

# Review of: "Identifying Psychological Distress Patterns during the COVID-19 Pandemic using an Intersectional Lens"

Jonathan Rothwell

**Potential competing interests:** No potential competing interests to declare.

One challenge with this paper is that much has been written on how the pandemic has affected social, economic, and health outcomes, and how these outcomes vary by socio-economic status. It would be difficult to make a novel contribution to this area of research, and I do not think the authors do so here.

More positively, I have not seen a paper use conditional inference trees or random forests, and these are potentially useful tools, but the authors did not convince me of their added-value. To convince me, they could run more traditional regression models and rank variable importance against the variables revealed by their method. Moreover, I do not think I am alone in finding these methods difficult to interpret in both their operational sense and in understanding outcomes. The variable importance scores are clear in terms of ranking, but entirely unclear about the scale, the degree of robustness, confidence intervals, and traditional ways of inferring relevance, relative or otherwise. This is a substantial weakness of these methods that must be justified by the novelty and insightfulness of the findings. I do think the findings--or data used to generate them--reach that threshold.

A more substantial challenge is that the authors lack experiential or behavioral data--or do not include it in their analysis, relying instead on pure demographic variables. Intersectionality theory, which the authors rely upon to justify their analytic approach, provides a superficial foundation for social science research, because the categories it considers--namely gender and race/ethnicity--are excessively broad and neither determinative of experiences nor behaviors. In short, there is tremendous variation within each group. The authors emphasize this fact as motivation for their machine-learning approach, but the variable importance scores are collapsed into these broad superficial categories (e.g. sex, race, age, education), with no overlap. By this point, many other papers have already been published that note how education, income, race/ethnicity, and gender predicted outcomes during the pandemic, and these papers often rely upon larger samples that cover longer time periods.

More sophisticated papers in this literature consider how job loss, public policy variation, politics/partisan identity, beliefs (including misinformed beliefs), and behaviors (such as social distancing or working from home) affect the outcomes studied here. None of this is acknowledged in the literature review or analysis.

As for the discussion section, finding greater instances of mental distress in certain groups does not imply that resources should be disproportionately devoted to these populations. The authors need to establish what it is about these groups that puts them at greater risk, and whether the pandemic affected them differently or whether these conditions preceded

it. Their theoretical section suggests discrimination, but they provide no analysis consistent with that theory, and they would need to confront contradictory evidence (e.g. Republican Party members and men were more likely to die of COVID than women, which is unlikely the result of discrimination against men or Republican Party members; boys have been more likely to commit suicide than girls, at least in recent years, despite higher self-reported depression and suicidal ideation among girls; non-Hispanic White Americans had higher rates of deaths of despair during most of the 21st century compared to Hispanic, Black Americans, and Asian Americans).

Finally, while we are in 2023, the authors use data from only two months of the pandemic. Many larger and better data sources are available, some of which are publicly available--including the Census Household Pulse. The paper could be strengthened by considering a much longer time series and plotting how socio-economic status--or other variables--varied in their importance throughout the pandemic and discussing why with greater attention to evidence and well-established theory.

In summary, I think the paper is a useful exercise for applying an interesting and sophisticated analytic approach, but does not reach the level of effort, novelty, and overall quality that I would hope to see in an academic publication, especially on a much-studied topic with abundant data.