Identifying Autism in Electronic Health Records: The Role of Natural Language Processing Algorithms

Beth A. Malow, Olivia J. Veatch*, Xinnan Niu, Kasey Fitzpatrick, Donald Hucks, Angie Maxwell-Horn, Lea Davis
Vanderbilt University Medical Center, and University of Kansas Medical Center*

Introduction

Electronic health records (EHR) are rich sources of data for conducting studies and identifying participants for clinical trials in autism spectrum disorder (ASD). While diagnostic codes are useful for defining cases, they may not capture all patients or be applied by subspecialists (sleep, GI). ICD codes may also be applied to non-ASD cases. We examined the value of natural language processing (NLP) to accurately identify ASD in a deidentified, EHR-derived cohort of patients

Methods

- Two cohorts, currently ages 20-25 years old, were retrieved from the Vanderbilt Synthetic Derivative (SD) implemented under turn++
- All patients had at least four notes containing key terms (autism, autistic, autism spectrum disorder, Asperger, pervasive developmental disorder, ASD, or PDD). In Cohort 1 only, patients also had at least one instance of an International Classification of Diseases, Ninth or Tenth Revision (ICD-9 or ICD-10) code for ASD.
- To validate and score notes within a patient's chart, NLP algorithms were applied. Data were pulled using SQL and validated using a UNIX bash shell script embedded with KnowledgeMap Concept Indexer.
- Notes were scored "1" for only positive, "0" for only negated ("does not have autism"), and "0.5" for only possible ("has possible autism"). Notes were scored "N/A" if no ASD key terms relevant to the patient were identified ("brother with autism"). If a combination of ASD key terms were present in a note, the average was determined (e.g., a note with a positive and a possible would receive a combined score of 0.75).

Work Supported by NIMH K18 MH122755, Career Enhancement Award to Dr. Malow

- To account for potential differences in the number of available notes per patient, a final score for each patient's chart was calculated by taking the mean of the scores for each individual note. Cases of ASD were defined based on a cutoff of ≥0.8.
- Sensitivity and specificity were calculated based on manual chart reviews performed on all notes using a standard rubric, previously developed by Dr. Davis, which classified patients into high-, mid-, or low-evidence for ASD, or exclusions for ASD.

Results

Cohort 1 had n=355 and Cohort 2 had n=44 potential ASD charts. For Cohort 1, NLP algorithm sensitivity was 100% for high-, and 97% for both mid- and low-evidence patients. For Cohort 2, sensitivity was 100% for mid- and 94% for low-evidence patients. Specificity was 27% for Cohort 1 and 33% for Cohort 2; this was due to a combination of factors, including how ASD is characterized, sentence structure, or evolution of ASD diagnosis over time.

Average score improved for the Cohort 1 low-evidence charts when limited to 4 or more notes (Figure). High and mid-evidence charts were not affected by note number.

Discussion

Our NLP methods identified ASD charts with high sensitivity, even in the absence of ICD codes. While specificity was low, we expect to achieve higher levels through NLP refinement, scoring adjustments, and standardization of EHR documentation. Limiting charts with 4 or more notes improved total score in the Cohort 1 low evidence level.

Had ICD codes been used as the sole criteria for ASD diagnosis, a sizable proportion of patients (11%) would have been missed. Our findings support the use of NLP in the identification of ASD, and other complex conditions.

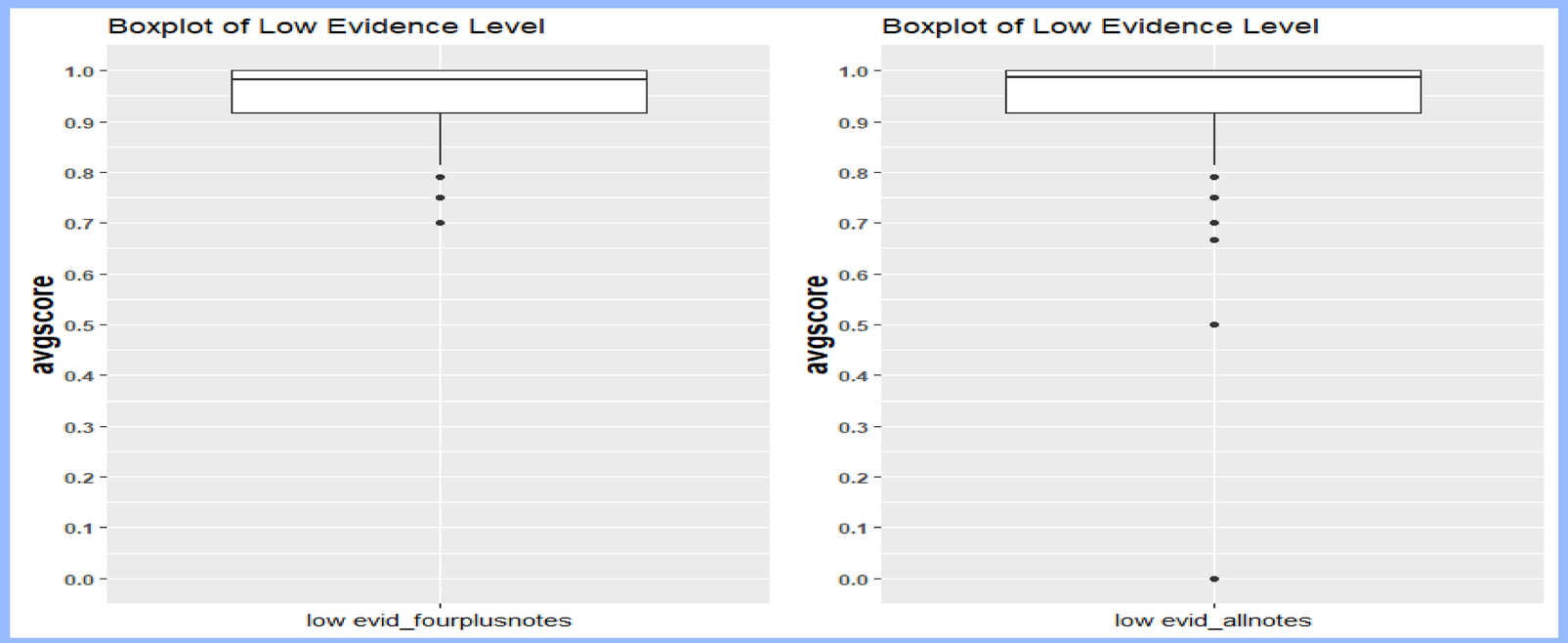

Figure. Mean scores for ASD low evidence level. Plotted are the average scores, on the y-axis, calculated for patients with low-evidence for ASD based on manual reviews. Scores improved when charts were limited to 4 or more notes. Black circles represent outliers.