Assessing the Validity of Prevalence Estimates in Double List Experiments

Appendix

Gustavo Diaz*

November 17, 2021

A. Sensitive item with different correlations to baseline lists

I thank Alex Coppock for noticing this. The main text mentions how the difference in correlations test assumes that the sensitive item has the same correlation with both baseline lists. If the sensitive item correlates differently with each baseline lists, we can have a non-zero test statistic without any violation of the list experiment assumptions. Researchers can minimize this possibility through research design, but it still remains an assumption.

Using similar notation to the main text, let $\boldsymbol{y_A(D)}$ and $\boldsymbol{y_B(D)}$ denote vectors of observed responses under treatment status $D = \{0, 1\}$. Let $\boldsymbol{X}$ denote a vector of whether a respondent reports to have the sensitive trait.

Assuming no liars and no design effects, for respondents who get the sensitive item in list A, we have

$$\begin{aligned} Cov(\boldsymbol{y_A(1)}, \boldsymbol{y_B(0)}) = \\ Cov(\boldsymbol{y_A(0)} + \boldsymbol{X}, \boldsymbol{y_B(0)}) = \\ Cov(\boldsymbol{y_A(0)}, \boldsymbol{y_B(0)}) + Cov(\boldsymbol{y_B(0)}, \boldsymbol{X}) \end{aligned} \tag{1}$$

Whereas for respondents who get the sensitive item in list B, we have

*Postdoctoral Fellow. Center for Inter-American Policy and Research. Tulane University. E-mail: gustavodiaz@tulane.edu

$$Cov(\boldsymbol{y_A(0)}, \boldsymbol{y_B(1)}) =$$
$$Cov(\boldsymbol{y_A(0)}, \boldsymbol{y_B(0)} + \boldsymbol{X}) = \tag{2}$$
$$Cov(\boldsymbol{y_A(0)}, \boldsymbol{y_B(0)}) + Cov(\boldsymbol{y_A(0)}, \boldsymbol{X})$$

So if $Cov(\boldsymbol{y_A(0)}, \boldsymbol{X}) \neq Cov(\boldsymbol{y_B(0)}, \boldsymbol{X})$, we have a non-zero test statistic even without any unexpected respondent behavior.

The main text mentions that researchers can minimize concerns over this possibility through research design practices such as shuffling the order of lists or choosing baseline lists with positive correlation. The simulations in the main text assume the sensitive item is independent from the responses to baseline lists. Let's call these "independent" experiments.

I simulate additional experiments varying this feature. Let $U \sim N(0,1)$ be a latent vector which probability distribution indicates the probability of individual respondents holding the sensitive trait Y_i^*. Independent experiments assume that potential outcomes without the sensitive trait $Y_{iA}(0)$ and $Y_{iB}(0)$ are uncorrelated with Y_i^*.

I simulate additional "correlated" and "uneven" experiments. In correlated experiments, all three of $Y_{iA}(0)$, $Y_{iB}(0)$, and Y_i^* take values based on the PDF of U, so that $Y_{iA}(0) \sim B(4, f(U))$, where f is a shorthand for the PDF. The analogous distribution applies to $Y_{iB}(0)$. This simulates a situation wherein the sensitive item has the same correlation with both lists.

In "uneven" experiments, everything stays the same, except that $Y_{iA}(0) \sim B(4, 0.5)$. This captures a situation where list A correlates with the sensitive item, but list B does not.

I set the deflation rate $\delta = 1$ so that there are no violations to the list experiment assumptions. I also consider both baseline lists correlate with each other with rank-correlation $\rho = \{0, 0.4, 0.8\}$. See the main text for additional details on the simulation setup.

Figure A1 shows the distribution of test-statistics based on $1,000$ replications for each parameter combination. These experiments have no deflation, so distribution of tests statistics centered away from zero suggests a tendency to report false positives. The figure shows that false positives are likely in the case of uneven experiments, but they are less frequent when the two baseline lists correlate positively with each other.

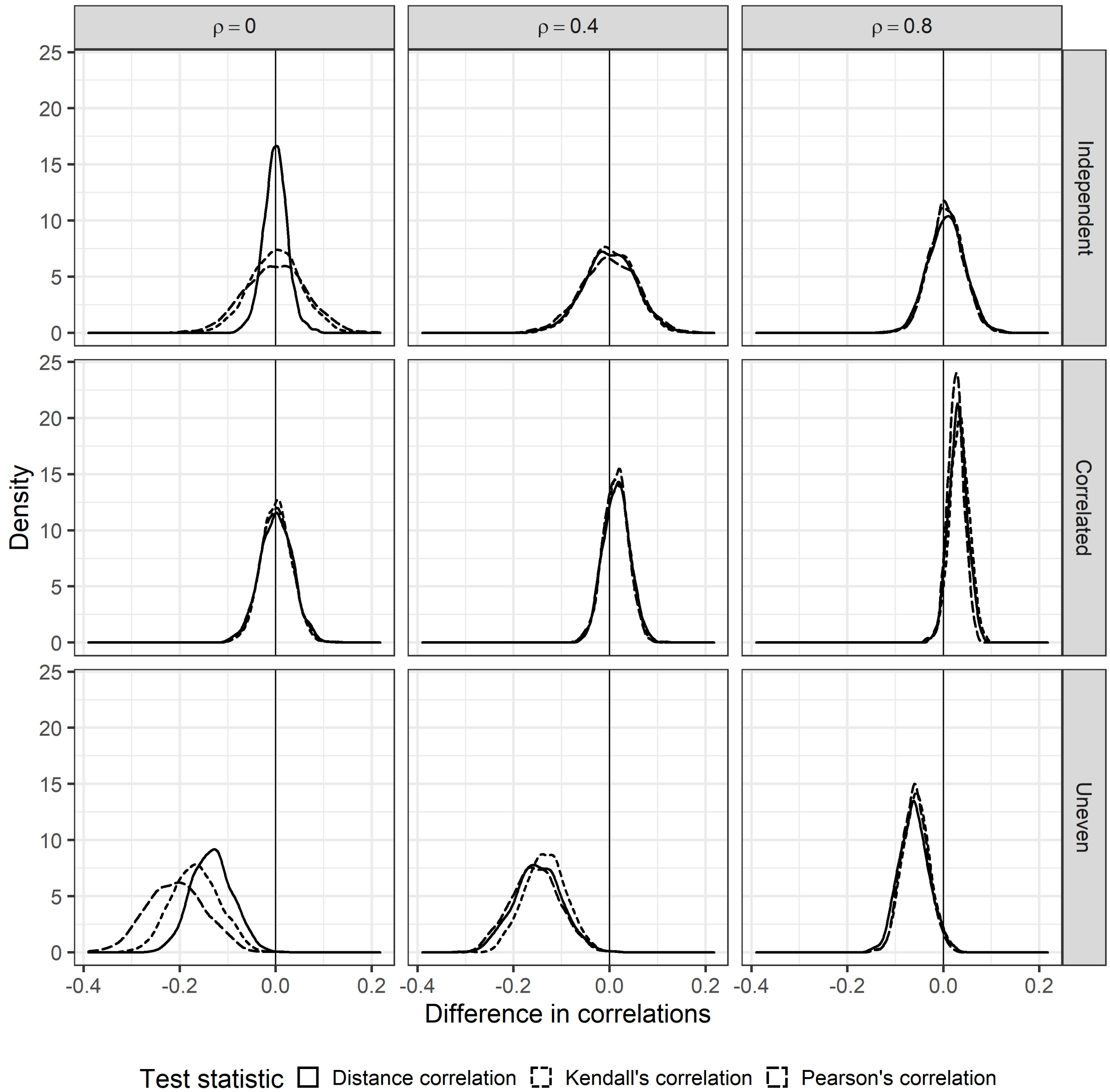

Figure A1: Distribution of test statistics for experiments without deflation under different correlation structures

Note: Each panel is based on 1,000 replications.

B. Different correlation structures

The main text shows the performance of the test statistics under simulated experiments assuming the sensitive item is independent from the distribution of potential responses to the baseline items. This section presents additional results considering alternative assumptions about the correlation structure.

I consider correlated and uneven experiments based on the description in the previous section. Figure B1 shows the performance of the test statistic for correlated experiments. This figure is analogous to Figure 3 in the main text.

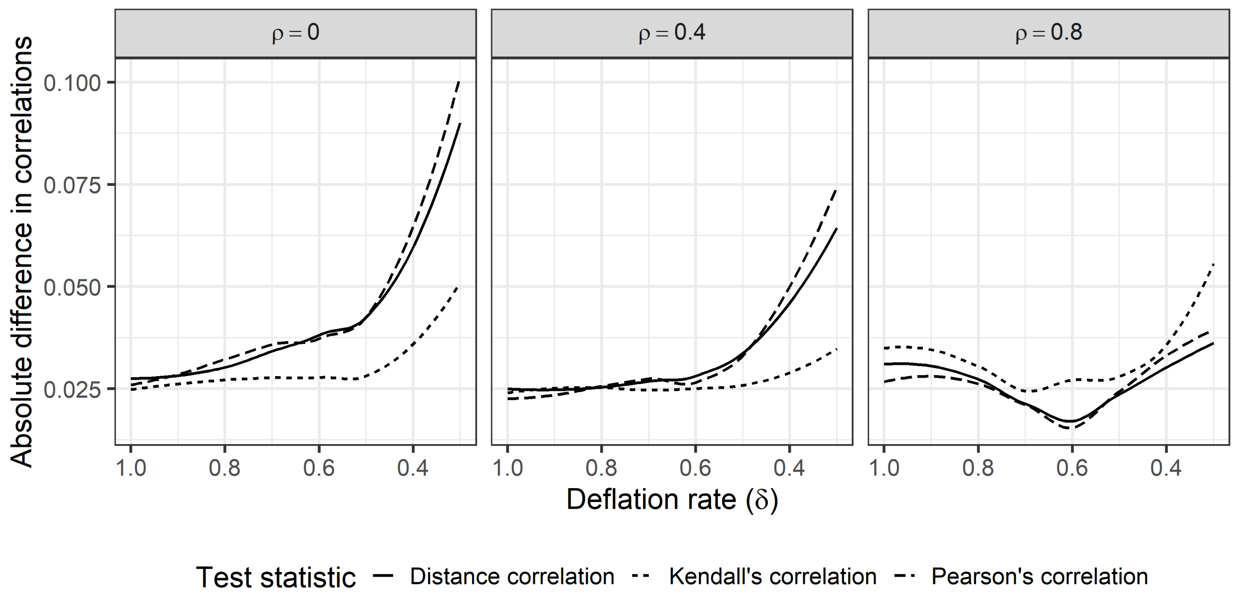

Figure A2: Difference in correlations test performance in correlated experiments (LOESS fit)

The takeaway is similar to Figure 3, except that test statistics are now also sensitive to deflation even if the baseline lists are not correlated with each other. In these stylized simulations, deflation introduces a negative shift in the test statistics, whereas having the sensitive item and both baseline lists drawn from the same latent variable to induce positive correlation among them tends to produce a positive test statistic. These two patterns tend to cancel out when we also introduce high correlation between baseline lists, which leads to a less sensitive test statistic. This pattern is mainly a construct of the simulation setup and I do not expect it to translate to applications [I think, still trying to figure this one out].

Figure B2 shows the performance of the difference in correlations test of uneven experiments. The lesson here is similar than in the previous section. The test is sensitive to false positives under uneven correlation structures, which translates to test statistics that are not sensitive to response deflation. Researchers can recover the ability to detect violations to the list experiments assumptions in this setting by inducing positive correlation between baseline lists.

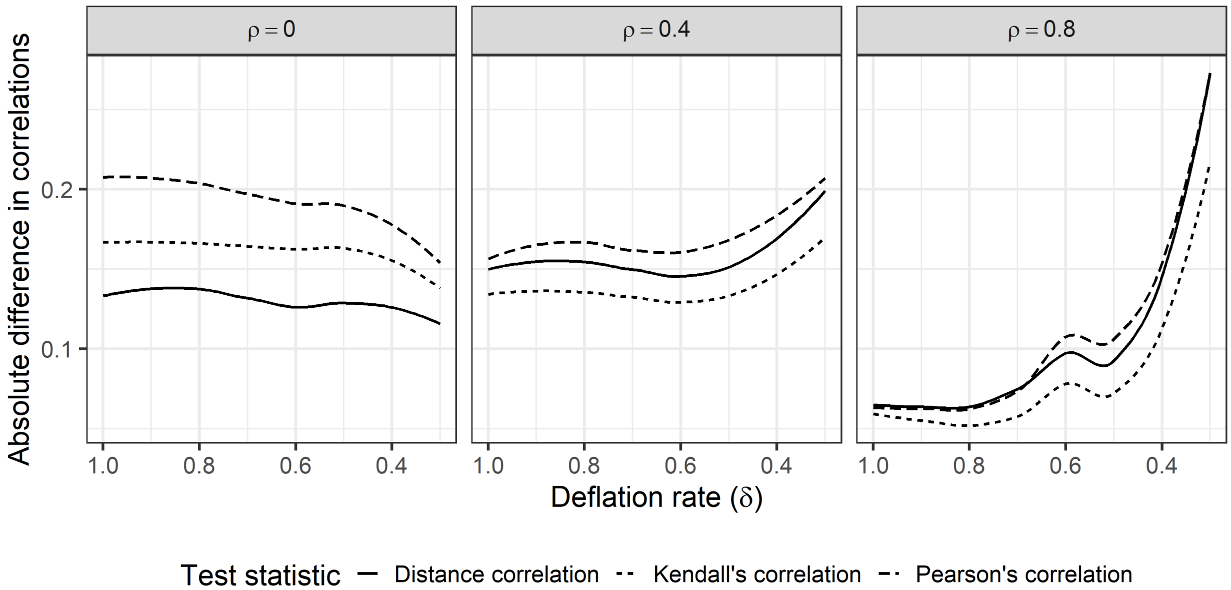

Figure A3: Difference in correlations test performance in uneven experiments (LOESS fit)

Glynn (2013) documents how inducing positive correlation between baseline lists also has the benefit of further reducing the variance of DLE estimates, so this is yet another benefit of adopting this research design practice.

References

Glynn, Adam N. 2013. "What Can We Learn with Statistical Truth Serum?" *Public Opinion Quarterly* 77 (S1): 159–72. https://doi.org/10.1093/poq/nfs070.