CREDO response to critique for multiple comparisons adjustment

There is no need in CREDO's analysis for a multiple comparisons adjustment such as a Bonferroni adjustment. Such adjustments were designed to eliminate the increased likelihood of committing a Type I statistical error (declaring a finding significant when it is not) when multiple comparisons are being made in a single analysis.

For a Type I error to occur in CREDO's analyses, several assumptions must be true. The primary assumption must be that the universal null hypothesis, none of the effects are real, is true. As long as one coefficient has a p-value of less than the Bonferroni adjusted α, then we have proven the universal null to be false. Due to the high power derived from the large data sets with which CREDO works, the vast majority of our significant coefficients are significant with p-values>0.000. The maximum Bonferroni adjustment based on our model sizes would be α=.001. That is to say we have proven the universal null hypotheses is false.[1]

This discussion, however, brings up the larger question of the appropriateness of adjusting for multiple comparisons in large sample multiple regressions. Making a Bonferroni adjustment would only be necessary if CREDO were conducting analyses on randomly generated numbers. Only in a set of random numbers would we expect the universal null hypothesis to be true. Since CREDO's analysis is based on observations of actual students and we are testing expected positive hypotheses (we have already proven the universal null hypothesis to be false in the previous paragraph), the idea that there is no association between any of the factors CREDO examines and academic growth is incorrect. Thus, including a Bonferroni adjustment in the tests of significance will unnecessarily increase the likelihood of committing a Type II error. There is no method which decreases the likelihood of a Type I error which does not automatically increase the likelihood of a Type II error. "And Type II errors are no less false than Type I errors."[2]

Finally, CREDO has conducted many analyses in recent years using multiple data sets which have produced consistent findings. The findings within the National Charter School Study 2013 and the Urban Study are consistent even though the data sets were not identical. The National Charter School Study 2013 contained data from 2005/06 through 2010/11. The Urban Study uses data from 2006/07 through 2011/12. This amounts to a 20% replacement of the data from one study to the next. If the results of the analyses were spurious, only the result of chance, such a large change to the data set should produce different sets of spurious findings. This has not happened.

While there are conditions where multiple comparisons adjustments are necessary, the analyses by CREDO do not meet those conditions. We believe the call for multiple comparison adjustments in our studies are not appropriate based on the size of our datasets and the hypotheses being tested. Instead, we will continue to provide detailed descriptive statistics of our datasets, thorough explanations of our analyses, and full disclosure of our results to demonstrate that we are respectful of the limitations of our data and analyses. We feel our current studies are well within these limitations.

1. Rothman, K.J. No adjustments are needed for multiple comparisons. Epidemiology 1990; Vol. 1, No. 1 : 43-46
1. Perneger, T.V. What's wrong with Bonferroni adjustments. BMJ 1998; 316 : 1236