

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

Yoshiro Tsutsui¹

¹ Osaka University

Potential competing interests: No potential competing interests to declare.

Reviewer's comments on

"Identifying Psychological Distress Patterns during the COVID-19 Pandemic using an Intersectional Lens"

written by Samantha Robinson and Brittany Hearne

This paper analyses psychological distress due to the COVID-19 pandemic using Conditional Inference Trees (ctree) and Random Forests. Since method is the core of science including social science, I welcome this paper.

This paper is simple. The authors apply ctree to the data of psychological distress, which is the sum of the answers consisted of 7 items on a 4-point Likert scale, and of determinants of mental health including age, gender, race/ethnicity, education, income, geographic area of residence and community attachment. They found that age and sex were the most important social determinants, which well separates people into high- and low-distress groups at the first and the second stages both in March and April, 2020. However, determinants at lower stages are different between low- and high-distress groups and between March and April. In addition, Random Forests suggests that age, gender, and income were most important in March, whereas race-ethnicity and age were most important in April.

There have been numerous analyses of how the Pandemic has affected people's well-being, anxiety, and other emotions. The key variables are individual perceptions of the risks of Covid-19 and individual attitudes toward the risks. The former is, for example, "the expected probability of being infected with SARS-CoV-2" and "the degree of symptoms if infected," while the latter (assuming preference stability) is "the individual's general risk-preference parameter." Furthermore, many studies believe that behind "subjective risk" there is an "objective risk" that brings it about, and they take the number of infected or deaths as the "objective risk." In these analyses, the individual's attributes (e.g., sex, age, marital status, education, occupation, income, health, religion, etc.) are usually added as control variables to these key variables. Although most studies assume rational individuals, it is also known that people have many irrational traits, so variables such as optimism, over-confidence, altruism, openness, and nervousness would also be considered.

From this approach, the analysis in this study focuses on what many studies treat as control variables. Then a question that arises is how to ascertain the influence of the key variables mentioned above. In regression analysis, if key variables are excluded and only attribute variables are regressed, the influence of key variables will be reflected in the coefficients

of each control variable through the correlation between key variables and each control variable. However, as this study does not include any of the above key variables that are usually assumed, it is extremely difficult to decipher the impact of key variables through attribute variables from the estimation results of this study. Therefore, interpretation of the estimation results is also difficult. For example, age and gender are shown to be the most important determinants, but it is not clear why this is so. This inadequacy, however, does not mean that ctree is inferior to regression analysis. The problem with this study is that variables such as “expected probability of being infected with coronavirus,” “degree of symptoms if infected,” and “number of cases or deaths in the area of residence” are not included in the determinants. Conversely, the same problem would arise in regression analysis if only attribute variables were considered.

So, what might one expect about the superiority of ctree and regression analysis? I have to admit that I know very little about ctree, so this is just a speculation. The most noteworthy results of this study are that age and gender splits at the first and second stage in both March and April and the third stage determinants, Race/Ethnicity, Education, Income, and Community attachment, in April are consistent with common sense, as explained in Discussion section. In addition, the absence of Race/Ethnicity and Education splits in March could be interpreted as the emergence of such social structure effects as Covid-19 became more severe. Of course, such an interpretation would need to be confirmed by a descriptive analysis. Such results might only be obtained by hierarchical analytical methods, which may be an analytical advantage of ctree. (A similar analysis may be possible with regression analysis, but I am not familiar with it.) On the other hand, in regression analysis, a researcher builds a model and tests the success or failure of the model. In contrast, ctree, with the exception of variable selection, is rather model-free and seems to be intended to give some information about what kind of model is valid, distancing it from the researcher’s subjectivity. If so, ctree and regression analysis are complementary rather than substitutable methods. Researchers who have used regression analysis may also find ctree analysis useful in setting up a model.