Creating Sex-Specific Phenome Risk Classifiers via Clinical Comorbidities to Identify Under-documented Cases of Developmental Stuttering in Electronic Health Records

Dillon Pruett[1], Doug Shaw[2], Lauren Petty[2], Hung-Hsin Chen[2], Hannah Polikowsky[2], Shelly Jo Kraft[3], *Jennifer Below[2] and *Robin Jones[1]

Affiliations:

[1]Hearing and Speech Sciences, Vanderbilt University, Nashville, TN

[2]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN

[3]Communication Sciences and Disorders, Wayne State University, Detroit, MI

**Senior authors made equal contributions*

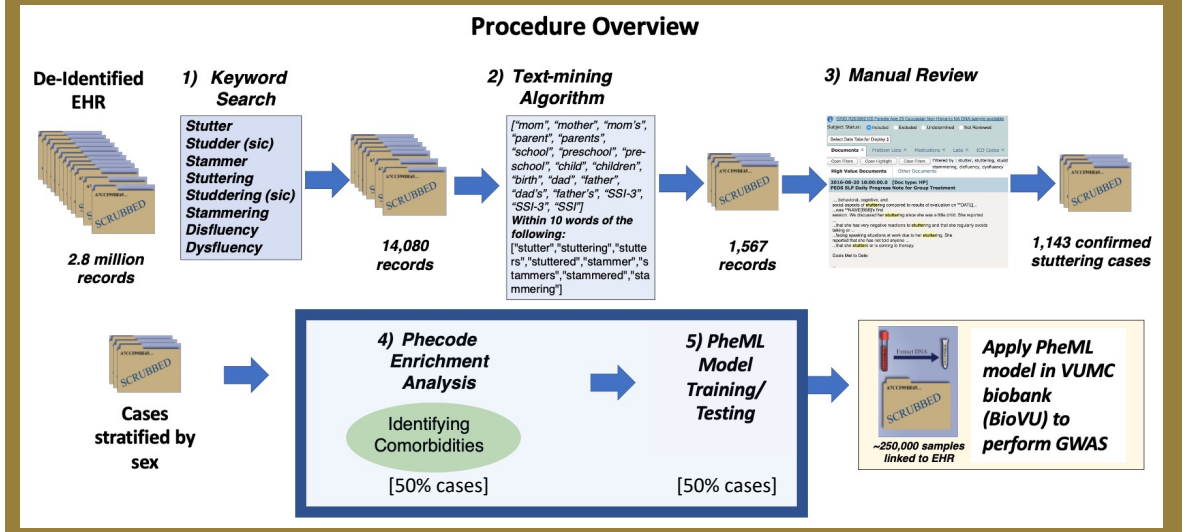
INTRO

- Stuttering is difficult to characterize in electronic health records (EHRs). Phenotyping using **keywords**, **text mining**, and **manual review** can identify far more cases of stuttering than ICD 9/10 codes alone.
- Once cases are identified, EHRs offer a unique tool for 1) examining stuttering comorbidities, which in turn can be used to 2) impute a stuttering phenotype to predict cases.

METHODS

- Cases (male = 430, female = 140) were compared to matched controls (~5 per case) in a **permutation-based phecode enrichment analysis** to examine **sex-specific comorbidities.**
- Enriched phecodes were used to build and test sex-specific, phecode driven Gini impurity-based **classification and regression tree phecode machine learning (PheML) models** to predict under-documented stuttering cases.

Males who stutter have ***more associated comorbidities*** than females who stutter; comorbidities can be used to impute the stuttering phenotype to identify additional under-documented cases in EHRs.



Questions? Comments? Email at Dillon.G.Pruett@Vanderbilt.edu

RESULTS

Male Phecode Enrichments of Previously Suggested Comorbidities

Phecode	Description	
Childhood onset fluency disorder		
315	Developmental delays and disorders	
315.2	Speech and language disorder	
315.1	Learning disorder	*
Pervasive developmental disorders and Adult onset fluency disorder		
313	Pervasive developmental disorders	
313.2	Tics and stuttering	
Hearing loss		
389	Hearing loss	*
389.2	Conductive hearing loss	*
Sleep disorders		
327	Sleep disorders	*
Atopic triad		
465	Acute upper respiratory infections	*
512.8	Cough	*
687.1	Rash and other nonspecific skin eruption	*
930	Allergic reaction to food	*
939	Atopic/contact dermatitis	
512	Other symptoms of respiratory system	*
519.9	Symptoms involving respiratory system and other chest symptoms	*

Female Phecode Enrichments of Previously Suggested Comorbidities

Phecode	Description	
Childhood onset fluency disorder		
315	Developmental delays and disorders	
315.2	Speech and language disorder	
Pervasive developmental disorders and Adult onset fluency disorder		
313	Pervasive developmental disorders	
313.2	Tics and stuttering	
Hearing loss		
382	Otalgia	*
Atopic triad		
939	Atopic/contact dermatitis due to other or unspecified	
949	Allergies; other	*

** Indicates comorbidity unique to that sex*

Male Phecode Enrichments of Novel Comorbidities

Phecode	Description	
Diagnostic testing and Infections		
1010	Other tests	
783	Fever of unknown origin	
79	Viral infection	*
472	Chronic pharyngitis and nasopharyngitis	*
381	Otitis media/Eustachian tube disorders	*
381.1	Otitis media	*
381.11	Unspecified otitis media	*
110	Dermatophytosis / Dermatomycosis	
110.1	Dermatophytosis	
369	Infection of the eye	
369.5	Conjunctivitis; infectious	
870	Open wounds of head; neck; and trunk	*
Neurological deficits		
292	Neurological disorders	
292.1	Aphasia/speech disturbance	
350	Abnormal movement	
350.2	Abnormality of gait	
350.3	Lack of coordination	*
Weight control		
278	Overweight; obesity and other hyperalimentation	
278.1	Obesity	
278.4	Abnormal weight gain	
1002	Symptoms concerning nutrition; metabolism	
264	Lack of normal physiological development	*
264.9	Lack of normal physiological development	*

Female Phecode Enrichments of Novel Comorbidities

Phecode	Description	
Diagnostic testing and Infections		
1010	Other tests	
369	Infection of the eye	
369.5	Conjunctivitis; infectious	
110	Dermatophytosis / Dermatomycosis	
110.1	Dermatophytosis	
783	Fever of unknown origin	
Neurological deficits		
292	Neurological disorders	
292.1	Aphasia/speech disturbance	
345	Epilepsy; recurrent seizures; convulsions	*
345.3	Convulsions	*
350	Abnormal movement	
350.2	Abnormality of gait	
788	Syncope and collapse	*
Weight control		
278	Overweight; obesity and other hyperalimentation	
278.1	Obesity	
278.4	Abnormal weight gain	
1002	Symptoms concerning nutrition; metabolism; and development	

PheML Model Confusion Matrices

Males	Confirmed Stuttering Cases	Matched Controls	
Cases classified as *high-likelihood*	62	16	PPV = 0.795
Cases classified as *low-likelihood*	31	517	

Females	Confirmed Stuttering Cases	Matched Controls	
Cases classified as *high-likelihood*	27	10	PPV = 0.730
Cases classified as *low-likelihood*	8	154	

DISCUSSION/FUTURE DIRECTIONS

- **Males had more significantly enriched phecodes** compared (38) than females (24) with a higher proportion of sex-specific enrichments (50%), **but broad clinical categories remained similar.**
- Testing the PheML model indicates it has **adequate positive predictive value for stuttering case acquisition** within biobank linked EHRs.
- PheML-predicted cases will be used for GWAS.