University of Reading

These are not the effects you are looking for: causality and the within-/between-person distinction in longitudinal data analysis

Article

Published Version

Creative Commons: Attribution-Noncommercial 4.0

Open Access

Rohrer, J. M. and Murayama, K. (2023) These are not the effects you are looking for: causality and the within-/between-person distinction in longitudinal data analysis. Advances in Methods and Practices in Psychological Science, 6 (1). pp. 1-14. ISSN 2515-2459 doi: 10.1177/25152459221140842 Available at https://centaur.reading.ac.uk/108581/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1177/25152459221140842

Publisher: SAGE

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

General Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Advances in Methods and Practices in Psychological Science
January-March 2023, Vol. 6, No. 1, pp. 1–14
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25152459221140842
www.psychologicalscience.org/AMPPS

SAGE

These Are Not the Effects You Are Looking for: Causality and the Within-/Between-Persons Distinction in Longitudinal Data Analysis

Julia M. Rohrer[1] and Kou Murayama[2,3,4]

[1]Wilhelm Wundt Institute for Psychology, Faculty of Life Sciences, Leipzig University, Leipzig, Germany; [2]Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany; [3]Research Institute, Kochi University of Technology, Kochi, Japan; and [4]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

Abstract

In psychological science, researchers often pay particular attention to the distinction between within- and between-persons relationships in longitudinal data analysis. Here, we aim to clarify the relationship between the within- and between-persons distinction and causal inference and show that the distinction is informative but does not play a decisive role in causal inference. Our main points are threefold. First, within-persons data are not necessary for causal inference; for example, between-persons experiments can inform about (average) causal effects. Second, within-persons data are not sufficient for causal inference; for example, time-varying confounders can lead to spurious within-persons associations. Finally, despite not being sufficient, within-persons data can be tremendously helpful for causal inference. We provide pointers to help readers navigate the more technical literature on longitudinal models and conclude with a call for more conceptual clarity: Instead of letting statistical models dictate which substantive questions researchers ask, researchers should start with well-defined theoretical estimands, which in turn determine both study design and data analysis.

Keywords

causal inference; longitudinal data; fixed-effects model, cross-lagged panel model; dynamic panel model

Received 3/5/22; Revision accepted 10/25/22

Daily diary studies, experience sampling, mobile sensing: Technological innovations have made it much easier for psychologists to collect longitudinal data from multiple participants. Accordingly, the number of studies making use of such data has increased steadily (e.g., Hamaker & Wichers, 2017), relevant statistical models have gained prominence, and interest in psychology as an idiographic science has been rekindled (Molenaar, 2004). That is not to say that the idea of assessing a person multiple times is a new one—"occasions" constitute one of the three axes of Cattell's (1952) well-known "data cube" (the other two being "persons" and "variables")—but empirical research is finally catching up with a dimension that has always been considered conceptually important.

With the increased amount of longitudinal data available, recent studies have paid particular attention to "disentangling" within- and between-persons associations (e.g., Curran & Bauer, 2011; Hamaker et al., 2005; Voelkle et al., 2014). For example, a positive association between talkativeness and subjective well-being may exist on the between-persons level—people who are (on average) more talkative than others are (on average) happier than others—or it may exist on the within-persons level—people who are more talkative today (than they usually are) are happier today (than they usually are). Between- and within-persons associations can be statistically independent (i.e., they can take on different values or even opposite signs), and it is the

Corresponding Author:
Julia M. Rohrer, Department of Psychology, Leipzig University
Email: julia.rohrer@uni-leipzig.de

Creative Commons NonCommercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/), which permits noncommercial use, reproduction, and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

latter within-persons associations that psychologists often deem more interesting because they are meant to inform about "within-person processes" (e.g., Molenaar & Campbell, 2009). In line with traditions of the field (Grosz et al., 2020), however, the psychological literature on within-persons data often shies away from explicitly interpreting such processes as causal effects, and only recently have authors tried to explicitly bridge the gap with the causal-inference literature (e.g., Gische et al., 2021; Lüdtke & Robitzsch, 2022; Voelkle et al., 2018). However, just because the term "causal" was not used does not mean that it was not implied all along. For example, Curran and Bauer (2011) invoked the following "within-person process": When an individual engages in effective coping, this mitigates the effects of stress for them. The most plausible reading of this "process" is that effective coping has a *causal* effect on various relevant outcomes for the individual.

In some parts of the psychological literature, concerns about the within/between-persons distinction have taken center stage. Here, we argue that researchers should change gears and put causal inference upfront when planning to collect and analyze longitudinal data. The within/between distinction plays only an instrumental role in this endeavor, and we make three points to clarify its substantive utility. First, it is not necessary to investigate within-persons associations to identify causal effects—other designs can do so, too. Second, it is not sufficient to investigate within-subjects associations to identify causal effects—confounding can still be an issue. Third, although longitudinal within-persons data are neither necessary nor sufficient for causal inference, they still can be tremendously helpful. They can aid causal identification, allowing researchers to relax some assumptions; they can inform about interindividual differences in causal effects; and they can give a more dynamic view of how effects unfold over time. We conclude with some recommendations for how to approach longitudinal data analysis from a causal-inference perspective. With this article, we aim to provide some general, nontechnical guidance for researchers who may have some initial experience with longitudinal-data analysis but are less familiar with the causal-inference literature, who would like to understand how those two topics are connected, and who are looking for entry points into the more technical literature on causal inference and statistical models of longitudinal data.

Within-Persons Data Are Not Necessary for Causal Inference

In the potential-outcomes framework (Holland, 1986; Rubin, 1974), individual causal effects are defined as differences in potential values of an outcome variable (Y) under different treatments (A). We start with an example that reflects a "typical" question that researchers may aim to answer with observational within-persons data: Does talkativeness, a behavior associated with the personality trait extraversion, increase subjective well-being? As a starting point for a formalization of this effect, we focus on a single unit (for our purposes, an individual), measured without error, and we consider the case of a binary independent variable to simplify matters (for a more comprehensive introduction, see West & Thoemmes, 2010). Here, Y may be an individual's subjective well-being at the end of today, and A may refer to the treatment of spending the day being talkative (A = 1) as opposed to spending the day being untalkative (A = 0; we return to the actual realization of such a treatment below). If the person was talkative, $Y^{a=1}$ would be observed; if the person was untalkative, $Y^{a=0}$ would be observed. These two values ($Y^{a=1}$, $Y^{a=0}$) are the individuals' potential outcomes. The individuals' causal effect of being talkative today on their subjective well-being at the end of the day is defined by the contrast between the two, $Y^{a=1} - Y^{a=0}$. This individual causal effect is unobservable. Today, the individual will have been *either* talkative or untalkative, so only one of the individual potential outcomes ($Y^{a=0}$ or $Y^{a=1}$) will be realized as Y and become observable. How could researchers possibly recover it from between-persons data?

In virtually all circumstances, researchers cannot. With between-persons data, the individual causal effect is out of reach. However, if researchers collect data from multiple people for a single day, it can become possible to estimate the average of their individual causal effects. This works best in randomized experiments, and most psychological researchers will be aware of the special status of this research design. Nonetheless, it is worth spelling out the details to clarify some terminology and crucial assumptions.

Consider the possibility of a randomized experiment in which researchers assign a large number of people to either spend the day being talkative or spend the day being untalkative. For the sake of the argument, pretend there was a psychological intervention that manipulates individuals' talkativeness in a highly reliable and targeted manner for the duration of a single day, rendering them either talkative or untalkative to an exactly prespecified degree without any unintended side effects.

Before this hypothetical intervention, individuals' "natural" talkativeness may be correlated with the potential outcomes. For example, people who are more talkative might be people who are happier generally, meaning that both of their potential outcomes are higher. But after the randomized treatment, the assigned talkativeness will not be correlated with the individuals'

potential outcomes $Y^{a=0}$ and $Y^{a=1}$; this means that the two treatment groups are exchangeable with respect to their potential outcomes. Thus,

$$E\left[Y^{a=1} - Y^{a=0}\right] = E\left[Y^{a=1}\right] - E\left[Y^{a=0}\right] \quad (1)$$

$$= E\left[Y^{a=1} \mid A = 1\right] - E\left[Y^{a=0} \mid A = 0\right] \quad (2)$$

$$= E\left[Y \mid A = 1\right] - E\left[Y \mid A = 0\right]. \quad (3)$$

The expected value of the individual-level causal effect across individuals is the arithmetic mean; the expected value of a difference is equivalent to the difference between the expected values (Equation 1). Because the groups are exchangeable with respect to their potential outcome, their expected potential outcomes will not systematically vary; one can thus substitute the expected potential outcomes across all individuals with the expected potential outcomes in the respective groups (Equation 2). In the two different groups, the respective potential outcomes are realized (Equation 3). The assumption to ensure the equivalence between Equation 2 and Equation 3 is called "consistency," and we discuss potential violations later.

Thus, by taking the difference between the group means of the outcome, researchers recover the average of the individual causal effects, which is often called the average treatment effect. This fundamental property is what renders randomization such a valuable tool. And it clearly demonstrates that between-persons data can inform about within-persons processes—albeit only in aggregate.[1]

Going beyond experimental data, average causal effects can also, at least in theory, be estimated with the help of nonexperimental between-persons data (including natural experiments). Such attempts require strong additional assumptions (for accessible introductions, see e.g., Elwert, 2013; Hernán & Robins, 2010; Pearl et al., 2016; Rohrer, 2018; Rosenbaum, 2017), which may range in their plausibility from defensible to untenable, depending on the application. How does this mesh with the often-emphasized fact that between- and within-persons associations are statistically independent (e.g., Schmitz & Skinner, 1993)? Although these associations are sometimes also labeled between- and within-persons "effects," they do not refer to causal quantities. The between- and within-persons associations are a combination of (a) noncausal associations between the two variables of interest (e.g., associations induced by confounders) and (b) causal associations (induced by effects flowing either way).

Considering noncausal associations, we argue that one of the reasons why between- and within-persons associations may diverge is that some confounders—time-invariant factors, such as stable sociodemographic variables but also stable personality traits—will affect only the between-persons associations but not within-persons associations. This is because within-persons associations are calculated on the basis of time-by-time fluctuations within a person, and time-invariant factors do not have any variance within persons.[2] In practice, this means that between-persons associations (e.g., between talkativeness and happiness) can often plausibly be explained away by time-invariant third variables (e.g., gender, age, childhood socioeconomic status, stable personality traits). In contrast, within-persons associations cannot be explained away by third variables that are stable and have constant effects for the duration of the data collection.[3]

Considering causal associations, we argue that another reason why within- and between-persons associations may diverge is scenarios in which causal associations between the variables of interest cannot be captured by within-persons associations. There may be a lack of within-persons variability in the independent variable of interest over the course of the study, or the causal effects may unfold over a time frame longer than the duration of the study. We return to the issue of time frame when discussing design parameters.

Within-Persons Data Are Not Sufficient for Causal Inference

We have shown that within-persons data are not *necessary* for causal inference. But they are not *sufficient* either: Longitudinal data on their own do not justify causal inferences. The reason for this is, once again, confounding. As discussed above, within-persons associations are not affected by time-invariant confounding factors with constant effects. However, they can still be influenced by time-varying confounding factors.

Assume researchers had intensive within-persons data of individuals' talkativeness and subjective well-being. This allows them to compare days on which they were talkative to days on which they were untalkative. But talkativeness was not randomized, so it is possible that the treatment (being talkative vs. untalkative) is correlated with the potential outcomes (potential well-being that day).

For example, social events (e.g., dates or parties) may affect both talkativeness and happiness. First consider the "contemporaneous" association between the reported level of talkativeness on a given day and happiness at the end of that day. This association will be confounded because talkative days are not exchangeable—they are days on which more social events happened, and those events alone may be sufficient to make one happier.

Next, consider the lagged association between the reported talkativeness on a given day (Day 1) and happiness at the end of the next day (Day 2). Researchers might think that confounding by social events is no longer an issue when they adjust for happiness on Day 1 because this may already capture the confounding influence of social events. However, this depends on the time course over which the causal effects of social events unfold. If the events immediately induce talkativeness (small talk at the party) but affect happiness more slowly over multiple days (the warm, ongoing glow after having reconnected with old friends), then they will end up confounding the lagged associations as well. And in this particular substantive example, the lagged effect may not be particularly informative; to begin with—if researchers assume that the affective benefits of talkativeness are reaped immediately, they would expect the benefits to be mostly captured in happiness on the same day rather than the next day.

In short, causal inferences based on within-persons data still rest on the assumption that all (time-varying) confounders have been appropriately adjusted for.

Within-Persons Data Can Be Very Helpful for Causal Inference

As we have explained, both between- and within-persons data require strong assumptions to warrant causal inference. However, the use of within-persons data allows one to relax certain assumptions. Thus, although researchers still need to think clearly about the remaining assumptions, within-persons data can aid causal inference.

As stated above, within-persons associations are in principle not affected by time-invariant confounders. Within-persons data thus have the potential to control for various types of between-persons confounding, including *unobservable* confounders. One of the simplest ways to do this is through so-called fixed-effects models (Hamaker & Muthén, 2020; Imai & Kim, 2019; McNeish & Kelley, 2019). Although fixed-effects models are not very commonly used in psychological research, the approach is conceptually equivalent to the procedure in which variables are mean-centered within persons before entering them into multilevel models (e.g., Hamaker & Muthén, 2020; McNeish & Kelley, 2019). In Box 1, we go into detail about the assumptions under which the fixed-effects model can identify contemporaneous causal effects. This model focuses on the contemporaneous effect of X on Y and not on their broader causal dynamics.

Psychologists are, of course, often interested in estimating precisely these causal dynamics, which may explain why models including reciprocal effects are so much more popular in the psychological literature. Can one identify such reciprocal causal effects in longitudinal data? Recent literature has suggested that the most widely used form of such models, the cross-lagged panel model (Box 2), does not sufficiently control for between-subjects confounding despite its use of longitudinal data (Hamaker et al., 2015).

The shortcomings of the cross-lagged panel model have prompted researchers to use modifications of the model to separate such between-subjects confounding from cross-lagged effects. In psychology, the random-intercept cross-lagged panel model (Hamaker et al., 2015) has become the most popular choice, but similar models have been proposed in other fields. In Box 3, we highlight one such model, the dynamic panel model, which is a combination of the fixed-effects model with the cross-lagged panel model.

Whether these modified approaches are sufficient to adjust for all time-invariant confounders still depends on additional assumptions about the precise nature of the confounding (e.g., Lüdtke & Robitzsch, 2022; Murayama & Gfrörer, 2022). Furthermore, models are usually unable to identify both contemporaneous and lagged effects simultaneously. This highlights that there is no "one-size-fits-all" procedure to enable causal inference. Instead, researchers need to be very clear about the type of causal effects they want to examine (e.g., lagged effect vs. contemporaneous effects) and to carefully evaluate the underlying assumptions. However, even if those assumptions may be deemed unrealistic, observational longitudinal data combined with an appropriate model may often provide answers that are "less wrong" (i.e., potentially less biased) than answers provided by observational cross-sectional data, all else being equal.

Helpful further readings

Maybe because of psychology's fraught relationship with causality (Grosz et al., 2020), the literature on the many models discussed in the field—such as varieties of change-score models, cross-lagged models, and latent-curve models—is unfortunately not always transparent with respect to the assumptions under which these models can successfully identify causal effects. However, more recently, researchers have tried to bridge the gap between longitudinal data modeling in psychology and causal inference.

Gische et al. (2021) introduced graphical causal models for researchers familiar with structural equation modeling and the cross-lagged panel design, and Voelkle et al. (2018) provided a more general discussion of the role of time for understanding psychological mechanisms, which they quite explicitly described as a sequence of causal effects. Usami et al. (2019) provided

Box 1. The Fixed-Effects Model

The fixed-effects approach (or alternatively, within-persons mean centering) can control for unobserved time-invariant confounders whose effects do not change over time. For example, when considering the effects of talkativeness on happiness, extraversion (a stable personality trait) may be such a confounder: Extraverted individuals are habitually more talkative, but extraverted individuals may also simply be dispositionally happier. Figure 1, which has been adapted from (Hamaker & Muthén, 2020, p. 367), shows the causal model underlying the standard fixed-effects model. Note that this model focuses only on the (contemporaneous) effects of X on Y, and X is treated as exogenous (i.e., the model does not impose constraints on the causal relationship between X variables and U).

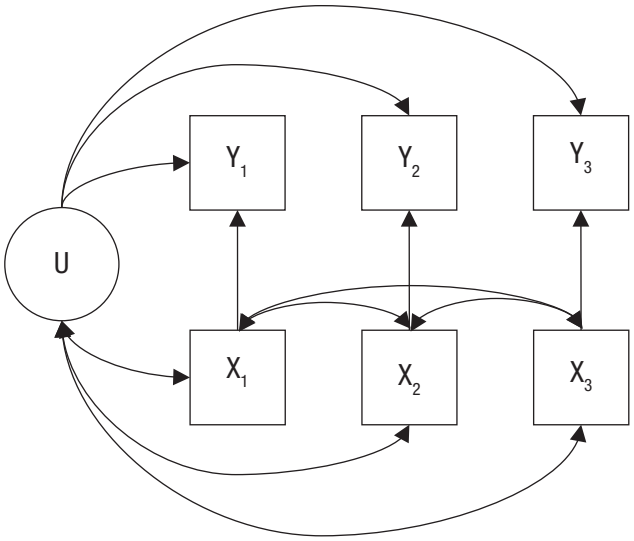

Fig. 1. Causal graph underlying the fixed.

When does this model successfully identify the effects of X on Y? As indicated by the *absence* of certain arrows in Model 1, one has to assume that there are no lagged causal dynamics; for example, past happiness does not affect current talkativeness (no cross-lagged paths from Y to X), past talkativeness does not affect current happiness (no cross-lagged paths from X to Y), past happiness does not affect current happiness (no autoregressive paths among the Y). Such dynamics will bias estimates, although the standard model can be modified to partially relax assumptions (Imai & Kim, 2019).

Another common scenario in which a fixed-effects model is biased occurs if people vary in their change over time (i.e., heterogeneous slopes) and if these differences are related to the effect of interest; this can be addressed by another modification of the model (Rüttenauer & Ludwig, 2020). Finally, one has to assume that any time-varying confounder has been included in the model, which is also the case for the following models (Box 2, Box 3).

The fixed-effects model considers only within-persons changes over time. In our example, the resulting effect estimate would be informed only by those people who actually do experience some changes in their talkativeness over the course of the study.

a discussion of the causal assumptions underlying popular models. Both Andersen (2021) and Lüdtke and Robitzsch (2022) discussed different classes of longitudinal models with respect to the conditions under which they recover the (cross-lagged) causal effects of interests and conditions under which they will result in equivalent results. Finally, Zyphur et al. (2020) developed a comprehensive general cross-lagged panel model as a generic approach to translate assumptions into a statistical model.

Of course, other fields have also tackled the issue of causal inference with longitudinal data. For example, the sociologists Elwert and Pfeffer (2019) developed an approach that uses future values of the independent variable to detect and reduce omitted variable bias. In epidemiology, the particularly promising approach of marginal-structural models (Cole & Hernán, 2008; Robins et al., 2000) has been developed. These models implement a multistep estimation procedure to control for time-varying confounding variables (Williamson & Ravani, 2017). The promise of such models for causal inference in psychology, however, has not yet been well recognized (Lüdtke & Robitzsch, 2020; Usami, 2020). Thoemmes and Ong (2016) provided an introduction to marginal-structural models in combination with inverse probability weighting as a means for third-variable adjustment in longitudinal data, including annotated SPSS and R code for psychologists. The tutorial by Bray et al. (2006) showcased an implementation in SAS and highlighted how this method can, unlike other common methods, successfully adjust for time-varying confounders. Finally, VanderWeele et al. (2020) developed a comprehensive template for so-called outcome-wide longitudinal designs in which the goal is to identify the

Box 2. The Cross-Lagged Panel Model

The cross-lagged panel model (Fig. 2a) has different aims than the fixed-effects model. First, it aims to identify *lagged* effects (not the contemporaneous effects examined in fixed effects model). Second, it usually aims to identify *reciprocal* effects (i.e., X influences Y *and* Y influences X). The model addresses so-called Granger causality (Granger, 1969), but Granger causality is a form of prediction—and such prediction implies causation only when certain assumptions are met. For example, the model provides biased causal estimates when there are contemporaneous causal effects (e.g., current happiness affects current talkativeness). This highlights trade-offs when trying to simultaneously consider contemporaneous and lagged effects that were also discussed by Imai and Kim (2019).

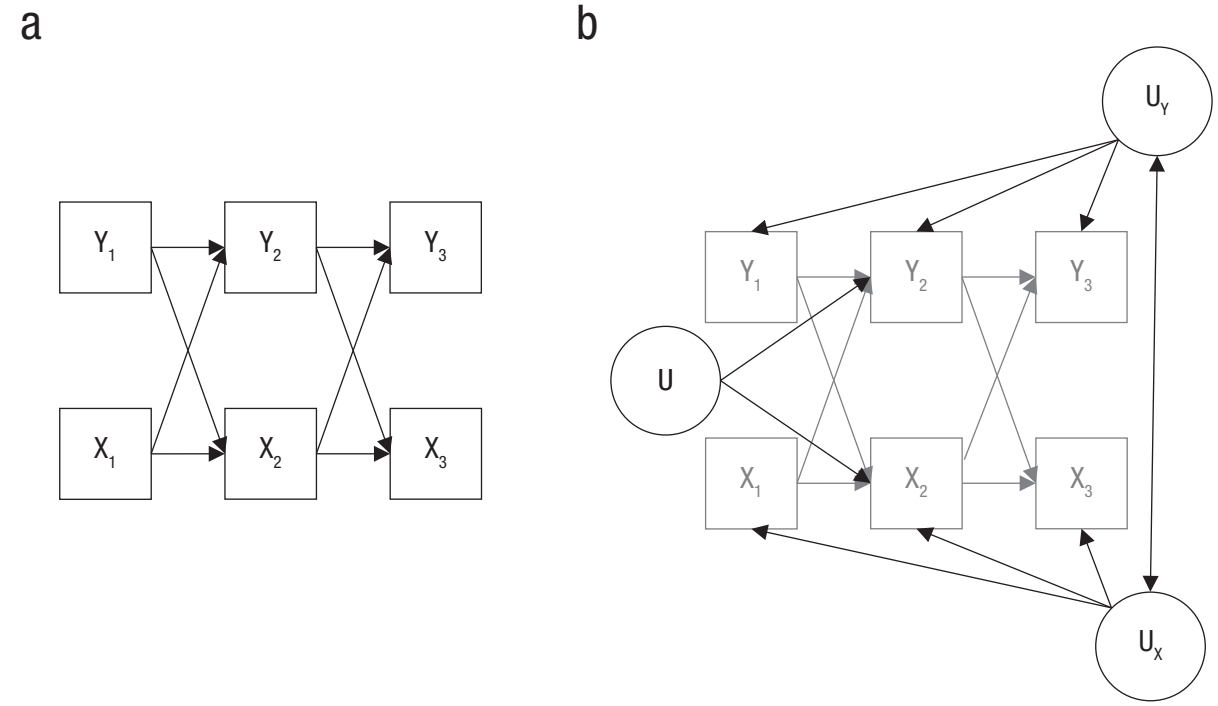

Fig. 2. (a) Causal graph underlying the cross-lagged panel model. (b) Different scenarios involving unobserved confounders.

The cross-lagged panel model can partly account for unobserved confounders. For example, when one is interested in the causal effect of X_2 on Y_3, the existence of the unobserved confounder U, which has only a temporal effect, is unproblematic (Fig. 2b): The confounding goes through Y_2, which is included in the model and is thus statistically accounted for. However, the model fails if the constructs are trait-like in nature. The existence of more stable effects U_X and U_Y (Fig. 2b) would be problematic because they directly open a confounding path between X_2 and Y_3. Thus, the resulting estimates of the cross-lagged paths would be biased.

causal effects of an independent variable on a number of outcome variables and longitudinal data are leveraged to reduce concerns about confounding.

Additional advantages of longitudinal data

Aside from causal identification in the narrow sense (getting rid of confounding), which often focuses on average effects, longitudinal data may also enhance causal inference for other reasons. First, longitudinal data can improve the understanding of how causal effects unfold over time (Voelkle et al., 2018). Second, they may provide the means to actually estimate individual-level causal effects. Causal effects may vary between individuals, and researchers can take into account such between-persons variability of causal effects with longitudinal data. The optimal approach to identify such effects are experiments in which one observes individuals repeatedly in different experimental conditions (see Fine Point 2.1, Hernán & Robins, 2020, p. 16), and there

Box 3. The Dynamic Panel Model

Dynamic panel models exist in several different versions, but we focus on the model depicted in Figure 3. Note that Figure 3 omitted some details for the purpose of simplicity, but a more detailed practical tutorial of the model can be found in Dishop and DeShon (2021). This dynamic panel model has the same goal as the cross-lagged panel model—it aims to identify lagged reciprocal causal effects. Like the fixed-effects model, it takes into account (constant) effects of time-invariant confounders; like the cross-lagged panel model, it allows for reciprocal lagged dynamics. Although the model can control confounders that the cross-lagged panel model cannot (i.e., time-invariant confounders), important assumptions of the more basic models (Box 1, Box 2) still apply: Like in the fixed-effects model, we need to make assumptions about the type of time-invariant confounders (i.e., no heterogeneous slopes). And like in the cross-lagged panel model, the existence of contemporaneous causal effects would bias our estimates of the cross-lagged causal effects.

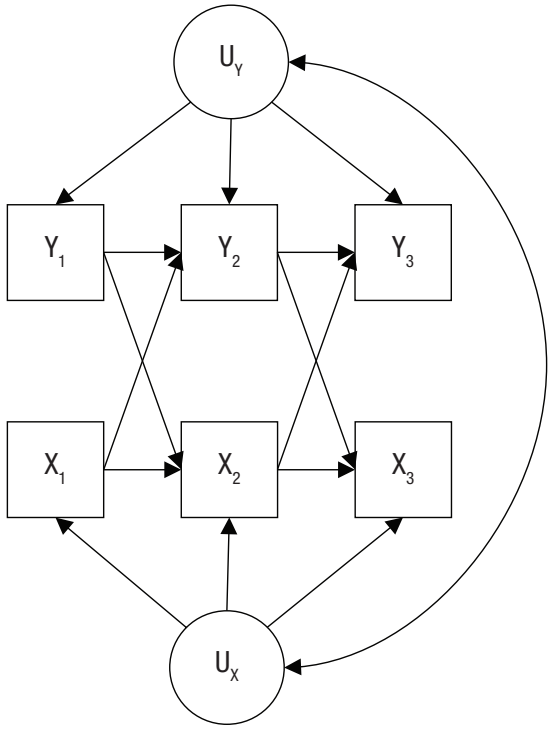

Fig. 3. Causal graph underlying the dynamic panel model.

have been some recent methodological developments to obtain a better causal estimate with this type of design (Schmiedek & Neubauer, 2020). Doing this with observational longitudinal data once again requires more and stronger assumptions, and this is an important avenue for future methodological work.

Making the Most of Within-Persons Data

In recent years, we have observed considerable enthusiasm for the within-persons approach in psychology, with various advanced statistical models proposed. In line with this enthusiasm, we believe that within-persons data are a promising way to advance causal inference. Yet we also feel like its promises have led people to put the technological and methodological cart before the conceptual horse. Researchers may decide to collect within-persons data with an experience sampling study because it is the innovative thing to do right now; they may decide to apply certain statistical models because they appear novel and highly sophisticated. Journals may further implicitly reinforce this style of research when they automatically dismiss studies that are "merely cross-sectional" or do not employ "sophisticated statistical modeling." An approach that we believe to be more productive, and which we describe in the following, puts the substantive question first. Although this may sound trivial—of course the substantive question should be the starting point of any empirical investigation—debates such as the one surrounding the age trajectory of happiness (Kratz & Brüderl, 2021) and the interpretation of the Many Analysts project (Auspurg & Brüderl, 2021; Silberzahn et al., 2018) highlight how arguments often focus on statistical aspects when in fact researchers do not even agree about which substantive question is being addressed.

Setting the analysis goal

Researchers should start by explicitly spelling out the theoretical estimand of interest in precise terms that exist outside of any statistical model (Lundberg et al., 2020). At this point, it may become clearer whether the research question targets causal quantities—but even noncausal endeavors, such as "description of developmental trajectories," require conceptual clarity. This estimand in combination with the additional assumptions researchers are (or are not) willing to make determine which research design is appropriate, be it experimental

or nonexperimental, cross-sectional or longitudinal, needing many time points or not.

What does such a well-defined estimand look like? Psychologists like to make claims about broad concepts (Yarkoni, 2020) and to address broad research questions ("What is the interplay between talkativeness and happiness?"). However, from a causal-inference perspective, things need to be broken down and taken more slowly (see also Rohrer et al., 2022). For example, a more tractable research question may concern the effect of being continuously talkative (vs. continuously untalkative) for a certain defined amount of time on well-being immediately after the episode. Formalization, for example with the help of the potential-outcomes model, makes it explicit that causal effects are defined by contrasts of specific treatments on specific outcomes. Treatments may be time-varying—there are many different sequences of talkativeness and untalkativeness that one could contrast to learn something about the effects of talkativeness on happiness (for an introduction to time-varying treatments, see Hernán & Robins, 2020, Chapter 19)—and outcomes can be evaluated at different points in time.

Thus, there is no such thing as "the" effect of talkativeness on happiness; there are many different possible theoretical estimands that may all be deemed informative with respect to the overarching question of whether and how talkativeness affects happiness. This nuance may get lost if researchers simply apply an out-of-the-box model, which may target a different estimand than the one they actually have in mind. Furthermore, insufficient clarity about estimands may lead to ostensible contradictions between empirical studies that in reality target different estimands.

Once researchers try to be more precise about the causal effects they have in mind, they may run into deeper issues. Theories in psychology are often quite underspecified—a topic that was addressed in a recent special issue of the journal *Perspectives on Psychological Science* (Volume 16, Issue 4, July 2021)—and may thus make only vague predictions about which causal effects are to be expected. But even if theories were more precise, causal effects involving psychological variables pose conceptual obstacles. Although these are neither unique nor central to the within/between distinction and thus outside of the focus of the present article, we provide more details in Box 4 for interested readers.

Identification strategy and design parameters

Once researchers have settled on an estimand, they can start thinking about appropriate identification strategies. Accessible articles provide some guidance on this step (Foster, 2010a; Grosz et al., 2020). Considerations may include whether a sufficiently targeted intervention is available to manipulate talkativeness (but see Eronen, 2020) and thus whether an experiment is plausible, whether a suitable natural experiment may exist (e.g., a situation that affects talkativeness in a plausibly random manner), and which time-invariant or time-varying confounders are deemed relevant.

If within-persons data turn out to be a productive way forward—for example, because time-invariant confounders are deemed particularly relevant and because researchers can assume that there is relevant within-persons variability in the independent variable of interest—the causal angle can clarify specific design parameters. Consideration of potential time-varying confounders tells researchers what needs to be measured. Consideration of the precise definition of the causal effect of interest tells researchers which time lag between assessments is sensible.

Discussions of the appropriate time lag in psychology often focus on attempts to uncover the true underlying dynamic system (Haslbeck & Ryan, 2021), which is, of course, unknown. Hence, one might conclude that the narrowest possible sampling is desirable because it still allows one to estimate effects with a wider lag (e.g., in the crudest case, one might just drop the measurement points in between). In practice, when it comes to time lags, pragmatic concerns need to be considered as well. High-frequency sampling can overburden participants, and there is a very real possibility that the assessment interferes with the causal system of interest. Self-reporting positive affect 100 times a day may influence mood; filling out a personality questionnaire over and over again may change the way individuals answer the items. Thus, the smallest possible time lag is not always advisable. But if researchers decide that they want to investigate a relatively well-defined specific causal effect, such as "the *immediate* effect of picking up one's smartphone on well-being" or "the effect of cumulative smartphone usage over the course of a day on well-being on the next day," the research question already implies how data need to be collected.[4]

Statistical estimation

If the theoretical estimand is set, the aim of the statistical analysis is to provide an actual empirical estimate. We have already extensively referred to the longitudinal modeling literature above and will thus just briefly emphasize a central concern. Psychological researchers have often relied on out-of-the-box longitudinal models such as cross-lagged panel models, which could, in principle, be applied to any pair of variables. Such default solutions have multiple shortcomings. First, we note that

Box 4. Hypothetical Interventions, Real-World Complications

Earlier, we alluded to a hypothetical intervention that fixes an individual's talkativeness at a given level without any side effects. Such an intervention does not exist—no psychological intervention will achieve precisely the desired level of talkativeness for everyone. Furthermore, the intervention may affect all sorts of other variables, and some of these may in turn explain any effect of the intervention on well-being. In general, psychological variables as causes pose challenges because interventions targeting them are often "fat-handed" (Eronen, 2020), meaning that they will affect multiple variables simultaneously. This not only constrains experimentation but also makes it hard to pin down which hypothetical states of the world one has in mind when estimating causal effects on the basis of observational data. For example, researchers likely do not encounter many situations in which individuals' talkativeness could have varied while all other psychological variables were held constant.

And there is a second concern that goes beyond this. The effects of talkativeness may depend on how talkativeness was induced (e.g., by an individual's genetic disposition, by a specific situation, or by a psychological intervention). In such a scenario, does it even make sense to talk about effects of talkativeness per se? An example from a different field of research may illustrate the matter more clearly. Does obesity shorten life? If one takes a particular individual, there may be many different ways to intervene on their body weight. For example, one may put them on a specific diet or a specific exercise regime or chop off a body part. Any of these will affect body weight, but how this change in body weight subsequently affects mortality may vary between interventions. Formally speaking, this violates the assumption of consistency (i.e., the assumption needed to ensure the equivalence of Equations 2 and 3 described earlier), and Hernán and Taubman (2008) went so far as to state the effects of body mass index (BMI) on mortality in observational data cannot be well defined. In contrast, the effects of specific interventions on BMI can be well defined and can at least potentially be recovered from observational data. In line with this, Hernán and Robins (2016) argued that researchers should conceptualize (hypothetical) target trials to precisely define which causal effects they are interested in.

However, this "interventionist" account of causality has been challenged. Pearl (2018) championed a structural account of causality in which causal relations exist independently of hypothetical interventions. In this account, consistency is not an assumption but a theorem. Actual interventions may often have side effects that need to be reckoned with, but these do not render causal effects inconsistent.[5]

We do not aim to solve this philosophical debate but would like to highlight that degree of concreteness in variables/constructs renders causal inference more or less challenging, and psychological variables tend to be less concrete. "Obesity" as an independent variable is much more concrete than concepts such as "talkativeness" or "subjective well-being" (Rohrer & Lucas, 2020). Assuming an interventionist account of causality, problems may arise because one cannot even come up with hypothetical targeted interventions or because consistency fails (the effects of talkativeness may vary depending on whether talkativeness is induced through drugs or through verbal encouragement), resulting in ill-defined causal effects. Assuming a structural account of causality, problems may arise because one lacks knowledge of the structure of the causal web linking psychological variables (e.g., talkativeness, extraversion, and other personality traits). In this case, effects are not necessarily ill defined, but estimating them may still be virtually impossible given the current state of knowledge.

currently, the psychological literature on within-persons associations is not well integrated with the causal-inference literature—thus, for at least some of the out-of-the-box solutions, it is unclear or at least untransparent which causal effect is targeted by the analysis. Second, the causal webs linking different sets of variables can look very different, and a model that is not tailored to the specific underlying causal web cannot recover the causal effects of interest. Third, many published implementations of these models pay less attention to including measured (time-invariant and time-varying) confounders, further limiting the chances that the causal effects of interest will be recovered.

Interpreting model results

Finally, once the model has been estimated, how does one interpret it? In the psychological literature, model coefficients associated with particular paths are often treated as the relevant analysis output. But even if researchers are lucky and their coefficients correctly identify the causal processes of interest, they may still

not provide a straightforward answer to their (causal) research question. For example, the model-implied effects of X at a given point in time on Y at a given point in time may vary between individuals because of either systematic heterogeneity (e.g., interactions) or nonlinearity (for a discussion of why in nonlinear models, "everything interacts," see Rohrer & Arslan, 2021). Furthermore, in a cross-lagged panel model, the effect of a predictor on the outcome at a later time point can be the sum of both direct and indirect effects (Lüdtke & Robitzsch, 2022) so that multiple coefficients need to be added up when contrasting different states of the world.

Here again, causal thinking can clarify how to summarize effects in complex models in an interpretable manner. We can once again consider a hypothetical intervention and use the model to predict how it would affect individuals' outcomes at a point in time we consider relevant. Gische et al. (2021) demonstrated how to work with such hypothetical interventions in the context of cross-lagged panel models and how to calculate both average and person-specific effects.

The general framework for using models to make predictions about the effects of various interventions are so-called marginal effects. Marginal effects have received comparatively little attention in psychology outside of methods journals; they are more common in, for example, sociology (e.g., Mize et al., 2019), possibly because the statistical software Stata, popular in that field, makes calculating them quite easy (Williams, 2012). However, there are now packages available that allow researchers to calculate marginal effects in R using structural equation models (Mayer, 2019; Mayer et al., 2016) and a vast number of other model classes, including multilevel models (Arel-Bundock, 2022; Lenth, 2022). As far as we know, no comprehensive primer to marginal effects for psychologists has been published to date, but recent blog posts have tried to provide a gentle introduction (Heiss, 2022; Rohrer, 2022).

Should researchers just do description instead?

Having read so far, readers may feel that inferring causal effects from (nonexperimental) within-persons data in an overwhelming task and may instead prefer a "descriptive" approach. Indeed, many longitudinal analyses claim to be descriptive in nature, although that term may be used in an ambiguous manner. This may partly be a strategic move to avoid the heightened scrutiny that results from overtly causal claims (Alvarez-Vargas et al., 2020; Grosz et al., 2020). We do believe that descriptive research is currently undervalued in psychology (see e.g., Scheel et al., 2020). But many models in psychology are too complex to produce good descriptions (Foster, 2010b), and this holds true for longitudinal models in which the explanation for how coefficients behave quickly turns opaque.

An actually informative descriptive analysis should involve much more basic description than researchers routinely encounter in studies analyzing longitudinal data. For example, a fruitful first step to describe associations in longitudinal data may consist of a fixed-effects model in its most basic specification or the equivalent multilevel model with within-persons centering. This tells the strength of the contemporaneous association between the variables after removing stable between-persons differences in the level of the predictor and the outcome. With sufficient data points per individual, researchers may even simply calculate the bivariate association for every individual in isolation. As we discussed earlier, any association in such analyses may still be confounded and thus does not necessarily provide a convincing causal estimate, but at least researchers will have narrowed down the range of confounders while sticking with estimates that still have a straightforward descriptive interpretation. Reporting results from such analyses before moving on to more complex models also mirrors established practices in cross-sectional studies in which authors routinely report bivariate correlations before moving on to more complex regression analysis.

In contrast, it is not easy to apply the "standard" longitudinal models for descriptive purposes. The literature is filled with many related "state-of-the-art" longitudinal models (e.g., the autoregressive-latent-trajectory model, Bollen & Curran, 2004; the random-intercept cross-lagged panel model, Hamaker et al., 2015; the stable trait/autoregressive trait/state model, Kenny & Zautra, 1995; the dual-change score model, McArdle & Hamagami, 2001), and these in turn can usually be specified in multiple ways. Assuming researchers chose the right model that correctly reflects the data-generating mechanism, they could elegantly capture the underlying causal within-persons dynamics. But in reality, researchers do not know which model generated the observed data, and presented with a daunting number of different models and little guidance on which nuances matter and which do not, they may resort to the standard that is accepted in the field—and that might not be optimal, as is known from the story of cross-lagged panel models (Hamaker et al., 2015). Trying to uncover the complete causal dynamics of a system is a more ambitious task than identifying a specific causal effect; once researchers fail to uncover the complete dynamics, the interpretation of any specific component of the model becomes questionable.

Putting causal inference upfront, researchers are still confronted with a challenging task, but one that is

potentially more tractable because there is at least a clear circumscribed analysis goal: recovery of a specific causal effect. This also opens the possibility to use available experimental evidence as a benchmark to evaluate observational longitudinal analysis—if a longitudinal model implies certain effects that contradict existing evidence from intervention studies targeting similar cause and effect, this can at least be taken as a warning sign (for an implementation of this logic, see Wan et al., 2021). The choice of the data-analytic model matters only insofar as it should map onto the assumptions about the underlying causal web that researchers are willing to make. These assumptions will often be strong and potentially unrealistic—there is no free lunch—but at least researchers are actually tackling the question of interest.

Consider, for example, the debate surrounding the age trajectory of happiness (Galambos et al., 2020). This actually seems to be one of the easier questions one could answer with longitudinal data, yet it has spawned a bloated literature and lots of confusion about how to specify the model. If one tackles the problem from a causal-inference perspective, as demonstrated by Kratz and Brüderl (2021), it becomes clear that some analytic decisions are just wrong (e.g., statistical adjustment for mediators, which makes sense only if researchers are trying to address a *different* research question), whereas others hinge on additional assumptions (e.g., about the existence and shape of period and cohort effects). This does not mean that the debate is automatically settled, but at least one can pinpoint where exactly analysts disagree and how to make progress on the research question.

We believe that a better understanding of causal inference and how it can be enhanced with the help of within-persons data has the potential to clarify other debates in psychological research as well, resulting in an overall improvement of the quality of our inferences.

Transparency

Action Editor: Pamela Davis-Kean
Editor: David A. Sbarra
Author Contributions

Julia M. Rohrer: Conceptualization; Writing – original draft; Writing – review & editing.

Kou Murayama: Conceptualization; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research is partly funded by Leverhulme Trust Research Leadership Award (RL-2016-030) and JSPS Kakenhi (18H01102) to Kou Murayama.

ORCID iD

Julia M. Rohrer https://orcid.org/0000-0001-8564-4523

Acknowledgments

We thank Henrik Andersen, Ruben Arslan, Tobias Debatin, Ellen Hamaker, Oliver Lüdtke, Brent Roberts, Anne Scheel, Stefan Schmukle, Felix Thoemmes, Satoshi Usami, and Manuel Voelkle for their feedback on a draft of this article. We also thank Hayley Jach for proofreading the manuscript.

Prior Versions

An earlier version of the manuscript was made available on PsyArXiv: https://psyarxiv.com/tg4vj/.

Notes

1. If we additionally assume that the causal effect is precisely the same for every individual, then this is a nonissue because the average causal effect will equal every single individual level causal effect (e.g., West & Thoemmes, 2010). However, many psychologists would probably object to this assumption.
2. Sometimes people define within-persons associations as computed by mean-centered data of multiple persons (see section, Within-Persons Data Can Be Very Helpful for Causal Inference, and Box 1). With this definition, within-persons associations can suffer from time-invariant confounders that have varying effects over time (Usami et al., 2019). In this scenario, within-persons associations are protected *only* from time-invariant confounders with constant effects but not from time-invariant confounders with time-varying effects.
3. This means that it can also be, in theory, possible to recover average within-persons associations from between-persons associations when the relevant time-invariant confounders are adjusted for, just like average causal effects can potentially be recovered from between-persons data (Murayama et al., 2017); but again, very strong assumptions are necessary.
4. Continuous time modeling (e.g., Driver et al., 2017) can be a useful tool to identify an underlying continuous process regardless of the particular time lags, and this information may be useful to determine an optimal lag for a new research design.
5. To render matters even more complex, there are more than two sides in this debate about the nature of causal inference. For example, Krieger and Smith (2016) highlighted the limits of both counterfactuals thinking (championed by Hernán) and directed acyclic graphs (as championed by Pearl) and proposed a broader and more flexible framework of "inference to the best explanation."

References

Alvarez-Vargas, D., Braithwaite, D. W., Lortie-Forgues, H., Moore, M. M., Castro, M., Wan, S., Martin, E. A., & Bailey, D. H. (2020). Hedges, Mottes, and Baileys: Causally ambiguous statistical language can increase perceived study quality and policy relevance. https://doi.org/10.31234/osf.io/nkf96

Andersen, H. K. (2021). Equivalent approaches to dealing with unobserved heterogeneity in cross-lagged panel models? Investigating the benefits and drawbacks of the latent curve

model with structured residuals and the random intercept cross-lagged panel model. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000285

Arel-Bundock, V. (2022). *marginaleffects: Marginal effects, marginal means, predictions, and contrasts*. https://vincentarelbundock.github.io/marginaleffects/

Auspurg, K., & Brüderl, J. (2021). Has the credibility of the social sciences been credibly destroyed? Reanalyzing the "Many Analysts, One Data Set" project. *Socius*, 7. https://doi.org/10.1177/23780231211024421

Bollen, K. A., & Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) Models: A synthesis of two traditions. *Sociological Methods & Research*, *32*(3), 336–383.

Bray, B. C., Almirall, D., Zimmerman, R. S., Lynam, D., & Murphy, S. A. (2006). Assessing the total effect of time-varying predictors in prevention research. *Prevention Science*, *7*(1), 1–17.

Cattell, R. B. (1952). The three basic factor-analytic research designs—their interrelations and derivatives. *Psychological Bulletin*, *49*(5), 499–520. https://doi.org/10.1037/h0054245

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, *168*(6), 656–664.

Dishop, C. R., & DeShon, R. P. (2021). A tutorial on Bollen and Brand's approach to modeling dynamics while attending to dynamic panel bias. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000333

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*(5). https://doi.org/10.18637/jss.v077.i05

Elwert, F. (2013). Graphical Causal Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 245–273). Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_13

Elwert, F., & Pfeffer, F. T. (2019). The future strikes back: Using future treatments to detect and reduce hidden bias. *Sociological Methods & Research*. https://doi.org/10.1177/0049124119875958

Eronen, M. I. (2020). Causal discovery and the problem of psychological interventions. *New Ideas in Psychology*, *59*, 100785. https://doi.org/10.1016/j.newideapsych.2020.100785

Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, *46*(6), 1454–1480. https://doi.org/10.1037/a0020204

Foster, E. M. (2010). The U-shaped relationship between complexity and usefulness: A commentary. *Developmental Psychology*, *46*(6), 1760–1766. https://doi.org/10.1037/a0020180

Galambos, N. L., Krahn, H. J., Johnson, M. D., & Lachman, M. E. (2020). The U Shape of Happiness Across the Life Course: Expanding the Discussion. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *15*(4), 898–912. https://doi.org/10.1177/1745691620902428

Gische, C., West, S. G., & Voelkle, M. C. (2021). Forecasting causal effects of interventions versus predicting future outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(3), 475–492.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The Taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *15*(5), 1243–1255. https://doi.org/10.1177/1745691620921521

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*(2), 207–233. https://doi.org/10.1207/s15327906mbr4002_3

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.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

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), 365–379.

Haslbeck, J. M. B., & Ryan, O. (2021). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research*, 1–32. https://doi.org/10.1080/00273171.2021.1896353

Heiss, A. (2022, June 30). *marginalia: A guide to figuring out what the heck marginal effects, marginal slopes, average marginal effects, marginal effects at the mean, and all these other marginal things are*. https://www.andrewheiss.com/blog/2022/05/20/marginalia/

Hernán, M. A., & Taubman, S. L. (2008). Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International Journal of Obesity*, *32*, (suppl_3), S8–S14. https://doi.org/10.1038/ijo.2008.82

Hernán, M. A., & Robins, J. M. (2016). Using Big Data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, *183*(8), 758–764.

Hernán, M. A., & Robins, J. M. (2020). Causal inference: What If. CRC Boca Raton, FL. https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

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

Imai, K., & Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data?: Unit fixed effects models for causal inference. *American Journal of Political Science*, *63*(2), 467–490.

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*(1), 52–59.

Kratz, F., & Brüderl, J. (2021). *The age trajectory of happiness*. *PsyArXiv*. https://doi.org/10.31234/osf.io/d8f2z

Krieger, N., & Smith, G. D. (2016). The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International Journal of Epidemiology*, *45*(6), 1787–1808. https://doi.org/10.1093/ije/dyw114

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

Lüdtke, O., & Robitzsch, A. (2020). Commentary regarding the section "Modelling the Effectiveness of Teaching Quality." *Zeitschrift Für Pädagogik Beiheft*, *1*, 210–223. https://doi.org/10.3262/ZPB2001210

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.

Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review*, *86*(3), 532–565. https://doi.org/10.1177/00031224211004187

Mayer, A. (2019). Causal effects based on latent variable models. *Methodology: European Journal of Research Methods for the Behavioral & Social Sciences*, *15*(S1), 15–28.

Mayer, A., Dietzfelbinger, L., Rosseel, Y., & Steyer, R. (2016). The EffectLiteR approach for analyzing average and conditional effects. *Multivariate Behavioral Research*, *51*(2–3), 374–391.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). American Psychological Association. https://doi.org/10.1037/10409-005

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. https://doi.org/10.1037/met0000182

Mize, T. D., Doan, L., & Long, J. S. (2019). A general framework for comparing predictions and marginal effects across models. *Sociological Methodology*, *49*(1), 152–189.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education—Current Trends: The Role of Competence Beliefs in Teaching and Learning*, 71–87.

Murayama, K., & Gfrörer, T. (2022). *Thinking clearly about time-invariant confounders in cross-lagged panel models: A guide for model choice from causal inference perspective*. PsyArXiv. https://doi.org/10.31234/osf.io/bt9xr

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Pearl, J. (2018). Does obesity shorten life? Or is it the soda? On non-manipulable causes. *Journal of Causal Inference*, *6*(2), Article 20182001. https://doi.org/10.1515/jci-2018-2001

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560.

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

Rohrer, J. M., & Lucas, R. E. (2020). *Causal Effects of Well-Being on Health: It's Complicated*. https://doi.org/10.31234/osf.io/wgbe4

Rohrer, J. M. (2022, May 27). Unleash your inner stats sparkle with this very non-technical introduction to marginal effects. *The 100% CI*. http://www.the100.ci/2022/05/27/%e2%9c%a8-unleash-your-inner-stats-sparkle-%e2%9c%a8-with-this-very-non-technical-introduction-to-marginal-effects/

Rohrer, J. M., & Arslan, R. C. (2021). Precise answers to vague questions: Issues with interactions. *Advances in Methods and Practices in Psychological Science*, *4*(2). https://doi.org/10.1177/2515245921100736

Rohrer, J. M., Hünermund, P., Arslan, R. C., & Elson, M. (2022). That's a lot to process! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science*, *5*(2). https://doi.org/10.1177/25152459221095827

Rosenbaum, P. (2017). *Observation and experiment: An introduction to causal inference*. Harvard University Press. https://doi.org/10.4159/9780674982697

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Rüttenauer, T., & Ludwig, V. (2020). Fixed effects individual slopes: Accounting and testing for heterogeneous effects in panel data or other multilevel models. *Sociological Methods & Research*. Advance online publication. https://doi.org/10.1177/0049124120926211

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2020). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*. https://doi.org/10.1177/1745691620966795

Schmiedek, F., & Neubauer, A. B. (2020). Experiments in the wild: Introducing the within-person encouragement design. *Multivariate Behavioral Research*, *55*(2), 256–276. https://doi.org/10.1080/00273171.2019.1627660

Schmitz, B., & Skinner, E. (1993). Perceived control, effort, and academic performance: Interindividual, intraindividual, and multivariate time-series analyses. *Journal of Personality and Social Psychology*, *64*(6), 1010–1028. https://doi.org/10.1037/0022-3514.64.6.1010

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E. E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Cervantes, I. F., . . . Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, *4*(1), 40–59. https://doi.org/10.1177/2167696815621645

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*(5), 637–657. https://doi.org/10.1037/met0000210

Usami, S. (2020). Within-person variability score-based causal inference: A two-step semiparametric estimation for joint effects of time-varying treatments. In arXiv [stat.ME]. *arXiv*. http://arxiv.org/abs/2007.03973

VanderWeele, T. J., Mathur, M. B., & Chen, Y. (2020). Outcome-wide longitudinal designs for causal inference: A new template for empirical studies. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *35*(3), 437–466. https://doi.org/10.1214/19-sts728

Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, *53*(6), 782–805.

Wan, S., Brick, T. R., Alvarez-Vargas, D., & Bailey, D. H. (2021). *Toward a causally informative fit index of longitudinal models: A within-study design approach*. PsyArXiv. https://doi.org/10.31234/osf.io/5cbkt

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*(1), 18–37. https://doi.org/10.1037/a0015917

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, *12*(2), 308–331.

Williamson, T., & Ravani, P. (2017). Marginal structural models in clinical research: When and how to use them? Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association. *European Renal Association*, *32*(suppl_2), ii84–ii90. https://doi.org/10.1093/ndt/gfw341

Yarkoni, T. (2020). The generalizability crisis. *The Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/S0140525X20001685

Zyphur, M. J., Voelkle, M. C., Tay, L., Allison, P. D., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods*, *23*(4), 688–716. https://doi.org/10.1177/1094428119847280