A Primer on Fixed-Effects and Fixed-Effects Panel Modeling Using R, Stata, and SPSS

Nicolas Sommet[1*] 0000-0001-8585-1274 &

Oliver Lipps[2] 0000-0001-9865-231

[1]LIVES Centre, University of Lausanne, Switzerland

[2]FORS, University of Lausanne, Switzerland

Submission date: November 18, 2024

Word count: 6,275 (excluding references, figures, tables, and boxes)

[*]Corresponding author: Nicolas Sommet, Bâtiment Géopolis, Bureau #5785, Quartier UNIL-Mouline, Université de Lausanne, Switzerland; +41 (0)21 692 32 55; nicolas.sommet@unil.ch

Acknowledgments. This work was funded by a SNSF Ambizione fellowship granted to N. Sommet (grant #PZ00P1_185979).

Conflict of Interest. The authors declare no competing interests.

OSF. https://osf.io/bf32c/?view_only=00477cc083b048ebaabde2ec5cc2f482

Abstract

Fixed-effects modeling is a powerful tool for estimating within-cluster associations in cross-sectional data and within-participant associations in longitudinal data. Although commonly used by other social scientists, this tool remains largely unknown to psychologists. To address this issue, we offer a pedagogical primer tailored for this audience, complete with R, Stata, and SPSS scripts. This primer is organized into three parts. In **PART 1**, we show how fixed-effects modeling applies to clustered cross-sectional data. We introduce the concepts of 'cluster dummies' and 'demeaning,' and provide scripts to estimate the within-school association between sports and depression in a fictional dataset. In **PART 2**, we show how fixed-effects modeling applies to longitudinal data, and provide scripts to estimate the within-participant association between sports and depression over time in a fictional four-wave dataset. In this part, we cover three additional topics. First, we explain how to calculate effect sizes and offer simulation-based sample size guidelines to detect median-sized within-participant effects with sufficient power. Second, we show how to test two possible interactions: between a time-constant and a time-varying predictor and between two time-varying predictors. Third, we introduce three relevant extensions: first-difference modeling (estimating changes from one wave to the next); time-distributed fixed-effects modeling (estimating changes before, during, and after an individual event); and within-between multilevel modeling (estimating both within- and between-participant associations). In **PART 3**, we discuss two limitations of fixed-effects modeling: time-varying confounders and reverse causality. We conclude with reflections on causality in nonexperimental data.

Keywords: *Primer*; *Fixed Effects*; *Fixed-Effects Panel*; *Dummies; Demeaning*; *Cluster; Two-Way Fixed-Effects; Interaction; First Difference; Time-Distributed Fixed-Effects; Within-Between Multilevel; Longitudinal; Causality*

A Primer on Fixed-Effects and Fixed-Effects Panel Modeling Using R, Stata, and SPSS

Fixed-effects modeling is a simple yet powerful tool for estimating: (i) within-cluster associations in clustered cross-sectional data, such as students nested in schools, employees in firms, or residents in countries, and (ii) within-person associations in longitudinal data, where observations are viewed as nested in participants (Allison, 2009). For instance, this analytical tool was recently used to estimate the within-country association between generosity and happiness (Araki, 2023) and the within-person change in personality over the life course (Seifert et al., 2024). While fixed-effects modeling is widely used in economics, sociology, political science, and medical fields, it has yet to gain widespread acceptance in psychology (Bauer & Sterba, 2011; McNeish et al., 2017; Petersen, 2008).[1] However, fixed-effects modeling often proves more practical than popular alternatives such as multilevel modeling, as it removes between-cluster heterogeneity and confounders while remaining within the ordinary least squares (OLS) framework (McNeish & Kelley, 2019). Note that the term "fixed effects" does not carry the same meaning as it does in multilevel modeling.

The siloed nature of social sciences may be one reason why many psychologists are unfamiliar with fixed-effects modeling. Our primer aims to break down this silo by providing a non-technical and clear introduction to fixed-effects modeling. It builds on two fictional examples of clustered and longitudinal data at the crossroads of education, clinical, and sports psychology to guide readers through the key equations, offer concrete explanations, and provide intuitive illustrations. The primer includes scripts for analyzing these datasets using R, Stata, and SPSS. It is organized into three parts. **PART 1** covers the fundamentals of fixed-effects modeling with clustered cross-sectional data; **PART 2** introduces fixed-effects panel modeling and its extensions for longitudinal data; **PART 3** discusses the strengths and limitations of fixed-effects modeling for causality.

[1] One exception may be personality psychology, which often uses fixed-effects panel modeling to examine personality traits in longitudinal designs (e.g., Krämer et al., 2024).

PART 1. Fixed-Effects Modeling Applied to Clustered Cross-Sectional Data

In this section, we show how fixed-effects modeling can be used to estimate within-cluster associations in cross-sectional data. Imagine you have collected a sample of $N = 200$ students across $K = 4$ schools (your four clusters). Students reported the number of hours they spent on sports activities in the previous week (your predictor). Then, they completed a depression screening tool and received a depression score ranging from 0 to 5 (your outcome). Your hypothesis is: "The more students do sports, the less depressed they are."[2]

Here, we go through four modeling strategies designed to pedagogically explain how fixed-effects modeling works and how it can be used to estimate the within-school association between sports and depression: (i) a simple pooled regression, (ii) a regression with cluster dummies, (iii) a regression with demeaned variables, and (iv) a fixed-effects regression with R, Stata, and SPSS. The synthetic clustered dataset, along with the Quarto (R), Stata, and SPSS scripts to run the analyses, are available on the OSF. We encourage you to work with these resources while reading this section. **Table 1** is a glossary of concepts and notation used throughout this section, and **Box 1** provides a concise summary in bullet points in the conclusion.

Space intentionally left blank

---

[2] In this and the subsequent example, we assume that the data meet the regression assumptions (e.g., normality, homoscedasticity). Also, note that the link between exercise and depression is well-established (Noetel et al., 2024), but that our simulated datasets show larger associations for clearer graphical representation.

Table 1. Glossary for Fixed-Effects Modeling with Clustered Cross-Sectional Data (Used to Estimate Within-Cluster Associations)

Concept	Definition	
Cluster	The type of unit in which participants are nested, such as a school, firm, or country.	
Fixed Effects	An estimator that eliminates all between-cluster differences to focus exclusively on within-cluster variations.	
Cluster Dummies	A set of $K - 1$ binary variables that eliminate the differences between K clusters (equivalent to fixed effects).	
Demeaning	Cluster-mean centering: Subtracting the cluster mean from each observation to produce scores relative to this mean.	
Main Notations	i	Subscript denoting participant 1, 2, 3, …, N
	j	Subscript denoting cluster 1, 2, …, K
	α_j	Cluster dummies (representing $K - 1$ dummy variables, that is, the fixed effects)
	$e_{w_{ij}}$	Within-cluster residual (variation for participant i within cluster j that is unexplained by the predictors)
	$\overline{var}_j$	The mean of a variable var_j for a cluster j
	$v\ddot{a}r_{ij}$	A demeaned variable var_{ij} (cluster-mean centered)

Model #1. Simple Pooled Regression

First, you run this simple pooled regression model:

$$depression_{ij} = B_0 + B_1 \times sport_{ij} + e_{ij} \quad (1)$$

…where i = 1, 2, …, 200 participants, j = 1, 2, 3, 4 schools, B_0 is the intercept (the mean depression score for those who did not do sports in the prior week), and e_{ij} is the residual

As shown in **Table 2**, your estimate of interest is B_1 = -0.01, SE = 0.07, p = .913. You may believe it accurately captures the association between doing sports and depression. However, this estimate is biased for two reasons. First, using a simple pooled regression with a clustered dataset *violates the independence assumption*, which requires that residuals associated with different participants be independent of one another (Snijders & Bosker, 2012). In your case, this assumption does not hold because data from students from the same school are likely more similar than data from students from different schools.

Second, using a simple pooled regression with a clustered dataset *fails to isolate the within-cluster associations*, as estimates reflect differences not only among participants in the same clusters but also across different clusters (Enders & Tofighi, 2007). In your case, comparing students from different schools introduces considerable noise. For instance, a school located in a less affluent area may have students who are more susceptible to depression. Similarly, a school that specializes in athletics may have students who are more active in sports. There are various tools for addressing the violation of the independence assumption (e.g., clustered standard errors; Cameron & Miller, 2015) and isolating within-cluster associations (e.g., cluster-mean centering in multilevel modeling; Bell et al., 2018). However, fixed-effects modeling arguably stands out as the simplest solution.

Model #2. Regression With Cluster Dummies

Second, you run the same regression as in Equation 1 but add school dummies:

$$depression_{ij} = B_1 \times sport_{ij} + \alpha_0 + \alpha_1 \times S1vsS2 + \alpha_2 \times S1vsS3 + \alpha_3 \times S1vsS4 + e_{w_{ij}} \quad (2)$$

…where $i = 1, 2, \ldots, 200$ participants, $j = 1, 2, 3, 4$ schools, and $e_{w_{ij}}$ is the within-school residual

In this model, you basically treat school identifiers as a categorical variable. You include $K - 1 = 3$ dummy-coded variables in the equation, with school S1 serving as the arbitrary reference category. This means that: (i) α_0 represents the mean depression score in School S1 when $sport_{i1} = 0$ (the intercept); (ii) α_1 contrasts this score with the intercept in School S2; (iii) α_2 contrasts it with the intercept in School S3; and (iv) α_3 contrasts it with the intercept in School S4. This approach is referred to as the least squares dummy variable (LSDV) model (Brüderl & Ludwig, 2015) and can be written more simply as follows:

$$depression_{ij} = B_1 \times sport_{ij} + \alpha_j + e_{w_{ij}} \quad (2')$$

In the context of fixed-effects modeling, the $K - 1$ dummies are represented by α_j and are referred to as "fixed effects." These dummies are not interpreted; their sole function is to eliminate all unobserved differences between clusters. Simply put, by using school dummies, your model becomes fully saturated at the cluster level, leaving no between-school differences to be explained. As such, if you were to add a school-related variable as an additional predictor (e.g., school size, school resources), its effect could not be estimated since all between-school variation is already accounted for. This addresses the violation of the independence assumption (McNeish & Kelley, 2019).[3]

As shown in **Table 2**, your estimate of interest is now $B_1 = -0.26$, $SE = 0.03$, $p < .001$. In this model, B_1 is uncontaminated by between-cluster differences, as these have been eliminated by the cluster dummies (i.e., the "fixed effects"). Therefore, B_1 is interpreted as the pooled within-school association between doing sports and depression. To be clear, students

[3] This assumes that residuals are normally distributed around the cluster means and are homogeneous diagonal with $e_{ij} \sim N(\mu_j, \sigma^2\mathbf{I})$. This is a plausible assumption in clustered cross-sectional data but not in longitudinal data due to serial correlation. We will revisit this point in **PART 2** when discussing panel-adjusted standard errors.

are only compared to their schoolmates (e.g., students from School S1 are only compared to their peers in School S1), meaning that B_1 reflects the average association for a given school. Following the same principle, the error term is now denoted $e_{w_{ij}}$ and represents the within-school residual. This approach effectively addresses the problem of isolating within-cluster associations. However, while this LSDV model is equivalent to a fixed-effects model, this is not how fixed-effects modeling works in your statistical software, which typically uses a technique known as "demeaning."

Model #3. Regression With Demeaned Variables

Third, you run the same regression as in Equation 1 but first subtract the mean of each school from each individual observation:

$$\left(depression_{ij} - \overline{depression}_j\right) = B_1 \times \left(sport_{ij} - \overline{sport}_j\right) + \left(e_{ij} - \overline{e}_j\right) \quad (3)$$

...where the diacritical lines (the bars) denote the school-specific mean (four possible values)

By subtracting the relevant cluster mean from each individual observation, you generate values that are relative to that cluster mean. For instance: (i) a negative score on the demeaned depression score indicates that the student is less depressed than their school average, (ii) a zero score indicates that they are as depressed as their school average, and (iii) a positive score indicates that they are more depressed than their school average. This procedure is known as demeaning (Wooldridge, 2010) and is similar to cluster-mean centering in multilevel modeling (Bell et al., 2018; Enders & Tofighi, 2007; Hamaker & Muthén, 2020). The equation can be written more simply as follows:

$$\ddot{depression}_{ij} = B_1 \times \ddot{sport}_{ij} + e_{w_{ij}} \quad (3')$$

…where the diacritical marks (the dots) indicate that the variable has been demanded

As illustrated by **Figure 1**, after demeaning your predictor and outcome, each school ends up with an average sport time and depression score of zero. In the left panel, the

uncentered variables suggest that School #1 has higher averages in both sport time and depression (perhaps this is an elite sports school where students feel intense pressure), whereas School #4 has lower averages for both sport time and depression (perhaps this is a school where sports teachers are understaffed). As seen in the other two panels, demeaning aligns the averages for both predictor and outcome across schools, thereby removing all between-school differences, and addressing the violation of the independence assumption.

As shown in **Table 2**, your estimate of interest is $B_1 = -0.26$, $SE = 0.02$, $p < .001$. The coefficient estimate is strictly identical to that of the regression using cluster dummies. The reason is that it is also uncontaminated by between-cluster differences, as these have been eliminated through demeaning. The coefficient estimate is again interpreted as the pooled within-school association between doing sports and depression. Note that the model no longer requires an intercept, as it is effectively zero, and that $e_{w_{ij}}$—which could also be denoted $\ddot{e}_{ij}$—again represents the within-school residual.

However, the standard error—and by extension, the inferential test, confidence interval, and *p*-value—differ from that in the regression using cluster dummies. The reason is that demeaning was performed manually, while standard errors should be adjusted to account for the degrees of freedom consumed by the estimation of cluster means (Judge et al., 1991). As such, these standard errors are incorrect, although the bias is minor since we only have $K = 4$ clusters (as K increases, the discrepancy will become larger, which can lead to erroneous conclusions). This issue will be resolved by your statistical software.

Model #4. Fixed-Effects Regression Using R, Stata, and SPSS

Finally, you run a fixed-effects regression using your favorite software:

R	`summary(plm(depression ~ sport, data = df_clustered, index = "school_id", model = "within")`
Stata	`xtreg depression sport, fe i(school_id)`
SPSS	`UNIANOVA depression BY school_id WITH sport /INTERCEPT=EXCLUDE /PRINT PARAMETER`

…where depression is the outcome, sport the predictor, and school_id the school identifier. R users should install the package plm (Croissant & Millo, 2008)

Technically, both plm() with model = "within" in R and xtreg, fe in Stata use demeaning while appropriately adjusting the standard errors (SPSS does not have a built-in command for fixed-effects modeling, and we used a workaround). Regardless, you can describe your model using either Equation 2′ (using cluster dummy notation) or Equation 3′ (using demeaning notation). As shown in **Table 2**, your estimate of interest is B_1 = -0.26, SE = 0.03, $p < .001$. This finding supports your hypothesis: For each additional hour of sports activity, students in a given school score approximately a quarter of a point lower on the depression screening tool. Importantly, this result eliminates any potential between-school confounders, as the analysis focuses exclusively on within-cluster variation

Strengths and Limitations of Fixed-Effects Modeling For Cross-Sectional Data

Fixed-effects modeling is an elegant and practical approach for estimating the pooled within-cluster association between two variables in clustered cross-sectional data. While psychologists often use multilevel modeling for such analyses, fixed-effects modeling offers several advantages (for a comparison between the two approaches, see Bell & Jones, 2015; McNeish & Kelley, 2019). For instance, unlike multilevel modeling, fixed-effects modeling does not require clusters to be sampled from a representative population of clusters (Bryan & Jenkins, 2015), and it performs well with a small number of clusters (even with K = 4 as in our example; Maas & Hox, 2005; McNeish & Stapleton, 2016). Moreover, while multilevel modeling relies on complex algorithms and optimizers, fixed-effects modeling remains within the OLS framework. This makes it simpler to calculate residual-based effect sizes (Olejnik & Algina, 2003) or conduct 1-1-1 mediation analysis (Hayes, 2013)[4].

[4] A 1-1-1 mediation model involves a predictor, a mediator, and an outcome, all located at the lowest level of analysis (the individual level in our example).

Despite these strengths, fixed-effects modeling has an important limitation: Since it focuses on within-cluster variation, it cannot be used to estimate the main effects of a cluster-related variable. However, we will see in the next part that fixed-effects modeling can be used to investigate whether the within-cluster association between two variables is moderated by a cluster-related variable. For instance, you could examine whether the within-school association between doing sports and depression depends on whether the school is public or private, though a much larger number of clusters would be required (Dieleman & Templin, 2014). As we will also see later, extensions of fixed-effects models, such as the so-called hybrid specification, can be used to estimate both within-cluster and between-cluster effects (Mundlak, 1978).

Finally, concluding that doing sports leads to reduced depression based on your data may be premature. Psychologists often overlook a fundamental assumption in regression: exogeneity (Gardiner et al., 2009). This assumption requires that regressors be uncorrelated with the residual, which is a sophisticated way of saying that your focal predictor must be a true independent variable, unaffected by between-participants confounders (for a discussion, see Maxwell et al., 2017, p. 62). In your case, it is plausible that more athletic students engage in more sports and are inherently less prone to depression, suggesting that the observed relationship documented in your analysis is not causal but is fully explained by a third-variable effect (Pearl, 2009). Addressing such concerns requires longitudinal analysis.

Space intentionally left blank

Table 2. Parameters From the Four Modeling Strategies Presented in PART 1. The table shows results from the simple pooled regression, the regression with cluster dummies, the regression with demeaned variables, and the fixed-effects regression run using your preferred software. The latter three models yield the pooled within-cluster association between the predictor and outcome (highlighted in yellow).

Model #1. Simple Pooled Regression[A]	*B*	*SE*	*p*	
Intercept, B_0	2.65	0.25	< .001	← Mean depression score for students not doing sports
Number of hours devoted to sports, B_1	-0.01	0.07	0.913	← Overall effect for students from similar *and* different schools
Residual (root mean squared error), e_i	1.32			← Overall residual
Model #2. Regression With Cluster Dummies[B]	*B*	*SE*	*p*	
Intercept, α_0	5.40	0.13	< .001	← Mean depression score for students not doing sports in School #1[ø]
Number of hours devoted to sports, B_1	-0.26	0.03	< .001	← **Pooled within-school association between sports and depression**
School dummies (i.e., fixed effects), α_i				
School #1 vs. School #2, α_1	-1.34	0.08	< .001	← Difference between intercepts in School #1 and School #2[ø]
School #1 vs. School #3, α_2	-3.51	0.07	< .001	← Difference between intercepts in School #1 and School #3[ø]
School #1 vs. School #4, α_3	-2.57	0.09	< .001	← Difference between intercepts in School #1 and School #4[ø]
Residual (idiosyncratic errors), $e_{w_{ij}}$	0.33			← Within-school residual
Model #3. Regression With Demeaned Variables[C]	*B*	*SE*	*p*	
Intercept	0.00	0.02	1.000	← Not informative, as demeaning results in a value of zero
Number of hours devoted to sports, B_1	-0.26	0.02	< .001	← **Pooled within-school association between sports and depression**
Residual, $\ddot{e}_{ij}$	0.32			← Within-school residual
Model #4. Fixed-Effects Regression[D]	*B*	*SE*	*p*	
Intercept	n/a	n/a	n/a	← Not estimated, as demeaning results in a value of zero
Number of hours devoted to sports, B_1	-0.26	0.03	< .001	← **Pooled within-school association between sports and depression**
Residual (idiosyncratic errors), $e_{w_{ij}}$(sometimes $\ddot{e}_{ij}$)	0.33			← Within-school residual

Note. Parameters that appear in green represent unbiased fixed-effects model parameters, whereas those in red are biased.

[A] This model is biased because it violates the independence assumption and fails to isolate within-cluster associations

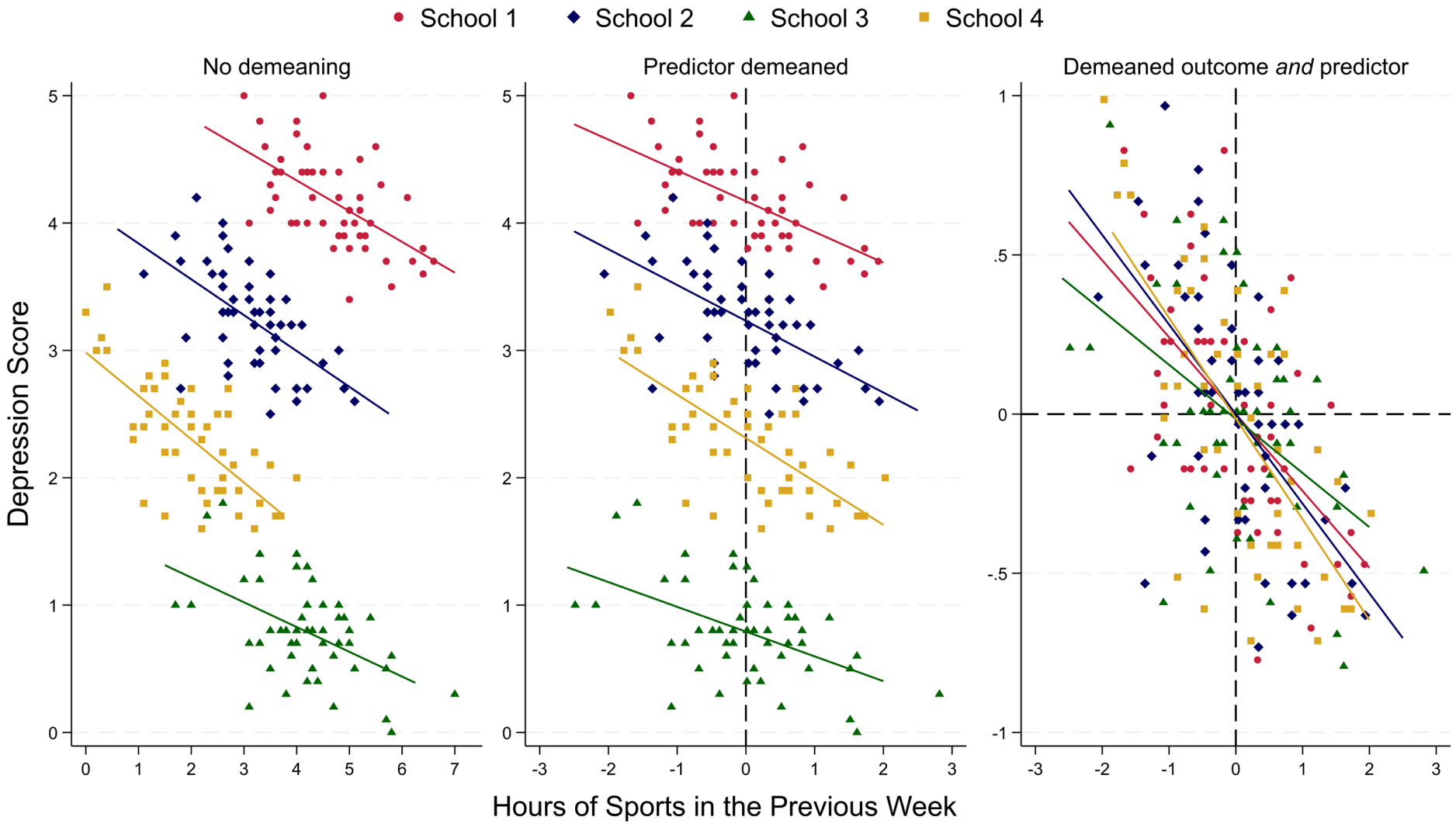
[B] This model is unbiased and is equivalent to a fixed-effects model

[C] This model isolates within-cluster associations, but the SEs are biased because it ignores the degrees of freedom used to estimate cluster means

[D] This model is unbiased and can be replicated using the R, Stata, and SPSS commands described in the text

[ø] This estimate is typically neither interpreted nor reported in regression results.

Figure 1. Graphical Representation of the Demeaning Procedure. This figure shows school-specific regression lines with uncentered variables (left panel), demeaned outcome (middle panel), and—as in fixed-effects modeling—demeaned outcome *and* predictor (right panel).



Box 1. Summary of PART 1.

- Fixed-effects modeling remains within the OLS framework, and effectively accounts for data clustering and isolates within-cluster associations between variables.
- Fixed-effects modeling is analogous to treating the cluster identifier as a categorical variable in standard regression, thereby eliminating all between-cluster variations.
- Fixed-effects modeling is also analogous to estimating the association between demeaned variables in standard regression, though SEs need to be adjusted for mean estimation.
- Fixed-effects modeling applied to cross-sectional clustered data exclusively focuses on within-cluster variations and eliminates all between-cluster confounders.

PART 2. Fixed-Effects Modeling Applied to Longitudinal Data

In this section, we show how fixed-effects modeling can be used to estimate within-participant associations in longitudinal data. Imagine you have followed $N = 800$ undergraduates over the course of $T = 4$ weeks (yielding a total of $800 \times 4 = 3{,}200$ within-participant observations). Each week, students recorded the number of hours they spent on sports activities (your predictor, with four data points per participant) and completed the same depression screening tool used in the cross-sectional study (your outcome, with four data points per participant). Your hypothesis is: "The more students increase their sports activities over time, the less depressed they become."

Here, we begin by showing how the classical two-way fixed-effects panel model can be used to estimate the within-participant association between sports and depression over time. Next, we show how to estimate two types of interaction that involve different combinations of time-constant and time-varying predictors. Finally, we introduce three extensions: (i) first-difference modeling, (ii) time-distributed fixed-effects modeling, and (iii) within-between multilevel modeling. The synthetic longitudinal dataset, along with the Quarto (R), Stata, and SPSS scripts to run the analyses, are available on the OSF. We encourage you to work with these resources while reading this section. **Table 3** provides a glossary of the jargon and notation used in this section.

Space intentionally left blank

Table 3. Glossary for Fixed-Effects Panel Modeling with Longitudinal Data (Used to Estimate Within-Person associations)

Concept	Definition	
Participant Fixed Effects	An estimator that eliminates all between-participant differences to focus exclusively on within-cluster variations	
Participant Dummies	A set of $N - 1$ binary variables that eliminate the differences between N participant (equivalent to fixed effects)	
Demeaning	Person-mean centering: Subtracting the participant mean from each observation to produce change scores	
Panel-Adjusted SEs	An estimator that makes standard errors robust to the violation of regression assumptions in longitudinal data	
Detrending	The removal of time trends in longitudinal data, with controlling for the linear effect being a simple approach	
Time Fixed Effects	An estimator that eliminates all between-wave differences to produce estimation net of period effects	
Two-Way Fixed Effects	A fixed-effects panel model that includes both participant and time fixed effects	
Time-Constant Variable	A participant trait-like feature that does not vary over time (e.g., biological sex at birth, family social class)	
Time-Varying Variable	A participant state-like feature that varies over time (e.g., income, health)	
Double Demeaning	Subtracting the participant mean from each observation to produce change scores	
Main Notations	i	Subscript denoting participant 1, 2, 3, …, N
	t	Subscript denoting wave 1, 2, …, T
	α_i	Participant dummies (representing $N - 1$ dummy variables, that is, the participant fixed effects)
	λ_t	Time dummies (representing $T - 1$ dummy variables, that is, the time fixed effects)
	$e_{w_{it}}$	Within-participant residual (variation for participant i at time t that is unexplained by the predictors)
	$\overline{var}_i$	The mean of a variable *var* for a participant i
	$värij$	A demeaned variable *var* (person-mean centered)

Main Effect in Fixed-Effects Panel Regression

The analysis of longitudinal data using fixed-effects modeling is similar to the approach described in PART 1. However, a key difference is that the clusters are now the participants themselves, rather than higher-level units such as countries (Allison, 2009; Brüderl & Ludwig, 2015; Wooldridge, 2010). In that case, the model is called a "fixed-effects panel model," and can again be conceptualized in two ways: (i) as a regression using participant identifiers as a categorical variable, which eliminates all between-participant differences by incorporating $N - 1$ dummy variables (also known as "fixed effects"), or (ii) as a regression using demeaned outcome and predictor(s), which eliminates all between-participant differences by subtracting each participant-specific mean from each individual observation (also known as "person-mean centering"). Before proceeding with how to run this model, we should consider two precautions.

Precaution 1: Using Panel-Adjusted Standard Errors

Scholars have pointed out that fixed-effects panel modeling does not entirely eliminate within-cluster correlation (Arellano, 1987) due to serial correlation (McNeish & Kelley, 2019). Consequently, many researchers advocate for the use of panel-adjusted standard errors to thoroughly address the problem of nonindependence of residuals in panel data (Angrist & Pischke, 2009; Brüderl & Ludwig, 2015; MacKinnon et al., 2023). To eliminate all sources of bias, we systematically incorporate panel-adjusted standard errors in our R, Stata, and SPSS commands (Cameron & Miller, 2015). Note that such panel-adjusted standard errors are biased in samples smaller than $N \approx 50$ participants (Kézdi, 2004), and that alternative variance formulas and bootstrap procedures have been recently proposed (Abadie et al., 2023).

Precaution 2: Taking Time Into Account

Another important consideration when analyzing longitudinal data is the inclusion of time in the model. There are two main strategies here. First, time can be treated as a

continuous covariate. This approach, often referred to as "detrending the time-varying outcome," conceptualizes time as a linear confounder (Wang & Maxwell, 2015). It enables analysts to partial out the variance attributable to the passage of time (Hoffman & Stawski, 2009; for a more in-depth discussion, see Falkenström et al., 2023). Second, time can be treated as a categorical covariate. This approach, known as "two-way fixed-effects panel modeling," incorporates both participant fixed effects (participant dummies) and time fixed effects (time dummies; Wooldridge, 2021). It enables analysts to isolate the estimates of interest net of period effects (Kropko & Kubinec, 2020). Technically, in such a model, observations are treated as cross-classified by participants and periods. While two-way fixed-effects panel modeling has been criticized for creating interpretational complexities (Hill et al., 2020; Imai & Kim, 2021; Kropko & Kubinec, 2020), it is a common strategy to account for time, and we will use it throughout the remainder of this primer.

Implementing a Two-Way Fixed-Effects Panel Model in R, Stata, and SPSS

In the conceptualization of fixed-effects modeling using participant dummies to focus on within-person variation, the "two-way fixed-effects panel modeling" equation is:

$$depression_{it} = B_1 \times sport_{it} + \alpha_i + \lambda_t + e_{w_{it}} \quad (4)$$

…where $i = 1, 2, \ldots, 800$ participants, $t = 1, 2, 3, 4$ waves, α_i are participant fixed effects ($N - 1$ dummies), λ_t are time fixed effects ($t - 1$ dummies), and $e_{w_{it}}$ is the within-participant residual

In the conceptualization of fixed-effects modeling using demeaning to focus on within-person variation, the "two-way fixed-effects panel modeling" equation is:

$$\ddot{depression}_{it} = B_1 \times \ddot{sport}_{it} + \lambda_t + e_{w_{it}} \quad (5)$$

…where the diacritical marks (the dots) indicate that the variable has been demanded (with SEs adjusted for the degrees of freedom used to estimate participant means)

The two equations above are equivalent. Below are the relevant commands in your preferred software with panel-adjusted standard error:

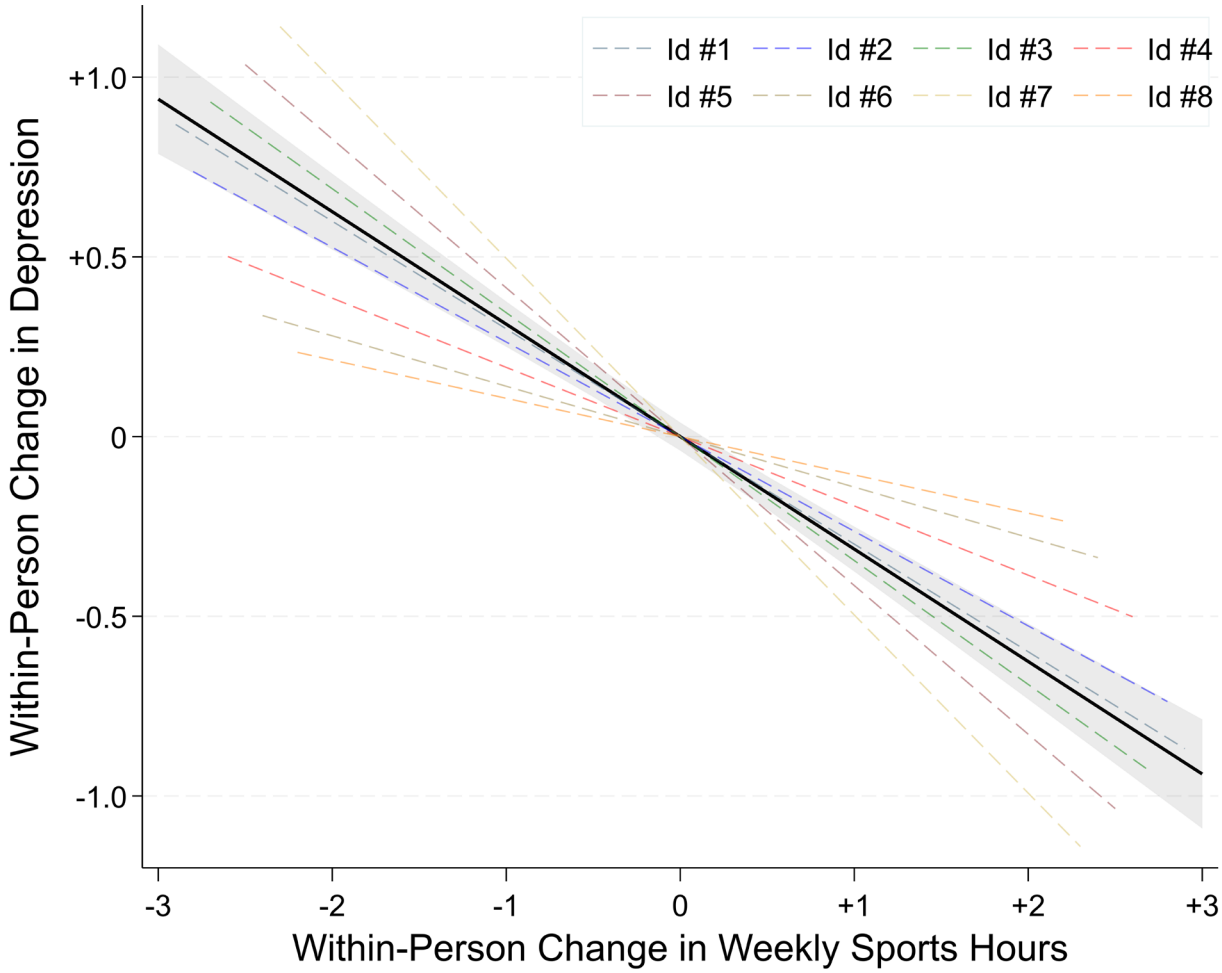
R	plm(depression ~ sport, data = df_longitudinal, index = c("part_id", "wave"), model = "within", effect = "twoways") %>% coef_test(vcov = "CR1", cluster = "part_id")
Stata	xtreg depression sport i.wave, fe i(part_id) vce(cluster part_id)
SPSS	UNIANOVA depression BY part_id wave WITH sport /INTERCEPT=EXCLUDE /PRINT PARAMETER /ROBUST=HC1 /DESIGN=part_id wave sport.

…where depression is the outcome, sport the predictor, part_id the participant identifier, and wave the time identifier. R users should install clubSandwich (Pustejovsky & Tipton, 2018). Commands may yield slightly different SEs as variance estimators differ.

Running either of these commands, you will find that your estimate of interest is $B_1 = -0.31$, $SE = 0.02$, $p < .001$. **Figure 2**, provides a graphical representation of this finding, showing the individual associations between the demeaned variables for the first eight participants for illustrative purposes. This finding supports your hypothesis: For each additional hour of sports activity per week, the depression score of a given participant decreases by a third of a point. Importantly, this estimate is free from what economists and sociologists refer to as "time-constant individual heterogeneity" (Brüderl & Ludwig, 2015). The fixed-effects estimator accounts for all observed and unobserved participant characteristics that do not vary over time (e.g., genetic dispositions, family social class, ethnicity), thereby removing any potential between-participant confounders.

Space intentionally left blank

Figure 2. Within-Participant Association in a Fixed-Effects Panel Model. This figure shows the pooled associations between sports and depression (solid line), with the 95% CI (shaded area), and individual associations for the first eight participants (dashed lines).



Calculation of Effect Sizes and Simulation-Based Statistical Power

As mentioned previously, effect sizes in fixed-effects modeling can be calculated in the same way as for any OLS regression. For example, the partial eta-square, which quantifies the proportion of variance explained by a dichotomous or continuous predictor, can be calculated using the following formula (Cohen, 1965):

$$\eta_p^2 = \frac{(B \div SE)^2}{(B \div SE)^2 + df_{residual}} \quad (6)$$

…where $df_{residual}$ represents the degrees of freedom for the residuals, calculated as $N - 1$ (the degrees of freedom consumed by the fixed effects) minus the number of regressors

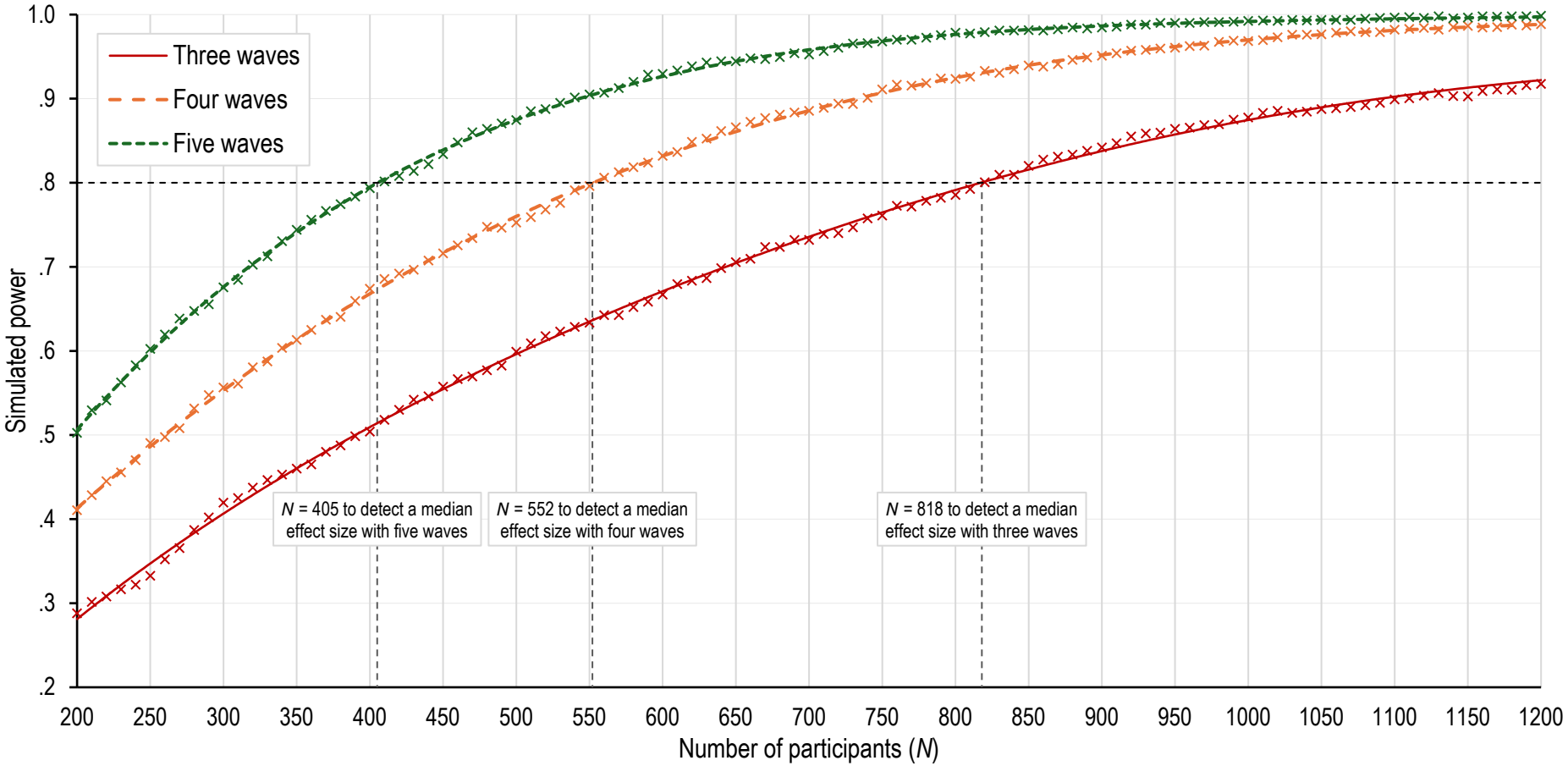
Importantly, effect sizes in longitudinal survey data are typically smaller than those in cross-sectional survey data (Adachi & Willoughby, 2015). For instance, Gignac and Szodorai (2016) collected 708 meta-analytic correlations from psychological studies using cross-sectional designs and found that the 25th, 50th, and 75th percentiles of the distribution for between-participant effect sizes were $\eta_p^2 \approx .01$ (smaller estimate), .04 (median estimate), and .08 (larger estimate), respectively. In contrast, Orth et al. (2024) collected 1,028 effect sizes from studies using longitudinal designs and found that the same percentiles for within-participant effect sizes were $\eta_p^2 \approx .001$ (smaller estimate), .005 (median estimate), and .01 (larger estimate), respectively. In other words, within-participant effects are 8–10 times smaller than between-participant effects.[5]

We conducted a series of simulations to estimate the sample size required to detect a median within-participant effect size of $\eta_p^2 \approx .005$ with 80% power (the script can be found on the OSF). We simulated 2,000 datasets for each of 100 *N* values across three-wave, four-wave, and five-wave designs, with the goal of identifying the sample size at which 80% of the

[5] Two points should be noted: (i) the benchmarks reported in Gignac and Szodorai (2016) and Orth et al. (2024) were expressed as *r*s, which we converted into η_p^2 (proportion of variable explained) and (ii) the within-participant effects analyzed in Orth et al. (2024) were collected from cross-lagged panel models, not fixed-effects panel models.

effects are detected using fixed-effects panel modeling with $\alpha = .05$. The analysis yielded three key findings: (i) a sample size of $N = 818$ is required to detect a median within-participant effect for a three-wave panel dataset; (ii) $N = 552$ is required for a four-wave panel dataset; and (iii) $N = 405$ is required for a five-wave panel dataset (see **Figure 3** for the power curve, and further details about the simulations). Importantly, these figures are only approximate estimates of the sample size typically required in fixed-effects panel analysis. The benchmarks used here apply to long-term longitudinal survey data (e.g., a yearly panel study) and do not apply to repeated measures experiments, which tend to show stronger effects than between-participant experiments (Flora, 2020; Onwuegbuzie & Levin, 2003; Sommet et al., 2023). Generally, these benchmarks should be viewed as heuristic rather than universally applicable tools, as effect sizes vary by subdiscipline, population, and measurement (Schäfer & Schwarz, 2019).

Figure 3. Power Curves for Detecting A Median-Sized Within-Person Association ($\eta_p^2 \approx .005$) With Fixed-Effects Modeling and $T = 3, 4, 5$



Note. These power curves were derived from simulations where $\bar{X} \sim N(0, 1)$ and $\bar{Y} \sim N(0, 1)$, with $\beta_{(\ddot{X}_{it}, \ddot{Y}_{it})} = 0.07$ ($\Leftrightarrow \eta_p^2 \approx .005$) and $e_{ij} \sim N(0, 1)$. We began by simulating a batch of 2,000 datasets with sample size N, and continued simulating batches in increments of $N + 10$ until we could identify the tipping points where over 80% of the datasets showed a significant effect ($\alpha = .05$) in a one-way fixed-effects panel model.

Interaction Effects in Fixed-Effects Panel Regression

Fixed-effects panel modeling allows analysts to test two types of second-order interactions: (i) the interaction between a time-constant and a time-varying predictor, and (ii) the interaction between two time-varying predictors.

Type #1. Interaction Between a Time-Constant and a Time-Varying Predictors

Fixed-effects panel modeling cannot be used to estimate the statistical effects of a time-constant predictor, as the fixed effects leave no residual between-participant variance to be explained (Hsiao, 2022). However, it can be used to estimate whether the within-participant effect of a time-varying variable varies as a function of a time-constant participant characteristic (Allison, 2009). In multilevel modeling, this would be referred to as to 'cross-level interactions' (Aguinis & Gottfredson, 2010). For example, here is how to test whether the association between sports and depression over time varies between men (coded -0.5) and women (coded +0.5):[6]

$$depression_{it} = B_1 \times sport_{it} + B_2 \times sport_{it} \times sex_i + \lambda_t + \alpha_i + e_{w_{it}} \quad (7)$$

Running the OSF-uploaded script for this equation, you will see that the interaction term is B_2 = -0.01, SE = 0.05, p = .907. This indicates that the pooled within-participant association between sports and depression over time does not significantly differ between men and women. The equation does not include the main effect of sex because—as with any time-constant predictor—it cannot be estimated.

Type #2. Interaction Between Two Time-Varying Predictors

Fixed-effects panel modeling can also be used to estimate whether the within-participant effect of a time-varying variable varies as a function of another time-varying variable (Schunck, 2013). However, if you test such interactions using participant dummies or

[6] For simplicity, we use biological sex at birth as an example of a binary time-constant moderator. Importantly, this does not account for intersex individuals or those whose gender identity differs from their biological sex.

the demeaned product term of the non-demeaned variables (as your software typically does), the estimation will be contaminated by between-participant differences (Giesselmann & Schmidt-Catran, 2022). To mitigate this bias, one should use 'double demeaning': One should demean the product term of the already demeaned predictors to ensure accurate estimation of the pooled within-participant interaction term (Balli & Sørensen, 2013). For example, you can test whether the association between sports and depression over time varies as a function of sleep quality during the week of data collection:

$$\ddot{depression}_{it} = B_1 \times \ddot{sport}_{it} + B_2 \times \ddot{sleep}_{it} + B_3 \times \ddot{sport}_{it} \ddot{\times} \ddot{sleep}_{it} + \lambda_t + e_{w_{it}} \quad (8)$$

…where the diacritical marks denote demeaned variables, and the stacked diacritical marks denote double demeaning with $\ddot{sport}_{it} \ddot{\times} \ddot{sleep}_{it} = \ddot{sport}_{it} \times \ddot{sleep}_{it} - \overline{\ddot{sport}_{it} \times \ddot{sleep}_{it}}$

Running the OSF-uploaded script for this equation, you will see that the interaction term is B_3 = -0.07, SE = 0.02, p < .001. This indicates that the pooled within-participant association between sports and depression over time differs according to sleep quality. Decomposing this interaction reveals that the association between sports and depression is weaker in weeks when the participant has poorer-than-usual sleep, B_{-1SD} = -0.07, SE = 0.03, p = .047, and stronger when they have better-than-usual sleep, B_{+1SD} = -0. 23, SE = 0.03, p < .001.

Three Extensions of Fixed-Effects Panel Modeling

Below, we briefly introduce three extensions of fixed-effects panel modelsing: (i) first-difference modeling, (ii) time-distributed fixed-effects modeling, and (iii) within-between multilevel modeling.

Extension #1. First-Difference Modeling

While fixed-effects panel modeling estimates the long-term statistical association between two variables across waves, first-difference modeling captures the short-term statistical association between changes in a predictor and changes in an outcome from one

wave to the next (Allison, 2009). When only $T = 2$ waves of data are available, a fixed-effects model is equivalent to a first-difference model (Schmidheiny, 2023; for a paper on longitudinal analyses with two waves of data, see Johnson, 2005). Here is the equation:

$$\Delta depression_{it} = B_1 \times \Delta sport_{it} + e_{w_{it}} \quad (9)$$

…where $i = 1, 2, \ldots, 800$ participants, $t = 2, 3, 4$ waves, Δ denotes the difference between the variable score at time t and $t - 1$, and $e_{w_{it}}$ is the within-participant residual.

Running the OSF-uploaded script, you will see that the focal estimate is $B_1 = -0.31$, $SE = 0.03$, $p < .001$.[7] As illustrated by **Figure 4** (left panel), this indicates that for each additional hour of sports activity in a given week compared to the previous week, the depression score of a given participant decreases by 0.31 points. Note that fixed effects (for participants and time) do not appear in Equation (9) as they cancel each other when the first differences are taken (Schmidheiny & Siegloch, 2019). Moreover, data from the first wave cannot be used in this approach, as the initial value of the predictor or outcome is unknown (i.e., $t = 2, 3, 4$).

Extension #2. Time-Distributed Fixed-Effects Modeling

While fixed-effects panel modeling estimates the gradual statistical effect of a continuous predictor, time-distributed fixed-effects modeling tracks changes in the outcome for each wave preceding or following a particular event (Ludwig & Brüderl, 2021). This approach—also known as 'dummy impact functions'—requires significant preliminary data management. In our example, we begin by dichotomizing the predictor and focusing on the participants who started doing sports at some point in the study. Here is the equation:

$$depression_{it} = B_1 \times T_{-2_{it}} + B_2 \times T_{-1_{it}} + B_3 \times T_{0_{it}} + B_4 \times T_{+1_{it}} + B_5 \times T_{+2_{it}} + \alpha_i + e_{w_{it}} \quad (10)$$

…where $i = 1, 2, \ldots, 175$ participants, $t = 1, 2, 3, 4$ waves, $T_{0_{it}}$ is a binary variable coded 1 the week the participant began sports (0 otherwise), $T_{\pm n_{it}}$ are coded 1 for the n^{th} week before/after

[7] SPSS users will need to install Huang et al.'s (2022) macro to calculate cluster-robust SEs.

the participant began sports (0 otherwise), and $e_{w_{it}}$ is the within-participant residual.

Running the OSF-uploaded script, you will see that B_1 and B_2 are null, whereas B_3 through B_5 are significantly negative. As illustrated by **Figure 4** (middle panel), this indicates that a participant who starts doing sports during a given week experiences an immediate reduction in depression scores, which persists for at least two weeks. Importantly, this analysis is confined to the subset of participants who started sports within the four-week window of the study (i.e., $N = 175$ out of 800). Generally speaking, this analysis may exclude additional data points because some participants may experience multiple spells of activity (e.g., resuming sports after a break), with analysts often focusing only on the initial spell.

Extension #3. Within-Between Multilevel Modeling

While fixed-effects panel modeling focuses solely on within-participant variation, within-between multilevel modeling partitions variance into within- *and* between-participant components (Allison, 2005). This approach—also known as 'hybrid modeling'—is one of several multilevel modeling strategies that use advanced centering techniques to estimate both within- and between-participant associations, including the contextual model (Mundlak, 1978), multilevel model with person-mean centered predictor (Kreft et al., 1995), and various SEM-based models using latent person means (Hamaker & Muthén, 2020). Specifically, the within-between multilevel model incorporates two distinct versions of the focal predictor: (i) the person-mean centered predictor (i.e., demeaned), which estimates the within-participant association, and (ii) the person-specific predictor mean (i.e., a single value per participant), which estimates the between-participant association. Here is the equation:

$$depression_{it} = B_{00} + B_{\text{within}} \times \ddot{sport}_{it} + B_{\text{between}} \times \overline{sport}_i + \lambda_t + u_{0i} + e_{w_{it}} \quad (11)$$

...where the diacritical marks denote the demeaned predictor, the diacritical line denotes the person-specific mean, B_{00} is the fixed intercept (mean outcome across participants), u_{0i} is the

random intercept (between-participant residual), and $e_{w_{it}}$ is the within-participant residual.

Running the OSF-uploaded script, you will see that $B_{\text{within}} = -0.31$, $SE = 0.02$, $p < .001$ and $B_{\text{between}} = -0.04$, $SE = 0.01$, $p < .001$. As illustrated by **Figure 4** (right panel), this indicates that (i) for each additional hour of sports per week, a given participant's depression score decreases by 0.31 points, whereas (ii) for each additional average weekly hour of sport, participants' depression scores are 0.04 points lower compared to other participants. Interestingly, B_{within} is identical to B_1 from the fixed-effects panel regression model (Eqs. 4-5), although minor decimal variations might occur as multilevel modeling uses maximum likelihood rather than OLS as methods of estimation

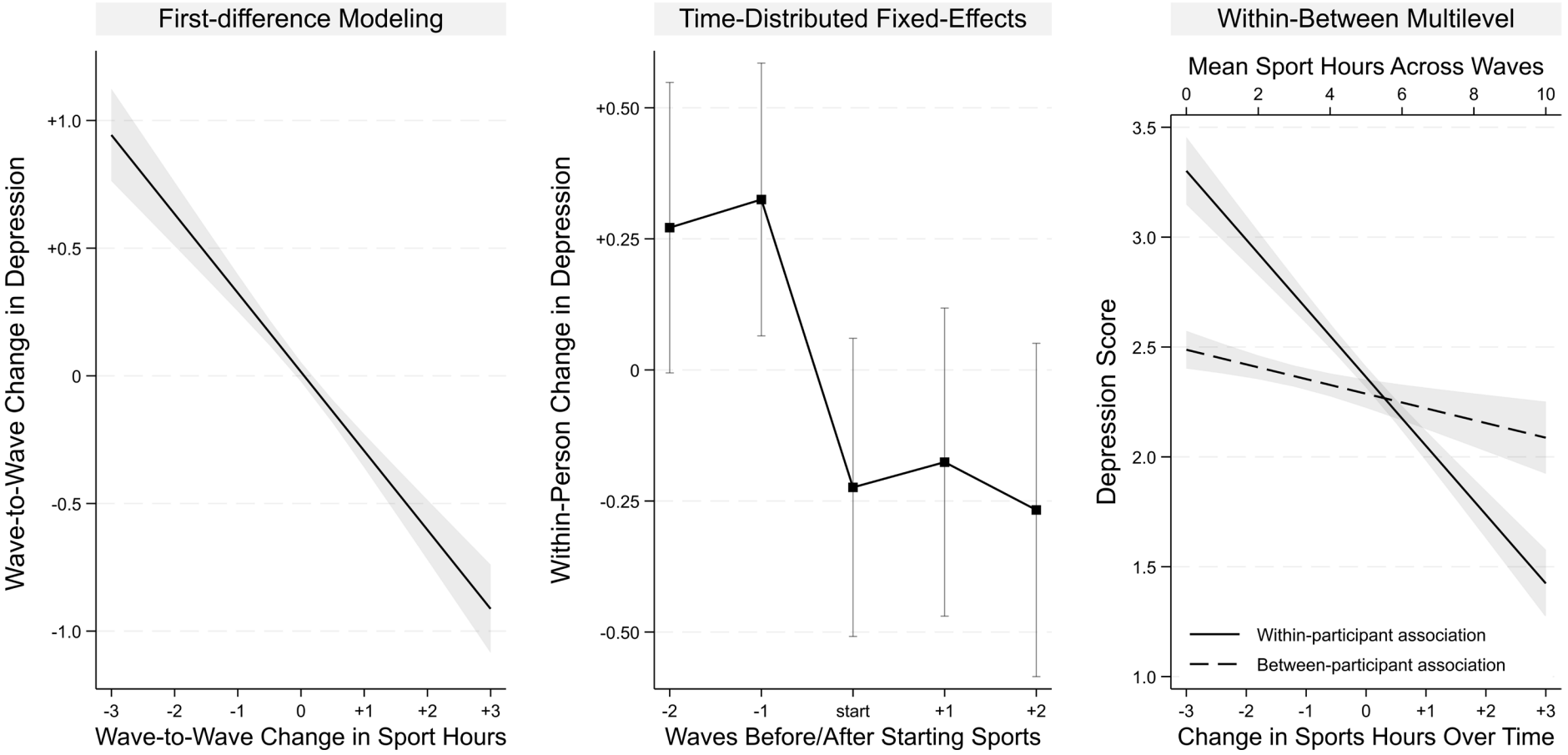
Given that the within-between multilevel model estimates within-participant associations while retaining the possibility of estimating between-participant associations, one might wonder why this model is not always preferred. The reason is twofold. First, the within-between multilevel model necessitates advanced centering strategies, which may increase the likelihood of errors and lead to interpretative ambiguities. Second, this model may lead to confusion in interpretation, as the effects of time-constant regressors should not be interpreted as causal (Brüderl & Ludwig, 2015). Third, this model leaves the OLS framework, which—as previously discussed—can complicate model comparisons and other statistical procedures.

Space intentionally left blank

Box 2. Summary of PART 2.

- Fixed-effects panel modeling eliminates all between-participant differences and isolates within-participant associations between variables
- It is advisable to estimate panel-adjusted standard errors to comprehensively address the problem of nonindependence of residuals.
- Time can be taken into account using either detrending (treating time as a continuous covariate) or two-way fixed-effects modeling (treating time as a categorical covariate).
- Fixed-effects panel modeling can be used to test time-constant × time-varying variable interactions; in this case, the main effect of the time-constant predictor is not estimated.
- Fixed-effects panel modeling can also be used to test time-varying × time-varying variable interactions; in this case, double demeaning should be used.
- First-difference modeling estimates the within-participant association between two variables from one wave to the next.
- Time-distributed fixed-effects modeling estimates within-participant changes in the outcome for each wave preceding or following an event.
- Within-between multilevel modeling combines centering strategies to estimate within-participant associations while preserving the ability to test between-participant differences.

Figure 4. Associations of Interest in First-Difference Modeling (Left Panel), Time-Distributed Fixed-Effects Modeling (Middle Panel), and Within-Between Multilevel Modeling (Right Panel). The three panels show (i) the associations between sports and depression from one wave to the next, (ii) the changes in depression scores before, during, and after beginning sports, and (iii) the within-participant and between-participant associations between sports and depression (the 95% CIs are indicated by the shaded areas or error bars).



PART 3. Fixed-Effects Panel Modeling and Causality

Causality with nonexperimental data is a key topic of discussion in other social sciences (e.g., Arkhangelsky & Imbens, 2024; Gebel, 2023; Grätz, 2022), but remains somewhat taboo in psychological science (Grosz et al., 2020). Fixed-effects panel modeling is often regarded as the "gold standard" for leveraging the structure of longitudinal data (Bliese et al., 2020; Osgood, 2010; Schurer & Yong, 2012) and inferring causality from within-participant effects (Halaby, 2004; for relevant research from psychologists, see Bailey et al., 2024; Quintana, 2021; Rohrer & Murayama, 2023). As such, we believe that fixed-effects panel modeling deserves a place in psychologists' analytical toolbox, whether for analyzing panel data (as in our example), repeated-trial experiments, or daily diary studies (i.e., using experience sampling methods; Hektner et al., 2007). However, it is important to be aware of two key limitations of fixed-effects panel modeling for causal inference (Bell & Jones, 2015; Collischon & Eberl, 2020; Hill et al., 2020).

Limitation #1. Time-Varying Confounders

Although fixed-effects panel models effectively eliminate all potential time-constant between-participant confounders, they remain vulnerable to time-varying within-participant confounders (Treiman, 2014). In our example, participants who experience temporary health issues, such as a cold, indigestion, or allergies, may spend less time engaging in sports and feel sadder due to physical limitations. As such, the association between sports and depression may be partially or even fully explained by health. Of course, one could add health and other relevant variables as time-varying covariates, but it will always be possible to imagine yet another unmeasured (or even unmeasurable) time-varying variable that could act as a confounder (for relevant research, see Cinelli et al., 2024). In other words, adding controls offers a limited solution to the problem that the focal predictor is not exogenous by design (as in an experiment), leaving residual doubt about whether its effect is truly causal. That said,

one should keep in mind that cross-sectional analyses are sensitive to both time-constant *and* time-varying confounders, meaning that fixed-effects modeling resolves what is arguably the bigger half of the problem (Collischon & Eberl, 2020).

Limitation #2. Reverse Causation

Fixed-effects panel models are vulnerable to reverse causation (Vaisey & Miles, 2014). In our example, participants who experience increased depression over the course of the study might lose motivation and reduce their sports participation. As such, it could be that depression decreases sports participation, rather than the reverse. To address this issue, analysts sometimes turn to cross-lagged panel models (CLPMs). In their traditional form, CLPMs use a structural equation modeling (SEM) framework to simultaneously estimate: (i) directional paths (doing sports at time t → being depressed at t+1), (ii) reciprocal paths (being depressed at time t → doing sports at t+1), and (iii) autoregressive paths (doing sports at time t → doing sports at t+1; being depressed at time t → being depressed at t+1; for foundational work, see Duncan, 1969; Finkel, 1995; Heise, 1970). However, traditional CLPMs are known to be biased because they fail to properly distinguish within-person dynamics from between-person trait-like differences (Hamaker et al., 2015). To address this limitation, several refined CLPMs have been developed, incorporating random intercepts, fixed effects, or other features, and requiring at least three waves of data (Leszczensky & Wolbring, 2022; Orth et al., 2021; Usami et al., 2019). Despite these refinements, the debate continues, as these advanced CLPMs have been shown to be prone to bias and inflated error rates (Lucas, 2023). That said, one should keep in mind that cross-sectional analyses are also sensitive to reverse causation; unfortunately, fixed-effects panel modeling does not resolve this issue, and CLPMs are no panacea.

Causality as a Continuum of Plausibility

In epidemiology, experimental work is rarely possible, and scholars often conceptualize

causality in probabilistic rather than deterministic terms (Parascandola, 2011). In our field, experiments are not always feasible either, and psychologists could also benefit from viewing causality in observational studies as a continuum of plausibility rather than as a simple binary. For example, in longitudinal studies using self-reported variables, pooled regression would be located at the lower end of such a continuum, as it is vulnerable to third-variable effects. Fixed-effects panel regression would rank higher, though not at the top of the continuum, as time-varying confounders and reverse causation remain threats. Triangulating longitudinal evidence using other tools can push the analysis further toward the upper end of the continuum. These tools might include those covered in this primer, such as first-difference modeling and advanced CLPMs, as well as others, such as matching procedures and growth curve modeling (Duncan & Duncan, 2009; Thoemmes & Kim, 2011). Ultimately, true causality is more a matter of study design than analytical tools, and only a longitudinal analysis testing the influence of exposure to an external shock could reach the uppermost part of the continuum (Grosz et al., 2024). In our example, tracking the depression scores of high-school students before and after an educational reform doubling the hours spent in sports activities would considerably enhance the quality of causal inference. However, such exogenous predictors are often not available, making fixed-effects panel modeling an essential tool for psychologists seeking to leverage longitudinal data to approach causality.

Disclosures

Author Contributions

Author #1: Conceptualization, Methodology, Software, Data Curation, Writing - Original Draft, Visualization. **Author #2:** Validation, Writing - Review & Editing, Visualization

Conflicts of Interest

The authors declare that they have no conflict of interest.

Funding

This work was funded by a SNSF Ambizione fellowship granted to the first author.

Prior Versions

Prior to submission, we uploaded a preprint of the article on the OSF.

Data, Materials, and Online Resources

The synthetic dataset, Quarto (R), Stata, and SPSS scripts used in this primer, as well as scripts to simulate the data, create all figures, and run the power analysis, are available on the OSF: https://osf.io/bf32c/?view_only=00477cc083b048ebaabde2ec5cc2f482

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics, 138*(1), 1-35.

Adachi, P., & Willoughby, T. (2015). Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science. *European Journal of Developmental Psychology, 12*(1), 116-128.

Aguinis, H., & Gottfredson, R. K. (2010). Best-practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior, 31*(6), 776-786.

Allison, P. D. (2005). *Fixed effects regression methods for longitudinal data using SAS*. SAS Institute, Inc.

Allison, P. D. (2009). *Fixed effects regression models*. SAGE Publications

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press

Araki, S. (2023). The societal determinants of happiness and unhappiness: Evidence from 152 countries over 15 years. *Social Psychological and Personality Science, 15*(5), 603-615.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics & Statistics, 49*(4), 431-434.

Arkhangelsky, D., & Imbens, G. (2024). Causal models for longitudinal and panel data: a survey. *The Econometrics Journal*.

Bailey, D. H., Jung, A. J., Beltz, A. M., Eronen, M. I., Gische, C., Hamaker, E. L., Kording, K. P., Lebel, C., Lindquist, M. A., & Moeller, J. (2024). Causal inference on human behaviour. *Nature Human Behaviour, 8*(8), 1448-1459.

Balli, H. O., & Sørensen, B. E. (2013). Interaction effects in econometrics. *Empirical Economics, 45*(1), 583-603.

Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological Methods, 16*(4), 373-390.

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods, 3*(1), 133-153.

Bell, A., Jones, K., & Fairbrother, M. (2018). Understanding and misunderstanding group mean centering: a commentary on Kelley et al.'s dangerous practice. *Quality & Quantity, 52*(5), 2031-2036.

Bliese, P. D., Schepker, D. J., Essman, S. M., & Ployhart, R. E. (2020). Bridging methodological divides between macro- and microresearch: Endogeneity and methods for panel data. *Journal of Management, 46*(1), 70-99.

Brüderl, J., & Ludwig, V. (2015). Fixed-effects panel regression. In H. Best & C. Wolf (Eds.), *The Sage handbook of regression analysis and causal inference* (pp. 327-357).

Bryan, M. L., & Jenkins, S. P. (2015). Multilevel modelling of country effects: A cautionary tale. *European Sociological Review, 32*(1), 3-22.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources, 50*(2), 317-372.

Cinelli, C., Forney, A., & Pearl, J. (2024). A crash course in good and bad controls. *Sociological Methods & Research, 53*(3), 1071-1104.

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 95-121). McGraw-Hill

Collischon, M., & Eberl, A. (2020). Let's talk about fixed effects: Let's talk about all the good things and the bad things. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 72*(2), 289-299.

Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software, 27*(2), 1-43.

Dieleman, J. L., & Templin, T. (2014). Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: A simulation study. *PLoS One, 9*(10), e110257.

Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin, 72*(3), 177-182.

Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass, 3*(6), 979-991.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods, 12*(2), 121-138.

Falkenström, F., Solomonov, N., & Rubel, J. (2023). To detrend, or not to detrend, that is the question? The effects of detrending on cross-lagged effects in panel models. *Psychological Methods*.

Finkel, S. E. (1995). *Causal analysis with panel data*. Sage

Flora, D. B. (2020). Thinking about effect sizes: From the replication crisis to a cumulative psychological science. *Canadian Psychology/Psychologie Canadienne, 61*(4), 318-330.

Gardiner, J. C., Luo, Z., & Roman, L. A. (2009). Fixed effects, random effects and GEE: What are the differences? *Statistics in Medicine, 28*(2), 221-239.

Gebel, M. (2023). Causal inference based on non-experimental data in health inequality research. In R. Hoffmann (Ed.), *Handbook of Health Inequalities Across the Life Course* (pp. 93-111). Elgar

Giesselmann, M., & Schmidt-Catran, A. W. (2022). Interactions in fixed effects regression models. *Sociological Methods & Research, 51*(3), 1100-1127.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74-78.

Grätz, M. (2022). When less conditioning provides better estimates: overcontrol and endogenous selection biases in research on intergenerational mobility. *Quality & Quantity, 56*(5), 3769-3793.

Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., Müller, S. R., Rieger, S., Zapko-Willmes, A., & Rohrer, J. M. (2024). Natural experiments: Missed opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science, 7*(1), 25152459231218610.

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.

Halaby, C. N. (2004). Panel models in sociological research: theory into practice. *Annual Review of Sociology, 30*(Volume 30, 2004), 507-544.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102-116.

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods, 25*(3), 365.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press

Heise, D. R. (1970). Causal inference from panel data. *Sociological Methodology, 2*, 3-27.

Hektner, J., Schmidt, J., & Csikszentmihalyi, M. (2007). *Experience Sampling Method*. SAGE Publications, Inc.

Hoffman, L., & Stawski, R. S. (2009). Persons as Contexts: Evaluating Between-Person and Within-Person Effects in Longitudinal Analysis. *Research in Human Development, 6*(2-3), 97-120.

Hsiao, C. (2022). *Analysis of panel data*. Cambridge university press

Huang, F. L., & Li, X. (2022). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods, 54*(3), 1181-1199.

Johnson, D. (2005). Two-wave panel analysis: Comparing statistical methods for studying the effects of transitions. *Journal of Marriage and Family, 67*(4), 1061-1075.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1991). *The theory and practice of econometrics*. John Wiley & Sons

Kézdi, G. (2004). Robust standard error estimation in Fixed-Effects panel models. *Hungarian Statistical Review, Special Number 9*, 95-116.

Krämer, M. D., Rohrer, J. M., Lucas, R. E., & Richter, D. (2024). Life events and life satisfaction: Estimating effects of multiple life events in combined models. *European Journal of Personality*.

Kreft, I. G., De Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*(1), 1-21.

Kropko, J., & Kubinec, R. (2020). Interpretation and identification of within-unit and cross-sectional variation in panel data models. *PLoS One, 15*(4), e0231349.

Leszczensky, L., & Wolbring, T. (2022). How to deal with reverse causality using panel data? Recommendations for researchers based on a simulation study. *Sociological Methods & Research, 51*(2), 837-865.

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.

Ludwig, V., & Brüderl, J. (2021). What you need to know when estimating impact functions with panel data for demographic research. *Comparative Population Studies-Zeitschrift für Bevölkerungswissenschaft, 46*, 453-486.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86.

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics, 232*(2), 272-299.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2017). *Designing experiments and analyzing data: A model comparison perspective*. Routledge

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods, 24*(1), 20-35.

McNeish, D., & Stapleton, L. M. (2016). Modeling Clustered Data with Very Few Clusters. *Multivariate Behavioral Research, 51*(4), 495-518.

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114-140.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*(1), 69-85.

Noetel, M., Sanders, T., Gallardo-Gómez, D., Taylor, P., del Pozo Cruz, B., van den Hoek, D., Smith, J. J., Mahoney, J., Spathis, J., Moresi, M., Pagano, R., Pagano, L., Vasconcellos, R., Arnott, H., Varley, B., Parker, P., Biddle, S., & Lonsdale, C. (2024). Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials. *BMJ, 384*, e075847.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods, 8*(4), 434-447.

Onwuegbuzie, A. J., & Levin, J. R. (2003). Without supporting statistical evidence, where would reported measures of substantive importance lead? To no good effect. *Journal of Modern Applied Statistical Methods, 2*(1), 12.

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology, 120*(4), 1013-1034.

Orth, U., Meier, L. L., Bühler, J. L., Dapp, L. C., Krauss, S., Messerli, D., & Robins, R. W. (2024). Effect size guidelines for cross-lagged effects. *Psychological Methods, 29*(2), 421.

Osgood, D. W. (2010). Statistical models of life events and criminal behavior. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 375-396). Springer

Parascandola, M. (2011). Causes, risks, and probabilities: probabilistic concepts of causation in chronic disease epidemiology. *Prev Med, 53*(4-5), 232-234.

Pearl, J. (2009). *Causality*. Cambridge university press

Petersen, M. A. (2008). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *The Review of Financial Studies, 22*(1), 435-480.

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics, 36*(4), 672-683.

Quintana, R. (2021). Thinking within-persons: Using unit fixed-effects models to describe causal mechanisms. *Methods in Psychology, 5*, 100076.

Rohrer, J. M., & Murayama, K. (2023). These Are Not the Effects You Are Looking for: Causality and the Within-/Between-Persons Distinction in Longitudinal Data Analysis. *Advances in Methods and Practices in Psychological Science, 6*(1), 25152459221140842.

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*, 813.

Schmidheiny, K. (2023). *Panel data: fixed and random effects*. https://www.schmidheiny.name/teaching/panel2up.pdf

Schmidheiny, K., & Siegloch, S. (2019). On event study designs and distributed-lag models: Equivalence, generalization and practical implications.

Schunck, R. (2013). Within and between Estimates in Random-Effects Models: Advantages and Drawbacks of Correlated Random Effects and Hybrid Models. *The Stata Journal, 13*(1), 65-76.

Schurer, S., & Yong, J. (2012). Personality, well-being and the marginal utility of income: What can we learn from random coefficient models? In *Health, Economics and Data Group*. Working Paper, University of York

Seifert, I. S., Rohrer, J. M., & Schmukle, S. C. (2024). Using within-person change in three large panel studies to estimate personality age trajectories. *Journal of Personality and Social Psychology, 126*(1), 150-174.

Sommet, N., Weissman, D., Cheutin, N., & Elliot, A. J. (2023). How many participants do i need to test an interaction? Conducting an appropriate power analysis and achieving sufficient power to detect an interaction. *Advances in Methods and Practices in Psychological Science, 6*.

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research, 46*(1), 90-118.

Treiman, D. J. (2014). *Quantitative data analysis: Doing social research to test ideas*. John Wiley & Sons

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24*(5), 637-657.

Vaisey, S., & Miles, A. (2014). What you can—and can't—do with three-wave panel data. *Sociological Methods & Research, 46*(1), 44-67.

Wang, L., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods, 20*(1), 63-83.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press

Wooldridge, J. M. (2021). *Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators*. http://dx.doi.org/10.2139/ssrn.3906345