20	20	–	10	10	–
6	4	1	4	10	20	–10	30	20	10	15	5	10

Y_{CE} is the observed collision count at each time point and Y^0_{CE} is the counterfactual collision count had the policy not been enacted under the first scenario of constant and homogeneous effect. Y_{HE} and Y^0_{HE} are the corresponding observed and counterfactual collision counts under the second scenario of heterogeneous effects. Y_{HEs} and Y^0_{HEs} are the corresponding collision counts for the third scenario of heterogeneous effects with secular trends.
Time represents the number of total measurements. Post an indicator for the post-intervention period. All ID undergoing the intervention at the same time are classified into identical cohorts.

Table 3. Comparative results of average treatment effects on the treated (ATTs) across different scenarios

ATTs	First scenario			Second scenario			Third scenario		
	True	Model 5[a]	Model 6	True	Model 5[a]	Model 6	True	Model 5[a]	Model 6
Post_T2 cohort 2	–10	NA	–10	–10	NA	–10	–10	NA	–9.7
Post_T3 cohort 2	–10	NA	–10	–10	NA	–10	–10	NA	–9.4
Post_T4 cohort 2	–10	NA	–10	–10	NA	–10	–10	NA	–9.1
Post_T3 cohort 3	–10	NA	–10	–3.5	NA	–3.8	–3.5	NA	–3.5
Post_T4 cohort 3	–10	NA	–10	–6.5	NA	–6.3	–6.5	NA	–5.5
Post_T4 cohort 4	–10	NA	–10	–4	NA	0.7	–4	NA	2.4
Averages									
All post cohort–time[b]	–10	–10	–10	–6.2	–2.4	–4.8	–6.2	–0.4	–3.9
Cohort 2	–10	NA	–10	–10	NA	–10	–10	NA	–9.4
Cohort 3	–10	NA	–10	–5	NA	–5	–5	NA	–4.5
Cohort 4	–10	NA	–10	–4	NA	0.7	–4	NA	2.4
Of cohort	–10	NA	–10	–6.3	NA	–4.8	–6.3	NA	–3.8
Post T2	–10	–10	–10	–10	–4.3	–10	–10	–2.9	–9.7
Post T3	–10	–10	–10	–5.7	–2.5	–5.9	–5.7	–0.6	–5.4
Post T4	–10	–10	–10	–5.8	–0.7	–3.4	–5.8	1.7	–2.2
Of post	–10	–10	–10	–7.2	–2.5	–6.4	–7.2	–0.9	–5.8

[a] The specification of Model 5 as described in text: $E(Y) = \alpha + \sum_{g=1}^{N} \beta_g Group + \beta_1 Time + \delta Post$ cannot estimate all the post cohort–time values individually, but only the average across all cohorts and time (Model 5 without Time and Post product term) or average by time across all cohorts (Model 5 with the product term between Time and Post product term).

[b] Weighted average of all cohort–post-time ATTs with weights proportional to the number of groups within a cohort under the intervention at each time point. For example, under Scenario 2, the true value of –6.2 is obtained by using the following formula: [((–10–10–10 – (2*3.5) – (2*6.5) – (3*4))/10].

NA, not applicable.

Post is an indicator for each of the post-intervention periods. For example, Post_T3 cohort 2 represents the post period Time 3 for all groups that received the intervention at Time 2. All groups that undergo the intervention at the same time are classified into identical cohorts.

introduced at Time 2; Cohort 3, comprising two cities that implemented the intervention at Time 3; and Cohort 4, consisting of three cities that adopted the intervention at Time 4. This results in six post-intervention cohort periods: three for Cohort 2, two for Cohort 3 and one for Cohort 4.

Models 5 and 6 were fitted in each scenario and, in order to obtain the ATTs on an absolute scale (number of collisions prevented by the posted speed limits intervention), the weighted displacement difference was calculated by using a Poisson distribution with an identity link function.[25] Several estimands were considered: first, the post-intervention cohort–time ATTs corresponding to the six post-intervention cohort periods; second, the weighted average of all post-intervention cohort–time ATTs with weights proportional to the number of groups within each cohort being treated at each time point; third, the average of the ATTs for each cohort and the overall average across all cohorts; and, finally, the average of the ATTs for each post-intervention time points and the overall average across all time points.

Results

The analysis of the simulated data, through the application of Models 5 and 6, reveals how the underlying scenario influences each model capacity to accurately estimate the intervention effects. It is important to note that Model 5 is limited to estimating only the weighted average of all post-intervention cohort–time ATTs and the average ATTs of each post-intervention time.

Under the first scenario of constant and homogeneous treatment effects, both Models 5 and 6 recovered the true values (Table 3). Under the second scenario of heterogenous treatment effects without secular trends in collision counts, Model 5 consistently exhibited bias. Conversely, Model 6 successfully estimated the true post-intervention cohort–time ATTs, with the sole exception being the final post-intervention period of Cohort 4. This specific bias (ATT estimates for Time 4 of Cohort 4) affected all aggregated summary measures. Under the third scenario of heterogenous treatment effects with secular trends in injury counts, neither Model 5 nor Model 6 was capable of accurately identified the true values. Nonetheless, Model 6 closely approximated the ATTs for the initial post-intervention period across all cohorts.

To assess the impact of multicollinearity, the last time point for all groups (city) was removed from the analysis, as all the cities were treated at that point. Subsequently, Models 5 and 6 were reapplied. Again, Model 5 successfully identified the true values exclusively in the first scenario whereas it exhibited biases in the remaining two scenarios (Table 4). In contrast, Model 6 now nearly recovered the true values in all scenarios. Another interesting finding was that, for Model 5, the bias was towards the null (Figure 1).

The treatment of time as a categorical variable instead of a continuous yields similar results, with Model 5 only able to recover the true values in Scenario 1 and Model 6 being valid if the time periods in which all group are treated were removed to deal with multicollinearity (Supplementary Table S1, available as Supplementary data at *IJE* online). Also, computation of the results on the relative (incidence rate ratios) instead of the absolute scale (weighted displacement differences) produces comparable findings (Supplementary Tables S2 and S3, available as Supplementary data at *IJE* online), but Model 6 exhibited a slight bias under a certain scenario because the secular trend was generated on a linear scale.

Discussion

In this study, we explored the limitations of conventional ITS model specifications in the context of staggered interventions. Staggered interventions are prevalent in settings in which ITS

Table 4. Comparative results of average treatment effects on the treated (ATTs) across different scenarios—without the time point at which all the groups are treated

	First scenario			Second scenario			Third scenario		
ATTs	**True**	**Model 5[a]**	**Model 6**	**True**	**Model 5[a]**	**Model 6**	**True**	**Model 5[a]**	**Model 6**
Post_T2 cohort 2	–10	NA	–10	–10	NA	–10	–10	NA	–10
Post_T3 cohort 2	–10	NA	–10	–10	NA	–10	–10	NA	–10
Post_T3 cohort 3	–10	NA	–10	–3.5	NA	–3.9	–3.5	NA	–4
Averages									
All post cohort–time[¥]	–10	–10	–10	–6.75	–5.6	–6.9	–6.75	–5.4	–7
Cohort 2	–10	NA	–10	–10	NA	–10	–10	NA	–10
Cohort 3	–10	NA	–10	–3.5	NA	–3.9	–3.5	NA	–4
Of cohort	–10	NA	–10	–6.75	NA	–6.9	–6.75	NA	–7
Post T2	–10	–10	–10	–10	–7.4	–10	–10	–7.1	–10
Post T3	–10	–10	–10	–5.7	–5	–5.9	–5.7	–5	–6
Of post	–10	–10	–10	–7.8	–6.2	–8	–7.8	–6	–8

[a] The specification of Model 5 as described in text: $E(Y) = \alpha + \sum_{g=1}^{N} \beta_g Group + \beta_1 Time + \delta Post$ cannot estimate all the post-intervention cohort–time values individually, but only the average across all cohort–time points (Model 5 without Time and Post product term) or average by post time points across all cohorts (Model 5 with the product term between Time and Post product term).

[b] Weighted average of all post-intervention cohort–time ATTs with weights proportional to the number of groups within a cohort under the intervention at each time point. For example, under Scenario 2, the true value of –6.75 is obtained by using the following computation: [((–10–10 – (2*3.5))/4].

NA, not applicable.

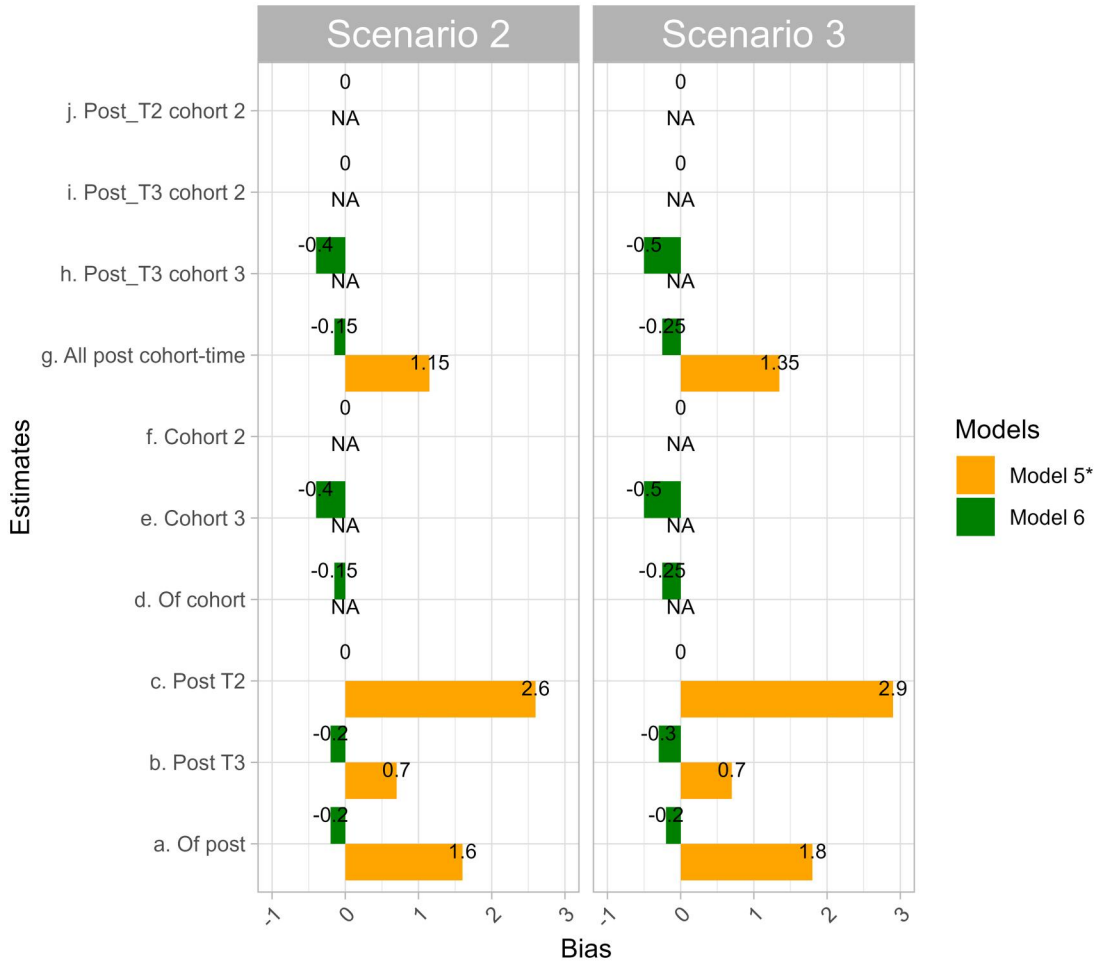

Figure 1. Biases magnitude and direction. The specification of Model 5 as described in the text: $E(Y) = \alpha + \sum_{g=1}^{N} \beta_g Group + \beta_1 Time + \delta Post$ cannot estimate all the post-intervention cohort–time values individually, but only the average across all cohort–time points (Model 5 without *Time* and *Post* product terms) or the average by post time points across all cohorts (Model 5 with the product term between *Time* and *Post* product terms). The last time point was removed for all groups

are used, such as interventions that are implemented over multiple areas in a sequential manner.[11–15] Our analysis revealed that standard ITS models are likely to yield biased estimates unless it is assumed that the intervention effects are homogenous and constant across groups and post-intervention times. Moreover, the bias seems to be towards the null (underestimation of actual effect).

Through the adoption of a model specification that was derived from the DiD literature for staggered interventions, we recovered the true values in all of the considered scenarios. It is worth noting that the removal of time points at which all groups are treated is necessary to deal with multicollinearity. Additionally, the alternative model specification provides more flexibility, as we can compute ATTs for each cohort–time period, by cohort, by time period and the average according to the relevant study question.

The methodological literature on staggered interventions is an active area of research,[4,7,23,26] with methods that are primarily concerned with DiD but are adaptable to contexts in which no control groups are available. The absence of control groups is prevalent in impact evaluation studies. Restriction of the study sample to the treated groups can be advantageous for positivity concerns if the factors behind the intervention assignments are not well understood. Common statistical packages for staggered DiD can accommodate the Model 6 specifications. For example, with the R *did* package,[27] the users will specify the control group as '*notyettreated*' and the time points at which all groups are treated will automatically be removed during the computation.

Several other aggregation schemes, beyond those presented in this study, could be employed. However, researchers must ensure that the selected scheme aligns with their research question. Furthermore, investigators should determine in case of heterogeneous and non-constant intervention effects if the provision of a single summary measure is of interest in the context of the research question. Other modelling strategies for ITS could be considered. For example, a two-stage ITS approach as follows: (i) develop and optimize a predictive model with pre-intervention data only, using various approaches including an autoregressive integrated moving average that will aim at fitting the observed data using time-varying variables that could not be impacted by the intervention; (ii) use this optimized model to predict counterfactual values (i.e. what would have happened without the intervention) for the post-intervention period and compare these predictions with actual post-intervention outcomes to infer effects. For multiple treated groups, repeat this process for each and aggregate the results by using methods from this paper or others, such as meta-analyses.[28,29]

This study has several limitations. We only considered a limited number of scenarios. In practice, the secular trend in outcomes might differ across cohorts or within the same cohort. In such cases, an alternative model specification might be necessary, potentially adding a product term between time and cohort and/or group.[30] Although we did not assess the mean squared error or confidence interval coverage, statistical properties of the proposed model specification have been extensively assessed in previous studies and have demonstrated good performance.[4,27] Moreover, the aim of this study was to illustrate how approaches from the DiD literature for staggered interventions can be applied in contexts without control groups.

Conclusion

This paper underscores methodological challenges that are inherent in staggered intervention studies without control groups and highlights the limitations of traditional impact evaluation models. By drawing from recent development in the DiD literature, we propose an alternative model specification to address these challenges. This work offers practical solutions for researchers who are facing the complexities of staggered interventions, thereby enhancing the robustness and reliability of findings in such settings.

Ethics approval

Ethics approval was not required for this research, as it did not involve human patients.

Data availability

The program code to replicate the simulation as well as the figure and tables are available at https://osf.io/tyfmz/?view_only=a189ed47d74d495dab400de16252a5e6.

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

B.B.: conceptualization, methodology, coding, visualization, writing—original draft preparation, writing—reviewing and editing. T.B.: conceptualization, methodology, writing—reviewing and editing.

Use of artificial intelligence (AI) tools

The authors did not use any AI tools in collecting and/or analysing data, producing images or graphical elements, or in writing the paper.

Funding

This work was supported by a grant from The Banting Research Foundation, through the Banting Discovery award.

Conflict of interest

None declared.

References

1. Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels. *J Clin Epidemiol* 2017;**89**:30–42.
2. Zhou H, Taber C, Arcona S, Li Y. Difference-in-differences method in comparative effectiveness research: utility with unbalanced groups. *Appl Health Econ Health Policy* 2016;**14**:419–29.
3. Goodman-Bacon A. Difference-in-differences with variation in treatment timing. *J Econom* 2021;**225**:254–77.
4. Goin DE, Riddell CA. Comparing two-way fixed effects and new estimators for difference-in-differences: a simulation study and empirical example. *Epidemiology* 2023;**34**:535–43.

5. Imai K, Kim IS. On the use of two-way fixed effects regression models for causal inference with panel data. *Polit Anal* 2021; **29**:405–15.
6. Ben-Michael E, Feller A, Stuart EA. A trial emulation approach for policy evaluations with group-level longitudinal data. *Epidemiology* 2021;**32**:533–40.
7. Callaway B, Sant'Anna PH. Difference-in-differences with multiple time periods. *J Econometr* 2021;**225**:200–30.
8. 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.
9. Lopez Bernal J, Cummins S, Gasparrini A. Difference in difference, controlled interrupted time series and synthetic controls. *Int J Epidemiol* 2019;**48**:2062–63.
10. Strezhnev A. *Generalized Difference-in-Differences Estimands and Synthetic Controls*. 2017.
11. Dowding DW, Turley M, Garrido T. The impact of an electronic health record on nurse sensitive patient outcomes: an interrupted time series analysis. *J Am Med Inform Assoc* 2012;**19**:615–20.
12. Fok CC, Henry D, Allen J. Research designs for intervention research with small samples II: stepped wedge and interrupted time-series designs. *Prev Sci* 2015;**16**:967–77.
13. Lavergne MR, Law MR, Peterson S *et al.* Effect of incentive payments on chronic disease management and health services use in British Columbia, Canada: Interrupted time series analysis. *Health Policy* 2018;**122**:157–64.
14. Rothman L, Macpherson A, Buliung R *et al.* Installation of speed humps and pedestrian-motor vehicle collisions in Toronto, Canada: a quasi-experimental study. *BMC Public Health* 2015; **15**:774.
15. Batomen B, Moore L, Strumpf E, Champion H, Nandi A. Impact of trauma centre accreditation on mortality and complications in a Canadian trauma system: an interrupted time series analysis. *BMJ Qual Saf* 2021;**30**:853–66.
16. Ryan AM, Burgess JF Jr, Dimick JB. Why we should not be indifferent to specification choices for difference-in-differences. *Health Serv Res* 2015;**50**:1211–35.
17. Donnelly NJ. The use of interrupted time series analysis to evaluate the impact of Pharmaceutical Benefits Scheme policies on drug utilisation in Australia [PhD Doctorate]: UNSW Sydney; 2005.
18. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol* 2017;**46**:348–55.
19. Bernal JL, Soumerai S, Gasparrini A. A methodological framework for model selection in interrupted time series studies. *J Clin Epidemiol* 2018;**103**:82–91.
20. Barnett AG, Van Der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 2004; **34**:215–20.
21. Xiao H, Augusto O, Wagenaar BH. Reflection on modern methods: a common error in the segmented regression parameterization of interrupted time-series analyses. *Int J Epidemiol* 2021; **50**:1011–15.
22. Woodward WA, Gray HL, Elliott AC, Stationary Time Series. In: *Applied Time Series Analysis with R*. 2nd ed. Boca Raton, FL, USA: CRC Press, 2017, pp. 5–6.
23. Borusyak K, Jaravel X, Spiess J, Revisiting event study designs: robust and efficient estimation. *Rev Econ Stud* 2024;rdae007.
24. World Health Organization. *SPEED MANAGEMENT: A road safety manual for decision-makers and practitioners:* World Health Organization; 2023. Report No.: 9240072497.
25. Wheeler AP, Ratcliffe JH. A simple weighted displacement difference test to evaluate place based crime interventions. *Crime Sci* 2018;**7**:11.
26. Wing C, Yozwiak M, Hollingsworth A, Freedman S, Simon K. Designing difference-in-difference studies with staggered treatment adoption: key concepts and practical guidelines. *Annu Rev Public Health* 2024;**45**:485–505.
27. Callaway B, Sant'Anna PH. *did: Difference in Differences*. R package version 2.1.2. 2021. https://bcallaway11.github.io/did/articles/did-basics.html (10 March 2024, date last accessed).
28. Batomen B, Macpherson A, Lewis J *et al.* Vulnerable road user injury trends following the COVID-19 pandemic in Toronto, Canada: an interrupted time series analysis. *J Safety Res* 2024; **89**:152–59.
29. Miratrix LW. Using simulation to analyze interrupted time series designs. *Eval Rev* 2022;**46**:750–78.
30. Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials* 2017;**18**:101.

International Journal of Epidemiology, 2024, 53, 1–7
https://doi.org/10.1093/ije/dyae137
Education Corner