Supplemental Material

A Traveler's Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions

Marco Del Giudice, Steven W. Gangestad

S1. Reliability of Composite Measures

If k indicators of a construct are standardized or otherwise scaled to the same variance, the reliability of their weighted composite (r_c) is

$$r_c = \frac{\sum_{i=1}^{k} w_i^2 r_i + \sum_{i=1}^{k} \sum_{j(\neq i)=1}^{k} w_i w_j r_{ij}}{\sum_{i=1}^{k} w_i^2 + \sum_{i=1}^{k} \sum_{j(\neq i)=1}^{k} w_i w_j r_{ij}} \qquad \text{(Eq. S1.1)}$$

where r_i is the reliability of each indicator, w_i is the weight of that indicator, and r_{ij} indicates the pairwise correlation between indicators i and j (Wang & Stanley 1970). If all the reliable variance is valid and the indicators are associated solely through the common construct, r_i is the square of the validity coefficient and r_{ij} corresponds to the product of the square roots of the reliabilities of the two indicators. If indicators are given equal weight in the composite, Eq. S1.1 simplifies to

$$r_c = \frac{\sum_{i=1}^{k} r_i + \sum_{i=1}^{k} \sum_{j(\neq i)=1}^{k} r_{ij}}{k + \sum_{i=1}^{k} \sum_{j(\neq i)=1}^{k} r_{ij}} . \qquad \text{(Eq. S1.2)}$$

In the special case in which indicators are given equal weight and have the same reliability r, the composite's reliability is given by the Spearman-Brown formula (see Revelle, 2015):

$$r_c = \frac{kr}{1+(k-1)r} . \qquad \text{(Eq. S1.3)}$$

Eq. S1.3 can be rearranged to yield the expected reliability of a new composite calculated from a different number of indicators (r_n), as for example the shortened version of an existing questionnaire. If r_c is the reliability of the current composite, k is the number of indicators in the current composite, and n is the number of indicators in the new composite, then

$$r_n = \frac{(n/k)r_c}{1+(n/k-1)r_c} . \qquad \text{(Eq. S1.4)}$$

References

Revelle, W. (2015). *An introduction to psychometric theory with applications in R*. http://personality-project.org/r/book/

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research, 40,* 663-705.

S2. Simultaneous Entry of Multiple Indicators

As we note in the main text, simultaneous entry of multiple indicators of a construct can substantially deflate the individual effect of each indicator, and hence reduce the corresponding statistical power. In multiple regression, the effect of an individual predictor X_i depends on the partial correlation between that predictor and the response variable Y, with variance of all other predictors controlled. When multiple indicators partly tap the same construct, the correlation between each indicator and the construct with all other indicators controlled for—that is, the partial validity coefficient—is necessarily less than the original validity coefficient.

For instance, suppose that we have two indicators, each with a validity coefficient of .50 (and no overlap in non-valid variance). With the other indicator partialed out, the validity coefficient of each indicator drops to .45. When three, four, and five indicators with .50 validity are entered simultaneously, the partial validity coefficients drop to .41, .38, and .35, respectively. The drop in validity is even greater if individual indicators have larger validity coefficients. If five indicators with .80 validity are entered simultaneously, each of them ends up with a partial validity coefficient of .42, barely half as large. For this reason, the unique effects of multiple indicators entered simultaneously can be expected to be much smaller than the effects of the same indicators entered individually (which, in turn, tend to be smaller than the effect of the corresponding composite).

Equally troubling, simultaneous entry not only reduces the validity of each indicator, but also changes their meaning in potentially non-obvious ways. By partialing out the shared variance, simultaneous entry increases the share of each indicator's variance that is unique to that indicator. This variance includes any reliable but invalid components of the indicator, which often reflect idiosyncratic content. To illustrate, indicators of social dominance may include (a) the ability to attract attention and (b) being perceived as self-confident by others. If the two indicators are entered simultaneously, the partial effect of (a) now taps the ability to attract attention *independently* of being perceived as self-confident—a quality that no longer reflects social dominance as normally understood. The meaning of individual indicators, then, changes when variables tapping a common construct are simultaneously entered. As a result, measurement non-equivalence is compounded by a form of effect non-equivalence (see the main text). Ironically, this problem is especially severe for highly valid indicators that share a large proportion of variance with one another.

S3. A Primer on Covariate Selection

In this primer, we unpack the generic concept of a "covariate" by reviewing three crucial roles that a variable can play in relation to an effect of interest ($X \rightarrow Y$), namely *mediator, confounder*, and *collider*. We also describe some common variations and extensions, e.g., scenarios in which a variable is a mediator of a confounder, or a descendant of a collider (Figure S3.1).

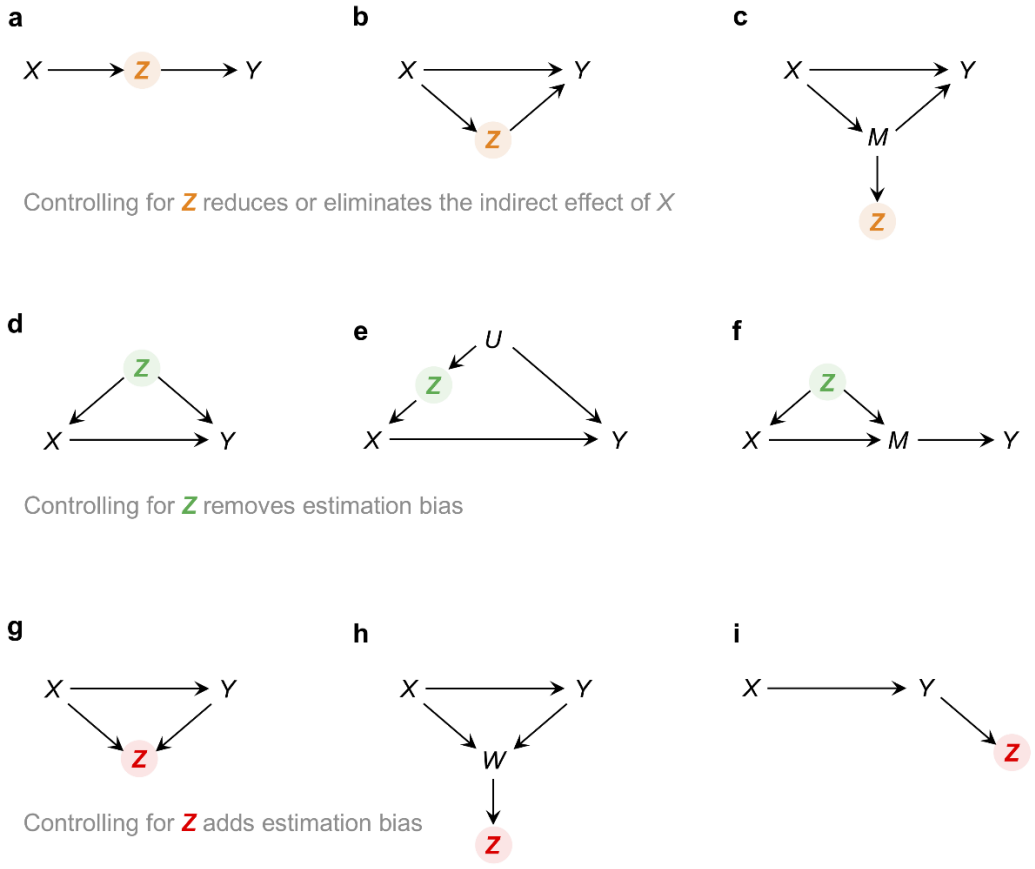

Figure S3.1. Simple causal models that illustrate the effects of covariate selection on the estimation of the effect of interest ($X \rightarrow Y$). In (a), (b), and (c), controlling for Z reduces or eliminates the indirect (mediated) effect of X on Y. In (d), (e), and (f), controlling for Z removes estimation bias by de-confounding the $X \rightarrow Y$ effect. In (g), (h), and (i), controlling for Z adds estimation bias to the $X \rightarrow Y$ effect.

Mediators

A mediator is a variable that lies on a causal path leading from X to Y, and thus serves as an intermediate step through which X affects Y. The effect of X may be fully mediated by other variables, as in Figure S3.1a; alternatively, X may also have a *direct* effect on Y that does not flow through any mediators (or at least not ones that have been measured), as in Figure S3.1b.

In the causal model of Figure S3.2, the effect of inflammation on depression is partly mediated by pain. If pain is included as a covariate, the path *inflammation* → *pain* → *depression* is blocked, and the statistical model estimates the direct effect of inflammation. If instead pain is excluded, the model estimates the total effect of inflammation, i.e., the sum of the direct and mediated effects. Both are potentially meaningful; which one should be the focus of the analysis depends on the theoretical background and goals of the study. If the direct effect is the focus of the analysis, failing to include mediators as covariates (or otherwise blocking the mediated paths) will bias the estimate (see Pearl et al., 2016; Rohrer, 2018). But if the quantity of interest is the total effect of *X*, mediators *must* be left out of the statistical model to avoid biasing the estimate.

Figure S3.1c illustrates a slightly more complex scenario, in which *Z* is not a mediator itself but a *descendant* of a mediator *M* (see Cinelli et al., 2019; Pearl et al., 2016). Because *Z* shares variance with *M*, including *Z* is equivalent to partially controlling for *M*. If the focus of the analysis is the total effect of *X* on *Y*, both *M* and *Z* must be excluded from the statistical model to prevent bias. Conversely, if the effect of interest is the direct effect of *X* on *Y*, including *Z* as a covariate does not completely remove bias, and *M* should be included instead.

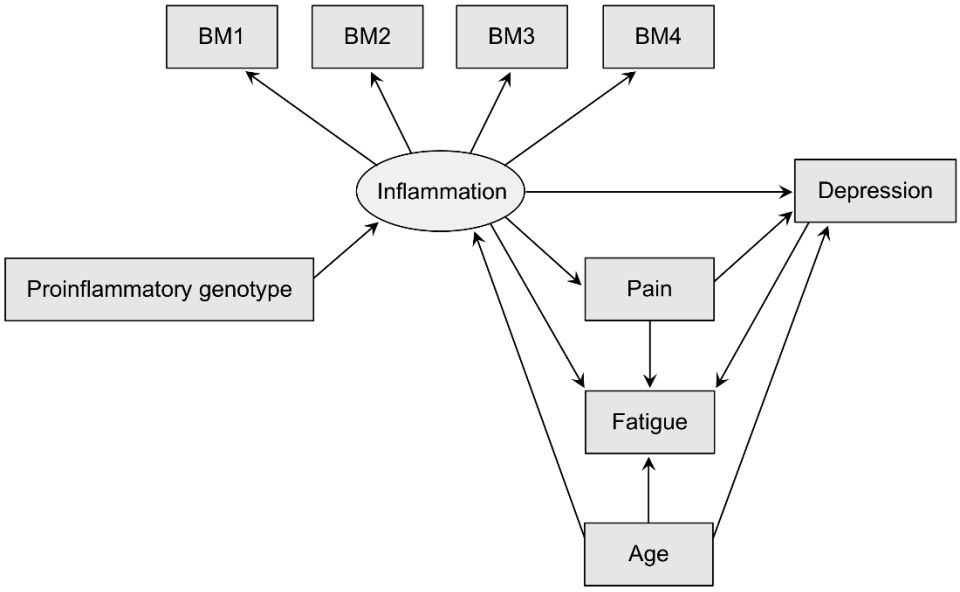

Figure S3.2. Causal model of a hypothetical study of the effect of inflammation on depression. Rectangles indicate observed variables; ellipses indicate unobserved latent constructs (same as Figure 2a in the main text).

Confounders

A confounder is a variable that affects both the predictor *X* and the response *Y*, as in Figure S3.1d. Being a common cause of *X* and *Y*, a confounder may spuriously inflate, deflate, or even reverse the *X* → *Y* effect. In the model of Figure S3.2, the effect of inflammation on depression is confounded by age, through the path *inflammation* ← *age* → *depression*. Unbiased estimates of the effect of interest require control of potential confounders by including them as covariates. Of course, if a confounder has been measured with error, including it as a covariate only partially corrects estimation bias (see Westfall & Yarkoni, 2016).

The causal model in Figure S3.1d shows the basic case of a confounder *Z* that directly affects *X* and *Y*. However, the effects of a confounder may also be mediated by additional variables, as illustrated in Figure S3.1e. In this example, *Z* mediates the effect of confounder *U* on the predictor *X*. Including either *Z* or *U* as a covariate in the statistical model blocks the confounding path $X \leftarrow Z \leftarrow U \rightarrow Y$ and corrects the estimation bias (Cinelli et al., 2019; Pearl et al., 2016). Figure S3.1f shows another variation on this theme. Here, *Z* is a common cause of the predictor *X* and of a variable *M* that mediates the effect of *X* on *Y*. The confounding effect of *Z* in this scenario is indirect but no less real, and *Z* must be controlled to avoid bias.

Colliders

A collider is the mirror image of a confounder—a common *effect* of both *X* and *Y* rather than a common cause (or, equivalently, a descendant of both *X* and *Y*; Figure S3.1g). In the model of Figure S3.2, both inflammation and depression affect fatigue, which plays the role of a collider. Whereas confounders add bias to estimation of the $X \rightarrow Y$ effect unless they are actively controlled for (or the confounding paths are otherwise blocked), colliders introduce bias if they *are* included as covariates ("conditioning on a collider;" see Elwert & Winship, 2014; Pearl et al., 2016; Rohrer, 2018). In Figure S3.2, including fatigue as a covariate would unblock the *inflammation* → *fatigue* ← *depression* path (as well as additional paths involving *pain*) and bias the estimated effect of inflammation on depression. Specifically, if both inflammation and depression increase fatigue, controlling for the level of fatigue introduces a spurious negative association between the two variables. The reason is that, at any fixed level of fatigue, a larger contribution from inflammation implies a smaller contribution from depression (and vice versa), all else being equal. This counterintuitive effect is also known as *Berkson's paradox* (Berkson, 1946; Snoep et al., 2014).

If a variable is a collider, it should not be included as a covariate in the statistical model, unless the biasing path is blocked again by the inclusion of other variables (e.g., a mediator of the effect of *X* or *Y* on the collider). The same applies if a variable is not a collider itself but a descendent of a collider, as illustrated in Figure S3.1h. Here, *Z* is a descendant of collider *W*; including *Z* as a covariate partly controls for *W*. Finally, Figure S3.1i depicts a scenario in which *Z* is a descendant of *Y*, but is not directly affected by *X*. Even in this seemingly neutral case, Z is a common effect of *X* (indirectly through *Y*) and *Y*, and can be expected to introduce estimation bias if included as a covariate (Cinelli et al., 2019).

Implications for precision

Even if a potential covariate is neutral with respect to estimation bias, it may still affect the *precision* of the estimate (Cinelli et al., 2019; Pearl et al., 2016). Figure S3.3 depicts three illustrative scenarios. In Figure S3.3a, variable *Z* has a causal influence on the predictor *X*, but no direct effect on the response variable *Y*. Including *Z* as a covariate does not affect bias on the $X \rightarrow Y$ effect, but reduces the variation of the predictor, and thus may decrease the precision of the estimated effect. In the model of Figure S3.2, this would correspond to including proinflammatory genotype as a covariate. (Note that genotype is a neutral control only if age has also been controlled for; if not, including genotype as a covariate *amplifies* the confounding effect of age. See Pearl [2012].)

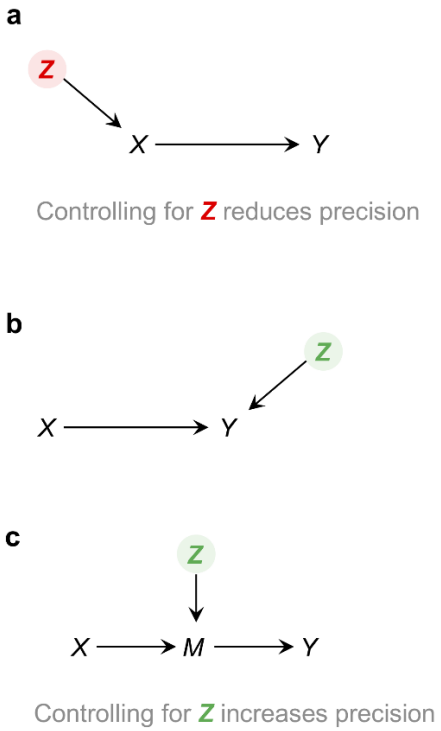

Figure S3.3. Simple causal models that illustrate the effects of covariates on the precision of the estimate of the effect of interest ($X \rightarrow Y$). In (a), controlling for Z reduces the precision of the estimate. In (b) and (c), controlling for Z increases the precision of the estimate.

In Figure S3.3b, variable Z has a causal effect on the response variable Y. Controlling for Z reduces the variation of the outcome that is not explained by X, and in doing so may increase the precision of the estimate. Likewise, controlling for Z in Figure S3.3c reduces the variation of mediator M that is not explained by X, with a positive effect on precision.

References

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin, 2,* 47–53.

Cinelli, C., Forney, A., & Pearl, J. (2019, August 14). A crash course in good and bad control. *Causal Analysis in Theory and Practice,* http://causality.cs.ucla.edu/blog/index.php/2019/08/14/a-crash-course-in-good-and-bad-control

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology, 40*, 31-53.

Pearl, J. (2012). On a class of bias-amplifying variables that endanger effect estimates. *arXiv,* 1203.3503.

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

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,* 27-42.

Snoep, J. D., Morabia, A., Hernández-Díaz, S., Hernán, M. A., & Vandenbroucke, J. P. (2014). A structural approach to Berkson's fallacy and a guide to a history of opinions about it. *International Journal of Epidemiology, 43,* 515-521.

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE, 11, e0152719.*

S4. Replicate Analyses

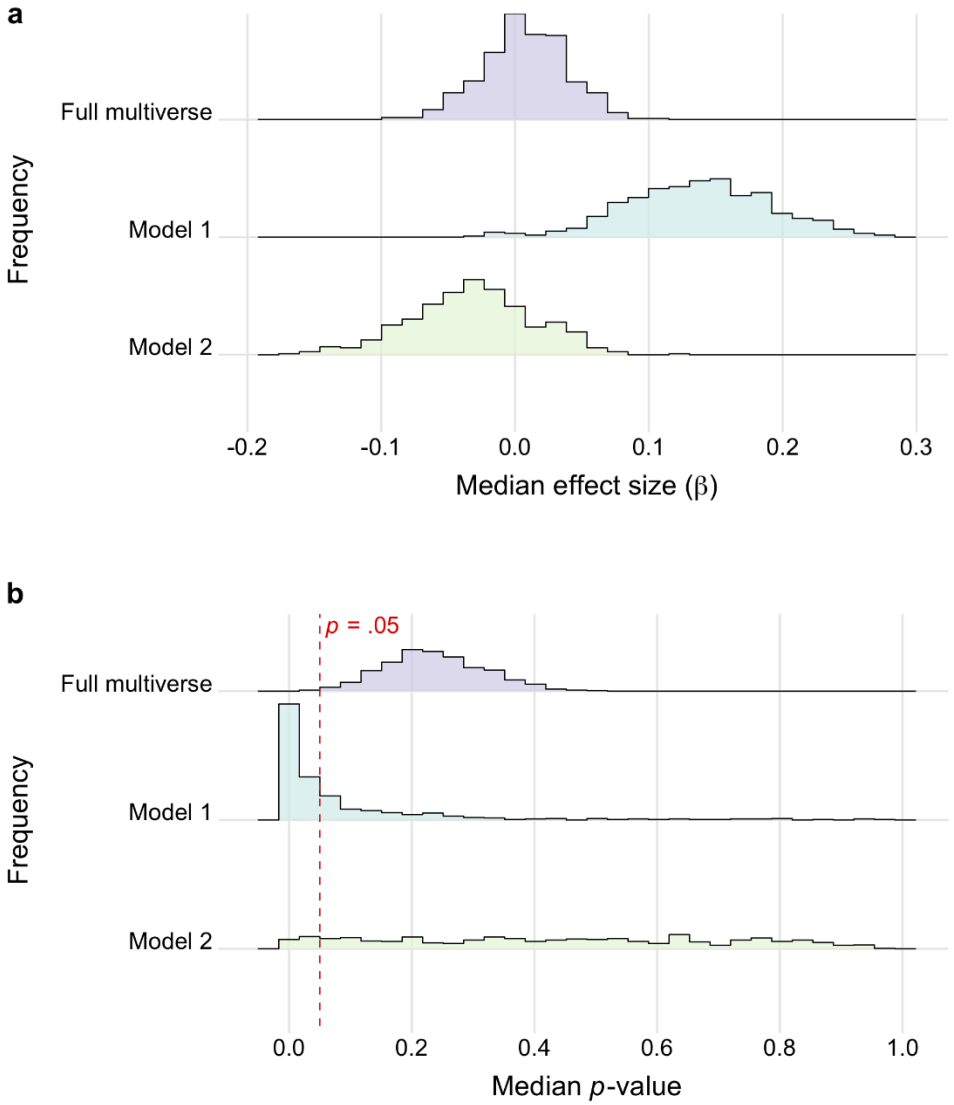

Figure S4.1. Summary statistics for 500 replicate analyses: (a) median effect size and (b) median *p*-value across specifications. The 500 samples were generated with the same simulation code used in the main analysis ($N = 300$ each).