



**Figure S1.** Example of variable thresholding. The number of claims is shown for the first 100 diagnostic variables within both cohorts (top) and only the ASD cohort (bottom). Dashed lines represent 2% threshold for both cohorts and ASD cohort, 2,477 and 25 claims, respectively. Variables with claims above the threshold (noted by patterned bars) are kept in the analysis. The y-axis begins at 11 in order to remain compliant with OptumLabs' de-identification policy.