**RR:C19 Evidence Scale** rating by reviewer:

- **Misleading.** Serious flaws and errors in the methods and data render the study conclusions misinformative. The results and conclusions of the ideal study are at least as likely to conclude the opposite of its results and conclusions than agree. Decision-makers should not consider this evidence in any decision.

************

**Review:**

Following the guidelines put forth by RR:C19 I classify this paper as “Misleading”. With a few changes to the econometrics it could be either “Reliable” or “Strong”, although the results may be different.

**Here are the reasons I do not consider the results reliable:**

1. The errors should be clustered at the level of the treatment, which is the state level (Abadie, et al. 2017; Bertrand, Duflo and Mullainathan 2004; Cameron and Miller 2015). Because the standard errors are clustered at the level of the outcome (household), they could be biased downward. This correction could render the results statistically insignificant.

2. Lockdowns are endogenous to cases, so running a regression with lockdowns, but not cases, as an explanatory variable could lead to the misattribution of lockdown effects to case effects. In order to isolate the lockdown effect from the case effect, all regressions that include lockdown as an explanatory variable should also include cases as an explanatory variable.

   It is also possible that cases are endogenous to lockdown, depending on when they are recorded. To estimate the impact of cases *net of lockdown*, which may or may not be of interest, the regressions that include cases as an explanatory variable should also include lockdown as an explanatory variable.

3. The interpretation of the results in Table 5 are not statistically supported. For instance, “…households in states recording high COVID-19 cases and with lockdown measures are hardest hit…”. This conclusion should be drawn from the difference between the coefficient estimates for high*lockdown and high*no lockdown. It appears there is no statistically significant difference between these two coefficients except for perhaps a *negative* impact on meal skipping (column 3). There is also no significant
effect of lockdown measures in low COVID states, and the coefficient is negative in 2 of 3 regressions. These results indicate the lockdown was not a significant driver of outcomes, which runs counter to the paper’s most important policy-relevant finding.

4. Table 9 is really problematic. CPI is an endogenous variable in the regressions for Table 8 but an exogenous variable for the regressions in Table 9. This analysis is misleading without the full set of interactions as controls: cases*post, CPI*post, and CPI*cases.

**The following could be done to substantially improve the credibility of the paper**

5. If there are multiple rounds of pre-COVID data, these should be exploited. At a minimum, the data could be used to test for parallel trends. If there are non-parallel trends, these could be controlled for them in the econometric analysis. This is a major missed opportunity, if such data exist.

**Other minor points**

6. It is unclear what regression is behind Table 10. One possibility is that the results are from single regression model with all the heterogeneity variables at once. Alternatively, the results could be from a series of regressions, each ran with a single heterogeneity variable. The choice of specification could affect the results and their interpretation.

7. The economic activity regressions should include interaction terms between lockdown and cases. Not only is this a way to look at interaction effects, but it facilitates comparison between the lockdown and cases impacts (see #2 above).

8. In Table 5 the “post” coefficient deserves some attention in the discussion. This is the outcome for low cases without lockdown. Clearly for meal skipping and running out of food the biggest change is associated with being in the COVID period regardless of cases or lockdown (although one cannot establish causality without a counterfactual).

**References**

Abadie, A., et al. 2017. "When Should You Adjust Standard Errors for Clustering?". National Bureau of Economic Research.
Bertrand, M., et al. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly journal of economics* 119(1):249-275.

Cameron, A.C., and D.L. Miller. 2015. "A Practitioner’s Guide to Cluster-Robust Inference."

*Journal of human resources* 50(2):317-372.