Appendix G. Propensity Scoring Results

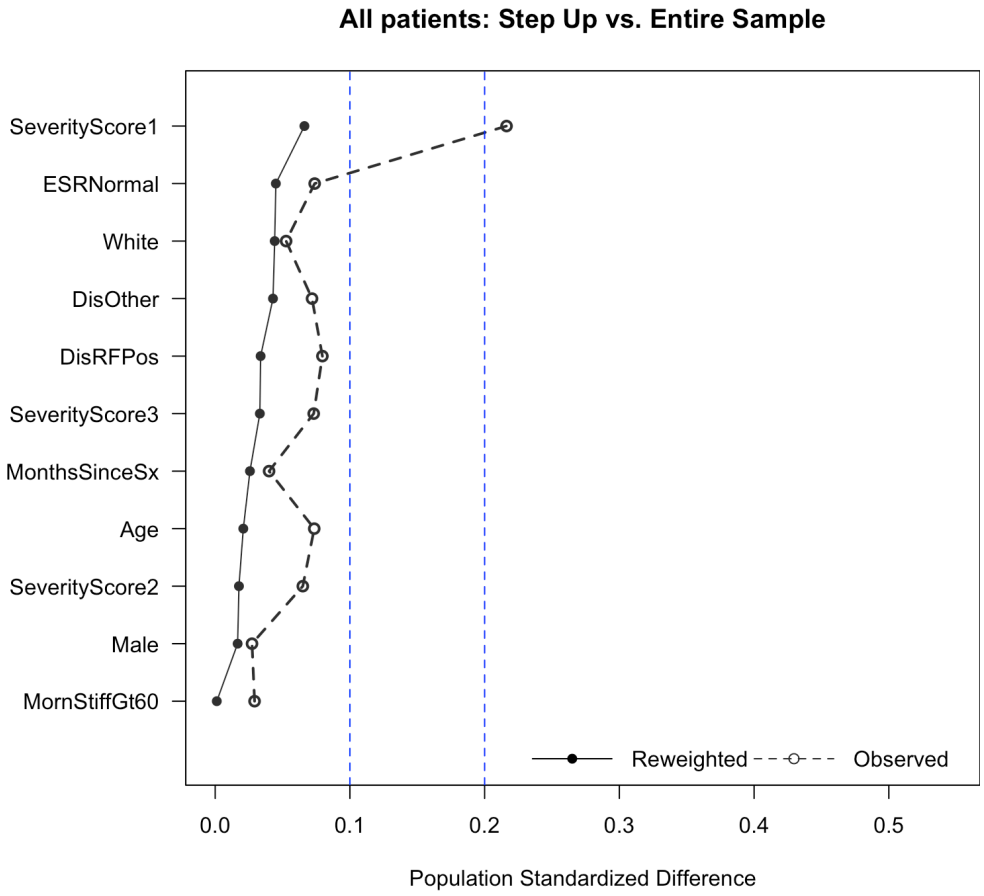
Appendix G: Propensity Scoring Results

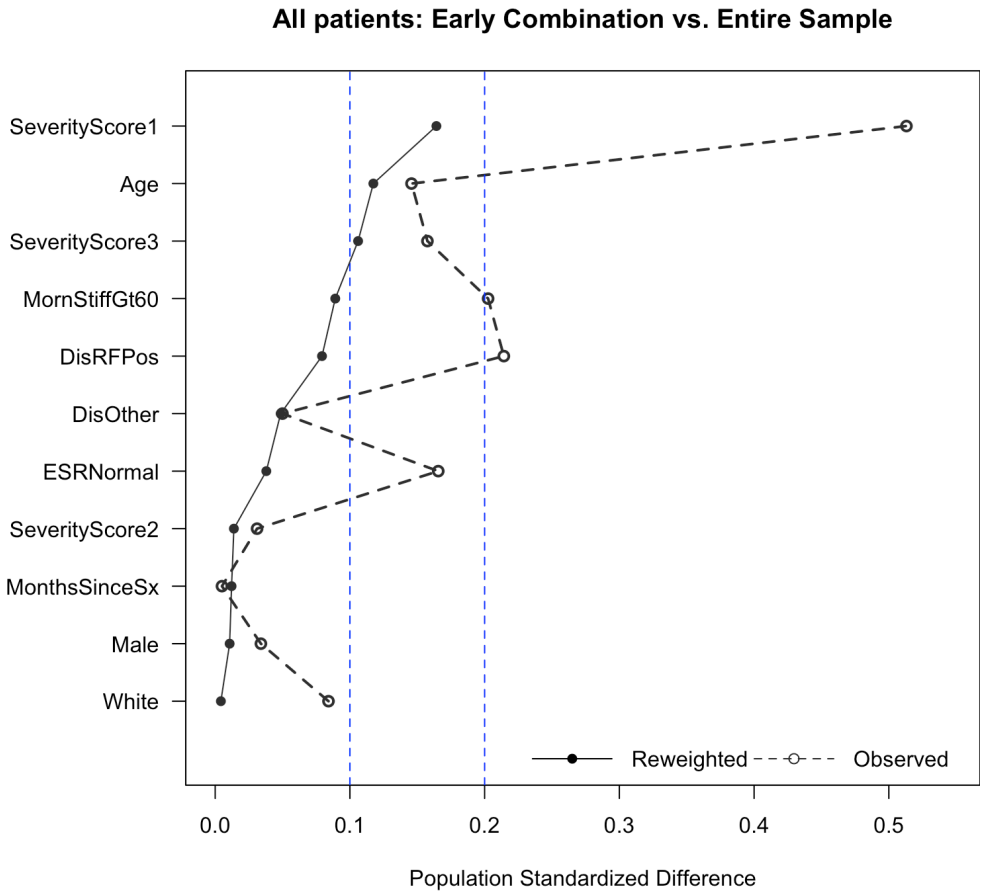
The propensity score (PS) model was built using the twang package (1) in R 3.6 following the methods described in McCaffrey (2). The moderate sample size meant that some care had to be exercised to include potential confounders, but not variables associated only with CTP, as this can increase variance without additional reducing bias. The PS model included variables that had a standardized difference larger than 0.1 between children achieving CID at 12 month and those that did not achieve CID or between those in remission (JADAS $\leq$ 2.4) and not in remission at 12 months as well as showing an average standardized difference between CTPs at baseline. In addition, as the main severity/disease activity variables appeared to differ between groups and the PS would need to balance on these variables to achieve face validity, the three severity scores (as described in the main methods of the report) was added to the PS, even though not all components met the 0.1 point standardized difference criterion.

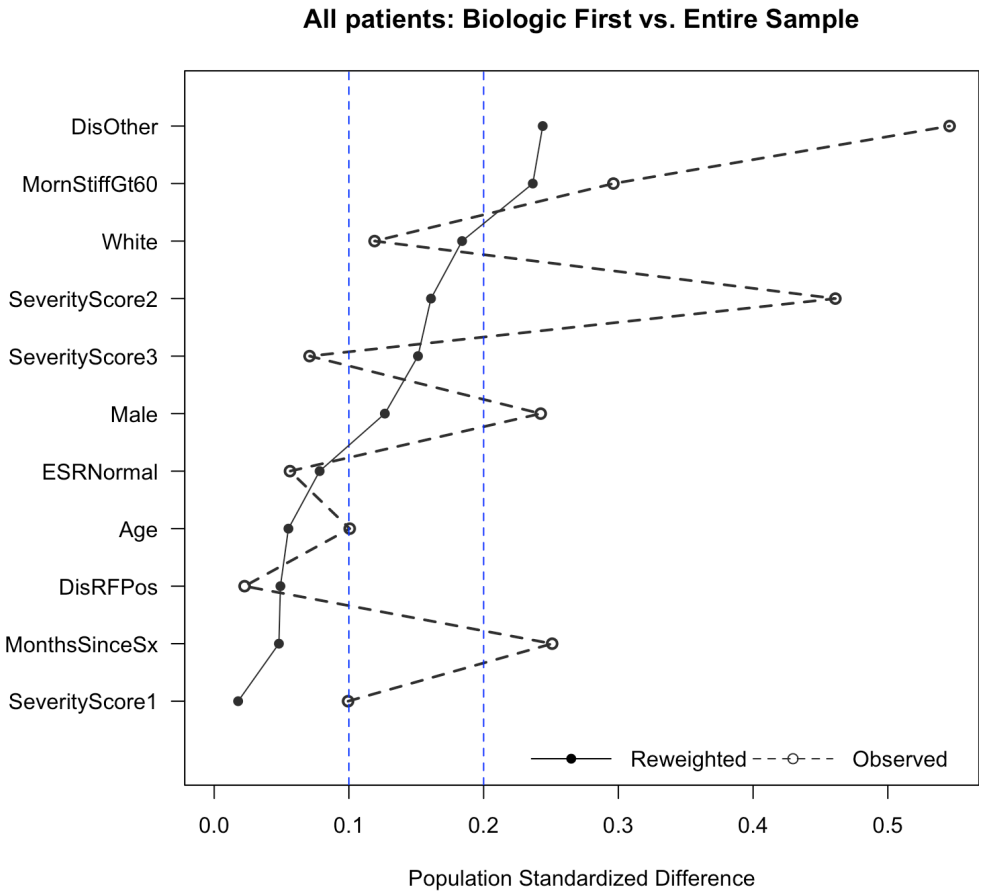
Three PS models were developed, each one calculating the propensity to be in a particular CTP group versus the other two. A single inverse probability of treatment weight (IPTW) was assigned to each participant as the reciprocal of the propensity to receive the treatment actually received. To limit excessive influence of any particular individual, IPTW values above the 99th percentile IPTW (11.8) were replaced by the 99th percentile value. This affected 4 (1%) of participants; the median IPTW was 1.5; 95% of IPTW were less than 4 and 95% were less than 5.5.

These IPTWs allow estimation of the average treatment effect (ATE). The ATE can be thought of as the effect of changing all patients from one CTP to another. We did not implement another common approach to use of the PS, matching. This first chooses a group that received one particular CTP (treatment A) and then for each patient in the group, finds a patient (or patients) receiving another CTP (treatment B) who have similar PSs. The resulting estimate of the effect of A vs B is called the average treatment effect in the treated (ATT), in this case, those treated with A. In this example, the estimated ATT is the one that would result from switching the type of patients that receive A to instead receive B.

To check balance on the variables in the PS, the IPTW were used to calculate standardized difference between each CTP and the overall group. The figures below show the standardized differences without weighting and with weighting.


All patients: Step Up vs. Entire Sample
SeverityScore1
ESRNormal
White
DisOther
DisRFPos
SeverityScore3
MonthsSinceSx
Age
SeverityScore2
Male
MornStiffGt60
Reweighted
Observed
0.0
0.1
0.2
0.3
0.4
0.5
Population Standardized Difference


All patients: Early Combination vs. Entire Sample
SeverityScore1
Age
SeverityScore3
MornStiffGt60
DisRFPos
DisOther
ESRNormal
SeverityScore2
MonthsSinceSx
Male
White
Reweighted
Observed
0.0
0.1
0.2
0.3
0.4
0.5
Population Standardized Difference



Comparisons of CID at 12-months used the IPTW as survey weights to estimate weighted mean proportions and weighted mean differences in proportions. Analyses used the survey package in R (4, 5).

To assess whether results for the between-group comparisons of CID at 12 months were sensitive to any remaining imbalance (i.e., any standardized differences > 0.1 in the weighted comparisons) identified in the figures above, the main IPTW model was refitted with these variables as covariates.

(1) Ridgeway G, McCaffrey D, Morral A, Griffin BA, Burgette L. (2017). twang: Toolkit for Weighting and Analysis of Nonequivalent Groups. R package version 1.5. https://CRAN.R-project.org/package=twang

(2) McCaffrey, D., Griffin, B., Almirall, D., Slaughter, M., Ramchand, R., Burgette, L. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Statistics in Medicine 32(19), 3388 - 3414.

(3) *Adelson, J., McCoach, D., Rogers, H., Adelson, J., Sauer, T.* (2017). Developing and Applying the Propensity Score to Make Causal Inferences: Variable Selection and Stratification. Frontiers in psychology 8(), 1413.

(4) T. Lumley (2019) "survey: analysis of complex survey samples". R package version 3.35-1.

(5) T. Lumley (2004) Analysis of complex survey samples. Journal of Statistical Software 9(1): 1-19