Title
Universal Difference-in-Differences for Causal Inference in Epidemiology.

Journal
Epidemiology, 35(1)

Authors
Tchetgen Tchetgen, Eric
Park, Chan
Richardson, David

Publication Date
2024

DOI
10.1097/EDE.0000000000001676

OPEN

Universal Difference-in-Differences for Causal Inference in Epidemiology

Eric J. Tchetgen Tchetgen,[a] *Chan Park,*[a] *and David B. Richardson*[b]

Abstract: Difference-in-differences is undoubtedly one of the most widely used methods for evaluating the causal effect of an intervention in observational (i.e., nonrandomized) settings. The approach is typically used when pre- and postexposure outcome measurements are available, and one can reasonably assume that the association of the unobserved confounder with the outcome has the same absolute magnitude in the two exposure arms and is constant over time; a so-called parallel trends assumption. The parallel trends assumption may not be credible in many practical settings, for example, if the outcome is binary, a count, or polytomous, as well as when an uncontrolled confounder exhibits nonadditive effects on the distribution of the outcome, even if such effects are constant over time. We introduce an alternative approach that replaces the parallel trends assumption with an odds ratio equi-confounding assumption under which an association between treatment and the potential outcome under no treatment is identified with a well-specified generalized linear model relating the pre-exposure outcome and the exposure. Because the proposed method identifies any causal effect that is conceivably identified in the absence of confounding bias, including nonlinear effects such as quantile treatment effects, the approach is aptly called universal difference-in-differences. We describe and illustrate both fully parametric and more robust semiparametric universal difference-in-differences estimators in a real-world application concerning the causal effects of a Zika virus outbreak on birth rate in Brazil.

Keywords: Difference-in-differences; Equi-confounding; Odds ratios; Selection bias; Generalized linear models; Extended propensity score

(Epidemiology 2024;35: 16–22)

A supplementary digital video is available at: http://links.lww.com/EDE/C90

Difference-in-differences (DiD) is a popular approach to account for unmeasured confounding in observational data. DiD is typically used when (i) pre- and postexposure outcome measurements are available and (ii) the parallel trends assumption is reasonable, that is, the absolute magnitude of the association between the unobserved confounder and the outcome is equal across treated and control groups and constant over time. Under parallel trends, an estimate of the causal effect of the treatment can be obtained by taking a difference between treated and control groups of the average change in outcome over time; see Caniglia and Murray[1] for an introduction to DiD and the next Section for a brief review. Despite its popularity, parallel trends may not be credible for a variety of reasons; see the eAppendix (http://links.lww.com/EDE/C80). Therefore, the development of alternative identifying conditions for DiD settings continues to be an active area of research.

In this article, we introduce an alternative identification strategy for DiD settings. Specifically, as described in the Universal DiD Section, our approach is based on the assumption that confounding bias for the causal effect of interest, defined as an association between exposure and the treatment-free potential outcome, can be identified under a generalized linear model (GLM) relating the pre-exposure outcome and the exposure. The proposed approach allows for investigators to proceed with a new approach in DID settings under a slightly different key identifying assumption (i.e., different from parallel trends) that, if it holds, does not require one to assume equal and additive effects of an uncontrolled confounder under models for binary or polytomous outcomes. An appeal of the framework is that it permits both familiar parametric models, as well as more robust semiparametric estimation approaches, namely (i) a GLM approach, followed by (ii) an extended propensity score approach, and finally, (iii) a doubly robust approach, which remain valid if either (i) or (ii) provides valid

Submitted February 1, 2023; accepted September 21, 2023

From the [a]Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA; and [b]Department of Environmental & Occupational Health, University of California, Irvine, CA.

E.J.T.T. was supported by NIH grants R01AI127271, R01CA222147, R01AG065276, and R01GM139926. D.B.R. was supported by grant R01OH011409 from the National Institute for Occupational Safety and Health of the Centers for Disease Control and Prevention, and R01CA242852 from the U.S. National Cancer Institute.

The most recent version of the paper is available at https://arxiv.org/pdf/2302.00840.pdf.

The data and the analysis code are attached as supplemental digital content and are also accessible on the GitHub repository located at http://github.com/qkrcks0218/UDiDEpi

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Chan Park, Department of Statistics and Data Science, University of Pennsylvania, 265 S 37th St, Philadelphia, PA 19104. E-mail: chanpk@wharton.upenn.edu.

ISSN: 1044-3983/24/351-1622
DOI: 10.1097/EDE.0000000000001676

inferences without a priori knowing which might be misspecified. Importantly, the proposed methods can be used to identify and estimate the causal effect of a hypothetical intervention in the presence of unmeasured confounding, on virtually any scale of potential interest, including nonlinear scales such as quantile causal effects. For this reason, the approach is aptly called universal difference-in-differences, hereafter referred to as universal DiD. In the Application and Discussion Sections, we apply our methods to evaluate the causal link between a Zika virus outbreak and birth rates in Brazil and discuss a sensitivity analysis approach to assess the impact of a violation of key assumptions.

The proposed methods are a special case of a more general approach proposed by Park and Tchetgen Tchetgen.[2] This article serves as an introductory resource on universal DiD methods for an epidemiology audience, with a significant emphasis on utilizing parametric models, interpreting identifying assumptions, and offering practical guidelines for implementation and evaluation in practical settings. Specifically, this article focuses on applying well-established GLMs and propensity score weighting methods to the universal DiD framework, thereby easing their adoption in epidemiologic studies. Readers interested in more technical details can consult.[2]

NOTATION AND A BRIEF REVIEW OF DID

Consider a study in which pre- and postexposure outcomes Y_0 and Y_1 are observed; let Y_t^a denote the potential outcome at time $t \in \{0,1\}$ under a hypothetical intervention that sets a binary exposure/treatment A to a value $a \in \{0,1\}$. Following a standard DiD model,[3] suppose that the treatment-free potential outcome is generated from the following model for $t = 0, 1$:

$$\begin{aligned} Y_t^{a=0} &= h(U_t, t), \\ h(u,t) &= u + \beta_T t, \\ U_t &= \beta_0 + \beta_A A + \varepsilon_t, \end{aligned} \qquad \text{(DiD Model)}$$

ε_t satisfies either

$$\begin{cases} \text{time independence}: & \varepsilon_1 | A \overset{D}{=} \varepsilon_0 | A \quad \text{or} \\ \text{treatment independence}: & \varepsilon_t | (A=0) \overset{D}{=} \varepsilon_t | (A=1) \end{cases}.$$

Here, ε_t is an unobserved error at time t that is independent of time or treatment, and U_t is therefore also unobserved. Note that allowing U_t to depend on t accommodates the factors predicting Y_0 to be distinct from those predicting Y_1. In (DiD Model), $Y_t^{a=0}$ is a deterministic function of U_t (in fact a linear function of the latter), but the exposure mechanism of A given U_t is unrestricted. In terms of the distribution of U_t, (DiD Model) assumes that the conditional distribution of ε_t given A is either stable over time given A or independent of A at each time; DiD strictly only requires that ε_t does not depend on A and t in a manner that they interact on the additive scale. Finally, (DiD Model) implies rank preservation which rules out any additive interaction between A and U_1 in causing Y_1. In the Universal DiD Section, we present an alternative structural model compatible with (Universal DiD Model), thus allowing for heterogeneity of the causal effects of A with respect to U_1, that is, h is allowed to be unrestricted.

(DiD Model) implies that $E(Y_t^{a=0}|A=a) = \beta_0 + \beta_A a + \beta_T t + E(\varepsilon_t|A=a)$, which further implies the so-called parallel trends assumption:

$$E\left(Y_1^{a=0} - Y_0^{a=0} | A=1\right) = E\left(Y_1^{a=0} - Y_0^{a=0} | A=0\right). \qquad (1)$$

Expression (1) states that, on average, the trajectory of the potential outcomes under an intervention that sets the exposure to its control value, is equal between exposed and unexposed groups. Hence, under no unmeasured confounding of the average additive effect of A on $Y_0^{a=0}$ and $Y_1^{a=0}$, respectively, both lefthand and righthand sides of the display above would be zero. This gives an alternative interpretation of parallel trends as an assumption of additive equi-confounding bias, such that the confounding bias for the effect of A on $Y_1^{a=0}$ though not null, is equal to the confounding bias for the causal effect of A on $Y_0^{a=0}$ on the additive scale.[4] Note that the latter is empirically identified under consistency and no causal anticipation assumptions, which we now state:

Assumption 1a *Consistency:* $Y_t = Y_t^{a=A}$ *almost surely for* $t = 0, 1$

Assumption 1b *No Causal Anticipation:* $Y_0^{a=1} = Y_0^{a=0}$ *almost surely.*

It is then straightforward to deduce identification of the additive average causal effect of treatment on the treated (ATT) for the follow-up outcome Y_1, that is, $\psi_{ATT} = E\left(Y_1^{a=1} - Y_1^{a=0} | A=1\right) = E\left(Y_1 - Y_0 | A=1\right) - E\left(Y_1 - Y_0 | A=0\right)$, justifying DiD.

Despite its popularity, parallel trends has several limitations: (i) it can be violated when dealing with naturally constrained outcomes such as binary or count variables, (ii) it restricts the exposure mechanism and time-varying properties of the outcomes, and (iii) it is scale-dependent; see the eAppendix (http://links.lww.com/EDE/C80) for details. In the next section, we describe an alternative to parallel trends that accommodates (i)–(iii).

UNIVERSAL DID

Identification via Odds Ratio Equi-confounding

We introduce a parametrization for a unit's contribution to the likelihood for the potential outcome $Y_t^{a=0}$ conditional on A and observed baseline covariates X, assuming independent and identically distributed (i.i.d.) sampling. Let

$$h_t(y,x) = f(Y_t^{a=0} = y | A=0, X=x) \qquad (2)$$

$$\beta_t(y,x) = \log \frac{\left\{ \begin{array}{l} f\left(Y_t^{a=0} = y | A=1, X=x\right) \\ \times f\left(Y_t^{a=0} = y_{\text{ref}} | A=0, X=x\right) \end{array} \right\}}{\left\{ \begin{array}{l} f\left(Y_t^{a=0} = y | A=0, X=x\right) \\ \times f\left(Y_t^{a=0} = y_{\text{ref}} | A=1, X=x\right) \end{array} \right\}} \qquad (3)$$

The function $h_t(y,x)$, referred to as a baseline density, represents the conditional distribution of the potential outcome $Y_t^{a=0}$ given X and $A=0$ (i.e., baseline treatment), and the function $\beta_t(y,x)$ is the log of the generalized odds ratio function[5,6] where $\beta_t(y_{\text{ref}},x)=0$, with y_{ref} a user-specified reference value; without loss of generality, we take $y_{\text{ref}}=0$. Thus, $\beta_t(y,x)$ encodes the association between $Y_t^{a=0}$ and A, evaluated at y given $X=x$. Therefore, $\beta_t(y,x)=0$ for all y encodes no unmeasured confounding given $X=x$, while $\beta_t(y,x)\neq 0$ quantifies the degree of unmeasured confounding bias at a distributional level.

These functions parametrize the conditional density of $Y_t^{a=0}$ given $(A=a,X=x)$ as $f(Y_t^{a=0}=y|A=a,X=x)\propto h_t(y,x)exp\{\beta_t(y,x)a\}$; here, $\propto$ stands for the lefthand side being proportional to the righthand side for fixed a and x, with proportionality constant equal to the normalizing constant $\sum_y h_t(y,x)\,exp\{\beta_t(y,x)\,a\}<\infty$ for all a and x, ensuring that the lefthand side is a proper density or probability mass function, and the symbol $\sum_y$ may be interpreted as an integral if y is continuous. This likelihood parametrization can in principle be used to represent any proper likelihood function one might encounter in practice, that is, the above formulation is fully unrestricted (or nonparametric).[5,6] Under consistency and no anticipation, $f\left(Y_0^{a=0}=y|A=a,X=x\right)=f\left(Y_0=y|A=a,X=x\right)$ for $a=0,1$ and $f\left(Y_1^{a=0}=y|A=0,X=x\right)=f\left(Y_1=y|A=0,X=x\right)$, respectively, establishing identification of $\beta_0(y,x)$, $h_0(y,x)$, and $h_1(y,x)$ from equations (2) and (3). In contrast, identification of $f\left(Y_1^{a=0}=y|A=1,X=x\right)$ and $\beta_1(y,x)$ cannot be obtained without an additional condition because $Y_1^{a=0}$ is not observed for units with $A=1$. Our approach relies on the key assumption:

Assumption 2 *Odds Ratio Equi-confounding:* $\beta_0(y,x)=\beta_1(y,x)$ *for all* (y,x).

The assumption states that the degree of confounding captured on the log-odds ratio scale is stable over time, an assumption first considered by Park and Tchetgen Tchetgen,[2] which they refer to as odds ratio equi-confounding. Under this assumption, it follows that $f\left(Y_1^{a=0}=y|A=1,X=x\right)\propto h_1(y,x)exp\{\beta_0(y,x)\}$, establishing nonparametric identification of $f\left(Y_1^{a=0}=y|A=1,X=x\right)$ in the sense that $\beta_0(y,x)$, $f(Y_0=y|A=0,X=x)$, and $f(Y_1=y|A=0,X=x)$ are unrestricted. Furthermore, one may identify the additive ATT with the expression:

$$\psi_{ATT}=E(Y_1|A=1)-E\left[\frac{E[Y_1 exp\{\beta_0(Y_1,X)\}|A=0,X]}{E[exp\{\beta_0(Y_1,X)\}|A=0,X]}|A=1\right] \tag{4}$$

as we demonstrate in the eAppendix (http://links.lww.com/EDE/C80).

We briefly review examples of special interest where universal DiD yields intuitive alternatives to DiD when parallel trends is violated. Suppressing X, the ATT satisfies:

$$\psi_{ATT}=\underbrace{\{E(Y_1|A=1)-E(Y_1|A=0)\}}_{\text{Crude estimand}}-\underbrace{\{E\left(Y_1^{a=0}|A=1\right)-E(Y_1|A=0)\}}_{\text{Debiasing term}}. \tag{5}$$

Here, the crude estimand is the difference between the conditional means of the outcome in treated and control groups in the postexposure period. The debiasing term is the difference between the ATT and the crude estimand, reflecting unmeasured confounding bias. If there is no unmeasured confounding the bias term vanishes and, consequently, the crude estimand identifies the ATT. Otherwise, suppose that $Y_t^{a=0}|(A=a)\sim N\left(\mu_t(a),\sigma_t^2\right)$. Then, the parallel trends and odds ratio equi-confounding assumptions in this model imply that the bias term is equal to $\mu_1(1)-\mu_1(0)=\mu_0(1)-\mu_0(0)$ and $\sigma_1^{-2}\{\mu_1(1)-\mu_1(0)\}=\sigma_0^{-2}\{\mu_0(1)-\mu_0(0)\}$, respectively, and the corresponding debiasing terms are as follows:

$$\begin{aligned}&(\text{Parallel trends})\\ &\Rightarrow\ \text{Debiasing term}=\mu_0(1)-\mu_0(0),\\ &(\text{Odds ratio equi-confounding})\\ &\Rightarrow\ \text{Debiasing term}=\frac{\sigma_1^2}{\sigma_0^2}\{\mu_0(1)-\mu_0(0)\}.\end{aligned}$$

Therefore, universal DiD reduces to standard DiD if $\sigma_1^2=\sigma_0^2$, that is, the scale of the outcome at $t=0$ matches that at $t=1$; however, universal DiD is more flexible than DiD in the sense that if $\sigma_1^2\neq\sigma_0^2$, universal DiD rescales the standard DiD debiasing term $\mu_0(1)-\mu_0(0)$ of the crude estimate to account for a potential difference of scales between $t=0$ and $t=1$. Additionally, the eAppendix (http://links.lww.com/EDE/C80) provides analogous comparisons on multiplicative and odds ratio scales.

At this juncture, it is instructive to consider a structural model for odds ratio equi-confounding analogous to (DiD Model), which we adopt following.[2] Suppressing covariates, consider the following model:

$$\begin{aligned}&Y_t^{a=0}\perp\!\!\!\perp A|U_t,\ t=0,1,\\ &A|(U_1=u)\overset{D}{=}A|(U_0=u),\ \forall u,\\ &U_1|(A=0,Y_1=y)\overset{D}{=}U_0|(A=0,Y_0=y),\ \forall y.\end{aligned} \tag{Universal DiD Model}$$

In (Universal DiD Model), the relationship between $Y_t^{a=0}$ and U_t is unrestricted, while the exposure mechanism of A given U_t is assumed not to depend on time. In addition, (Universal DiD Model) assumes that the conditional distribution of U_t evaluated at u given $(A=0,Y_t)$ is stable over time but otherwise unrestricted. Unlike (DiD Model), (Universal DiD Model) is scale-invariant in that any monotone transformation of an outcome that satisfies (Universal DiD Model) remains in the model. Of note, unlike (DiD Model), (Universal DiD

Model) is agnostic about the presence of an additive interaction between the treatment and U_t. In the eAppendix (http://links.lww.com/EDE/C80), we establish that (Universal DiD Model) implies odds ratio equi-confounding. There, we also describe alternative data generating processes for odds ratio equi-confounding that may be of independent interest.

For estimation and inference, one may posit GLMs for the outcome process by specifying parametric models for $h_t(y,x)$ and $\beta_t(y,x)$. The parameters for these functions can be estimated by standard maximum likelihood theory, which can be easily performed using off-the-shelf software, such as geex package in R.[7] As this approach essentially amounts to methods previously described by Wooldridge[8] and Taddeo et al[9] details are relegated to the eAppendix (http://links.lww.com/EDE/C80). An alternative semiparametric approach that obviates the need to specify a likelihood for the outcome process is given next.

Universal DiD Estimation via Extended Propensity Score Weighting

In many real-world applications, GLMs for the outcome can be misspecified, particularly when the distribution of the outcome, such as zero-inflated or truncated outcomes, poses significant challenges for maximum likelihood estimation approaches developed under standard GLM framework. To resolve this issue, we provide an alternative approach that uses an extended propensity score model. The approach generalizes the standard propensity score model for the treatment[10] to accommodate unmeasured confounding by incorporating the treatment-free potential outcome in the propensity score model.

The approach is motivated by the invariance property of odds ratios, which provides the following alternative interpretation of $\beta_t(y,x)$ to that given in equation (3), namely

$$\beta_t(y,x) = log\frac{\left\{\begin{array}{l} Pr\left(A=1|Y_t^{a=0}=y,X=x\right) \\ \times Pr\left(A=0|Y_t^{a=0}=0,X=x\right)\end{array}\right\}}{\left\{\begin{array}{l} Pr\left(A=0|Y_t^{a=0}=y,X=x\right) \\ \times Pr\left(A=1|Y_t^{a=0}=0,X=x\right)\end{array}\right\}}$$

$$\Leftrightarrow \quad log\frac{\pi_t(y,x)}{1-\pi_t(y,x)} = \delta_t(x) + \beta_t(y,x),$$

$$\delta_t(x) = log\frac{Pr\left(A=1|Y_t^{a=0}=0,X=x\right)}{Pr\left(A=0|Y_t^{a=0}=0,X=x\right)}.$$

where the function $\pi_t(y,x) = Pr\left(A=1|Y_t^{a=0}=y,X=x\right)$ is referred to as the extended propensity score function. This alternative obviates the need for specification of a model for $h_t(y,x)$. Specifically, under odds ratio equi-confounding, one can posit a model for the log-odds ratio functions as $\beta_1(y,x) = \beta_0(y,x) = \alpha_0^\intercal S_0(y,x)$ where $S_0(y,x)$ is a user-specified sufficient statistic for the confounding odds ratio parameter. For instance, let $S_0(y,x) = (y, yx^\intercal)^\intercal$, that is, the log-odds of the extended propensity score have a linear relationship in y given x, with the corresponding parameter α_0. Also, specify a parametric model for $\delta_t(x)$, say $\delta_t(x) = (1,x^\intercal)\eta_t$, with similar interpretation. Then, (η_0, α_0) can be estimated via a standard logistic regression of A on $(1, X, S_0(Y_0, X))$. Next, η_1 can be identified as the solution to the population moment equation:

$$E\left\{\frac{1-A}{1-\pi_1(Y_1,X)}\right\} = 1 \Leftrightarrow$$

$$0 = E\left[\binom{1}{X}\left[(1-A)\exp\left\{(1,X^\intercal)\eta_1 + \beta_0(Y_1,X)\right\} - A\right]\right]. \quad (6)$$

Therefore, η_1 can be estimated using the empirical analogue to the above equation. Propensity score weighting universal DiD estimation of the ATT based on the estimated extended propensity score follows from the following identifying expression:

$$\psi_{ATT} = E(Y_1|A=1) - \frac{E\left[(1-A)\,Y_1 exp\left\{(1,X^\intercal)\eta_1 + \alpha_0^\intercal S_0(Y_1,X)\right\}\right]}{E\left[(1-A)\,exp\left\{(1,X^\intercal)\eta_1 + \alpha_0^\intercal S_0(Y_1,X)\right\}\right]}. \quad (7)$$

A proof of this claim is included in the eAppendix (http://links.lww.com/EDE/C80).

The proposed GLM-based and propensity score weighting universal DiD methods rely on (i) modeling the association between covariate and outcome in a GLM, and (ii) modeling the association between covariate and treatment in the extended propensity score model. However, this might introduce a concern that such modeling might introduce specification bias. To address this, we construct a doubly robust[11–13] estimator for the ATT, in the sense that the estimator is consistent and asymptotically normal if the conditional odds ratio function encoding the association between the treatment-free potential outcome and the treatment is correctly specified conditional on covariates, and either (i) the outcome conditional models for baseline and follow-times, or (ii) the treatment mechanism conditional on the treatment-free potential outcome at its reference value at baseline and follow-up, but not necessarily both, are correctly specified. Details of the doubly robust approach can be found in the eAppendix (http://links.lww.com/EDE/C80). Furthermore, to enhance the robustness of the odds ratio function specification particularly for continuous exposure, a flexible yet simple approach might be to posit a model for a discretized version of the outcome, similar to a histogram estimator of a density. In brief, the approach involves converting the outcome into M dummy variables based on the $100(m/M)$ th percentiles ($m = 1, 2, \ldots, M-1$) of the empirical distribution of pre-exposure outcome values. Using this discretized outcome, one then can specify the odds ratio function in the universal DiD framework; see the eAppendix (http://links.lww.com/EDE/C80).

APPLICATION: ZIKA VIRUS OUTBREAK IN BRAZIL

In 2015, a Zika virus outbreak occurred in Brazil, resulting in more than 200,000 cases by 2016.[14] Zika virus infection during pregnancy can affect fetal brain development and lead to severe brain defects such as microcephaly.[15] Consequently,

TABLE 1. Summary of Data Analysis

Estimator		Statistic		
		Estimate	SE	95% CI
Universal DiD under odds ratio equi-confounding	GLM-based	–1.487	0.340	(–2.153, –0.821)
	Propensity score weighting	–0.831	0.377	(–1.570, –0.091)
	Doubly robust	–0.974	0.342	(–1.645, –0.303)
Standard DiD under parallel trends	GLM-based	–1.171	0.151	(–1.467, –0.875)
	Propensity score weighting	–1.091	0.143	(–1.371, –0.811)
	Doubly robust	–1.124	0.159	(–1.435, –0.813)

Values in "Estimate" column represent the additive ATT of the Zika outbreak on the birth rate within the Pernambuco region, that is, the difference between the observed average birth rate of Pernambuco to a forecast of what it would have been had the Zika outbreak been prevented. Values in "SE" and "95% CI" columns represent the standard errors (SEs) associated with the estimates and the corresponding 95% CIs, respectively. The reported values are expressed as births per 1000 persons.

previous investigators[16,17] conjecture that the Zika virus epidemic might have resulted in a decrease in the birth rate, attributed to individuals' tendency to postpone pregnancy or increase the likelihood of abortion due to the fear of Zika virus-related microcephaly.

We illustrate universal DiD with a reanalysis of a study of the effects of an outbreak of Zika virus on birth rate in Brazil originally published in Taddeo et al.[9] Our study required no ethics approval as the dataset is publicly available and does not contain any personal information. Specifically, we consider 2014 and 2016 as pre- and postexposure periods, respectively, and municipalities in the northeastern state Pernambuco and those in the southernmost state Rio Grande do Sul as study units. According to a report from the Brazilian Ministry of Health,[18] the epidemic was more severe in the northeastern region of Brazil compared with the southern region. In addition, out of 1248 microcephaly cases that occurred in Brazil as of November 28, 2015, 646 (51.8%) cases were reported in Pernambuco,[19] whereas less than 10 cases of Zika-related microcephaly were reported in Rio Grande do Sul.[20] Based on the information, individuals in northeastern states may have been more concerned about Zika virus-related microcephaly than those in southern states, which could have contributed to a possible decrease in the birth rate in northeastern states. In contrast, individuals in southern states in Brazil may have experienced minimal behavioral changes compared with those in other regions, as these states were least impacted by the Zika epidemic. Moreover, given the substantial geographical separation of over 2000 kilometers between the 2 regions, it is plausible that the behavioral influence stemming from the Zika virus outbreak in Pernambuco had limited spillover effects on the population in Rio Grande do Sul; see Taddeo et al[9] for a related discussion. Therefore, we categorize Pernambuco as the treated group and Rio Grande do Sul as the control group.

As the pre- and postexposure outcomes, we use birth rates in 2014 and 2016, respectively, where the birth rate is defined as the total number of live births per 1000 persons. We treat birth rate as a normally distributed variable in the parametric GLM formulation. We focus on 673 municipalities with complete data on the pre- and postexposure outcomes and treatment, where 185 municipalities belong to Pernambuco and 488 to Rio Grande do Sul. To further address variation across municipalities due to population differences, we further adjusted for population size, population density, and proportion of females as covariates. The crude mean difference $E(Y_1|A=1)-E(Y_1|A=0)=3.384$ births per 1000 persons, indicating that Pernambuco showed a higher birth rate than Rio Grande do Sul in 2016 despite the Zika virus outbreak.

Based on the proposed approach, we obtained estimates of the additive ATT; see the eAppendix (http://links.lww.com/EDE/C80) for details on the specific steps used for estimation. The baseline densities for the outcomes are specified to follow normal distributions $Y_t^{a=0}|(A=0,X=x)\sim N(\mu_t(x),\sigma_t^2)$ where $\mu_t(x)$ is specified as $\mu_t(x)=(1,x^\intercal)\tau_t$. Therefore, under an odds ratio parametrization, the odds ratio function is represented as $\beta_0(y,x)=(y,yx^\intercal)\alpha_0$. Maximum likelihood estimators of $(\tau_0,\tau_1,\sigma_0^2,\sigma_1^2,\alpha_0)$ are obtained by maximizing the log-likelihood function of the pre-exposure data, i.e., (Y_0,A,X), and the postexposure data under control, i.e., $(Y_1,A=0,X)$, implemented with geex.[7] The extended propensity score is specified as $\pi_t(y,x)/\{1-\pi_t(y,x)\}=\exp\{\delta_t(x)+\beta_0(y,x)\}=\exp\{(1,x^\intercal)\eta_t+(y,yx^\intercal)\alpha_0\}$. Again, maximum likelihood estimators of (η_0,α_0) are obtained by maximizing the log-likelihood function of the pre-exposure data using the same software. We can then estimate η_1 by solving the empirical analogue of (6) with the estimated odds ratio function using the same software. Using these specifications, we obtain six estimates from GLM-based, propensity score weighting, and doubly robust universal DiD approaches. We compare these estimates to those derived under parallel trends obtained from att_gt function implemented in did R package.[21]

The Table 1 summarizes the data analysis results. The GLM-based and propensity score weighting universal DiD estimates show the largest and smallest effect estimates of -1.487 and -0.831 births per 1000 persons, respectively. In addition, effect estimates under parallel trends are of a similar value as those obtained from the universal DiD approaches, and the corresponding confidence intervals (CI) overlap with each other. Compared with the crude estimate of 3.384 births

per 1000 persons, the negative effect estimates suggest the presence of substantial confounding bias. This analysis provides compelling evidence that the Zika virus outbreak led to a decline in the birth in Brazil, corroborating similar findings in the literature,[9,16,17] and further indicating that parallel trends and odds ratio equi-confounding estimates are of similar magnitude (noting that CI are wider for universal DiD under odds ratio equi-confounding than for standard DID under parallel trends) and thus, estimates of the magnitudes of effect are not particularly sensitive to the specific nature of equi-confounding assumption used for causal identification.

In the eAppendix (http://links.lww.com/EDE/C80), we provide additional data analysis results for a more flexible discretized odds ratio function. The universal DiD estimates using discretized odds ratio are much closer to each other than those reported in the Table 1, demonstrating less model dependence, with substantially tighter CI.

Additional sensitivity analyses inspecting the extent to which empirical findings are sensitive to violation of identifying assumptions are given in the eAppendix (http://links.lww.com/EDE/C80).

DISCUSSION

In this article, we have described universal DiD as an alternative to standard DiD that can accommodate outcomes of any type and causal effect estimands possibly defined on nonlinear scales. For universal DiD inference, we have described three alternative approaches targeting the average effect of treatment on the treated, the first involves modeling the pre- and postexposure outcome process, while the second involves positing a model for the extended propensity score, and the third carefully combines both approaches to produce an estimator which possesses a desirable double robustness property of remaining unbiased for the treatment effect, if either outcome model or treatment model is correctly specified.

While standard DiD relies for validity on parallel trends, the validity of universal DiD invokes an assumption of odds ratio equi-confounding, that the degree of confounding bias encoded with an odds ratio association between the treatment and the treatment-free potential outcome at follow-up is exactly equal to that with the pre-exposure outcome. Realistically, this assumption might not hold exactly in all applications but can often be expected to be approximately correct if the time between the pre-exposure and postexposure outcomes is not too large, so that, though changing, the magnitude of confounding bias may be expected to evolve smoothly over time at a relatively slow rate. From this perspective, similar to unconfoundedness (a structural assumption that rules out unmeasured confounding), odds ratio equi-confounding can be logically understood as a structural conditional independence assumption about the distribution of U_t over time while accommodating the potential for unmeasured confounding. This interpretation provides a natural anchoring from which a sensitivity analysis can be initiated. Specifically, one might entertain a sensitivity analysis to the odds ratio equi-confounding assumption in which the odds ratio function $\beta_1(y,x)$ is set to $\beta_1(y,x) = \beta_0(y,x) + \Delta(y,x)$ where $\Delta(y,x)$ is a user-specified nonidentifiable sensitivity function that encodes a potential departure from this assumption. One would then proceed by repeating the proposed analyses and reporting various updated estimates of the causal effect of interest over various choices of Δ, thus providing an evaluation of the sensitivity of inferences to possible violations of the odds ratio equi-confounding condition; see the eAppendix (http://links.lww.com/EDE/C80) for details of such a sensitivity analysis and an application to the Zika study.

Although many DiD methods, including universal DiD, are developed assuming no interference,[22] it is important to acknowledge that interference may be plausible in various applications, particularly in the context of infectious diseases. For example, in our data analysis, interference could have occurred if the Zika virus outbreak in Pernambuco had resulted in substantial changes in behavior among individuals in Rio Grande do Sul. Hence, to properly use DiD methods, it is crucial in practice to assess the possibility of interference in the context in view. In the event that interference becomes a concern, one might consider DiD methods that are explicitly designed to account for interference.[23,24] In addition, given its relevance in epidemiological applications, this article has mainly focused on the additive ATT, and the comparison between odds ratio equi-confounding and parallel trends. Nonetheless, odds ratio equi-confounding has the capacity to accommodate nonlinear treatment effects, such as quantile treatment effects on the treated, as long as the treatment effect in view is uniquely defined as the solution to a moment equation. As a result, comparing odds ratio equi-confounding and identifying assumptions for nonlinear treatment effects in DiD settings could provide useful insights. We refer interested readers to Park and Tchetgen Tchetgen[2] for details.

An important potential generalization of universal DiD concerns settings where richer longitudinal data might be available for each unit. In such settings, one might be able to leverage past outcomes to either validate or relax odds ratio equi-confounding; a possibility we plan to explore in the future. In addition, in panel data settings, staggered treatment initiation might occur, in which case various generalizations of odds ratio equi-confounding might be possible, thus effectively extending recent developments under parallel trends to handle such complex study designs.[25–27]

REFERENCES

1. Caniglia EC, Murray EJ. Difference-in-difference in the time of cholera: a gentle introduction for epidemiologists. *Curr Epidemiol Rep*. 2020;7:203–211.
2. Park C, Tchetgen Tchetgen E. A universal difference-in-differences approach for causal inference. *arXiv*; 2022.
3. Athey S, Imbens GW. Identification and inference in nonlinear difference-in-differences models. *Econometrica*. 2006;74:431–497.

4. Sofer T, Richardson DB, Colicino E, Schwartz J, Tchetgen Tchetgen EJ. On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Stat Sci*. 2016;31:348–361.
5. Chen HY. A semiparametric odds ratio model for measuring association. *Biometrics*. 2007;63:413–421.
6. Tchetgen Tchetgen EJ, Robins JM, Rotnitzky A. On doubly robust estimation in a semiparametric odds ratio model. *Biometrika*. 2010;97:171–180.
7. Saul BC, Hudgens MG. The calculus of M-estimation in R with geex. *J Stat Softw*. 2020;92:10.
8. Wooldridge JM. Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*. 2023;26:C31–C66.
9. Taddeo MM, Amorim LD, Aquino R. Causal measures using generalized difference-in-difference approach with nonlinear models. *Stat Interface*. 2022;15:399–413.
10. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
11. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc*. 1999;94:1135–1120.
12. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med*. 2004;23:2937–2960.
13. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005;61:962–973.
14. Lowe R, Barcellos C, Brasil P, et al. The zika virus epidemic in Brazil: From discovery to future implications. *Int J Environ Res Public Health*. 2018;15:96.
15. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. Zika virus and birth defects–reviewing the evidence for causality. *N Engl J Med*. 2016;374:1981–1987.
16. Castro MC, Han QC, Carvalho LR, Victora CG, França GVA. Implications of Zika virus and congenital Zika syndrome for the number of live births in Brazil. *Proc Natl Acad Sci USA*. 2018;115:6177–6182.
17. Diaz-Quijano FA, Pelissari DM, Chiavegatto Filho ADP. Zika-associated microcephaly epidemic and birth rate reduction in Brazilian cities. *Am J Public Health*. 2018;108:514–516.
18. da Saúde M. *Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vrus zika até a semana epidemiológica*:52; 2017.
19. European Centre for Disease Prevention and Control. *Rapid risk assessment: Zika virus epidemic in the Americas: Potential association with microcephaly and Guillain-Barré syndrome*. *European Centre for Disease Prevention and Control*; 2015.
20. Gregianini TS, Ranieri T, Favreto C, Marion Alves Nunes Z, Tumioto Giannini GL, Sanberg ND. Emerging arboviruses in Rio Grande do Sul, Brazil: chikungunya and zika outbreaks, 2014-2016. *Rev Med Virol*. 2017;27:e1943.
21. Callaway B, Sant'Anna PHC. *did: Difference in differences*. R package version 2.1.2; 2021.
22. Cox DR. *Planning of Experiments*. New York: Wiley; 1958.
23. Clarke D. Estimating difference-in-differences in the presence of spillovers; 2017.
24. Butts K. Difference-in-differences estimation with spatial spillovers. *arXiv*; 2021.
25. Callaway B, Sant'Anna PHC. Difference-in-differences with multiple time periods. *J Econometrics*. 2021;225:200–230.
26. Dukes O, Richardson D, Shahn Z, Tchetgen Tchetgen E. Semiparametric bespoke instrumental variables. *arXiv*; 2022.
27. Shahn Z, Dukes O, Richardson D, Tchetgen Tchetgen E, James R. Structural nested mean models under parallel trends assumptions. *arXiv*; 2022.