Uncovering Shared and Distinct Seizure Phenotypes in Rare Diseases Using RARE-X PRO Data

Zohreh Talebizadeh PhD, Alexander Farnum PhD, Charlene Son Rigby MBA

Global Genes, 1012 14th Street NW, Washington, DC 20005, USA

BACKGROUND

Seizures & Rare Diseases: Seizures are a common complication in many rare diseases and often contribute to diagnostic delays and management challenges. Epilepsy, characterized by recurring seizures due to abnormal neuronal activity, is frequently linked to underlying genetic mutations. Its presentation varies widely across rare conditions, often co-occurring with other neurodevelopmental symptoms. Identifying patient-level patterns of seizure-related comorbidities is critical to improving understanding and care strategies across these diverse populations.

RARE-X: RARE-X, the research program of Global Genes, collects patient-reported outcomes (PRO) data based on symptom profiles rather than diagnosis alone. With over 9,000 participants representing 80+ rare conditions, including many neurodevelopmental disorders, RARE-X enables hypothesis-agnostic, cross-condition research. The platform's focus on shared and unique symptom patterns supports discovery across traditionally siloed diseases.

The RARE-X survey structure begins with a general health survey called "Head-to-Toe," adapted from and expanded upon the GenomeConnect Registry developed by ClinGen to capture comprehensive baseline health information. Participants then complete Level 2 domain-specific surveys, enabling standardized and scalable assessment across rare diseases. Surveys are assigned using a combination of branching logic and domain-specific selection, which tailor the content based on information already provided by participants.

METHODS

RARE-X data from 10 rare disorders were analyzed using both aggregate and patient-level approaches. Data preprocessing applied sequential inclusion criteria: filtering for diseases of interest, retaining the latest complete record per participant, excluding missing/"unsure" responses for features of interest, and restricting the sample to individuals who completed both the Head-to-Toe survey and the Level 2 Brain & Nervous System survey. This resulted in 555 participants with complete data (20 symptoms as features; seizure status as the target).

Aggregate analyses assessed prevalence differences and Chi-square associations, while patient-level models (random forest and logistic regression) evaluated feature contributions using accuracy, precision, recall, and F1 scores. Hierarchical clustering using Hamming distances examined symptom co-occurrence patterns. All boolean features were coded 1/0 for consistency.

RESULTS

Data were evaluated using complementary methods:

- Aggregate analyses identified symptoms with notable frequency differences and statistically significant associations with seizure status.
- Patient-level machine learning models highlighted features that contributed most to accurate classification, with performance metrics indicating where misclassifications occurred.
- Hierarchical clustering revealed symptom patterns that commonly co-occur across individuals, showing how features relate within and across disorders.

RESULTS

Table 1. Conditions and Participant Counts

Rare Diseases	Total Enrolled	Head to Toe Survey Completed
Leber hereditary optic neuropathy (LHON)	274	171
STXBP1-related disorders	245	135
Wiedemann-Steiner syndrome (WSS)	234	120
FOXP1 syndrome	273	97
CACNA1A-related disorders	147	71
SYNGAP1-related disorders	138	67
CHD2-related disorders	200	55
Malan syndrome	109	60
Classical homocystinuria	82	46
Kleefstra syndrome	145	43

Figure 1. RARE-X Data Collection Program

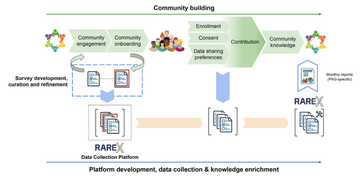


Figure 2. Symptom Prevalence Across Seizure Groups

Average Yes/No symptom responses were compared between participants with and without seizures to identify group-level differences. Features are ordered by the size of these differences, with those showing the largest gaps and therefore the strongest symptom prevalence contrast appearing on the top.

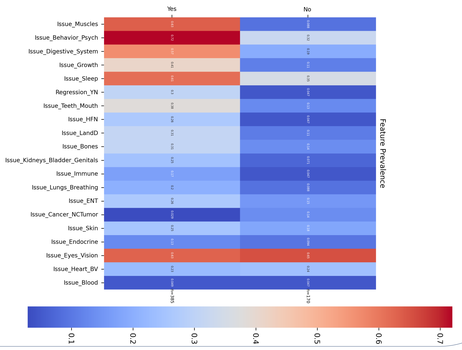


Figure 3. Symptom Associations With Seizure Status

A Chi-square test evaluated whether each symptom's Yes/No frequency differed between seizure and non-seizure groups. Features are sorted by p-value.

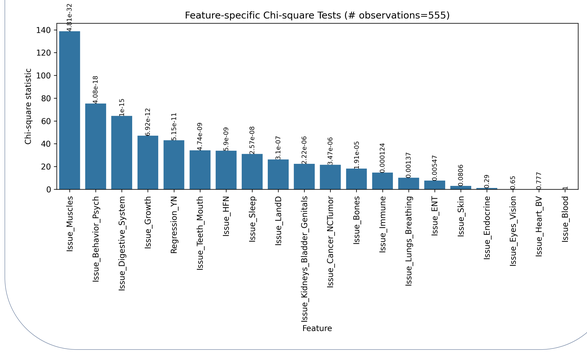


Figure 4. Random Forest Feature Importance (Patient-Level Analysis)

A stratified 5-fold random forest classifier was trained to predict seizure status from individual symptom patterns, and feature importance scores identify the symptoms that contributed most to accurate classification.

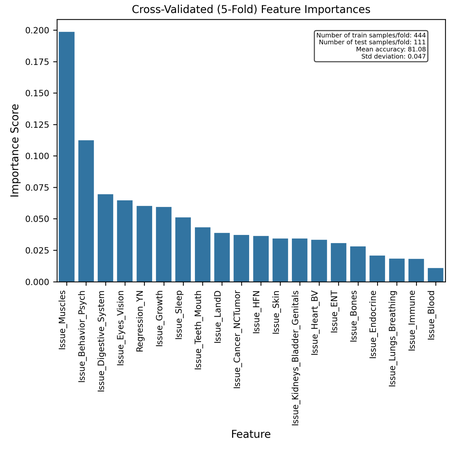


Figure 5. Hierarchical Clustering of Symptoms (Patient-Level Analysis)

Symptoms were clustered using Hamming distances to examine how often they co-occur across participants. Features that merge at lower points in the dendrogram share more similar reporting patterns, suggesting clinically relevant symptom groupings.

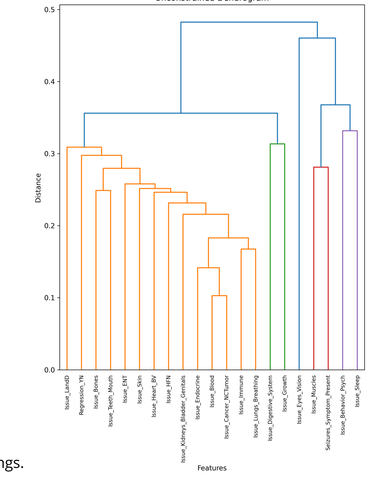
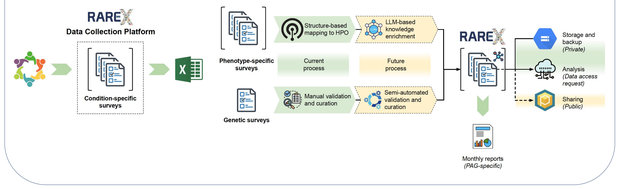


Figure 6. RARE-X Data Processing and Future Directions



DISCUSSIONS & ACKNOWLEDGEMENTS

1. Several symptom features consistently differentiate participants with seizure history across diverse rare disorders.
2. Converging evidence across aggregate statistics, machine learning models, and clustering supports these features as meaningful indicators of seizure-related burden.
3. The multi-method approach demonstrates how symptom-based data can reveal cross-disorder patterns in heterogeneous rare disease populations.

Future analyses should evaluate these feature patterns longitudinally, assess their relevance within specific diagnostic subgroups, and explore whether integrating clinician-reported or genetic data can enhance seizure-related phenotype modeling.

Acknowledgments: We thank the PAGs, RARE-X participants, sponsors, the Scientific Advisory Board, and the RARE-X team for their support and contributions.

Reference: RARE-X: A Patient-Driven Approach for Collecting Symptom and Patient-Reported Outcome Data in Rare Diseases. *Genetics in Medicine* 2025 (In Press)