



Start even Smaller, and then more Random. Comment on “Start Small, not Random: Why does Justifying your Time-Lag Matter?” by Yannick Griep, Ivana Vranjes, Johannes M. Kraak, Leonie Dudda, & Yingjie Li

Christian Dormann^{1,2} 

¹ Johannes Gutenberg-Universität Mainz (Germany)

² University of South Australia (Australia)

Abstract. The focal article by Griep and colleagues raises some highly important issues and it is timely to further advance organizational behavior (OB) and work and organizational psychology (WOP). It comes at the right time because the last two decades can be characterized by two opposing trends. On the one hand, there have been exciting developments in statistical methods to appropriately model time in statistical analyses, while on the other hand, the vast majority of studies have not considered time in analyses, or have not done so appropriately. For the sake of brevity, I use ‘temporal design’ and ‘temporal analysis’ as umbrella terms. As Griep et al. succinctly claim, many opportunities for better temporal designs and temporal analyses have been missed and there is a strong need to do better in the future. In this commentary, I add to some of the important issues raised and call for changes in future research.

Received 7 March 2022; Revised 19 May 2022; Accepted 20 May 2022

Keywords: diary, longitudinal, panel, time lags.

My comment on the paper of Griep et al. (2021) is organized along four of their major arguments. First, Griep et al. note that diary studies allow investigating within-person processes, but authors of such studies yet mainly failed to elaborate on justifying appropriate time lags. (1) I fully agree, and there is little to argue about this point. Therefore, I will rather add on to this by elaborating how diary studies have probably even become the major reason *why* our scientific community failed elaborating on appropriate time lags, and that we should stop analyzing diary data as has been typically done. Griep et al. further note that we are in need theories of time lags. (2) Again, I fully concur with Griep et al., but I would even go one step further insofar as we are in need for two types of theories: *Substantive and statistical theories of time (lags)*, the latter perhaps being even more important than the former. Griep et al. summarized their subsequent review by concurring with

Cole and Maxwell (2003) that convenience and tradition have been the major reasons for choosing particular time lags. (3) This point, too, can hardly be doubted, and to further highlight that convenience and tradition are problematic, I will add two comments. Firstly, I will outline that such traditions have consequently blocked scientific progress, and secondly, as possible causes of such traditions, I propose an epistemological fallacy and Dormann’s (this issue) Gestalt Psychology “Law of Nice Numbers”. Griep et al. then reviewed possible effect trajectories over time and derived two conclusions, and I will comment on their first one, which is the need for “a clear description of what is meant by optimal time-lags”. (4) I fully agree, but I will argue that one of the most cited current suggestions of choosing optimal time lags (Dormann & Griffin, 2015) is not optimal chosen. Griep et al. closed with a focused discussion of their previous arguments made, and they propose four

Correspondence concerning this article should be addressed to Christian Dormann. Johannes Gutenberg-Universität Mainz. Fachbereich Rechts- und Wirtschaftswissenschaften. Johannes Welter Weg 9. RLP 5512 Mainz (Germany).

E-mail: cdormann@uni-mainz.de.

Conflicts of Interest: None.

Funding Statement: This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

How to cite this article:

Dormann, C. (2022). Start even Smaller, and then more Random. Comment on “Start Small, not Random: Why does Justifying your Time-Lag Matter?” by Yannick Griep, Ivana Vranjes, Johannes M. Kraak, Leonie Dudda, & Yingjie Li. *The Spanish Journal of Psychology*, 25, e20. Doi:10.1017/SJP.2022.16

important requirements to be considered in time-sensitive theories. Their fourth and most comprehensive requirement is the development of research designs allowing falsification/re-formulation of time-related propositions in cause-effect relations. (5) I believe this is hardly ever possible in single studies, and I will call for common efforts to succeed by (intelligent) random variation in time lags.

(1) *Diary studies and why we failed to develop temporal theories.* So-called diary studies are conceptually identical to classical longitudinal studies, but diary studies typically have smaller sample sizes, and data are measured using (much) shorter time lags between (much) more measurement occasions. In several OB and WOP areas, diary designs have become the dominant methodological approach. Data gathered in diary designs are frequently analyzed using variants of multi-level models (MLM), whereas longitudinal designs are typically analyzed using variants of cross-lagged panel models (CLPM). Principally, both types of statistical analysis could be used to analyze data from both types of designs. In most instances that I encountered in the literature, MLM had a couple of problems. First, the time order of measurements is ignored in most analyses, that is, the results would have been identical even if the order of measurement occasions would be randomly shuffled (there are few exceptions, e.g., Sonnentag, 2001). In fact, many applications of MLM regard diary data as (within-person) cross-sectional data. I believe the cross-sectional nature of most MLM is a major reason why the development of temporal theories has not yet flourished; there has been no pressure for authors doing so because they did, in fact, cross-sectional analyses.

I would like to note three further issues with the cross-sectional nature of many MLM. First, in MLM analyses, researchers typically fail to test if a theoretically supposed cause (X causes Y) could rather be the effect (Y cause X). Thus, using MLM for analysis misses the opportunity to detect yet unknown reversed or even reciprocal effects. Reciprocal effects are particularly interesting from a temporal perspective because they could 'carry' effects that only have a short delay time across much longer periods of time (which is why we need a statistical theory of time, see below). Second, in the few instances in which time is considered in MLM analyses, this is typically done in terms of (latent) growth curves, where time is used to predict the trajectory of a variable across measurement occasions. Recall Kenny (1979), who stated that "For X to cause Y, X must precede Y in time. (p. 3)". Since in (latent) growth curve models, time is used as a predictor, this definition consequently turns into "For time to cause Y, time must precede Y in time". Since time cannot precede itself in time, growth curve models are inappropriate from a temporal perspective (Voelkle & Oud, 2015). Third, advocates of MLM typically reason that MLM should

be preferred because MLM allow investigating within-person processes. While this is entirely true, CLPM could do so, too! (cf. Hamaker et. al., 2015; Hamaker & Muthen, 2019). Thus, to trigger development of temporal theories we need better temporal evidence on cause-effect relations, and within-person CLPM are better suited for this endeavor than MLM.

(2) *Substantive and Statistical Theories of Time.* It is essentially true that few authors have explicitly proposed theories of time lags. Still, I believe that most authors have implicit theories, and I also believe that many of these implicit theories are long-lag theories. For instance, such a theory could be "since time pressure does not cause burnout immediately, one needs long time lags to show that time pressure causes burnout". While the first part of the sentence is true, the conclusion drawn in the second part is not (see my next comment) - The conclusion is wrong because virtually no temporal study has ever tried to demonstrate that time pressure causes burnout. A little yet very important word is missing twice: Temporal studies investigate if a *change* in time pressure causes a *change* in burnout. So we should replace our implicit theories with explicit propositions, which are likely to imply much shorter time lags than our implicit theories: When people's time pressure increase from yesterday to today, is it plausible that their burnout level today is also increased compared to yesterday?

I should add that even if the answer to the last question would be yes, authors just developed the first bit of a full theory of time lags. This first piece is essential for a *substantive theory of time* as it proposed the *minimal substantive lag* for a phenomenon under study. Shorter time lags do not make much sense then. However, authors should add at least one second substantive piece. This second part is also necessary to define the *scope* of a study. For instance, if authors want to derive 'longer-term' conclusions of time pressure on the development of burnout, they should elaborate on possibly relevant psychological processes that might revert changes in the dependent variable. For example, in the case of burnout, weekend recovery could represent such an important process. Weekend recovery may fully restore increased levels of burnout back to uncritical levels. Thus, combining the first piece (change-to-change effects can be observed within a single day) with the second piece (weekends have to be included), imply a substantive theory of one week time lags.

Shorter time lags than suggested by the substantive theory are not sensible, but longer ones usually are! As Dormann and Griffin (2015) demonstrated, a substantive theory can be used to conduct 'shortitudinal studies', where the significance of findings is not too relevant. Rather, the parameters estimated using such a shortitudinal study could be used to derive optimal time lags (see below). For this endeavor, authors need a

statistical theory of time (Dormann & Griffin used only one out of many). OB and WOP psychologist have spent little effort on this issue. For instance, this involves questions such as “do daily changes in burnout caused by daily changes in stressors decay as rapidly as burnout decays” (cf. Zyphur, Allison, et al., 2020, and Zyphur, Voelkle, et al., 2020). Note that this is not an empirical question. Rather, it is a theoretical statement culminating in a statistical theory of time that can then be used to design temporal studies.

(3) *Causes of consequences of preferred time lags.* Ten years ago, I got the impression that researchers’ choices of time lags were guided by a kind of Gestalt Psychology law of nice numbers. I extracted the length of time lags from abstracts of longitudinal job stress studies. Interestingly, most abstracts did not mention the time lags at all, supporting Griep et al.’s claim that time is not paid the necessary attention. In those abstracts in which time lags were mentioned, their frequencies were particularly high for 6 months, 1 year, 2 years, 3 years, 5 years, 7 years, and 10 years. Why not, e.g., 10 months or 6 years? Some numbers seem to be nicer than others. Some early studies may have used such nice numbers, and since one of the most frequent justifications of time lags in empirical articles probably is implicitly or explicitly referring to previous authors who “used the same time lag”, nice numbers for time lags have spread in our disciplines. Unfortunately, this has blocked the scientific progress because it has resulted in replications rather than extensions. Usually, when researchers in one field do the same as others did earlier, the added value of such a study is heavily challenged, but using the same time lags as previous studies did seems to be best practice. This is science upside down! Thus, eventually, we probably know little about temporal dynamics of processes yet because of researchers’ preferences for nice numbers.

At the time I did the above-mentioned analysis, a second observation was that 1, 2, and 3 years were nice numbers, but 1, 2, and 3 months were obviously not. My hunch is that this is an example of an epistemological fallacy. Like the well-known proverb “little strokes fell big oaks” (across a long time lags), researchers possibly believe that, for example, for many small stressors to create severe burnout symptoms, one needs to wait for a long time (until a 2nd, 3rd etc. measurement). While this is possibly correct, it is not what researchers then typically do in their studies, in which they related possibly small *changes* in stressors to possibly small *changes* in burnout. So, we should be careful in not letting our unquestioned beliefs governing our decisions for particular time lags. We typically theorize about X and Y, but then we analyze *changes* in X and *changes* in Y. Thus, we better should start theorizing about how fast *changes* in X cause *changes* in Y.

(4) *Definition of optimal time lags is not optimal.* Griep et al. (2022) refer to Dormann and Griffin (2015, p. 3), who defined an optimal time-lag as “the lag that is required to yield the maximum effect of X predicting Y at a later time, while statistically controlling for prior values of Y”. Noted in passing by Guthier et al., (2020), this definition is probably not optimal, but I feel it is worth some further elaboration. Since most researchers aim at finding significant cross-lagged effects in longitudinal studies, and since effect sizes vary across different time lags, it is intuitively appealing trying to catch the time lag that produces the *biggest* effect. However, typically some not-so-big effects exist that have larger test statistics (i.e., are ‘more’ significant). Technically speaking this is because the test statistic depends on the effect size *plus* its standard error, which should be small. The standard error becomes smaller if the overall R^2 increases, and the overall R^2 depends on the size of cross-lagged effects *plus* the size of the autoregressive effect. The size of the autoregressive effect becomes larger if time lags become shorter. Thus the ‘most’ significant cross-lagged effect is smaller than the maximum cross-lagged effect and it can be at observed across time lags that are even shorter than the optimal lag as defined by Dormann and Griffin (2015). Thus, yes, this comment is another plea for shortitudinal studies.

(5) *Falsification/re-formulation of temporal propositions.* Once researchers have determined their minimal lag using substantive theory, the scope of their theory, and their statistical theory of time, they could be able to derive optimal time lags and conduct their studies. The question then is, however, what exactly researchers could falsify when performing significance testing. Failures to find significant effects could indicate false substantive theories, wrong scopes, or false statistical theories. Thus, a unique single study is unlikely to advance science very much. Our combined efforts are in demand to make causal analyses in OB and WOP succeed, and this must be reflected in studies where substantive reasoning (and measures of constructs), time lags, and theories of time vary, instead of being fixed to not much more than a handful of parameters.

I hope readers will benefit much from the very important issues raised by Griep et al. (2021) and, hopefully, that readers also benefit from my five arguments above. Five is a nice number, too, but seemingly only for arguments and not yet for time lags measured in weeks or months units. Use five, and four, or six! Randomly vary them: Between temporal studies, within temporal studies, and within participants within temporal studies (cf. Voelkle & Oud, 2013). I am worried that we are going to stay in our traditional research methods bubble, but there are promising ways out. I hope that following these new ways will make our scientific community succeeding.

References

- Cole, D. A., & Maxwell, S. E.** (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 558–577. <https://doi.org/10.1037/0021-843X.112.4.558>
- Dormann, C., & Griffin, M. A.** (2015). Optimal time lags in panel studies. *Psychological Methods, 20*, 489–505. <http://doi.org/10.1037/met0000041>
- Griep, Y., Vranjes, I., Kraak, J. M., Dudda, L., & Li, Y.** (2021). Start small, not random: Why does justifying your time-lag matter? *The Spanish Journal of Psychology, 24*, Article e45. <http://doi.org/10.1017/SJP.2021.42>
- Guthier, C., Dormann, C., & Voelkle, M. C.** (2020). Reciprocal effects between job stressors and burnout: A continuous time meta-analysis of longitudinal studies. *Psychological Bulletin, 146*, 1146–1173. <https://doi.org/10.1037/bul0000304>
- Hamaker, E. L., & Muthén, B.** (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods, 25*, 365–379. <https://doi.org/10.1037/met0000239>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P.** (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. <https://doi.org/10.1037/a0038889>
- Kenny, D. A.** (1979). *Correlation and causality*. Wiley.
- Sonnentag, S.** (2001). Work, recovery activities, and individual well-being: A diary study. *Journal of Occupational Health Psychology, 6*, 196–210.
- Voelkle, M. C., & Oud, J. H. L.** (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical and Statistical Psychology (2013), 66*, 103–126. <https://doi.org/10.1111/j.2044-8317.2012.02043.x>
- Voelkle, M. C., & Oud, J. H. L.** (2015). Relating latent change score and continuous time models. *Structural Equation Modeling: A Multidisciplinary Journal, 22*, 366–381. <http://doi.org/10.1080/10705511.2014.935918>
- Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E.** (2020). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods, 23*(4), 651–687. <https://doi.org/10.1177/1094428119847278>
- 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>