We want to thank Leif, Uri, and Joe for writing this blog post about meaningless meta-analysis which we see as an important complement to our own comment on Mertens et al.

We agree with the main message and most of the specific comments. ("Meaningless means" were the main reason that in the systematic review some of us conducted on nudging a few years ago, we did not conduct a meta-analysis). At the time of drafting our comments on Mertens et al., given the word constraint of PNAS comments, some of the co-authors of our comment even hesitated between focusing on two possible options: meaningless means or the size and variability of nudges. The reason we chose the latter is that we assumed that the average reader of the PNAS paper will be someone who is considering using nudges and is a non-expert in methodology. That way, setting the expectations about the effect size and variability of the effect sizes seemed to be more important than a more general point about meta-analyses.

Although we do not think that publication bias corrections are without limitations, we believe that the corrected estimates are more realistic. Furthermore, based on the available evidence, we do not think that it is likely that the average effect of the nudge interventions is 0. But again, the interesting question is not the average effect, but whether for a given problem, for a given context, what effect can we expect applying a given intervention.

In this vein, we think that one solution would be to conduct meta-analyses about a more specific and defined instance of nudging. For this to be more informative, it would be necessary that many details of both the nudge and the outcome of interest remain similar across the included studies. For instance, even a meta-analysis of "default effects'" might still not be informative enough, as there might be many possible differences across studies, including the context, the operationalizations of the default effect, and the outcomes of interest. Deciding which studies to include in the meta-analysis will still be a matter of scientific judgment.

Alternatively, a study of the same scope could be conducted, but with a focus on understanding and explaining variations in the effects of nudge interventions. This would involve coding a variety of moderators that can help capture the variation in populations under study, types of nudges (and what they are compared to), outcomes measured, and settings. With these coded, meta-regression models could focus on the extent to which the effects of these interventions can be explained by these observed moderators, and the conditions under which nudge interventions are most (versus least) effective. These reviews would not have to be only quantitative, but might also include discussions of outlying cases. Overall, this approach would focus from the outset on understanding heterogeneity, instead of a single average effect.

Finally, we support your conclusion saying that "many nudges undoubtedly exert real and meaningful effects on behavior. But you won't learn that – or which ones – by computing a bunch of averages or adjusting those averages for publication bias. Instead, you have to read the studies and do some thinking."

Barnabas Szaszi, Anthony Higney, Ignazio Ziano, Dan Goldstein, Balazs Aczel, Elisabeth Tipton