Predicting Potential Candidacy for IBD Research Trials Using Integrated Data Science Modeling in Community-Based GI Practices

Donald J. Lazas, MD, FACP, AGAF[1]; Christian Bonnell, MSc[1]; Josh O'Rourke, MBA[1]; Colleen Hoke[1]; Anjali K. Morey, MD, PhD[1]; Vinay Patel, MD[1]; Matt Bachinski, MD[1]; Pankaj A. Patel, MD, MSc[1]; Senthil K. Raghavan, MD, MPH[1]
[1]ObjectiveHealth, Franklin, TN

INTRODUCTION

IBD patients often relapse and require a change in treatment (Rx). Our goal was to assess patients' therapeutic journey and identify those who may be candidates for research trials based on AI algorithm derived data that quantifies propensity for current Rx failure.

METHODS

We analyzed 2,712 patients with a confirmed diagnosis of IBD (Crohn's and UC). For each we compiled a list of related ICD10 symptom codes, partitioning into 11 distinct sets corresponding to symptom category. Similarly, we compiled a list of IBD medication codes prescribed to each patient, assigning each to one of six drug categories (e.g., 5ASA, steroids, Biologics). We constructed a feature set for each patient by summing the number of diagnoses and prescriptions which was then transformed into a 2-dimensional feature space using TSNE. This feature space was partitioned using HAC clustering with Ward linkage. Analysis indicated the optimal number of clusters to be 8 (Table 1).

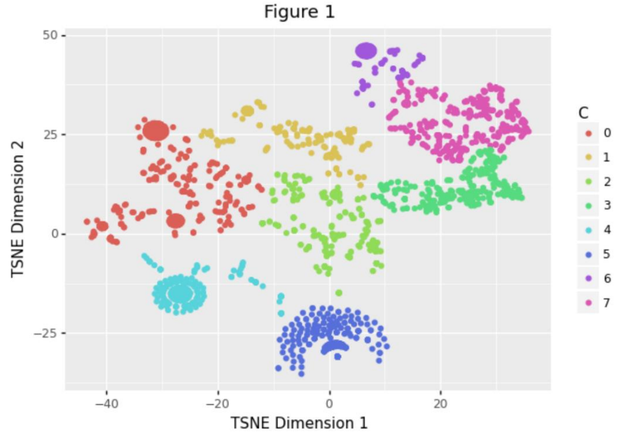
RESULTS

Analysis of the base feature set for each cluster (C) indicated a high degree of uniformity, and the overall picture confirmed the anecdotal evidence from physician experts for treating IBD. In Fig. 1 we see the clusters identified by color. C5 is the null group, for whom we have no data, or only a single secondary symptom of IBD, and C0 includes patients for whom we have more significant symptom data, but no treatment data. C 4, 2 and 3 form a continuum where C4 includes patients who have been treated with 1-2 steroid courses, C2 contains patients with >= 2 steroid courses together with at least one other drug category, and the right side of C3 contains patients with multiple steroid courses together with at least 4 other Rx categories. C6 and C7 form a similar continuum with similar trends but centered around ASA treatments. C1 contains patients with only antibiotic Rx and no other Rx types and blends into the corticosteroid spectrum of C2 and 3. Biologic therapy was not the defining feature of any specific cluster but does distinguish C2 from C3 and C6 from C7. We conclude that C3 and C7 contain strong candidates for research.

Table 1

Clusters	Calinsky-Harabasz Index
4	284.108512
5	373.804537
6	306.331922
7	284.240381
8	428.465493
9	383.481193
10	345.519132
11	320.023012

Figure 1



CONCLUSION

Our AI model successfully confirms the anecdotal data provided by our physician experts and drug trial sponsors. The tool can identify patients who are currently good candidates for IBD drug trials in advance of a clinic visit and predict those who are likely to become good candidates sometime in the future.

Presenter:
Jim Cremins, MD
ObjectiveHealth
Email: Jim.Cremins@Objective.Health
Phone: 240-291-2366
Website: www.objective.health