

Methodology and Research Practice

The Unfulfilled Promise of Longitudinal Designs for Causal Inference

Wen Wei Loh^{1a}, Dongning Ren^{2,3}

¹ Department of Quantitative Theory and Methods, Emory University, Atlanta, GA, USA, ² Department of Psychology, Emory University, Atlanta, GA, USA, ³ Department of Social Psychology, Tilburg University, Tilburg, The Netherlands

Keywords: Causal diagrams, Collider stratification bias, Lagged regression, Non-equivalent group design, Confounding adjustment

<https://doi.org/10.1525/collabra.89142>

Collabra: Psychology

Vol. 9, Issue 1, 2023

Longitudinal designs are frequently used in psychological research. An intuitive analytic approach is to adjust for previous measurements to bolster the validity of causal conclusions when estimating the effect of a focal predictor (i.e., treatment) on an outcome. This approach is routinely applied but rarely substantiated in practice. What are the implications of adjusting for previous measurements? Does it necessarily improve causal inferences? In this paper, we demonstrate that answers to these questions are far from straightforward. We explain how adjusting for previous measurements can reduce or induce bias in common longitudinal scenarios. We further demonstrate, in scenarios with less stringent causal assumptions, adjusting or not adjusting for previous measurements can induce bias one way or the other. Put differently, adjusting or not adjusting for a previous measurement can simultaneously strengthen and undermine causal inferences from longitudinal research, even in the simplest scenarios. We urge researchers to overcome the unwarranted complacency brought on by using longitudinal designs to test causality. Practical recommendations for strengthening causal conclusions in psychology research are provided.

Causality is central to psychology research. Randomized experiments offer the most persuasive evidence for causality but are often practically unfeasible or unethical. Hence, in many realistic scenarios, researchers turn to longitudinal data to address causal questions. A common practice to fortify causal conclusions when using longitudinal data is to adjust for previous measurements. For example, when estimating the effect of microaggression at time 1 on depression at time 2, a previous measurement of depression at time 1 is often included as a statistical control. In this paper, we raise the question: is this practice valid?

In this article, we demonstrate that the answer to this question is far from straightforward. We clarify the nuanced causal conditions – routinely unspecified or unexamined in practice – for drawing valid causal inferences in longitudinal designs. We argue that longitudinal designs can create an unwarranted complacency for drawing causal inferences, leading to potential pitfalls and erroneous conclusions. By focusing on the intuitive analytic practice of adjusting for previous measurements, we highlight an often-overlooked conundrum in longitudinal designs: adjusting for a variable

can simultaneously strengthen and undermine causal conclusions.

We will draw on concepts from the established causal diagram framework (M. M. Glymour, 2006; Greenland et al., 1999; Lee, 2012) to visualize the causal assumptions and characterize their consequences in a manner accessible to applied researchers.¹ Causal diagrams benefit from relying on readily accessible yet formally rigorous graphical rules for assessing biases due to non-causal associations. Crucially, they make no distributional or functional form assumptions about the statistical relations between the variables: the causal conclusions using causal diagrams are entirely nonparametric and not subject to assumptions of linear regression models (Pearl, 2013). We hope this article will empower researchers investigating causality in longitudinal studies to be more cognizant of the complexities of confounding adjustment, conduct thoughtful examinations of which causal assumptions are likely to hold, and make informed analytic decisions to fortify causal conclusions.

a Correspondence: wenwei.loh@outlook.com

1 A brief summary of causal diagrams is provided in Appendix A.

Nuances of adjusting for previous measurements in longitudinal designs

In longitudinal designs, deciding whether to adjust or not to adjust for² any given variable can be contradictory. We pay particular attention to the routine practice of adjusting for previous measurements. Intuitively, this approach should bolster causal inferences: previous measurements are often either predictive of or share common causes with both treatment and outcome, so pre-existing (or pre-treatment) stable differences in the outcome can be obviated by adjusting for such measures. But the causal assumptions underpinning valid inferences are largely unexamined and routinely overlooked in practice. What are the implications of adjusting for previous measurements when analyzing longitudinal data? In this section, we aim to delve into this inquiry by using typical scenarios encountered in longitudinal studies. As our illustrations will show, the implication of adjusting for previous measurements depends on the specific data-generating causal structures. Routinely adjusting for previous measurements can all too easily lead to incorrect causal inferences.

We consider a minimal example with longitudinal data in two waves. Suppose a researcher is interested in drawing inferences about the causal effect of being a victim of microaggression, such as being treated as irrelevant and invisible (non-randomized treatment X) on the development of depression symptoms (outcome Y). Variables recorded for each participant include relevant baseline time-invariant covariates, such as racial or ethnic identities, socioeconomic status, and unemployment (denoted collectively by C for simplicity), the experience of invisibility at time 1 (X_1), depression symptoms within the same time point or wave at time 1 (Y_1), and a follow-up measure of depression symptoms after a delay (time 2; Y_2).

To illustrate the causal assumptions in this example, we use causal diagrams to visualize plausible data-generating scenarios which cannot be ruled out empirically without imposing additional restrictions using theoretical knowl-

edge.³ In all causal diagrams, we denote participants' experience of perceived invisibility by X and depression symptoms by Y , with subscripts denoting the measurement time point. Measured common causes, such as being a racial or ethnic minority, socioeconomic status, and unemployment, are jointly denoted by C .⁴ We adopt the convention of using a round node to denote a hidden or unmeasured variable. For example, a hidden common cause of the contemporaneously measured X_1 and Y_1 is denoted simply by D . Repeated outcome measurements are likely to be (auto)correlated due to hidden common causes or underlying processes; these are denoted by U .⁵ Because D and U are unmeasured, they are ruled out from adjustment.⁶ Throughout this paper, we focus on the average (total) causal effect of X_1 on Y_2 . Therefore, we seek to close or block all non-causal paths with treatment X_1 and outcome Y_2 as the endpoints.

Adjusting for a previous outcome measurement can eliminate bias

We acknowledge that adjusting for a previous outcome measurement can strengthen causal conclusions. We illustrate this point using two possible scenarios in [Figure 1](#).

In [Figure 1\(a\)](#), within the same wave at time 1, experiencing invisibility (X_1) is correlated with participants' depression symptoms (Y_1) due to a hidden common cause D , such as being denied a promotion opportunity at work. Participants' depression symptoms at time 1 (Y_1) may have an autoregressive effect on their depression symptoms at time 2 (Y_2). Here, we make a stringent assumption that this effect of Y_1 on Y_2 is unconfounded, as indicated by the absence of unmeasured common causes shared by Y_1 and Y_2 . Under this scenario, adjusting for all baseline covariates (e.g., C) and the previous outcome measurement (Y_1) suffices to block all non-causal paths linking X_1 and Y_2 . Therefore, adjusting for the previous measurement Y_1 is necessary for valid causal inference.

We now turn to a different scenario. In [Figure 1\(b\)](#), we relax the causal assumptions regarding the depression

2 We thank Reviewer 1 for encouraging us to be clearer with the term “adjust for.” When estimating the effect between two variables (e.g., X_1 on Y_2), there is a wide variety of techniques by which one may adjust for (or statistically control for, or condition on) some third variable or set of variables (e.g., C or Y_1). These techniques all aim to eliminate non-causal (or “spurious”) associations generated by the latter. Examples of these techniques include outcome regression-based modeling, propensity score-based methods like inverse weighting or matching, stratification, or restriction to a subset with the same value of the covariate(s), among others. Hernán and Robins (2020), Imbens and Rubin (2015), Morgan and Winship (2015), and Rosenbaum (2002) offer book-length presentations of these techniques. Hence, the term “adjust for” in this paper refers to the broad procedure rather than any particular technique.

3 To simplify discussions of causal diagrams in this article, we will consider different measurements of the same variable, such as Y_1 and Y_2 , as distinct variables shown as different nodes (Hernán & Robins, 2020; Pearl, 2009).

4 We further assume no unmeasured common causes of treatment (e.g., X_1) and the later outcome (e.g., Y_2) beyond those included in C (Hernán & Robins, 2020; Imbens & Rubin, 2015; Morgan & Winship, 2015; Pearl, 2009). In practice, a rich selection of baseline common causes can be put together by including relevant covariates based on existing theoretical knowledge and external empirical information or in discussion with subject matter experts (Steiner et al., 2010).

5 While we have assumed C and U to be independent for simplicity, one can readily relax this assumption by further including a directed arrow or a hidden common cause between them in the causal diagram. Nonetheless, the arguments presented in this paper are maintained even when such an additional association is permitted because U is a non-collider on any path linking C via U to either Y_1 or Y_2 .

6 Note that D and U must be assumed as independent for the effect of X_1 on Y_2 to be consistently estimated. If D and U are associated due to an effect or a hidden common cause between them, then it is impossible to consistently estimate the effect of X_1 on Y_2 without adjusting for either D or U , regardless of whether Y_1 is adjusted for or not adjusted for.

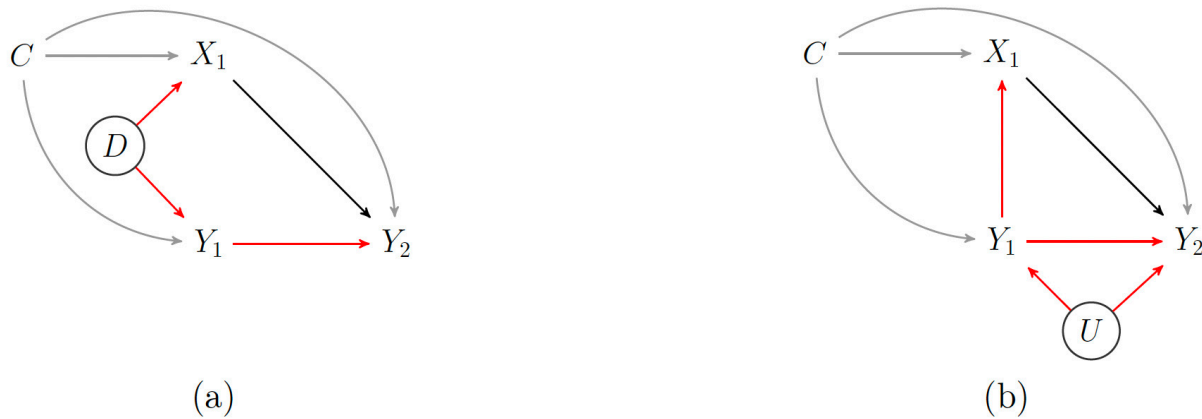


Figure 1. Causal diagrams with longitudinal data in two waves, where adjusting for the previous outcome measurement eliminates confounding bias.

Note. The non-randomized treatment (X_1) and baseline outcome (Y_1) were recorded at wave 1; the final outcome (Y_2) was recorded at wave 2. Subscripts denote the wave the measurements were recorded. The treatment effect on the outcome is drawn in black, while differences between the causal diagrams are marked in red; all other arrows are drawn in gray. Round nodes denote unmeasured or hidden variables. For visual clarity, the observed covariates that confound the treatment-outcome relations are collectively denoted by C and represented by a single node; the effects of (each covariate in) C on the other variables can have different strengths.

symptoms at times 1 and 2. For example, suppose participants' depression symptoms at time 1 (Y_1) may not only affect their depression symptoms at time 2 (Y_2), but they are simultaneously correlated due to unmeasured common causes U , such as limited access to medical care.⁷ Another possible reason why Y_1 and Y_2 are (auto)correlated is that they are instantiations of the same latent process at two different times. But now, suppose that within the same wave at time 1, participants' depression symptoms (Y_1) make them less likely to be engaging conversation partners, which contributes to their risks of being treated as invisible (X_1). This causal directionality – justifiable using theoretical knowledge and by measuring Y_1 before X_1 within the same wave – is indicated by the arrow from Y_1 to X_1 . Furthermore, suppose that this causal effect of Y_1 and X_1 is unconfounded (so that the hidden common cause D can be ruled out). Under this scenario, adjusting for the previous measurement Y_1 is necessary for valid causal inference.

Adjusting for a previous outcome measurement can introduce bias

Adjusting for a previous outcome measurement can be counterproductive and undermine causal conclusions when only a few minor alterations in the data-generating process are made. We illustrate this point using different scenarios in Figure 2.

In Figure 2(a), within the same wave at time 1, experiencing invisibility (X_1) is correlated with participants' depression symptoms (Y_1) due to a hidden common cause D ,

as in Figure 1(a). Here, we assume that an autoregressive effect of Y_1 on Y_2 can be precluded, as indicated by the absence of a $Y_1 \rightarrow Y_2$ arrow. Participants' depression symptoms at time 1 (Y_1) and at time 2 (Y_2) are correlated merely due to unmeasured common cause(s) U , such as limited access to medical care. Under this scenario, adjusting only for C suffices to block all non-causal paths linking X_1 and Y_2 . Crucially, adjusting for the previous measurement Y_1 is counterproductive: doing so opens a non-causal path ($X_1 \leftarrow D \rightarrow Y_1 \leftarrow U \rightarrow Y_2$) and induces “collider (stratification) bias” (Cole et al., 2009; Elwert & Winship, 2014; Greenland, 2003; Griffith et al., 2020). Therefore, the previous outcome measurement Y_1 should not be adjusted for (Foster, 2010; Morgan & Winship, 2015).

We now turn to a different scenario. In Figure 2(b), within the same wave at time 1, experiencing invisibility (X_1) can now be justified to contribute to participants' depression symptoms (Y_1), which in turn influences their depression symptoms at time 2 (Y_2). The depression symptoms at both time points also share unmeasured common causes U , such as limited access to medical care. Under this scenario, adjusting for the previous measurement Y_1 is counterproductive: doing so not only changes the causal effect of interest (to the direct effect that bypasses Y_1 ; Elashoff, 1969) but also induces collider bias in the estimator (Schisterman et al., 2009). For related discussions, see Ananth and Schisterman (2017), Montgomery, Nyhan, and Torres (2018) and Pearl (2016).

⁷ See Newsom (2015, p. 117) for an example in a different context of how an association between Y_1 and Y_2 can be generated by both an autoregressive effect ($Y_1 \rightarrow Y_2$) and other common causes ($Y_1 \leftarrow U \rightarrow Y_2$) simultaneously.

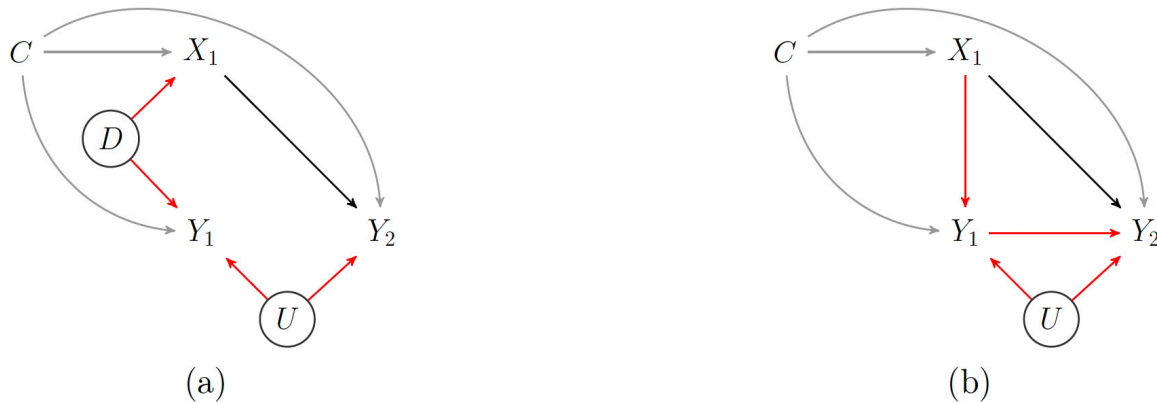


Figure 2. Causal diagrams with longitudinal data in two waves, where adjusting for the previous outcome measurement induces collider bias.

Note. The non-randomized treatment (X_1) and baseline outcome (Y_1) were recorded at wave 1; the final outcome (Y_2) was recorded at wave 2. Subscripts denote the wave the measurements were recorded. The treatment effect on the outcome is drawn in black, while differences between the causal diagrams are marked in red; all other arrows are drawn in gray. Round nodes denote unmeasured or hidden variables. For visual clarity, the observed covariates that confound the treatment-outcome relations are collectively denoted by C and represented by a single node; the effects of (each covariate in) C on the other variables can have different strengths.

Adjusting or not adjusting for a previous outcome measurement can, either way, lead to bias

Using the scenarios presented in Figures 1 and 2, we have explained how it can be clear-cut whether a previous outcome measurement Y_1 should be adjusted for (Figure 1) or not adjusted for (Figure 2) when estimating the effect of X_1 on Y_2 . But things are rarely so simple. The causal diagrams up to this point represent stringent causal assumptions valid only in specific circumstances: respectively, an unconfounded effect of Y_1 on Y_2 (Figure 1(a)); an unconfounded effect of Y_1 on X_1 (Figure 1(b)); no autoregressive effect of Y_1 on Y_2 (Figure 2(a)); and an unconfounded effect of X_1 on Y_1 (Figure 2(b)). Relaxing these causal assumptions – which are empirically untestable in practice – in a longitudinal design can quickly lead to an unavoidable conundrum. We will elaborate on this next.

In Figure 3(a), within the same wave at time 1, experiencing invisibility (X_1) is correlated with participants' depression symptoms (Y_1), similar to Figures 1(a) and 2(a). Furthermore, we impose fewer assumptions about participants' depression symptoms at the two time points. Specifically, not only are participants' depression symptoms at time 1 (Y_1) allowed to have an autoregressive effect on their depression symptoms at time 2 (Y_2), but they are also allowed to simultaneously be autocorrelated due to unmeasured common causes U , such as limited access to medical care – as in Figures 1(b) and 2(b). Under this scenario in

Figure 3(a), Y_1 adopts two conflicting roles simultaneously: It is a non-collider on one path ($X_1 \leftarrow D \rightarrow Y_1 \rightarrow Y_2$) and a collider on another path ($X_1 \leftarrow D \rightarrow Y_1 \leftarrow U \rightarrow Y_2$).⁸ Hence, not adjusting for Y_1 induces confounding bias, but adjusting for Y_1 induces collider bias. Simply put, consistent estimation of the causal effect of X_1 and Y_2 requires simultaneously adjusting for Y_1 and not adjusting for Y_1 (this conundrum is discussed in, e.g., Pearl & Robins, 1995).⁹

This conundrum is not unique to Figure 3(a). In Figure 3(b), suppose within the same wave at time 1, experiencing invisibility (X_1) can be justified to not only contribute to participants' depression symptoms (Y_1) – as in Figure 2(b) – but they also share a hidden common cause – as in Figures 1(a) and 2(a). These causal relations are indicated by the $X_1 \rightarrow Y_1$ and $X_1 \leftarrow D \rightarrow Y_1$ paths, respectively. Hence, for the same reasons as in Figure 2(b), Y_1 should not be adjusted for. But Y_1 is a non-collider on one path ($X_1 \leftarrow D \rightarrow Y_1 \rightarrow Y_2$) and should be adjusted for. Therefore, Y_1 must simultaneously be adjusted for and not be adjusted for to avoid biases when estimating the total effect of X_1 on Y_2 .

Summary

Under this minimal scenario with just two waves, we explained why simple advice to adjust or not for a previous outcome measurement is credible only under strict assumptions about the data-generating process. Following a routine rule-of-thumb to either adjust or not adjust for a

⁸ This setting is a simplification of a causal structure where a covariate is simultaneously a non-collider (in particular, a common cause of treatment and outcome) on one non-causal path and a collider on another non-causal path, such that adjusting for the covariate induces “butterfly bias” (Ding & Miratrix, 2015; Thoemmes, 2015).

⁹ In principle, researchers can determine the relative strengths of associations generated along each non-causal path and seek to minimize the bias. But this is unlikely feasible because it demands intricate knowledge rarely available in practice and is limited to narrow statistical assumptions.

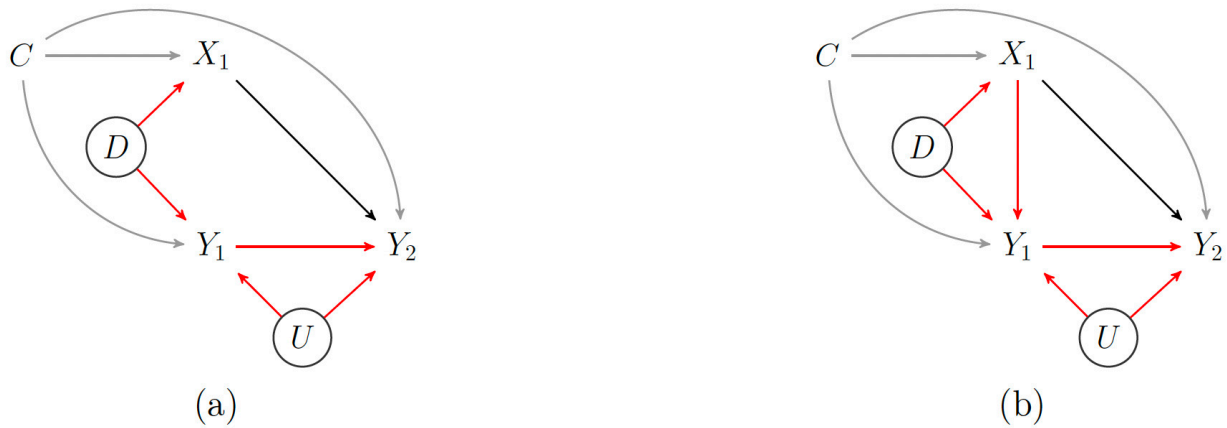


Figure 3. Causal diagrams with longitudinal data in two waves, where adjusting for the previous outcome measurement simultaneously eliminates confounding bias and introduces collider bias.

Note. The non-randomized treatment (X_1) and baseline outcome (Y_1) were recorded at wave 1; the final outcome (Y_2) was recorded at wave 2. Subscripts denote the wave the measurements were recorded. The treatment effect on the outcome is drawn in black, while differences between the causal diagrams are marked in red; all other arrows are drawn in gray. Round nodes denote unmeasured or hidden variables. For visual clarity, the observed covariates that confound the treatment-outcome relations are collectively denoted by C and represented by a single node; the effects of (each covariate in) C on the other variables can have different strengths.

previous measurement can easily lead to inconsistent estimators in other scenarios. Critically, previous measurements can adopt conflicting roles simultaneously on different non-causal paths in longitudinal designs, leading to unavoidable biases.

We focused on simple probative scenarios with just two time points to ease the exposition. But the core issues we have raised apply more generally (exacerbated with more time points and other time-varying measured covariates) that encompass the scenarios described in this paper. Please see Appendix B for an illustration using an example with three waves.

Simulation study

We carried out a Monte Carlo simulation study to empirically demonstrate the biases arising from inappropriate adjustment for previous outcome measurements.¹⁰ We generated data according to each of the six scenarios shown in Figures 1, 2, and 3. We used a sample size of 10,000 to demonstrate that these structural biases were not due to chance associations and persist even in large samples. For simplicity, we assumed no covariates in C . We used lavaan (Rosseel, 2012) to fit two regression models: one regressing Y_2 on X_1 and Y_1 (hence adjusting for Y_1)¹¹, and another regressing Y_2 on X_1 only (hence not adjusting for Y_1). Our

focus was on the (total) effect of X_1 on Y_2 . Results for 100 datasets are displayed in Table 1.

As shown in Table 1, adjusting for Y_1 yielded unbiased estimates only under the data-generating scenarios depicted in Figure 1. Under the scenarios in Figure 2, adjusting for Y_1 yielded severely biased estimates: the estimate could be either smaller or larger than the true effect. Finally, when data was generated under less restrictive assumptions, such as those depicted in Figure 3, adjusting or not adjusting for the previous outcome measurement led to incorrect causal inferences one way or the other. Under such scenarios, the estimate not only deviated from the true effect in magnitude but had opposing signs.

Practical Recommendations

Causal structures in longitudinal data analysis are perpetually unspecified or unexamined in practice. Yet, as we demonstrated in the preceding sections, this step is essential to drawing valid causal conclusions. Routine analytic approaches, such as adjusting for previous measurements, are predicated on stringent causal assumptions. They are, therefore, prone to possibly severe bias when these assumptions are – unbeknownst to the researchers – violated.

The routine advice to improve causality using observational data is to adjust for confounders and merely avoid

¹⁰ The full R (R Core Team, 2021) script with the data-generating process, analysis, and summarizing of the results is available in the Supplemental Online Materials.

¹¹ This analytic method is commonly termed ANCOVA (van Breukelen, 2013) or a basic lagged regression model (Newsom, 2015, p. 107), and is used in non-equivalent (control) group designs (Denny et al., 2023; Reichardt et al., 2023).

Table 1. Average estimates and relative biases of the effect of X1 on Y2, from either adjusting or not adjusting for Y1, under each of the six data-generating scenarios in Figures 1, 2, and 3.

Scenario	Mean estimate		Relative bias (%)	
	Adjusting for Y_1	Not adjusting for Y_1	Adjusting for Y_1	Not adjusting for Y_1
Figure 1(a)	0.30	0.65	0	117
Figure 1(b)	0.30	1.01	0	237
Figure 2(a)	0.04	0.30	-86	-1
Figure 2(b)	0.74	0.30	148	0
Figure 3(a)	0.04	0.65	-86	116
Figure 3(b)	-0.22	1.00	-172	234

Note. The value of the true effect is 0.30. The relative bias is calculated as the ratio (in %) of the bias over the true effect.

mediators and colliders. But as we have demonstrated in this paper, in longitudinal settings¹², this advice is inadequate and potentially misleading. Instead, we encourage researchers to look beyond common causes of treatment and outcome when analyzing longitudinal data. For example, researchers should consider possible common causes of the repeated outcome measurements (such as U in [Figure 3](#)).

We, therefore, cannot offer blanket advice on whether or not to adjust for previous measurements because there is no one-size-fits-all panacea. Researchers seeking to reap the benefits of longitudinal designs for drawing causal conclusions should carefully construct – preferably at the initial stages of a research project – a causal diagram that best represents theoretical knowledge and the underlying data-generating process. For example, recent tutorials offer concrete advice on how to construct and justify a causal diagram in practice (Barnard-Mayers et al., 2022; Digitale et al., 2022; Ferguson et al., 2019; Grosz et al., 2020; Tennant et al., 2020). As shown in this paper, causal diagrams are an excellent research tool that lets researchers “draw their assumptions before their conclusions” (Hernán & Robins, 2020). They are especially beneficial for guiding analytic choices and improving the understanding of possible sources of structural biases.

With a defensible postulated causal diagram in hand, further examining many non-causal paths and determining whether each variable should or should not be adjusted for – as we have done in this paper – can be challenging and seemingly impossible in practice with many waves of data. Consequently, researchers may feel helpless and discouraged when faced with such a proposition. We encour-

age researchers to use the open-source and freely available DAGitty tool (Textor et al., 2017). The DAGitty tool facilitates the crucial task of clarifying and checking posited causal assumptions. To help researchers visualize their postulated causal diagrams using DAGitty, we have provided an example of [Figure B1](#) at <http://dagitty.net/mei2SaP>. Researchers can use this as a starting point to modify and adapt the causal diagram for their unique substantive contexts. Crucially, researchers can then use DAGitty to automatically determine for a focal causal effect whether all non-causal associations can be eliminated by adjusting for a (minimal) subset of covariates. Researchers need not enumerate each path as we have done in this paper. This is achieved using the so-called “back-door criterion” that provides a set of sufficient graphical conditions for determining whether all non-causal paths (specifically “back-door” paths with an arrow pointing to treatment) linking treatment and outcome can be blocked by adjusting for a minimal set of variables (Pearl, 2009, chap. 3).¹³ Continuing our example in [Figure B1](#), a minimal adjustment set for the effect of X_1 on Y_2 is $\{C, U, Y_1\}$; whereas a minimal adjustment set for the effect of X_2 on Y_3 is $\{C, U, X_1, Y_1, Y_2\}$.

We recommend that a best practice is to submit the posited causal diagram (e.g., using DAGitty), and selected covariates for confounding adjustment, for peer review as part of a *Stage 1 Registered Report submission* (Kiyonaga & Scimeca, 2019). This practice utilizes the collective substantive expert knowledge of editors and reviewers to fortify the defensibility of the postulated causal structure and the adequacy of the selected (and omitted) confounders. Using a carefully constructed and rigorously justified causal diagram that clearly and honestly explicates the causal as-

12 In this paper, we focused on a simple scenario where the outcome Y was repeatedly measured, but the treatment X was not. We utilized this simplest probative case to highlight the complexities and nuances of causal inferences in longitudinal data. The issues we raised apply more generally in longitudinal designs with more time points, such as in settings where interest is in estimating reciprocal effects of X and Y over time using cross-lagged panel models (Berry & Willoughby, 2017; Hamaker et al., 2015; Lucas, 2023; Lüdtke & Robitzsch, 2022). But the core causal diagrammatic arguments in this paper apply similarly to such settings that encompass this simple case, which we focused on for expository reasons of being easier to understand. We thank Reviewer 2 for raising this point.

13 More precisely, the back-door criterion for a subset is satisfied if: (i) all back-door paths linking treatment and outcome are closed after adjusting for the subset, and (ii) no variable in the subset is causally affected by treatment (possibly indirectly via a causal path from treatment).

sumptions – before data collection and analysis – can foster more principled causal inferences (Shpitser et al., 2021).

Finally, there are further complications in longitudinal data analysis we did not detail in this article. In psychology research, treatments are often time-varying: e.g., a person may experience invisibility at time 1 but not at time 2. Longitudinal confounders are similarly bound to be affected by earlier treatments. *Treatment-dependent* (variously termed *post-treatment*, *time-varying*, or *treatment-induced*) *confounding* poses severe threats to valid causal inferences of the effects of a time-varying treatment (Daniel et al., 2012; Thoemmes & Ong, 2015). Conventional estimation methods, such as a single regression model for the outcome given all treatments and covariates, cannot avoid undue adjustment for measured post-treatment confounders that induce spurious associations (Rosenbaum, 1984). In the presence of measured time-varying confounding, we recommend researchers utilize the well-established “g-methods” framework (where the “g” stands for “generalized”). G-methods have been recently introduced to the psychology literature (Loh & Ren, 2023a, 2023b, 2023c). This broad class of methods pioneered by James Robins has deep roots in causal inference research and is widely used in (bio)statistics, epidemiology, and medical sciences to assess time-varying treatment effects in longitudinal data when treatment-dependent confounding is present (Clare et al., 2018; Wijn et al., 2022).

Conclusion

Psychology researchers commonly use longitudinal data to answer causal questions. A widely adopted analytic approach is adjusting for previous measurements. But valid causal conclusions rely on stringent causal assumptions routinely unexamined and overlooked in practice. So should previous measurements be adjusted for? We demon-

strate in this paper that the answer is nuanced and far from clear-cut. The advice of simply adjusting for common causes and merely avoiding adjusting for mediators and colliders must be revised, especially in longitudinal data. In particular, we highlight a conundrum in longitudinal designs: the routine analytic practice of adjusting for previous measurements can simultaneously eliminate and introduce non-causal associations, inadvertently leading to an inability to draw valid causal conclusions. We encourage researchers to make informed analytic decisions by conducting thoughtful examination and deliberate reflection of the causal assumptions. With this article, we hope to contribute to ongoing conversations on strengthening causal inferences from longitudinal data in psychological science.

Competing Interests

We have no conflicts of interest to disclose.

Contributions

Contributed to conception: WWL, DR
 Drafted and/or revised the article: WWL, DR
 Approved the submitted version for publication: WWL, DR

Funding

Wen Wei Loh was partially supported by the University Research Committee Regular Award of Emory University, Atlanta, GA.

Submitted: February 24, 2023 PST, Accepted: October 09, 2023 PST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Ananth, C. V., & Schisterman, E. F. (2017). Confounding, causality, and confusion: The role of intermediate variables in interpreting observational studies in obstetrics. *American Journal of Obstetrics and Gynecology*, 217(2), 167–175. <https://doi.org/10.1016/j.ajog.2017.04.016>
- Barnard-Mayers, R., Kouser, H., Cohen, J. A., Tassiopoulos, K., Caniglia, E. C., Moscicki, A.-B., Campos, N. G., Caunca, M. R., Seage, G. R. S., & Murray, E. J. (2022). A case study and proposal for publishing directed acyclic graphs: The effectiveness of the quadrivalent human papillomavirus vaccine in perinatally HIV infected girls. *Journal of Clinical Epidemiology*, 144, 127–135. <https://doi.org/10.1016/j.jclinepi.2021.12.028>
- Berry, D., & Willoughby, M. T. (2017). On the Practical Interpretability of Cross-Lagged Panel Models: Rethinking a Developmental Workhorse. *Child Development*, 88(4), 1186–1206. <https://doi.org/10.1111/cdev.12660>
- Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*, 004912412210995. <https://doi.org/10.1177/00491241221099552>
- Clare, P. J., Dobbins, T. A., & Mattick, R. P. (2018). Causal models adjusting for time-varying confounding—a systematic review of the literature. *International Journal of Epidemiology*, 48(1), 254–265. <https://doi.org/10.1093/ije/dyy218>
- Cole, S. R., Platt, R. W., Schisterman, E. F., Chu, H., Westreich, D., Richardson, D., & Poole, C. (2009). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*, 39(2), 417–420. <https://doi.org/10.1093/ije/dyp334>
- Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G., & Sterne, J. A. C. (2012). Methods for dealing with time-dependent confounding. *Statistics in Medicine*, 32(9), 1584–1618. <https://doi.org/10.1002/sim.5686>
- Deffner, D., Rohrer, J. M., & McElreath, R. (2022). A causal framework for cross-cultural generalizability. *Advances in Methods and Practices in Psychological Science*, 5(3), 251524592211063. <https://doi.org/10.1177/25152459221106366>
- Denny, M., Denieffe, S., & O’Sullivan, K. (2023). Non-equivalent control group pretest–posttest design in social and behavioral research. *The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences*, 314–332. <https://doi.org/10.1017/9781009010054.016>
- Digitale, J. C., Martin, J. N., & Glymour, M. M. (2022). Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology*, 142, 264–267. <https://doi.org/10.1016/j.jclinepi.2021.08.001>
- Ding, P., & Miratrix, L. W. (2015). To adjust or not to adjust? Sensitivity analysis of m-bias and butterfly-bias. *Journal of Causal Inference*, 3(1), 41–57. <https://doi.org/10.1515/jci-2013-0021>
- Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 6(3), 383–401. <https://doi.org/10.3102/00028312006003383>
- Elwert, F. (2013). Graphical causal models. *Handbooks of Sociology and Social Research*, 245–273. https://doi.org/10.1007/978-94-007-6094-3_13
- Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40(1), 31–53. <https://doi.org/10.1146/annurev-soc-071913-043455>
- Ferguson, K. D., McCann, M., Katikireddi, S. V., Thomson, H., Green, M. J., Smith, D. J., & Lewsey, J. D. (2019). Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): A novel and systematic method for building directed acyclic graphs. *International Journal of Epidemiology*, 49(1), 322–329. <https://doi.org/10.1093/ije/dyz150>
- Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, 46(6), 1454–1480. <https://doi.org/10.1037/a0020204>
- Glymour, C. (2001). *The Mind’s Arrows*. The MIT Press. <https://doi.org/10.7551/mitpress/4638.001.0001>
- Glymour, M. M. (2006). Using causal diagrams to understand common problems in social epidemiology. In *Methods in social epidemiology*. (pp. 393–428). Jossey-Bass/Wiley.
- Greenland, S. (2003). Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology*, 14(3), 300–306. <https://doi.org/10.1097/01.ede.0000042804.12056.6c>
- Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. *Epidemiology*, 10(1), 37–48. <https://doi.org/10.1097/00001648-199901000-00008>
- Griffith, G. J., Morris, T. T., Tudball, M. J., Herbert, A., Mancano, G., Pike, L., Sharp, G. C., Sterne, J., Palmer, T. M., Davey Smith, G., Tilling, K., Zuccolo, L., Davies, N. M., & Hemani, G. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature Communications*, 11(1), 5749. <https://doi.org/10.1038/s41467-020-19478-2>
- Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, 15(5), 1243–1255. <https://doi.org/10.1177/1745691620921521>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102–116. <https://doi.org/10.1037/a0038889>
- Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall CRC.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139025751>

- Kiyonaga, A., & Scimeca, J. M. (2019). Practical considerations for navigating registered reports. *Trends in Neurosciences*, 42(9), 568–572. <https://doi.org/10.1016/j.tins.2019.07.003>
- Lee, J. J. (2012). Correlation and causation in the study of personality. *European Journal of Personality*, 26(4), 372–390. <https://doi.org/10.1002/per.1863>
- Loh, W. W., & Ren, D. (2023a). A tutorial on causal inference in longitudinal data with time-varying confounding using g-estimation. *Advances in Methods and Practices in Psychological Science*, 6(3). <https://doi.org/10.1177/25152459231174029>
- Loh, W. W., & Ren, D. (2023b). Estimating time-varying treatment effects in longitudinal studies. *Psychological Methods*. In print. <https://doi.org/10.1037/met0000574>
- Loh, W. W., & Ren, D. (2023c). *G-formula: what it is, why it matters, and how to implement it in lavaan*. <https://doi.org/10.31234/osf.io/m37uc>
- Lucas, R. E. (2023). Why the cross-lagged panel model is almost never the right choice. *Advances in Methods and Practices in Psychological Science*, 6(1), 25152459231158378. <https://doi.org/10.1177/25152459231158378>
- Lüdtke, O., & Robitzsch, A. (2022). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(6), 888–907. <https://doi.org/10.1080/10705511.2022.2065278>
- Moerkerke, B., Loeys, T., & Vansteelandt, S. (2015). Structural equation modeling versus marginal structural modeling for assessing mediation in the presence of posttreatment confounding. *Psychological Methods*, 20(2), 204–220. <https://doi.org/10.1037/a0036368>
- Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), 760–775. <https://doi.org/10.1111/ajps.12357>
- Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.
- Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge. <https://doi.org/10.4324/9781315871318>
- Pearl, J. (2009). *Causality: Models, reasoning and inference*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511803161>
- Pearl, J. (2012). *The causal foundations of structural equation modeling* (pp. 68–91). Defense Technical Information Center. <https://doi.org/10.21236/ada557445>
- Pearl, J. (2013). Linear models: A useful “microscope” for causal analysis. *Journal of Causal Inference*, 1(1), 155–170. <https://doi.org/10.1515/jci-2013-0003>
- Pearl, J. (2016). Lord’s Paradox Revisited – (Oh Lord! Kumbaya!). *Journal of Causal Inference*, 4(2), 20160021. <https://doi.org/10.1515/jci-2016-0021>
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons Ltd.
- Pearl, J., & Robins, J. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 444–453.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reichardt, C. S., Storage, D., & Abraham, D. (2023). Quasi-experimental research. In A. L. Nichols & J. Edlund (Eds.), *The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences* (Vol. 1, pp. 292–313). Cambridge University Press. <https://doi.org/10.1017/9781009010054.015>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, 147(5), 656–666. <https://doi.org/10.2307/2981697>
- Rosenbaum, P. R. (2002). *Observational Studies*. New York: Springer.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*, 20(4), 488–495. <https://doi.org/10.1097/ede.0b013e3181a819a1>
- Shpitser, I., Kudchadkar, S. R., & Fackler, J. (2021). Causal inference from observational data: It is complicated*. *Pediatric Critical Care Medicine*, 22(12), 1093–1096. <https://doi.org/10.1097/pcc.0000000000002847>
- Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, 15(3), 250–267. <https://doi.org/10.1037/a0018719>
- Tate, C. U. (2015). On the overuse and misuse of mediation analysis: It may be a matter of timing. *Basic and Applied Social Psychology*, 37(4), 235–246. <https://doi.org/10.1080/01973533.2015.1062380>
- Tennant, P. W. G., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., Harrison, W. J., Keeble, C., Ranker, L. R., Textor, J., Tomova, G. D., Gilthorpe, M. S., & Ellison, G. T. H. (2020). Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: Review and recommendations. *International Journal of Epidemiology*, 50(2), 620–632. <https://doi.org/10.1093/ije/dyaa213>
- Textor, J., van der Zander, B., Gilthorpe, M. S., Liškiewicz, M., & Ellison, G. T. H. (2017). Robust causal inference using directed acyclic graphs: the R package ‘dagitty’. *International Journal of Epidemiology*, 45(6), dyw341. <https://doi.org/10.1093/ije/dyw341>

- Thoemmes, F. (2015). M-bias, butterfly bias, and butterfly bias with correlated causes – a comment on Ding and Miratrix (2015). *Journal of Causal Inference*, 3(2), 253–258. <https://doi.org/10.1515/jci-2015-0012>
- Thoemmes, F., & Ong, A. D. (2015). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, 4(1), 40–59. <https://doi.org/10.1177/2167696815621645>
- van Breukelen, G. J. P. (2013). ANCOVA versus CHANGE from baseline in nonrandomized studies: The difference. *Multivariate Behavioral Research*, 48(6), 895–922. <https://doi.org/10.1080/00273171.2013.831743>
- Vowels, M. J. (2023). Prespecification of structure for the optimization of data collection and analysis. *Collabra: Psychology*, 9(1), 71300. <https://doi.org/10.1525/collabra.71300>
- Wijn, S. R. W., Rovers, M. M., & Hannink, G. (2022). Confounding adjustment methods in longitudinal observational data with a time-varying treatment: A mapping review. *BMJ Open*, 12(3), e058977. <https://doi.org/10.1136/bmjopen-2021-058977>

Appendices

Appendix A: A brief summary of causal diagrams

Causal diagrams, also known as *graphical causal models* or *causal Directed Acyclic Graphs (DAGs)*, are widely used to represent theorized causal relations and to establish a set of graphical rules sufficient for drawing valid causal inferences. This framework has been extensively introduced and explained elsewhere in the behavioral, health, and social sciences literature; see, e.g., Digitale et al. (2022), Elwert (2013), C. Glymour (2001), M. M. Glymour (2006), Grosz, Rohrer, and Thoemmes (2020), Hernán and Robins (2020), Lee (2012), Moerkerke, Loeys, and Vansteelandt (2015, Figure 2), Morgan and Winship (2015), Pearl (2012), Pearl, Glymour M., and Jewell (2016), and Rohrer (2018). In [Table A1](#),

we summarize basic graph-theoretic language relevant to discussing the issues raised in this article. In this paper, we will assume that a causal diagram can be substantively and defensibly justified as accurately representing the underlying data-generating processes based on established theoretical knowledge and rigorous experimental evidence. Researchers should further exploit information from the measurement of the variables in their study design to support the posited causal structure.¹⁴ For example, the timings can be used to establish temporal precedence and rule out reverse causation based on temporal-logical constraints, and the spacings can be optimized to allow sufficient time for the causal effects to manifest (Cinelli et al., 2022; Deffner et al., 2022; Tate, 2015; Vowels, 2023).

Table A1. Basic terminology for causal diagrams.

Terminology	Description
Node or vertex	Variable (either measured or unmeasured) ^a
Single-headed arrow or uni-directed edge ^b	Causal effect exerted by the variable the arrow emanates from on the variable the arrow enters
Path	Sequence of distinct (i.e., non-recurring) variables connected by arrows pointing in possibly different directions
Causal or directed path	Path with all arrows oriented in the same direction
Non-causal path	Path with at least two arrows pointing in different directions
Open non-causal path	Generates a non-causal (or spurious) association between the endpoints ^c ; variously termed as unblocked, active, or d-connected
Closed non-causal path	Removes the spurious association – generated along this path when it is open – between the endpoints; variously termed as blocked, inactive, or d-separated
Collider on a path	Variable on a path with two arrows pointing directly at it ^d
Collider (stratification) bias	Bias produced when adjusting for a collider on a non-causal path pries the path open ^e

Note. ^aA node may also represent a set of variables all having the same causal relations with other nodes in the causal diagram.

^bThe presence of an arrow permits the possibility of a causal effect of an unknown magnitude that may even be absent empirically; in contrast, the absence of an arrow represents the (more severe) assumption ruling out such a possibility.

^cThe spurious association generated along an open path induces a statistical dependence between the endpoints and renders bias when estimating the causal effect of one endpoint on the other.

^dWe emphasize that a collider is not a variable-specific role but a path-specific role. That is, a variable that is a collider on one path can be a non-collider on another path, with both paths having the same endpoints. Moreover, a collider need not be causally affected by both endpoints on the path.

^eAdjusting for a collider on a path does not necessarily lead to bias. Adjusting for the collider(s) opens a path only if all non-colliders on the same path are unadjusted for. In other words, the path can be closed or blocked by adjusting for a non-collider on the same path.

¹⁴ We thank Reviewer 1 for raising this point.

Appendix B: Biases can be exacerbated when treatment is repeatedly measured

In this section, we explain how the unavoidable biases described above can quickly be exacerbated in longitudinal studies with repeatedly measured treatments and outcomes. To illustrate, we use a slightly more complex example with longitudinal data collected across three waves. Suppose that at time 1, a non-randomized treatment (X_1), an initial measurement of the outcome (Y_1), and baseline time-invariant covariates (C) are recorded. At time 2, the non-randomized treatment (X_2) and the outcome (Y_2) are recorded. At time 3, the outcome (Y_3) is recorded. A causal diagram corresponding to such a setting is shown in [Figure B1](#).

To simplify elucidating the challenges, we will focus on the lag one causal effects of treatment (e.g., X_t) on the outcome at the next wave (e.g., Y_{t+1} , for $t = 1, 2$). In the main text, we discussed the challenges of adjusting for the earlier outcome measurement Y_1 when estimating the effect of a single treatment X_1 on the later outcome Y_2 in a setting with two waves. We now explain how these challenges escalate when the treatment is repeatedly measured by focusing on the effect of the intervening treatment (X_2) on the final outcome (Y_3).

First, note that the same arguments in the previous section can be applied here, with $\{C, X_2, Y_2, Y_3\}$ taking the place of $\{C, X_1, Y_1, Y_2\}$ in [Figure 3\(a\)](#), to realize that Y_2 must simultaneously be adjusted and not adjusted for to block all non-causal paths linking X_2 and Y_3 . Therefore, we will inspect whether the initial treatment and outcome measurements (X_1 and Y_1) should or should not be adjusted for. Non-causal paths linking X_2 and Y_3 via either X_1 or Y_1 are displayed in [Table B1](#). On any given path, X_1 can be either a non-collider or a collider intersecting it; similarly, Y_1 can be either a non-collider or a collider on a path intersecting it. In other words, X_1, Y_1 , and Y_2 must each be simultaneously adjusted for and not adjusted for when targeting the effect of X_2 on Y_3 . In causal diagrammatic terminology, there is no subset of the measured variables $\{C, X_1, Y_1, Y_2\}$ that, when adjusted for, suffices to block all non-causal paths linking X_2 and Y_3 . Therefore, the causal structure in [Figure B1](#) rules out consistent and unbiased estimation of the $X_2 \rightarrow Y_3$ effect.

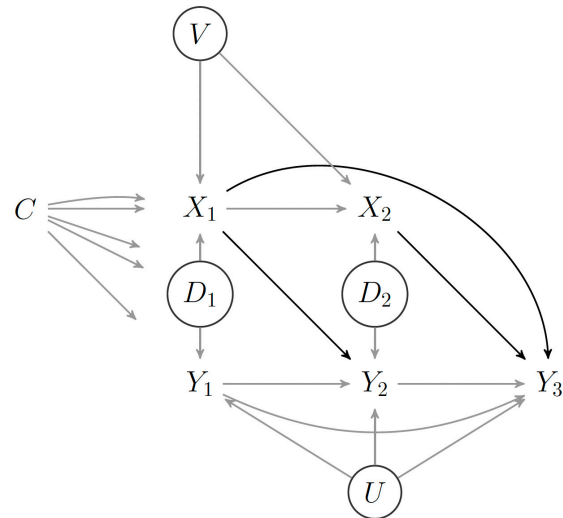


Figure B1. Causal diagram for an example of longitudinal data with three waves.

Note. The non-randomized treatment was recorded at waves 1 (X_1) and 2 (X_2); the outcome was recorded at waves 1 (Y_1), 2 (Y_2), and 3 (Y_3). At each time point $t = 1, 2$, the pair of contemporaneous treatment and outcome measurements (X_t and Y_t) share a hidden common cause (D_t). An unmeasured common cause or process underlying the repeated treatment measurements is denoted simply by V . For the sole purpose of simplifying this illustration of whether or not to adjust for X_1 or Y_1 when targeting the effect of X_2 on Y_3 , the lag one effect of Y_1 on X_2 was assumed to be absent. Descriptions of the nodes and arrows are provided in the caption of [Figure 3](#). The arrows emanating from C to the measured variables have been truncated to reduce visual clutter. This figure is available on DAGitty at <http://dagitty.net/mei2SaP>.

Table B1. Non-causal paths linking X_2 and Y_3 via either X_1 or Y_1 (but not C) in the causal diagram of Figure B1, and whether each of X_1 and Y_1 is a non-collider or a collider on each path.

Non-causal path	X_1	Y_1
$X_2 \leftarrow X_1 \leftarrow D_1 \rightarrow Y_1 \leftarrow U \rightarrow Y_3$	Non-collider	Collider
$X_2 \leftarrow X_1 \leftarrow D_1 \rightarrow Y_1 \leftarrow U \rightarrow Y_2 \rightarrow Y_3$	Non-collider	Collider
$X_2 \leftarrow X_1 \leftarrow D_1 \rightarrow Y_1 \rightarrow Y_2 \leftarrow U \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow X_1 \leftarrow D_1 \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow X_1 \leftarrow D_1 \rightarrow Y_1 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow X_1 \rightarrow Y_2 \leftarrow Y_1 \leftarrow U \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow X_1 \rightarrow Y_2 \leftarrow Y_1 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow X_1 \rightarrow Y_2 \leftarrow U \rightarrow Y_1 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow X_1 \rightarrow Y_2 \leftarrow U \rightarrow Y_3$	Non-collider	N.A.
$X_2 \leftarrow X_1 \rightarrow Y_2 \rightarrow Y_3$	Non-collider	N.A.
$X_2 \leftarrow X_1 \rightarrow Y_3$	Non-collider	N.A.
$X_2 \leftarrow V \rightarrow X_1 \leftarrow D_1 \rightarrow Y_1 \leftarrow U \rightarrow Y_3$	Collider	Collider
$X_2 \leftarrow V \rightarrow X_1 \leftarrow D_1 \rightarrow Y_1 \leftarrow U \rightarrow Y_2 \rightarrow Y_3$	Collider	Collider
$X_2 \leftarrow V \rightarrow X_1 \leftarrow D_1 \rightarrow Y_1 \rightarrow Y_2 \leftarrow U \rightarrow Y_3$	Collider	Non-collider
$X_2 \leftarrow V \rightarrow X_1 \leftarrow D_1 \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_3$	Collider	Non-collider
$X_2 \leftarrow V \rightarrow X_1 \leftarrow D_1 \rightarrow Y_1 \rightarrow Y_3$	Collider	Non-collider
$X_2 \leftarrow V \rightarrow X_1 \rightarrow Y_2 \leftarrow Y_1 \leftarrow U \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow V \rightarrow X_1 \rightarrow Y_2 \leftarrow Y_1 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow V \rightarrow X_1 \rightarrow Y_2 \leftarrow U \rightarrow Y_1 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow V \rightarrow X_1 \rightarrow Y_2 \leftarrow U \rightarrow Y_3$	Non-collider	N.A.
$X_2 \leftarrow V \rightarrow X_1 \rightarrow Y_2 \rightarrow Y_3$	Non-collider	N.A.
$X_2 \leftarrow V \rightarrow X_1 \rightarrow Y_3$	Non-collider	N.A.
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow X_1 \leftarrow D_1 \rightarrow Y_1 \leftarrow U \rightarrow Y_3$	Non-collider	Collider
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow X_1 \leftarrow D_1 \rightarrow Y_1 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow X_1 \rightarrow Y_3$	Non-collider	N.A.
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow Y_1 \leftarrow D_1 \rightarrow X_1 \rightarrow Y_3$	Non-collider	Non-collider
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow Y_1 \leftarrow U \rightarrow Y_3$	N.A.	Non-collider
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow Y_1 \rightarrow Y_3$	N.A.	Non-collider
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow U \rightarrow Y_1 \leftarrow D_1 \rightarrow X_1 \rightarrow Y_3$	Non-collider	Collider
$X_2 \leftarrow D_2 \rightarrow Y_2 \leftarrow U \rightarrow Y_1 \rightarrow Y_3$	N.A.	Non-collider

Note. A variable which is absent on a path is denoted by "N.A."

Supplementary Materials

Response Letter

Download: https://collabra.scholasticahq.com/article/89142-the-unfulfilled-promise-of-longitudinal-designs-for-causal-inference/attachment/184252.pdf?auth_token=xfRb6Av-pA5iB0zR1fIM

Peer Review History

Download: https://collabra.scholasticahq.com/article/89142-the-unfulfilled-promise-of-longitudinal-designs-for-causal-inference/attachment/184253.docx?auth_token=xfRb6Av-pA5iB0zR1fIM
