Reply to Data Colada #122

Oeindrila Dube, University of Chicago
Sandy Jo MacArthur, LAPD (Retired)
Anuj K. Shah, Princeton University

We thank Uri Simonsohn for taking the time to engage with our work, and we appreciate the chance to reply. This post suggests that there was too much flexibility in our pre-registrations and that this raises concerns about our analyses. It also suggests that there is insufficient transparency in our reporting and that this makes it difficult for readers to evaluate our work.

We think both of these concerns are based on an incomplete accounting of what's in our paper.

While this post focuses on the results presented in Table III, our paper's conclusions rest on a wider range of results drawn from two sources of data: (i) an endline assessment completed 4 months after the intervention and (ii) administrative outcomes from the field. Analyses of both data sources are guided by two pre-analysis plans (PAPs), which we link to on the same page where the data are introduced. We report on our analyses comprehensively and transparently in the main text and Appendix. We think our paper should be evaluated based on the whole set of results presented. But below, we focus on just replying to the main points in this post.

1) This post takes issue with the 4-month time period for the analysis in Table III, but it does not explain to readers why we begin with this analysis. Our results section in the paper starts with an analysis of an endline assessment conducted 4 months after the training was completed. We therefore constrained our initial analysis of field outcomes to 4 months after training, to assess whether the effects observed on the endline assessment correspond with changes in the field over the same time period. We then go on to report results from additional time periods, including the full 12 months of data, in the main text and Appendix. We are clear about what constrained our analysis, and we report transparently on additional time periods. In discussing these results, we ourselves make the point that the field effects fade out over the course of 12 months, and we discuss possible reasons for this fadeout in the paper.

2) The post also takes issue with the use of force analysis reported in Table III, but it does not give readers the context for why we begin with this analysis (or how the levels of force differ). As we discuss in our paper, lethal uses of force (Level 3 incidents) are very rare in our sample and we are not powered to detect changes in them. That leaves only one primary outcome for uses of force: All nonlethal uses of force (Levels 1 and 2).

   Moreover, we are transparent about what was pre-registered. We write (p. 17): "We prespecified examining not just nonlethal force (levels 1 and 2) but also higher levels of force (levels 2 and 3) and all levels of force together (levels 1, 2, and 3)" and we refer readers to the relevant Appendix section for those analyses.

   Thus, we are once again clear about the constraints that informed our analyses, and we report transparently on the broader scope of outcomes.

3) This post also suggests that our discretionary arrests result harms credibility. But it is not clear to us why the effect on a distinct pre-registered primary outcome would harm credibility, particularly since the results in Table III are not the only results that we present in the paper. Rather, these results go hand-in-hand with a set of other results from the field and from the endline assessment.

4) The post takes issue with us presenting and interpreting an analysis of officer injuries, suggesting that our inclusion of an outcome that was not pre-registered further raises concerns about p-hacking. We pursued data on this outcome after we were asked by others about impacts on officer safety. Moreover, when we introduce this measure in the paper, we transparently discuss that this outcome was not pre-registered (p. 19): "officer injuries were not included in the PAP. However, we present these in our main table examining key administrative outcomes because officer activity and injuries are important for contextualizing and interpreting the reductions in uses of force and discretionary arrests."

Thus, for each of the concerns raised in this post, there is context and transparency in our paper that the post omits (at least in the version we reviewed), particularly when it comes to the range of results presented in the paper. The data we discuss in our paper are indeed quite rich. It would be very problematic to present just a subset of these results and hide the remaining analyses. We don't do that. We report our results comprehensively and transparently.

The results from Table III are not the only results in our paper (or even the only results discussed in the abstract). Before presenting the field outcomes in this table, we first present analyses of our endline assessment outcomes. And for both endline and field outcomes, we note which outcomes are primary, secondary, and not pre-registered. We discuss where an outcome goes against our pre-registered predictions. We discuss null results. And, we discuss how we account for multiple comparisons (though this post suggests that this is insufficient). Some of our discussion is handled in more detail in the Appendix, but all of it is referenced in the main text and reported transparently for readers.

Given this post's concerns about flexibility, we felt it was important for readers to know the context and constraints that informed our analyses. And given this post's concerns about how results are reported, we thought readers should know how comprehensive and transparent our reporting is. We think readers will be better positioned to reach a well-informed conclusion about the work if they read the full reporting of results in our paper.