Figure 1 shows the ROC curves, with areas under the curves, for the SCQ and the PDDST-II using the imperfect gold standard of parental self-report of ASD diagnosis and/or autism special education classification. The discriminating ability of both tools is similar in this sample, with the SCQ performing only slightly better. The estimates in Table 2 suggest that, based on analyses in this common sample, at recommended cutpoints the SCQ is the more specific, and the PDDST-II the more sensitive screening tool. The data from the subsample with ADI-R scores (Figure 2 and Figure 3) supports this. An optimal cutpoint based on Youden’s J seeks to maximize both sensitivity and specificity (irrespective of prevalence in the sample) while the Index of validity maximizes the number correctly classified (will favor specificity when prevalence is low). In this example, the optimal cutpoints for both instruments perform similarly (not unexpected, given the similar ROC curves).

Discussion

In the single sample used in this study, the SCQ and the PDDST-II had very different performance characteristics based on their recommended cutpoints. However, ROC curve and optimal cutpoint analyses suggest that the tools actually perform quite similarly. The difference in performance of the tools when used as recommended is clearly a function of the cutpoints. The selection of a recommended cutpoint reflects both the preference of the developer and the characteristics of the populations used in initial validation studies. Developers select cutpoints based on the setting in which they intend the tool to be used, judgments about the relative costs associated with false negatives and positives, and empirical data available from initial validation studies on test performance. Although reports on the development of a tool rarely give all the specifics regarding these decisions, it appears that the SCQ and the PDDST-II had similar general objectives and where initially validated using similar study designs. Our analysis shows that they will, in fact, perform similarly when the same criteria are used to select cutpoints and a common validation sample is used.