Title: Using Random Forests to identify the most Diagnostically-Relevant Items on the Social Responsiveness Scale

Authors: Aaron J Kaat PhD, Cristan A Farmer PhD, Audrey Thurm PhD, Vanessa Hus PhD, & Somer Bishop PhD

Introduction: The Social Responsiveness Scale (SRS; Constantino, 2011) is one of the most commonly used rating scales for autism spectrum disorder (ASD). It has served a dual purpose of diagnostic screening and rating ASD severity. However, it has been criticized for inconsistent results as a level-2 diagnostic screener (Hus, Bishop, Gotham, Huerta, & Lord, 2013; Norris & Lecavalier, 2010). Efforts have been made to improve it, most recently by creating a 16-item short form using item response theory (Sturm, Kuhfeld, Kasari, & McCracken, 2017). This short form represents a significant improvement. However, limitations persist, especially when focusing exclusively on either diagnostic screening or on ASD severity (Kaat & Farmer, 2017). The purpose of this study was to identify the most diagnostically-relevant items on the full-length SRS, and compare those items to the short form proposed by Sturm et al. (2017) and a diagnostic screener proposed by Duda, Ma, Haber, and Wall (2016).

Method: We used the Simons Simplex Collection (SSC; n=4903) and a clinic-referred sample (n=1183) to compare individuals with best clinical estimates of nonASD or ASD. In order to identify the most salient items, we used random forests—a machine learning technique (Breiman, 2001). We split the SSC sample into a training set (75% of cases) and a validation set (25%). In the training set, we tuned the random forest to find the ideal number of variables to randomly split at each tree node using 500 classification trees, then grew a random forest of 100 trees. We cross-validated it in the SSC validation dataset and the clinic-referred sample. Examination of item usage within the forest, and item importance (accuracy and reduced Gini impurity) were used to reduce the number of items, which were then fit in a logistic regression model to estimate sensitivity and specificity.

Results: The complete random forest had a 2.6% out-of-bag error rate and a 3.1% error rate in the training and cross-validation datasets, respectively. The two importance measures and the item usage values largely supported a five-item subset as optimal. One item had low face-validity for diagnostic screening and so a 4-item diagnostic screener was suggested. Using only these items in the SSC full sample resulted in a sensitivity and specificity of 0.96. However, in the clinic-referred sample, the error rate increased, resulting in a sensitivity of 0.77 and specificity of 0.78 for the full random forest. Using only the optimal 4 items reduced specificity to 0.37, but increased sensitivity to 0.93.

Discussion: This study suggests that the 65-item SRS could be reduced to a 4-item rating scale if it were used only for diagnostic purposes. Three of these items also appear on the recently-proposed 16-item short form (Sturm et al., 2017) and all four appear on the screener proposed by Duda et al. (2016). The reduced specificity suggests that it may be most useful in identifying who needs a more thorough evaluation, as there will be a higher than acceptable number of false positives. However, these results are promising for future research.

References/Citations:
- Duda, M., Ma, R., Haber, N., & Wall, D.P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Translational Psychiatry, 6*, e732. DOI: 10.1038/tp.2015.221