NCO-Calibrated DID Analysis: Addressing Unmeasured Confounding in Difference-in-Differences Analyses Using Negative Control Outcomes Experiments

Dazheng Zhang[1,2,*], MS, Bingyu Zhang[1,2,*], MS, Huiyuan Wang[1,2,*], PhD, Charles J. Wolock[1,2], PhD, Yiwen Lu[1,2], Yong Chen[1,2], PhD

[1] The Center for Health AI and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA,

[2] Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA.

[*] Contributed equally.

Background

Difference-in-differences (DiD) analysis is a statistical method used to estimate causal effects by comparing changes in outcomes over time between an intervention group and a control group[1,2]. In political science and biomedical research, DiD analysis has been used to assess the impact of interventions[3,4] or policy changes[5,6].

A critical challenge in DiD analysis is the reliance on the parallel trends assumption. This assumption states that, in the absence of intervention, the intervention and control groups would, on average, have followed a parallel trajectory over time. However, a frequent issue arises when the effect of the unmeasured confounders varies with time[7,8]. Such variation can lead to diverging trends between the control and intervention groups and, if not accounted for, can introduce a systematic bias to the estimated effects of the intervention.

In recent years, negative control outcome (NCO) experiments[9–11], which assume no intervention effect on the NCO, have been used to calibrate the systematic bias such as the unmeasured confounding bias. For example, R package *EmpiricalCalibration* in OHDSI is equipped with such an empirical calibration method to calibrate the systematic bias[12]. However, existing methods for NCO experiments are generally limited to the regression analysis rather than DiD model. Developing methods to incorporate NCOs in DiD analysis is of paramount importance to extend the empirical calibration framework from the OHDSI community.

We propose a paradigm-shifting novel framework for DiD analysis, termed NC-DiD, including a hypothesis testing procedure for possible violation of the parallel trends assumption and a data-driven calibration process utilizing the NCO experiments. To the best of our knowledge, it is the first time to incorporate multiple NCOs within DiD analysis. We applied our method to a study focused on racial/ethnic differences in health outcomes following COVID-19 infection.

Methods

We consider measuring outcomes Y over two time periods, denoted by T, where $T = 0$ represents the pre-intervention period and $T = 1$ represents the post-intervention period. We aim to use the DiD method to estimate the intervention effect τ, which is the average discrepancy in $T = 1$ between the potential outcome under intervention and potential outcome under control, within the group that received the intervention (known as the average treatment effect on the treated). The DiD method analyzes the differences in changes of Y between the intervention group $A = 1$ and the control group

$A = 0$. The DiD model requires a parallel trends assumption. Typically, the assumption of parallel trends is violated, often when the effect of the unmeasured confounders is time-varying. This introduces a systematic bias, denoted b, to the model. To calibrate such systematic bias, we introduce our proposed method, NCO-Calibrated DiD (NC-DiD), which assumes the linearity and transferability between NCO and outcomes of interest.

Figure 1 demonstrates the steps of the proposed NC-DiD method. In the first step, we used the propensity score to match the treatment group to the control group. To implement the DiD method, we apply the log-linear model to the matched cohort:

$$\log(E(Y_i|A_i, T_i)) = \beta_0 + \beta_1 A_i + \beta_2 T_i + \beta_3 A_i T_i,$$

where β_0 is a constant, and β_1, β_2, and β_3 are coefficients of A, T, and their interaction, respectively. Notably, β_3 represents the intervention effect in risk ratio (RR), which may be affected by systematic bias. The estimated $\hat{\beta}_3$ is derived from this model. In the second step, we repeat this procedure using the NCOs, assuming that the intervention does not affect these outcomes. Applying this procedure to the NCOs provides an estimate $\hat{b}$ of the systematic bias. If $b = 0$, this suggests that the parallel trends assumption holds. Based on $\hat{b}$, we derive a test statistic $T = |\hat{b}|/\mathrm{SE}(\hat{b})$ and corresponding two-sided test of the null hypothesis, $H_0: b = 0$. In the last step, we calibrate $\hat{\beta}_3$ by subtracting the estimated bias, yielding the calibrated estimator $\hat{\tau} = \hat{\beta}_3 - \hat{b}$.

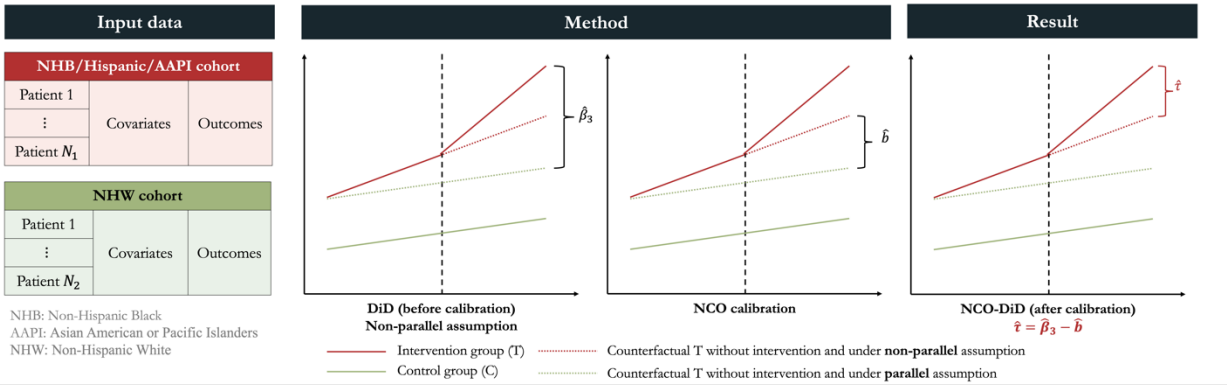

Figure 1. Workflow for the proposed method NC-DiD. It contains three steps. The first step is to implement the DiD method without the calibration and to obtain the estimated $\hat{\beta}_3$ of the intervention effect. The second step is to conduct the negative control outcome experiments and to estimate the systematic bias b. The third step is to calibrate the process and to obtain the calibrated intervention effect $\hat{\tau}$.

Results

Post-acute sequelae of SARS-CoV-2 infection (PASC), include symptoms like fatigue, shortness of breath, and cognitive dysfunction persisting beyond 28 days after the initial infection. We investigate whether the infection of SARS-CoV-2 increases racial/ethnic differences in the PASC symptoms and conditions. We additionally grouped the PASC symptoms and conditions into systematic (conditions) and syndromic (symptoms) categories[13]. We defined documented moderate to severe COVID-19 infection as cases where patients had positive PCR/antigen tests or a COVID-19 diagnosis, accompanied by moderately severe conditions such as gastroenteritis, dehydration, and pneumonia, or severe conditions requiring ICU admission or mechanical ventilation[14]. To illustrate our proposed method, we utilize a synthetic dataset including any patients under the age of 21 who had at least one documented moderate or severe COVID-19 infection. A total of 15,373 children and adolescents were included. The NCO is selected by a literature

review process.

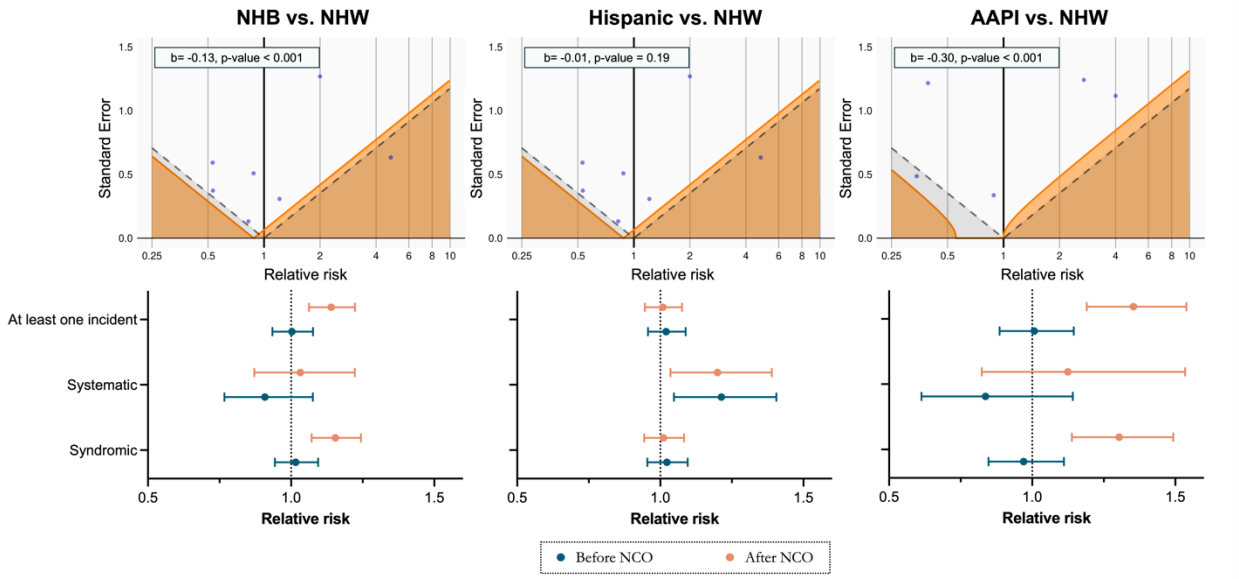

Figure 2. Results for the scientific question of investigating the racial/ethnic disparities after the COVID-19 infection. (a) showed the systematic bias of the NHB, Hispanic, and AAPI compared with NHW. (b) demonstrated the calibrated results of the NC-DiD.

Figure 2 showed the test of parallel trends assumption for NHB versus NHW was rejected (p-value <0.001), and the estimated systematic bias was -0.13. Similarly, the estimated systematic bias for AAPI versus NHW and Hispanic versus NHW were -0.30 (two-sided test p-value <0.001) and 0.01 (p-value =0.19), respectively. We also show the results from the DiD method (before calibration) and the proposed method (after calibration). For example, among the AAPI group, we observe significant evidence of racial/ethnic differences due to COVID-19 infection after calibration (RR 1.35, 95% confidence interval (CI) 1.19 to 1.54). Interestingly, the results before calibration show moderate, though not significant, evidence for a difference in the same direction (RR 1.01, 95% CI 0.89 to 1.15). Results before calibration show a greater association though not statistically significant Comparing the NHB and NHW groups, we found evidence of racial/ethnic differences for any visits to PASC symptoms and conditions after the COVID-19 infection by using the proposed calibration method. Specifically, the RR for the prevalence of any visits with PASC symptoms and conditions after the COVID-19 infection is 1.14 for NHB compared to NHW (95% CI 1.06 to 1.22).

Conclusion

We present a novel framework that directly addresses the significant challenge in DiD analyses, that is, the parallel trends assumption may not hold due to the impact of time-varying unmeasured confounders. Through rigorous testing and calibration methods incorporated in our NC-DiD framework, we demonstrate that our approach not only identifies when the parallel trends assumption does not hold but also adjusts for this, providing a more accurate estimation of causal effects.

References

1. Callaway B, Sant'Anna PHC. Difference-in-differences with multiple time periods. *J Econom* 2021; 225: 200–230.
2. Abadie A. Semiparametric difference-in-differences estimators. *Rev Econ Stud* 2005; 72: 1–19.
3. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA* 2014; 312: 2401–2402.
4. Wing C, Simon K, Bello-Gomez RA. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health* 2018; 39: 453–469.
5. Hansen SW. Polity Size and Local Political Trust: A Quasi-experiment Using Municipal Mergers in D enmark. *Scan Polit Stud* 2013; 36: 43–66.
6. Neureiter M. Evaluating the effects of immigrant integration policies in Western Europe using a difference-in-differences approach. *J Ethn Migr Stud* 2019; 45: 2779–2800.
7. Pazzagli L, Linder M, Zhang M, et al. Methods for time-varying exposure related problems in pharmacoepidemiology: an overview. *Pharmacoepidemiol Drug Saf* 2018; 27: 148–160.
8. Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *J Clin Epidemiol* 2017; 87: 23–34.
9. Schuemie MJ, Ryan PB, DuMouchel W, et al. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med* 2014; 33: 209–218.
10. Schuemie MJ, Hripcsak G, Ryan PB, et al. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Sciences* 2018; 115: 2571–2577.
11. Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *The Lancet* 2019; 394: 1816–1826.
12. OHDSI/EmpiricalCalibration: An R package for performing empirical calibration of observational study estimates, https://github.com/OHDSI/EmpiricalCalibration (accessed 3 October 2024).
13. Rao S, Lee GM, Razzaghi H, et al. Clinical features and burden of postacute sequelae of SARS-CoV-2 infection in children and adolescents. *JAMA Pediatr* 2022; 176: 1000–1009.
14. Forrest CB, Burrows EK, Mejias A, et al. Severity of acute COVID-19 in children< 18 years old March 2020 to December 2021. *Pediatrics* 2022; 149: e2021055765.