Journal of Behavioral Data Science, 2026, *6* (2), 1–12.
DOI: https://doi.org/10.35566/jbds0208

Difference-in-Differences and Treatment × Pretest Interactions: A Critique of Dual-Centered ANCOVA

Byunghoon Jeon[1 [0009−0001−3345−5576]] and
Yongnam Kim[1,2* [0000−0001−6731−7123]]

[1]Department of Education, Seoul National University, Seoul, Republic of Korea
[2]Education Research Institute, Seoul National University, Seoul, Republic of Korea
fineday8545@snu.ac.kr, ykims@snu.ac.kr

Abstract. This paper critically examines dual-centered ANCOVA, proposed by Lin and Larzelere (2020) and Larzelere and Lin (2025), which is claimed to accommodate a Treatment × Pretest interaction while preserving the difference-in-differences (DiD) estimate. We take issue with their claims on four grounds. First, DiD without an interaction term does not assume constant treatment effects; under the parallel trends assumption, it identifies the average treatment effect on the treated even when treatment effects vary across pretest scores. Second, ANCOVA requires correctly modeling the conditional expectation of the outcome in order to identify causal effects, whereas DiD does not, so the concern about omitting an interaction term applies to ANCOVA but not to DiD. Third, dual-centered ANCOVA is not a distinct ANCOVA-type method but merely a re-expression of DiD, sharing the same estimand and identification conditions. Fourth, the claimed innovation of presenting both estimates—the DiD estimate of the treatment effect and the ANCOVA estimate of the interaction term—within a single model offers little practical advantage, because obtaining correct standard errors requires a second analysis. Overall, we argue that dual-centered ANCOVA is based on a misunderstanding of what DiD assumes and is better understood as a re-expression of existing analytic methods rather than a methodological innovation.

Keywords: Difference-in-Differences · Dual-Centered ANCOVA · Parallel Trends Assumption · Effect Heterogeneity · Causal Inference

1 Introduction

In two-group, two-period settings, two approaches have been frequently used to estimate treatment effects: Analysis of Covariance (ANCOVA) and difference-in-differences (DiD). ANCOVA regresses the posttest score on the treatment

indicator while controlling for the pretest score as a covariate. DiD regresses the gain score (i.e., posttest minus pretest) on the treatment indicator. Although both approaches use the same data, they target different estimands, rely on different assumptions (Holland & Rubin, 1983; Kim & Steiner, 2021; Lüdtke & Robitzsch, 2025), and often produce different—even contradictory—treatment effect estimates (Lord, 1967; Lüdtke & Robitzsch, 2025; Maris, 1998).

ANCOVA and DiD have different strengths and limitations. ANCOVA is statistically efficient: adjusting for the pretest reduces residual variance and increases the precision of the treatment effect estimate (Senn, 2006; van Breukelen, 2006). However, in nonrandomized settings, the treatment effect estimate may be biased if pretest adjustment is insufficient to remove all confounding, or if the pretest is measured with error (Allison, 1990; Huitema, 2011; Jamieson, 2004). DiD, by contrast, is typically less efficient but, under the parallel trends assumption, can identify the treatment effect without bias even in the presence of time-invariant confounders (Callaway & Sant'Anna, 2021; Lechner, 2011). In addition, DiD has also been shown to be relatively robust to pretest measurement error, bias amplification, and collider bias (Kim & Steiner, 2021).

As an attempt to combine the strengths of the two approaches, Lin and Larzelere (2020) proposed a method called *dual-centered ANCOVA*, in which both the pretest and posttest scores are centered around the pretest group means before the analysis. They claim that this makes the treatment effect estimate equal to the DiD estimate while retaining an ANCOVA-type regression specification. More recently, Larzelere and Lin (2025) extended this approach and claim that dual-centered ANCOVA can accommodate a Treatment × Pretest interaction term while preserving the DiD estimate.

The purpose of the present paper is to critically examine these claims. We show that their claims are problematic and limited in four respects: (i) DiD does not assume constant treatment effects, contrary to what Larzelere and Lin (2025) maintain; (ii) ANCOVA requires correctly modeling the conditional expectation of the outcome, whereas DiD does not rely on such outcome modeling; (iii) dual-centered ANCOVA is not a distinct method but DiD itself; and (iv) the claimed innovation of displaying both estimates in a single model is undermined by the need for a second analysis to obtain correct standard errors.

The remainder of this paper is organized as follows. Section 2 reviews the claims of Larzelere and Lin (2025), including the relevant ideas from Lin and Larzelere (2020). Section 3 presents our critical examination of their claims. Section 4 concludes the paper.

2 Review of Larzelere and Lin (2025)

Larzelere and Lin (2025) extend their earlier proposal of dual-centered ANCOVA (Lin & Larzelere, 2020) to examine whether the treatment effect varies across levels of the pretest. They show that adding the pretest main effect and the Treatment × Pretest interaction directly to a DiD equation transforms the estimate of treatment effect into the ANCOVA estimate, and they propose dual-

centered ANCOVA as a way to avoid this. More specifically, their argument is as follows. Consider the ANCOVA model,

$$\mathbb{E}[Y \mid A, P] = a_0 + a_1 A + a_2 P,$$

and the DiD model,

$$\mathbb{E}[Y - P \mid A] = b_0 + b_1 A,$$

where A is the treatment indicator, P is the pretest, and Y is the posttest. The treatment effect estimates a_1 and b_1 generally differ when the pretest means are unequal across groups, as is often the case in nonrandomized settings. The basic dual-centered ANCOVA model, presented by Lin and Larzelere (2020), is given by

$$\mathbb{E}[Y - \tilde{P} \mid A, P] = c_0 + c_1 A + c_2 (P - \tilde{P}),$$

where $\tilde{P}$ is the group-specific pretest mean. Because both the pretest and the posttest are centered around $\tilde{P}$, the centered pretest has mean zero in each group, so that the pretest means are equal across groups by construction. Lin and Larzelere (2020) show that the coefficient on treatment in the dual-centered ANCOVA model, c_1, is equal to b_1, the DiD estimate. On this basis, they describe dual-centered ANCOVA as a "modification of quasi-ANCOVA" (p. 138), presenting it as a new method introduced in their study (see their discussion on p. 136), thereby treating it as distinct from DiD, even though it "makes the estimated effect equivalent to the estimate from difference-score analyses [i.e., DiD]" (p. 138).

Building on this earlier proposal, the main claimed contribution of Larzelere and Lin (2025) is an extension designed to test whether the treatment effect varies across pretest levels while preserving the DiD estimate. They show that, when both the pretest main effect and the Treatment × Pretest interaction are added to a DiD model,

$$\mathbb{E}[Y - P \mid A, P] = d_0 + d_1 A + d_2 P + d_3 AP,$$

the resulting equation becomes mathematically equivalent to the corresponding ANCOVA model with the same interaction,

$$\mathbb{E}[Y \mid A, P] = d_0 + d_1 A + (d_2 + 1) P + d_3 AP,$$

with the only difference being the coefficient on the pretest term. They describe this result as showing that the coefficient on treatment, d_1, is transformed from the DiD estimate into the ANCOVA estimate. On this basis, they treat this as a limitation of DiD, namely, that DiD cannot natively accommodate such an interaction without changing the treatment effect estimate. Furthermore, they claim that "without Treatment × Pretest interactions, difference-in-differences are limited to assuming that the estimated treatment effects are *identical* at every pretest score, an untenable assumption without sufficient evidence" (p. 61, emphasis added).

Larzelere and Lin (2025) then present dual-centered ANCOVA as a method that can overcome this limitation. They claim: "The lack of a parallel way to

test Treatment × Pretest interactions in difference-in-differences appears to be a limitation in such analyses, one that can be overcome after centering all data on the pretest group means" (pp. 61–62). The dual-centered ANCOVA model with an interaction term proposed by Larzelere and Lin (2025) is

$$\mathbb{E}[Y - \tilde{P} \mid A, P] = w_0 + w_1 A + w_2(P - \tilde{P}) + w_3 A(P - \tilde{P}). \tag{1}$$

They describe it as an "innovation" (p. 61) and state that "the dual-centered data can be analyzed with ANCOVA to test a Treatment × Pretest interaction in a model duplicating the treatment effect from difference-in-differences" (pp. 52–53), that is, $w_1 = b_1$, and $w_3 = d_3$.

Larzelere and Lin (2025) further claim that "To our knowledge, there is no generally accepted method of testing a Treatment × Pretest interaction within difference-in-differences without changing the main effect of treatment to ANCOVA's estimate" (p. 54), and repeat the same point in their discussion: "We do not know of a better way to test Treatment × Pretest interactions in difference-in-differences" (p. 61). They therefore claim that their proposed dual-centered ANCOVA makes it possible to obtain, within a single model, both estimates—the DiD estimate of the treatment effect and the ANCOVA estimate of the coefficient on the interaction term.

3 Examinations of Dual-Centered ANCOVA

3.1 Does DiD Assume Constant Effects?

As reviewed in Section 2, Larzelere and Lin (2025) treat the possibility of including an interaction term in the regression equation as equivalent to allowing treatment effect heterogeneity, leading them to claim that DiD necessarily assumes a *constant* treatment effect (see their discussion on p. 61). This, however, does not follow.

Admittedly, some studies in the DiD literature assume a constant treatment effect for convenience and simplicity of exposition (e.g., Kim & Steiner, 2021; van Breukelen, 2013), but such an assumption is not necessary for applying DiD. Even if there is no interaction term in the analytic model, DiD does not assume that the treatment effect is the same across pretest scores. As long as the parallel trends assumption holds—together with the standard causal conditions such as the stable unit treatment value assumption (SUTVA)—DiD can identify the ATT even when treatment effects are heterogeneous across units (Lechner, 2011).

To make this concrete, consider the following outcome data-generating process (DGP), which is assumed here to underlie the analytic model used by Larzelere and Lin (2025).[1]

$$Y = \alpha + \beta P + (\gamma_0 + \gamma_1 P) \cdot A + \varepsilon, \quad \varepsilon \perp (A, P), \tag{2}$$

where ε is a structural error term. Let $Y(1)$ and $Y(0)$ denote the potential posttest outcomes under treatment ($A = 1$) and control ($A = 0$), respectively (Holland, 1986; Rubin, 1974). From the DGP in Equation (2), the potential outcomes are

$$Y(1) = \alpha + \beta P + \gamma_0 + \gamma_1 P + \varepsilon,$$
$$Y(0) = \alpha + \beta P + \varepsilon,$$

so that the individual-level causal effect, defined as the difference between the two potential outcomes, is

$$Y(1) - Y(0) = \gamma_0 + \gamma_1 P. \tag{3}$$

When $\gamma_1 \neq 0$, treatment effects are heterogeneous across units, as the effect varies with the level of the pretest score P.

Now consider the DiD estimator, which is the treatment group difference in posttest-pretest differences:

$$\mathbb{E}[Y - P \mid A = 1] - \mathbb{E}[Y - P \mid A = 0].$$

Note that under consistency, $Y = Y(1)$ for units with $A = 1$ and $Y = Y(0)$ for units with $A = 0$ (Holland, 1986; Rubin, 1974). In contrast, because P is measured before treatment, P is unaffected by A, so $P(1) = P(0) = P$. The DiD estimator can then be written as

$$\mathbb{E}[Y(1) - P \mid A = 1] - \mathbb{E}[Y(0) - P \mid A = 0].$$

Adding and subtracting $\mathbb{E}[Y(0) - P \mid A = 1]$ gives

$$\Big\{\mathbb{E}[Y(1) - P \mid A = 1] - \mathbb{E}[Y(0) - P \mid A = 1]\Big\}$$
$$+ \Big\{\mathbb{E}[Y(0) - P \mid A = 1] - \mathbb{E}[Y(0) - P \mid A = 0]\Big\}.$$

The second term is the difference in no-treatment trends between groups. Under the parallel trends assumption, the second term equals zero.[2] Thus, the DiD

[1] It should be noted that Equation (2) imposes no restriction on the relationship between A and P and therefore permits treatment assignment to depend on the pretest, including through $P \to A$. From an identification perspective, DiD does not require the absence of $P \to A$ as a separate condition; rather, it requires the parallel trends assumption to hold under the actual treatment assignment mechanism. Thus, the parallel trends assumption can, in principle, hold even when $P \to A$ (see Footnote 2 for the corresponding mathematical conditions). In practice, however, such dependence may make the assumption less plausible and more difficult to assess (Kim & Steiner, 2021).

[2] Under the DGP specified in Equation (2), the parallel trends assumption is equivalent to $(\beta - 1)\{\mathbb{E}[P \mid A = 1] - \mathbb{E}[P \mid A = 0]\} = 0$. This equality holds if either $\beta = 1$ or $\mathbb{E}[P \mid A = 1] = \mathbb{E}[P \mid A = 0]$. Here, under the stipulated DGP, $\beta = 1$ means

estimator becomes the treated-group average difference in potential outcomes:

$$\begin{aligned}&\mathbb{E}[Y - P \mid A = 1] - \mathbb{E}[Y - P \mid A = 0] \\ &\quad = \mathbb{E}[Y(1) - P \mid A = 1] - \mathbb{E}[Y(0) - P \mid A = 1] \\ &\quad = \mathbb{E}[Y(1) - Y(0) \mid A = 1].\end{aligned}$$

Substituting Equation (3) into the final expression gives

$$\begin{aligned}\mathbb{E}[Y - P \mid A = 1] - \mathbb{E}[Y - P \mid A = 0] &= \mathbb{E}[\gamma_0 + \gamma_1 P \mid A = 1] \\ &= \gamma_0 + \gamma_1 \cdot \mathbb{E}[P \mid A = 1].\end{aligned} \tag{4}$$

Whatever heterogeneity exists in the treatment effect—here captured by $\gamma_1 P$—is automatically averaged over the treated group, whether or not the analyst intends this to happen. Therefore, under the parallel trends assumption, what standard DiD—which, according to Larzelere and Lin (2025), does not include an interaction term—identifies is the *average effect* for the treated group, *not a constant effect* at every value of the pretest. Whether one wishes to model such heterogeneity explicitly with interaction terms is a separate analytic question, which is taken up in the next section.

3.2 ANCOVA Requires Outcome Modeling; DiD Does Not

If DiD does not assume constant treatment effects, then what kind of problem are Larzelere and Lin (2025) actually pointing to? Our argument is that their concern about a Treatment × Pretest interaction is relevant to ANCOVA, but not to DiD. The reason is that ANCOVA requires correctly modeling the conditional expectation of the outcome, whereas DiD does not. The following discussion makes this point more concrete.

As established earlier, under the DGP in Equation (2) and the parallel trends assumption, DiD identifies, as shown in Equation (4),

$$\gamma_0 + \gamma_1 \cdot \mathbb{E}[P \mid A = 1],$$

which is the ATT. Importantly, no interaction term—or any outcome model—is required to arrive at this result.

---

that a one-point increase in P shifts the conditional mean of the untreated potential posttest outcome $Y(0)$ by one point, so that the between-group mean difference in $Y(0)$ equals the between-group mean difference observed at pretest. Alternatively, the parallel trends assumption holds when $\mathbb{E}[P \mid A = 1] = E[P \mid A = 0]$. This condition is satisfied, for example, when treatment assignment is independent of P, as in RCTs. We emphasize that these conditions are not additional identifying assumptions but rather the model-specific algebraic form of the parallel trends assumption under the stipulated DGP; they should not be interpreted as implying either that DiD is generally appropriate whenever $P \to A$ or that $\beta = 1$ is generally plausible in nonrandomized settings.

ANCOVA, by contrast, requires correct outcome modeling. To see this, recall that under the DGP in Equation (2), the pretest-specific treatment effect, $\tau(P)$, is given by Equation (3),

$$\tau(P) = \gamma_0 + \gamma_1 P.$$

If the model correctly reflects the DGP in Equation (2),

$$\mathbb{E}[Y \mid A, P] = a_0 + a_1 A + a_2 P + a_3 AP,$$

it can then estimate $\tau(P)$ at each pretest value, from which the ATE can be recovered via standardization (or g-computation) (Hernán & Robins, 2020; Robins, 1986; Snowden, Rose, & Mortimer, 2011):

$$\mathbb{E}[\tau(P)] = \gamma_0 + \gamma_1 \mathbb{E}[P]. \tag{5}$$

If the ANCOVA model is misspecified by omitting the interaction,

$$\mathbb{E}[Y \mid A, P] = a'_0 + a'_1 A + a'_2 P,$$

the coefficient on treatment, a'_1, generally recovers a weighted average of $\tau(P)$, with weights induced by the residual variation in treatment after adjusting for P, rather than the ATE in Equation (5) (Angrist, 1998; Morgan & Winship, 2015; Słoczyński, 2022). This means that, in ANCOVA, misspecifying the functional form or omitting relevant interactions can cause the treatment effect estimate to miss its causal target, even when the key confounder (i.e., the pretest) is included.

In this sense, whether to add an interaction term is not simply a matter of analytic preference in ANCOVA; it is directly tied to whether the estimator can reach its causal target. This is why it is essential in ANCOVA to examine interaction terms carefully and specify the correct functional form. However, this requirement does not apply to DiD.

3.3 Is Dual-Centered ANCOVA Distinct from DiD?

Lin and Larzelere (2020) characterize dual-centered ANCOVA as a newly introduced ANCOVA-type method, effectively presenting it as distinct from DiD (see also Larzelere & Lin, 2025). However, according to our discussion, although dual-centered ANCOVA appears on the surface to represent a different analysis, from the viewpoint of obtaining the treatment effect estimate, it is in fact DiD itself. We demonstrate this algebraically and through an intuitive illustration.

Dual-centered ANCOVA analyzes Equation (1). Using the notation $Y^* = Y - \tilde{P}$ and $P^* = P - \tilde{P}$, this model can be written as

$$\mathbb{E}[Y^* \mid A, P^*] = w_0 + w_1 A + w_2 P^* + w_3 AP^*.$$

Because P^* is centered within each group by construction,

$$\mathbb{E}[P^* \mid A = 1] = \mathbb{E}[P^* \mid A = 0] = 0. \tag{6}$$

Furthermore, since $A = 1$ and $A = 0$ imply $AP^* = P^*$ and $AP^* = 0$, respectively,

$$\mathbb{E}[AP^* \mid A = 1] = \mathbb{E}[P^* \mid A = 1] = 0 \quad \text{and} \quad \mathbb{E}[AP^* \mid A = 0] = 0. \tag{7}$$

Equations (6) and (7) together imply

$$\mathrm{Cov}(A, P^*) = 0 \quad \text{and} \quad \mathrm{Cov}(A, AP^*) = 0.$$

Because both P^* and AP^* are uncorrelated with A, whether or not they are included as regressors does not affect the coefficient on A. Therefore, the coefficient on treatment in dual-centered ANCOVA, w_1, equals w_1' in the reduced model,

$$\mathbb{E}[Y^* \mid A] = w_0' + w_1' A. \tag{8}$$

Since $P^* = P - \tilde{P}$, we have $\tilde{P} = P - P^*$. Then,

$$Y^* = Y - \tilde{P} = Y - (P - P^*) = Y - P + P^*. \tag{9}$$

Substituting Equation (9) into Equation (8) and using Equation (6) to compute the conditional means for each group gives

$$\begin{aligned}
\mathbb{E}[Y^* \mid A = 1] &= \mathbb{E}[Y \mid A = 1] - \mathbb{E}[P \mid A = 1] = w_0' + w_1', \\
\mathbb{E}[Y^* \mid A = 0] &= \mathbb{E}[Y \mid A = 0] - \mathbb{E}[P \mid A = 0] = w_0'.
\end{aligned}$$

Thus, w_1' is exactly the DiD estimator:

$$\begin{aligned}
w_1' &= \Big\{\mathbb{E}[Y \mid A = 1] - \mathbb{E}[P \mid A = 1]\Big\} - \Big\{\mathbb{E}[Y \mid A = 0] - \mathbb{E}[P \mid A = 0]\Big\} \\
&= \mathbb{E}[Y - P \mid A = 1] - \mathbb{E}[Y - P \mid A = 0].
\end{aligned}$$

As the derivation above shows, it is not merely that the two methods—dual-centered ANCOVA and DiD—yield the same estimate, but that they target the same estimand. Therefore, with respect to the treatment effect estimate, dual-centered ANCOVA is *not* a new ANCOVA-type method, but DiD itself.

The same point can be seen more intuitively by looking at what dual-centering does to the data. Dual-centered ANCOVA analyzes data after subtracting each group's pretest mean from both the pretest and posttest scores of individuals in that group. This means that the treatment-group and control-group line plots in Figure 1(a), which represent the usual DiD setting, are both shifted vertically so that the pretest mean becomes zero in each group, as shown in Figure 1(b). Figure 1(b) corresponds to Equation (8). As a result, the data are transformed into the form shown in Figure 1(b), where the between-group mean difference in Y^* directly corresponds to the DiD estimate. Dual-centered ANCOVA should therefore be understood as DiD itself, conducted after transforming the data so that the two groups share the same pretest mean.

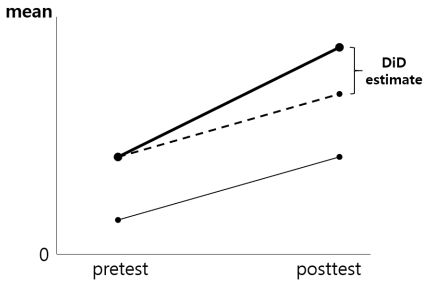

Figure 1(a). Usual DiD estimation: the thick solid line indicates the treatment group, the thin solid line indicates the control group, and the dashed lines represent the parallel trends assumption. The difference shown in the graph is calculated as the DiD estimate.

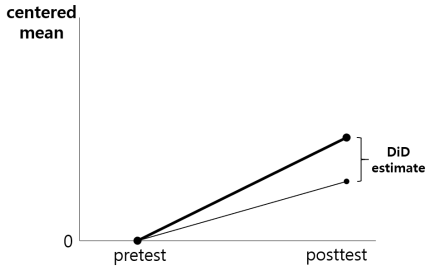

Figure 1(b). Estimation in dual-centered ANCOVA: by transforming the data, the graph in Figure 1(a) is vertically translated as above so that the between-group mean difference in Y^* is directly calculated as the DiD estimate.

3.4 Presenting DiD and ANCOVA Together: How Novel Is That?

Despite the three points raised above, one might still maintain that dual-centered ANCOVA constitutes an "innovation" in the limited sense that it delivers, within a single model, both the DiD estimate of the treatment effect and the ANCOVA estimate of the coefficient on the interaction term (see their discussion on pp. 61–62). However, this claim is undermined on closer examination: obtaining correct standard errors requires a second analysis, regardless of how the regression equation is written.

Brorsen, Lin, and Larzelere (2025) show that group-mean centering creates a generated regressor problem, because the centered pretest term, $P - \tilde{P}$ in Equation (1), is not directly observed but is estimated from the data. As a result, standard statistical packages report a standard error of the treatment effect estimate that is too small. To obtain the correct standard error, one must use either the 2SLS correction or a separate DiD analysis.

Larzelere and Lin (2025) also note this problem; nevertheless, their acknowledgment calls into question the claimed innovation of dual-centered ANCOVA with the interaction term. The correct standard error of the coefficient on the interaction term can be obtained directly from the OLS fit of the dual-centered ANCOVA model, but the correct standard error of the treatment effect estimate still requires either the 2SLS correction or a separate DiD analysis. In other words, although dual-centered ANCOVA appears to deliver both point estimates from a single analysis, obtaining correct inferences in fact requires *two* analyses in practice. At that point, the procedure offers little practical advantage over simply running DiD and ANCOVA separately—the very approach that dual-centered ANCOVA was meant to improve upon. What was claimed as a one-analysis convenience therefore turns out to require two analyses, leaving little that is genuinely novel.

4 Conclusion

Analytic methods should be distinguished by the estimand they target and their identification logic—including the assumptions it requires—rather than by the surface form of the regression equation (Dahabreh & Bibbins-Domingo, 2024; Petersen & van der Laan, 2014). Viewed from this perspective, dual-centered ANCOVA shares the same estimand and identification conditions as DiD—both target the ATT under the parallel trends assumption—and what differs is only the way the regression equation is written. Indeed, Lin and Larzelere (2020) themselves note that "the consistent causal estimates from dual-centered ANCOVA were unbiased only if the original difference-score analysis was unbiased" (p. 143)—which implies that the causal justification of the treatment effect estimate ultimately rests on the identification conditions of DiD. The role of centering, in turn, is not to introduce a new identification strategy. Rather, it is merely an algebraic device that rewrites the regression equation so that the treatment effect estimate already obtainable from DiD appears in an ANCOVA-type regression equation. Dual centering simply alters how the regression equation looks, but not what it identifies.

This is not merely a terminological dispute. Presenting dual-centered ANCOVA as a way of testing Treatment × Pretest interactions *within DiD* risks reinforcing the misconception that DiD itself requires interaction terms for valid causal identification—a misconception evident in Larzelere and Lin (2025)'s claim that DiD without such interactions is "limited to assuming that the estimated treatment effects are identical at every pretest score" (p. 61). Such a view places unnecessary modeling demands on DiD and casts unwarranted doubt on standard DiD estimates that already identify the ATT under heterogeneous treatment effects. Being explicit about what each method—DiD or ANCOVA—requires and what it identifies is therefore not merely a matter of conceptual tidiness, but a safeguard against carrying modeling concerns from one framework into another where they do not apply.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5A8028864). Correspondance should be sent to Yongnam Kim, Department of Education, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea. E-mail: ykims@snu.ac.kr

References

Allison, P. D. (1990). Change Scores as Dependent Variables in Regression Analysis. *Sociological Methodology*, *20*, 93–114. doi: https://doi.org/10.2307/271083

Angrist, J. D. (1998). Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants. *Econometrica*, *66*(2), 249–288. doi: https://doi.org/10.2307/2998558

Brorsen, B. W., Lin, H., & Larzelere, R. E. (2025). Critique of enhanced power claimed for Quasi-ANCOVA and Dual-Centered ANCOVA. *PLOS ONE*, *20*(1), e0317860. doi: https://doi.org/10.1371/journal.pone.0317860

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, *225*(2), 200–230. doi: https://doi.org/10.1016/j.jeconom.2020.12.001

Dahabreh, I. J., & Bibbins-Domingo, K. (2024). Causal Inference About the Effects of Interventions From Observational Studies in Medical Journals. *JAMA*, *331*(21), 1845–1853. doi: https://doi.org/10.1001/jama.2024.7741

Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960. doi: https://doi.org/10.1080/01621459.1986.10478354

Holland, P. W., & Rubin, D. B. (1983). On Lord's Paradox. In H. Wainer & S. Messick (Eds.), *Principals of Modern Psychological Measurement: A Festschrift for Frederic M. Lord* (pp. 3–25). Hillsdale, NJ: L. Erlbaum Associates.

Huitema, B. E. (2011). *The Analysis of Covariance and Alternatives: Statistical Methods for Experiments, Quasi-Experiments, and Single-Case Studies* (2nd ed.). Hoboken, NJ: Wiley. doi: https://doi.org/10.1002/9781118067475

Jamieson, J. (2004). Analysis of Covariance (ANCOVA) with Difference Scores. *International Journal of Psychophysiology*, *52*(3), 277–283. doi: https://doi.org/10.1016/j.ijpsycho.2003.12.009

Kim, Y., & Steiner, P. M. (2021). Gain scores revisited: A graphical models perspective. *Sociological Methods & Research*, *50*(3), 1353–1375. doi: https://doi.org/10.1177/0049124119826155

Larzelere, R., & Lin, H. (2025). An Innovation to Test Treatment X Pretest Interactions within Difference-in-Differences. *Journal of Behavioral Data Science*, *5*(1), 51–66. doi: https://doi.org/10.35566/jbds/larzelere

Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends® in Econometrics*, *4*(3), 165–224. doi: https://doi.org/10.1561/0800000014

Lin, H., & Larzelere, R. E. (2020). Dual-centered ANCOVA: Resolving contradictory results from Lord's paradox with implications for reducing bias in longitudinal analyses. *Journal of Adolescence*, *85*, 135–147. doi: https://doi.org/10.1016/j.adolescence.2020.11.001

Lord, F. M. (1967). A Paradox in the Interpretation of Group Comparisons. *Psychological Bulletin*, *68*(5), 304–305. doi: https://doi.org/10.1037/h0025105

Lüdtke, O., & Robitzsch, A. (2025). ANCOVA versus Change Score for the Anal-

ysis of Two-Wave Data. *The Journal of Experimental Education*, *93*(2), 363–395. doi: https://doi.org/10.1080/00220973.2023.2246187

Maris, E. (1998). Covariance Adjustment versus Gain Scores—Revisited. *Psychological Methods*, *3*(3), 309–327. doi: https://doi.org/10.1037/1082-989X.3.3.309

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (2nd ed.). Cambridge: Cambridge University Press. doi: https://doi.org/10.1017/CBO9781107587991

Petersen, M. L., & van der Laan, M. J. (2014). Causal Models and Learning from Data: Integrating Causal Modeling and Statistical Estimation. *Epidemiology*, *25*(3), 418–426. doi: https://doi.org/10.1097/EDE.0000000000000078

Robins, J. M. (1986). A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect. *Mathematical Modelling*, *7*(9–12), 1393–1512. doi: https://doi.org/10.1016/0270-0255(86)90088-6

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, *66*(5), 688–701. doi: https://doi.org/10.1037/h0037350

Senn, S. J. (2006). Change from Baseline and Analysis of Covariance Revisited. *Statistics in Medicine*, *25*(24), 4334–4344. doi: https://doi.org/10.1002/sim.2682

Słoczyński, T. (2022). Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights. *The Review of Economics and Statistics*, *104*(3), 501–509. doi: https://doi.org/10.1162/rest_a_00953

Snowden, J. M., Rose, S., & Mortimer, K. M. (2011). Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique. *American Journal of Epidemiology*, *173*(7), 731–738. doi: https://doi.org/10.1093/aje/kwq472

van Breukelen, G. J. P. (2006). ANCOVA versus Change from Baseline Had More Power in Randomized Studies and More Bias in Nonrandomized Studies. *Journal of Clinical Epidemiology*, *59*(9), 920–925. doi: https://doi.org/10.1016/j.jclinepi.2006.02.007

van Breukelen, G. J. P. (2013). ANCOVA versus CHANGE from baseline in nonrandomized studies: The difference. *Multivariate Behavioral Research*, *48*(6), 895–922. doi: https://doi.org/10.1080/00273171.2013.831743