Acta Psychologica 264 (2026) 106468

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Multiverse analyses can be used to evaluate cross-lagged panel network models: An example with psychological flexibility

Kimmo Sorjonen [a,*], Bo Melin [a], Marika Melin [a], Björn N. Persson [a]

[a] *Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden*

A R T I C L E I N F O

Keywords:
Cross-lagged panel network (CLPN) model
meta-analytic averaging
Multiverse analyses
Psychological flexibility and inflexibility
Spurious effects
Statistical rigor and transparency

A B S T R A C T

Cross-lagged panel network (CLPN) models estimate prospective effects between several nodes (e.g., symptoms) while adjusting for initial scores on the outcome variables. However, it is well established that such adjusted cross-lagged effects may be spurious due to correlations with residuals and regression toward the mean. Here, we recommend that researchers conduct multiverse analyses, where cross-lagged effects are estimated with alternative models. Then, conclusions can be based on meta-analytic averaging of the estimated effects. Multiverse analyses will add rigor and transparency to analyses by acknowledging and incorporating, instead of ignoring, uncertainty due to the analyzed model. In an application of this methodology, we found that most cross-lagged effects between indicators of psychological flexibility and inflexibility, reported in a recent study, did not survive scrutiny.

1. Introduction

The cross-lagged panel network (CLPN) model is a version of the traditional cross-lagged panel model (CLPM). In CLPN, effects of several initial nodes (e.g., symptoms) on subsequent nodes are estimated while adjusting for initial scores on the outcome nodes, alternatively for all initial nodes (Wysocki et al., 2025). If adjusting only for the initial outcome node, the CLPN model corresponds to analyzing a traditional CLPM for each pairwise combination of the nodes. If adjusting for all initial nodes, the CLPN model could be described as a "super-adjusted" CLPM for each pairwise combination of the nodes. For example, Qi et al. (2025) reported that fusion, committed action, and self-as-context were influential nodes in a network of indicators of psychological flexibility and inflexibility (see the Method section below for descriptions of these indicators). Moreover, based on their findings, Qi et al. suggested that high levels of fusion tend to lead to low levels of defusion and this effect is mediated by other indicators of flexibility. Qi et al. argued that they had identified potential loci for interventions and that their findings, therefore, had clinical relevance. For example, they proposed that promotion of committed action and reduction of inaction may benefit overall flexibility.

However, similar to traditional cross-lagged panel models, effects in the CLPN model may be spurious due to correlations with residuals and regression to the mean (Sorjonen, Melin, & Nilsonne, 2025). For example, due to a negative correlation, we may suspect that among individuals with the same initial score on committed action (an indicator of psychological flexibility), those with a higher initial score on inaction (an indicator of psychological inflexibility) have received a higher score on committed action compared with their true score, i.e., a more positive residual, while those with a lower initial score on inaction have received a more negative residual in the initial measurement of committed action. However, residuals tend to regress toward a mean value of zero between measurements. Therefore, we may expect a negative, but spurious, association between initial inaction and subsequent change in committed action among individuals with the same initial score on committed action. Such a combination of correlations with residuals and regression toward the mean could possibly explain findings in various cross-lagged network analyses, e.g., in the one reported by Qi et al.

When analyzing observational (i.e., non-experimental) data, it can often be difficult to know which model is the most appropriate and which potential confounders one should or should not adjust for. In psychological research the norm appears to be that researchers pick one model that they assume, hope, or are told is the best and ignore alternative models. Whether or not the model tends to result in statistically significant findings might also have an impact on the probability for researchers to use it (Muñoz & Young, 2018). In multiverse analyses, on the other hand, researchers fit several alternative reasonable models on data and base conclusions on a juxtaposition of findings (Steegen et al.,

* Corresponding author at: Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, 171 77, Stockholm, Sweden.
E-mail address: kimmo.sorjonen@ki.se (K. Sorjonen).

https://doi.org/10.1016/j.actpsy.2026.106468
Received 2 October 2025; Received in revised form 30 January 2026; Accepted 10 February 2026
0001-6918/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Abbreviations

CLPN	cross-lagged panel network
CLPM	cross-lagged panel model
PMA	present moment awareness
ACC	acceptance
DEF	defusion
SCX	self-as-context
CA	committed action
VAL	values
LPM	lack of present moment awareness
EA	experiential avoidance
FU	fusion
SCN	self-as-content
INA	inaction
LCV	lack of contact with values

2016, see also Olsson-Collentine et al., 2023, and Girardi et al., 2024). This means that multiverse methodology acknowledges and incorporates uncertainty in findings due to the possibility to use different models.

We have suggested that cross-lagged effects of an initial score on a predictor (X_1) on a subsequent score on an outcome (Y_2) while adjusting for an initial score on the outcome (Y_1) may be evaluated through multiverse analyses (Sorjonen, Melin, & Melin, 2025b, see also Sorjonen, Melin, & Melin, 2025a). Here, we suggest that the same methodology may be applied on cross-lagged network analyses and the objective of the present study was to show how this can be accomplished through reanalyses of data used by Qi et al.

2. Theoretical section

2.1. Multiverse analyses

As mentioned above, researchers often have to make several decisions on how to prepare and analyze data. These decisions could concern treatment of missing values and outliers, how to combine variables into indices and factors, transformations, which potential confounders to control for, which statistical techniques to use, etc. In multiverse methodology, instead of making tough decisions on which road to choose at different forks in the analytic space, researchers walk down all reasonable roads, collect useful information along the way and base conclusions on a juxtaposition of findings (Steegen et al., 2016, see also Olsson-Collentine et al., 2023, and Girardi et al., 2024).

Multiverse methodology resembles crowdsourcing and many-analysts methodology, where several researchers or research groups are invited to analyze the same data with the operationalizations, statistical models, and/or covariates they find appropriate to answer a specific research question (Bastiaansen et al., 2020; Silberzahn et al., 2018; Silberzahn & Uhlmann, 2015). Crowdsourced studies have indicated that subjective researcher decisions may result in radically different outcomes, e.g., statistically significant findings in opposite directions (Breznau et al., 2022; Landy et al., 2020; Schweinsberg et al., 2021). Multiverse analyses could possibly be characterized as employing “single team crowdsourcing methodology”.

Other analytic methodologies acknowledging that researcher choices may affect findings, and that this uncertainty should be incorporated into analyses and conclusions rather than swept under the rug, include estimating vibration of effects (Klau et al., 2023; Patel et al., 2015), specification curve analysis (Simonsohn et al., 2020), single-paper (or mini) meta-analyses (Goh et al., 2016; McShane & Böckenholt, 2017), metastudies (Baribault et al., 2018; DeKay et al., 2022), Bayesian model averaging (Fragoso et al., 2018; Hinne et al., 2020; Hoeting et al., 1999), and the computational model robustness framework (Muñoz & Young, 2018; Young & Holsteen, 2017).

2.2. Cross-lagged effects

Adjusted cross lagged effects refer to an association between an initial score on a predictor (X_1) and a subsequent score on an outcome (Y_2) while adjusting for an initial score on the outcome (Y_1). A cross-lagged effect of X_1 on Y_2 when adjusting for Y_1 is identical to the adjusted effect of X_1 on the Y_2-Y_1 difference score. This is revealed by some algebra, because:

$$\text{if } Y_2 = a_0 + a_1X_1 + a_2Y_1 + e \tag{1}$$

$$\text{then } Y_2 - Y_1 = (a_0 + a_1X_1 + a_2Y_1 + e) - Y_1 \tag{2}$$

$$\text{and } Y_2 - Y_1 = a_0 + a_1X_1 + (a_2 - 1)Y_1 + e \tag{3}$$

Hence, if adjusting for Y_1, the effect of X_1 on the Y_2-Y_1 difference score in Eq. (3) (a_1) is identical to the effect of X_1 on Y_2 in Eq. (1). However, the effect of X_1 on the Y_2-Y_1 difference score may also be estimated when adjusting for Y_2 (Eq. (4)) and when not adjusting for Y_1 or Y_2 (Eq. (5)).

$$Y_2 - Y_1 = b_0 + b_1X_1 + b_2Y_2 + e \tag{4}$$

$$Y_2 - Y_1 = c_0 + c_1X_1 + e \tag{5}$$

Eqs. (3)–(5) answer different questions: (1) What is the association between X_1 and subsequent change in Y among individuals with the same score on Y_1 (Eq. (3))? (2) What is the association between X_1 and subsequent change in Y among individuals with the same score on Y_2 (Eq. (4))? This corresponds to asking: What is the association between X_1 and subsequent change in Y among individuals when correcting for (potential) differences in Y_2. This question, and the corresponding analysis, is in line with Campbell and Kenny's (1999) recommendation to use time-reversed analyses as a tool to identify statistical artifacts. (3) What is the association between X_1 and subsequent change in Y when not adjusting/controlling for individuals' score on Y_1 or Y_2 (Eq. (5))? Hence, all three equations assess the association between X_1 and subsequent change in Y. Traditionally, researchers in psychology have used Eq. (1) (which gives the same answer as Eq. (3)) to assess this association, but without specifying that their findings are conditional on individuals having the same score on Y_1. In our opinion, if the goal is to answer the more general question “Is there an association between X_1 and subsequent change in Y?”, Eq. (1)/Eq. (3) is not superior to Eq. (4) or Eq. (5). Answering the question “Is there an association between X_1 and subsequent change in Y among individuals with the same score on Y_1?” does not, in our opinion, automatically have a higher scientific value than answering, for example, the question “Is there an association between X_1 and subsequent change in Y among individuals with the same score on Y_2?”. Hence, if the goal is to answer the more general question, rather than one of the three more specific questions (see above), we propose, in line with the logic of multiverse analyses, to estimate all three associations (i.e., coefficients a_1, b_1, and c_1 in Eqs. (3)–(5)) and to juxtapose findings.

Juxtaposition can be achieved through meta-analytic averaging of the three estimated associations (Sorjonen, Melin, & Melin, 2025b). This is not a traditional meta-analysis, which estimates a weighted mean of an association estimated in different samples. However, our methodology has some resemblance with traditional meta-analyses. a_1 in Eq. (3) could be seen to estimate the association between X_1 and subsequent change in Y in subsamples with the same score on Y_1, b_1 in Eq. (4) estimates the same association in subsamples with the same score on Y_2, and c_1 in Eq. (5) estimates the association in the full sample. A meta-analytic averaging of these associations gives the expected association between X_1 and subsequent change in Y in the present sample if assessing the association with one of the three models (Eqs. (3)–(5)), randomly selected. The 95% confidence interval of this meta-analytic association indicates what association to expect (with 95% probability) if estimating the association in another sample, of the same size and drawn from the same population, with one of the three models, randomly selected (given that the meta-analytic point estimate is the true expected association in the population). This confidence interval incorporates uncertainty both due to sample variation and due to a possibility to assess the association with different models.

3. Method

3.1. Data

Here, we reanalyzed data used and made publicly available by Qi et al. (2025). According to Qi et al., "The present study was approved by the ethical committee of the authors' university. Participants were recruited through Credamo and provided informed consent digitally before completing the survey" (p. 3). No additional ethical approvals or informed consent were required for the present reanalyses. We refer to Qi et al. for more comprehensive information on the study sample, used instruments, procedures, etc. In short, a sample of Chinese adults (N = 447, mean age = 31.2 years (SD = 6.4), 58% females) completed the Multiple Psychological Flexibility Inventory (MPFI), brief version, on two occasions, one week apart. The MPFI includes six indicators of psychological flexibility: (1) present moment awareness (PMA); (2) acceptance, i.e., willingness to experience unwanted thoughts and feelings (ACC); (3) defusion, i.e., not being engulfed by negative thoughts (DEF); (4) self-as-context, i.e., being able to act as an observer of oneself (SCX); (5) committed action in accordance with goals and values (CA); (6) awareness of goals and values (VAL). The MPFI also includes six indicators of psychological inflexibility: (1) lack of present moment awareness (LPM); (2) problematic avoidance of unwanted experiences (EA); (3) fusion, i.e., not being able to question one's own thoughts (FA); (4) self-as-content, i.e., focusing on internal experiences (SCN); (5) inaction, i.e., being unable to overcome internal barriers (INA); (6) lack of contact with important goals and values (LCV). Qi et al. have made the data they used publicly available (https://osf.io/6v2kg/).

3.2. Analyses

We followed the methodology outlined above and estimated all three effects (coefficients a_1, b_1, and c_1 in Eqs. (3)–(5)) and a random-effects meta-analytic average of these three effects for each of the 12 × 11 = 132 pairwise combinations of the 12 indicators of flexibility and inflexibility. The meta-analytic averaging was conducted on Fisher's r-to-z transformed coefficients. However, the meta-analytic effects were back-transformed (z-to-r) for the presentations below. Researchers employing the cross-lagged panel network (CLPN) model usually adjust for all initial nodes, i.e., not only the initial outcome node Y_1, when estimating the cross-lagged association between X_1 and Y_2. Here, we did not follow this convention, mainly for three reasons: (1) We do not appreciate why it would be essential to adjust for all nodes (symptoms/items) that happen to be included in the used instrument but not for anything else. Surely, we could construct additional items measuring the same phenomenon. Hence, if the argument is that it is essential to adjust for all other possible items/symptoms, then all analyses with CLPN could be deemed as invalid. In the other direction, if we constructed a short version of the instrument, containing only the two items X and Y, would it then, suddenly, be admissible to adjust only for Y_1 when estimating the association between X_1 and Y_2? It would seem arbitrary if "what it is essential to adjust for" is equated with "items that happen to be included in the instrument"; (2) Adjustment for a large number of nodes/symptoms correlated with the predictor X_1 increases the likelihood for collinearity and nonsensical estimates; (3) To estimate coefficients a_1, b_1, and c_1 in Eqs. (3)–(5) is sufficient for making the point that associations in the CLPN, just like associations in the traditional CLPM, may be biased and untrustworthy. We see no reason to make the analyses more convoluted and opaque than necessary by adding a large number of additional covariates into the models.

The analyses were conducted with R 4.3.3 statistical software (R Core Team, 2025) employing the osfr (Wolen et al., 2020), metafor (Viechtbauer, 2010), and qgraph (Epskamp et al., 2012) packages. The analytic script, which also downloads data used by Qi et al. and reanalyzed by us, is available at the Open Science Framework at https://osf.io/gdf7q/. The analytic script includes a function (multinet) that takes two datasets (time 1 and time 2) as input and delivers results similar to those reported below.

4. Results

In this section, we label positive (and statistically significant) effects between the same type of symptoms (flexibility and inflexibility, respectively) and negative effects between different types of symptoms as congruent effects. Contrarily, we label negative effects between the same type of symptoms and positive effects between different types of symptoms as incongruent effects. It should be noted that we use the term "effect" in the regression analysis sense, without making causal claims.

In agreement with findings by Qi et al. (2025), when adjusting for an initial score on the outcome variable and thus examining a_1 in Eq. (3), a higher number of effects were congruent than incongruent (68 vs. 8 effects, and 132–68-8 = 56 effects were statistically non-significant, Fig. 1A). However, when adjusting for a subsequent score on the outcome variable and thus examining b_1 in Eq. (4), a lower number of effects were congruent than incongruent (10 vs. 90 effects, and 32 non-significant, Fig. 1B). Congruence was less common than incongruence also when not adjusting for an initial or a subsequent score on the outcome symptom and thus examining c_1 in Eq. (5) (2 vs. 17 effects, and 113 non-significant, Fig. 1C). Most (127 of 132, i.e., 96%) of the meta-analytic averages of these three effects were statistically non-significant and of the remaining five effects, 1 was congruent while 4 were incongruent. It can be noted that all significant meta-analytic effects involved acceptance (ACC). Initial acceptance predicted a subsequent decrease in awareness of goals and values (VAL) (β = −0.098 [−0.160; −0.035], p = 0.002) and a subsequent increase in lack of present moment awareness (LPM) (β = 0.079 [0.025; 0.132], p = 0.004). On the other hand, initial lack of present moment awareness (β = 0.065 [0.011; 0.118], p = 0.018) and lack of contact with important goals and values (LCV) (β = 0.069 [0.015; 0.122], p = 0.012) predicted a subsequent increase in acceptance while initial problematic avoidance of unwanted experiences (EA) (β = −0.081 [−0.138; −0.023], p = 0.006) predicted a subsequent decrease in acceptance (Fig. 1D).

5. Discussion

The objective of the present study was to show how multiverse analyses with meta-analytic averaging of effects can be used to evaluate findings from cross-lagged panel network (CLPN) models. As mentioned in the introduction, in multiverse methodology, instead of making tough decisions on which road to choose at different forks in the analytic space, researchers walk down all reasonable roads, collect useful information along the way and base conclusions on a juxtaposition of findings (Steegen et al., 2016, see also Olsson-Collentine et al., 2023, and Girardi et al., 2024). Here, reanalyses of data used by Qi et al. (2025) showed, in agreement with findings by Qi et al., that initial indicators of psychological flexibility tended to have positive cross-lagged associations with subsequent change in other indicators of flexibility, and negative associations with subsequent change in indicators of inflexibility, when adjusting for an initial score on the outcome variable, and vice versa. This could be taken to suggest that aspects of flexibility longitudinally promote other aspects of flexibility and counteract aspects of inflexibility while, on the other hand, aspects of inflexibility longitudinally

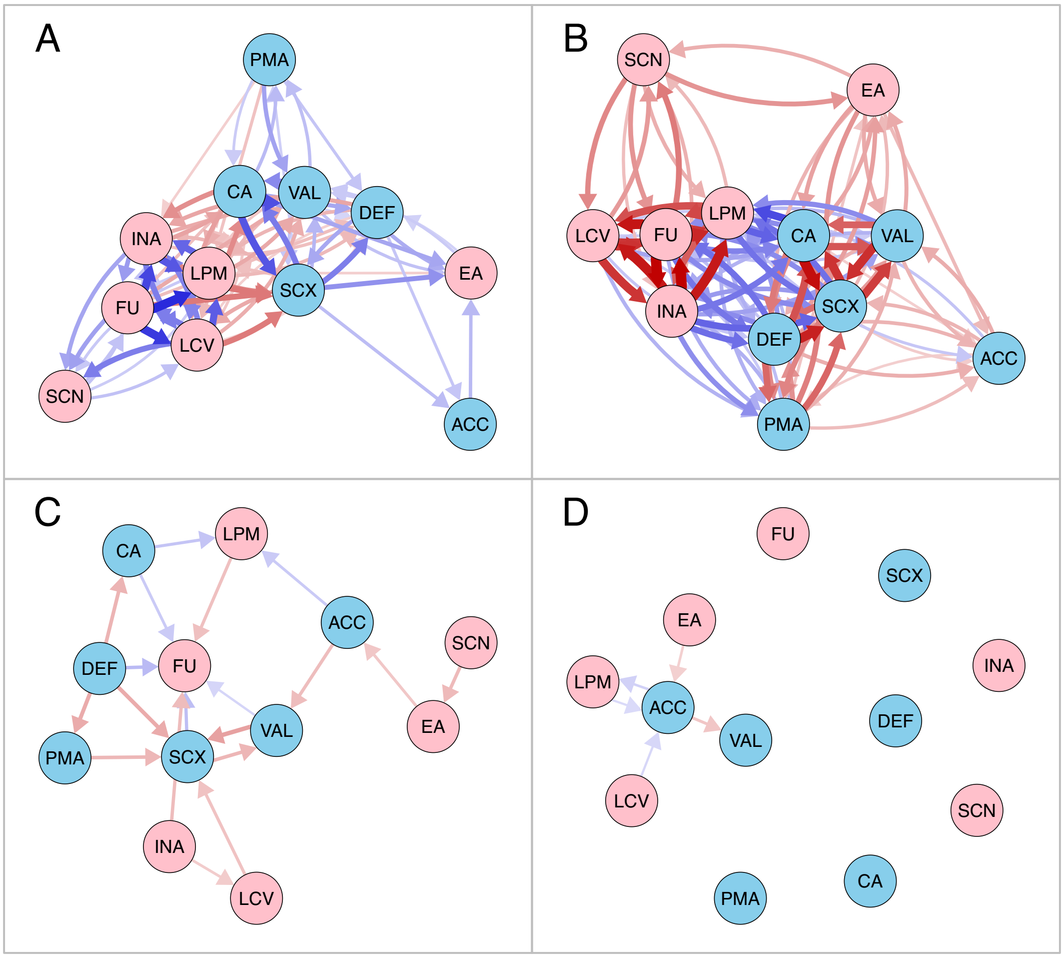

Fig. 1. Effects of initial indicators of psychological flexibility and inflexibility on subsequent change in other indicators of flexibility and inflexibility when adjusting for: (A) An initial score on the outcome variable (coefficient a_1 in Eq. (3)); (B) A subsequent score on the outcome variable (coefficient b_1 in Eq. (4)); (C) Neither an initial nor a subsequent score on the outcome variable (coefficient c_1 in Eq. (5)); (D) Meta-analytic averaging of effects A-C. Note: Blue and red arrows for positive and negative effects, respectively; Width and color saturation of arrows correspond to the strength of the effect, from −0.44 (INA on FU in panel B) to 0.37 (FU on LPM in panel A), i.e., when adjusting for subsequent fusion, the subsequent - initial fusion difference score was expected to decrease by 0.44 for every increase in initial inaction by one standard deviation (panel B) and when adjusting for initial lack of present moment awareness, the subsequent - initial lack of present moment awareness difference score was expected to increase by 0.37 for every increase in initial fusion by one standard deviation (panel A); Statistically non-significant ($p > 0.05$) effects not shown; Blue nodes for indicators of flexibility (PMA = present moment awareness, ACC = acceptance, DEF = defusion, SCX = self-as-context, CA = committed action, VAL = values); Pink nodes for indicators of inflexibility (LPM = lack of present moment awareness, EA = experiential avoidance, FU = fusion, SCN = self-as-content, INA = inaction, LCV = lack of contact with values). Values for the effects are reported in supplementary Tables S1–S4 at https://osf.io/gdf7q/. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

promote other aspects of inflexibility and counteract aspects of flexibility. However, contradicting this conclusion, when adjusting for a subsequent score on the outcome variable, effects tended to reverse and suggest that initial flexibility longitudinally counteracted other aspects of flexibility and promoted aspects of inflexibility and vice versa. Effects that change sign depending on made adjustments have been called "Janus effects" and may be expected if the evaluated crude effect is weak (Patel et al., 2015). The same pattern of effects materialized when not adjusting for an initial or a subsequent score on the outcome variable, although in these analyses many more effects were statistically non-significant. When meta-analytically averaging these three effects, only 5 of 132 (3.8%) effects were statistically significant, all involving acceptance (an aspect of flexibility).

Based on analyses of the same data as in the present study, Qi et al. suggested, for example, that high levels of fusion tend to lead to low levels of defusion and that promotion of committed action and reduction of inaction may benefit overall flexibility. However, the present divergent findings, differing between analyzed models, and lack of meta-analytic effects, suggest that the results reported by Qi et al. may have been spurious due to correlations with residuals and regression toward the mean. Hence, conclusions by Qi et al. appear premature and can be challenged. If the goal is to answer the more general question "Is there an association between X_1 and subsequent change in Y?", rather than more specific questions about associations in subgroups with the same score on initial or subsequent Y, we recommend that researchers conduct, as we did here, multiverse analyses with meta-analytic averaging when estimating cross-lagged effects, e.g., in network models, and base their conclusions on the meta-analytic effect. This will add rigor and transparency to the analyses as the multiverse methodology incorporates uncertainty due to model rather than sweeping it under the rug. The multinet-function, available in the analytic script (https://osf.io/gdf7q/), can be used for such analyses.

5.1. Limitations

As we used the same data, the present study suffers from some of the same limitations as the scrutinized study by Qi et al. For example, the study sample consisted of predominantly younger and well-educated individuals in China. It remains an open question if the present finding, that there tends to be no prospective effects between indicators of psychological flexibility and inflexibility, generalizes to other social, cultural, and economic contexts.

Moreover, the measures of flexibility and inflexibility were all self-reported and were conducted only one week apart, which may not have been optimal. However, it is important to bear in mind that such aspects were constant across the analyzed models and can, therefore, not explain why the models indicated diametrically different prospective effects.

The present findings should not be used to conclude, once and for all, that there are no prospective effects between aspects of psychological flexibility and inflexibility. The more limited conclusion to be drawn is that the observational data used by Qi et al., and reanalyzed by us here, do not allow such inferences.

5.2. Conclusions

Cross-lagged effects may be diametrically different depending on the analyzed model, e.g., including different covariates. Instead of ignoring it, multiverse analyses acknowledge and incorporate this uncertainty. We recommend researchers to use multiverse analyses when estimating cross-lagged effects, e.g., in cross-lagged panel network (CLPN) models. Then, conclusions may be based on meta-analytic averagings of the effects. This will add rigor and transparency to the analyses. In an application of this methodology, we found that most cross-lagged effects between indicators of psychological flexibility and inflexibility, reported in a recent study, did not survive scrutiny.

CRediT authorship contribution statement

Kimmo Sorjonen: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bo Melin:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Marika Melin:** Writing – review & editing, Validation, Resources, Methodology, Investigation, Conceptualization. **Björn N. Persson:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The analytic script, which also downloads data used by Qi et al. and reanalyzed by us, is available at the Open Science Framework at https://osf.io/gdf7q/.

References

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., … Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences, 115*(11), 2607–2612. https://doi.org/10.1073/pnas.1708285114

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., Chen, M., … Bringmann, L. F. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of Psychosomatic Research, 137*, Article 110211. https://doi.org/10.1016/j.jpsychores.2020.110211

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H. V., Adem, M., Adriaans, J., … Żółtak, T. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences, 119*(44), Article e2203150119. https://doi.org/10.1073/pnas.2203150119

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. Guilford Press.

DeKay, M. L., Rubinchik, N., Li, Z., & De Boeck, P. (2022). Accelerating psychological science with metastudies: A demonstration using the risky-choice framing effect. *Perspectives on Psychological Science, 17*(6), 1704–1736. https://doi.org/10.1177/17456916221079611

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4). https://doi.org/10.18637/jss.v048.i04

Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review, 86*(1), 1–28. https://doi.org/10.1111/insr.12243

Girardi, P., Vesely, A., Lakens, D., Altoè, G., Pastore, M., Calcagnì, A., & Finos, L. (2024). Post-selection inference in multiverse analysis (PIMA): An inferential framework based on the sign flipping score test. *Psychometrika, 89*(2), 542–568. https://doi.org/10.1007/s11336-024-09973-6

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Hinne, M., Gronau, Q. F., Van Den Bergh, D., & Wagenmakers, E.-J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science, 3*(2), 200–215. https://doi.org/10.1177/2515245919898657

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science, 14*(4), 382–417. https://doi.org/10.1214/ss/1009212519

Klau, S., Felix, Patel, C. J., Ioannidis, J. P. A., Boulesteix, A.-L., & Hoffmann, S. (2023). Comparing the vibration of effects due to model, data pre-processing and sampling uncertainty on a large data set in personality psychology. *Meta-Psychology, 7*. https://doi.org/10.15626/MP.2020.2556

Landy, J. F., Jia, M.(L.), Ding, I. L., Viganola, D., Tierney, W., Dreber, A., … Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin, 146*(5), 451–479. https://doi.org/10.1037/bul0000220

McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research, 43*(6), 1048–1063. https://doi.org/10.1093/jcr/ucw085

Muñoz, J., & Young, C. (2018). We ran 9 billion regressions: Eliminating false positives through computational model robustness. *Sociological Methodology, 48*(1), 1–33. https://doi.org/10.1177/0081175018777988

Olsson-Collentine, A., van Aert, R. C. M., Bakker, M., & Wicherts, J. (2023). Meta-analyzing the multiverse: A peek under the hood of selective reporting. *Psychological Methods.*. https://doi.org/10.1037/met0000559

Patel, C. J., Burford, B., & Ioannidis, J. P. A. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology, 68*(9), 1046–1058. https://doi.org/10.1016/j.jclinepi.2015.05.029

Qi, W.(G.), Wen, J., Zhou, Z., & Miao, M. (2025). Longitudinal associations between psychological flexibility and inflexibility dimensions: A cross-lagged panel network analysis. *Journal of Contextual Behavioral Science, 35*, Article 100877. https://doi.org/10.1016/j.jcbs.2025.100877

R Core Team. (2025). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., … Luis Uhlmann, E. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes, 165*, 228–249. https://doi.org/10.1016/j.obhdp.2021.02.003

Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature, 526*(7572), 189–191. https://doi.org/10.1038/526189a

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., … 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

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Sorjonen, K., Melin, B., & Melin, M. (2025a). *Multiverse analyses reveal no prospective effects between self-compassion and psychological richness: A simulated reanalysis and comment on Liu et al. (2025)*. Center for Open Science. https://doi.org/10.31234/osf.io/dt4fp_v1

Sorjonen, K., Melin, B., & Melin, M. (2025b). *Multiverse analyses with meta-analytic aggregation can be used to evaluate cross-lagged effects: Examples with anxiety, depression, and aggression*. https://doi.org/10.31234/osf.io/4ua2b_v1

Sorjonen, K., Melin, B., & Nilsonne, G. (2025). Cross-lagged network models do not prove causality and may be evaluated through triangulation. *Acta Psychologica, 260*, Article 105562. https://doi.org/10.1016/j.actpsy.2025.105562

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3). https://doi.org/10.18637/jss.v036.i03

Wolen, A., Hartgerink, C., Hafen, R., Richards, B., Soderberg, C., & York, T. (2020). osfr: An R interface to the open science framework. *Journal of Open Source Software, 5*(46), 2071. https://doi.org/10.21105/joss.02071

Wysocki, A., McCarthy, I., van Bork, R., & Cramer, A. O. J. (2025). Cross-lagged panel networks. *Advances.in/Psychology, 2*, Article e739621. https://doi.org/10.56296/aip00037

Young, C., & Holsteen, K. (2017). Model uncertainty and robustness: A computational framework for multimodel analysis. *Sociological Methods & Research, 46*(1), 3–40. https://doi.org/10.1177/0049124115610347 doi:10.56296/aip00037