Delineating the Clinical and Genetic Spectrum of Speech Disorders in 52,143 Individuals

Jan Magielski[1,2,3,4], Julie Xian[2,3,4], Shridhar Parthasarathy[2,3,4], Sarah Ruggiero[2,3], Peter Galer[2,3,4,6], Shiva Ganesan[2,3,4], Amanda Back[2,3], Jillian McKee[2,3], Alex Gonzalez[4], Joseph Donaher[7,8], Ingo Helbig[2,3,4,5]

[1] University of Pennsylvania (COL 2024) ; [2] Division of Neurology, Children's Hospital of Philadelphia; [3] The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hospital of Philadelphia; [4] Department of Biomedical and Health Informatics (DBHi), Children's Hospital of Philadelphia; [5] Department of Neurology, University of Pennsylvania Perelman School of Medicine; [6] Center for Neuroengineering and Therapeutics, University of Pennsylvania; [7] Center for Childhood Communication, Children's Hospital of Philadelphia; [8] Department of Otorhinolaryngology, University of Pennsylvania Perelman School of Medicine

BACKGROUND

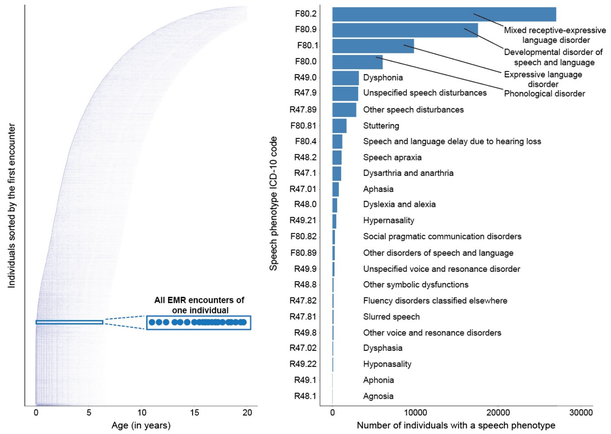
- Speech and language disorders are known to have a substantial genetic contribution. Although frequently examined as components of other conditions, research on speech differences as separate phenotypic subgroups has been limited so far.
- Previous studies suggest a strong genetic component to speech and language disorders.
- No studies to date have mapped longitudinal clinical histories of speech and language disorders through an integrative approach, leveraging both in-depth genomic data and clinical information captured over time.
- A variety of genetic etiologies *(GRIN2A*, *FOXP2*, *STXBP1)* have been suggested to be associated with neurological disruptions of speech and language, but these studies often lack the statistical support that is now available through our increased understanding of population genetics and the development of human genome databases.

METHODS

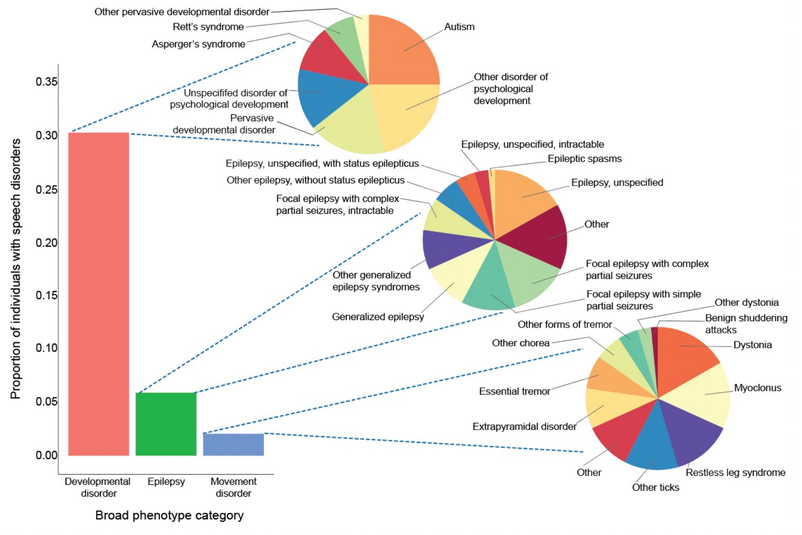
- We performed an in-depth characterization of speech disorders in 52,143 individuals, reconstructing clinical histories using a large-scale data mining approach of the Electronic Medical Records (EMR) from a large pediatric healthcare network.
- We selected a group of the relevant ICD-10 codes to define a broad neurological cohort. Subsequently, we compiled a list of ICD-10 codes describing speech phenotype-related diagnoses (F80, R47-R49) to delineate our speech cohort.
- In the sub-cohort comprised of individuals from the Epilepsy Genetics Research Project (EGRP), we were able to access charts from all encounters, assign relevant Human Phenotype Ontology (HPO) terms, and analyze raw whole-exome sequencing data.

RESULTS

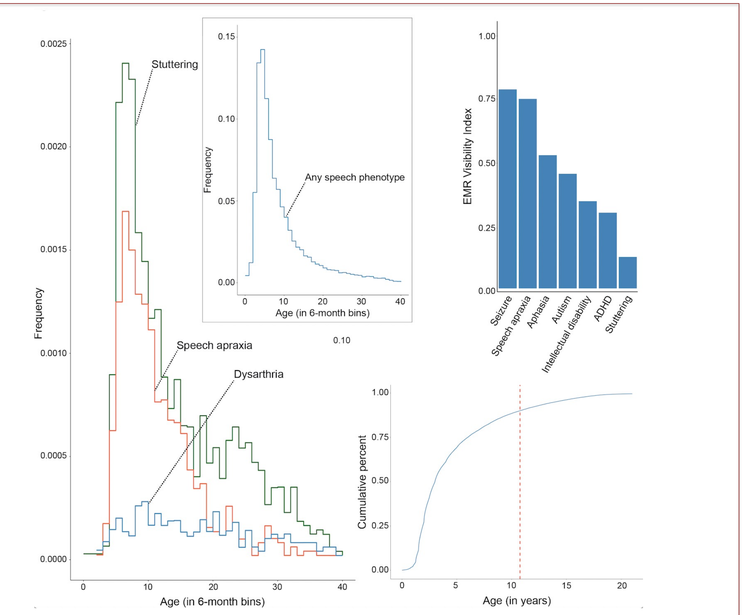
Speech phenotype	Genetic diagnosis	N	*P*-value	OR	95% CI	Frequency
Aphasia	*STXBP1*	9	8.57×10^{-12}	50.23	18.62-130.39	0.43
	POLG	2	0.0013	65.87	4.77-898.38	0.50
Speech apraxia	*GRIN2A*	3	3.30×10^{-4}	34.06	4.98-201.11	0.43
	NAA10	2	0.0014	90.60	4.71-5110.56	0.67
	MT-TI1	2	0.0014	90.52	4.71-5106.15	0.67
Speech delay due to hearing loss	*MYO7A*	3	1.24×10^{-5}	Inf	17.46-Inf	1
Other developmental disorders of speech	*MECP2*	2	9.81×10^{-4}	54.02	5.45-284.24	0.22



- We identified 1,671,257 encounters across 52,143 individuals with speech and language disorders, spanning a total of 203,150 patient-years .
- The most common speech-related diagnoses for this cohort were mixed receptive-expressive language disorder developmental disorder of speech and language, and expressive language disorder.



- Among broad phenotype categories, speech disorders most frequently co-occur with developmental conditions, with autism being the most common comorbidity.
- Speech disorders also commonly overlap with epilepsy and movement disorders.
- In the subgroup with comorbid speech and epilepsy diagnoses, 90% of individuals had a speech diagnosis documented by 14.6 years, while for those in the speech cohort without an epilepsy diagnosis, that age was 10 years.
- In the speech-neurodevelopmental sub-cohort, for individuals with co-occurring speech and neurodevelopmental disorders, 90% of individuals received their speech diagnosis at 10.2 years, in comparison to 10.7 years for the individuals presenting with a speech phenotype, but without a neurodevelopmental disorder.
- The most frequent genetic diagnoses in the speech cohort include *STXBP1*, *PTEN*, *CACNA1A*, *SCN2A*, and *SYNGAP1*, known monogenic causes of epilepsy and developmental disorders.



- Speech phenotype-related diagnoses were most prevalent in the second year of life, and the majority of speech diagnoses were made between ages 2 and 5.
- In individuals who have stutter, speech apraxia, and dysarthria, we observed that the highest frequency still occurs within the 2-5 years old window, but slightly later than in the case of pediatric speech phenotypes at large.
- Analyzing exome sequencing data showed that missense variants in *GRM2*, a gene encoding a metabotropic glutamate receptor 2 without an established phenotype, is associated with aphasia.

CONCLUSIONS

- There is a considerable heterogeneity in the landscape of speech disorders. Further refining of key time points in speech conditions progression should allow for earlier detection by clinicians.
- Distinct speech phenotypes are associated with specific genetic diagnoses, which shows an important facet of known monogenic developmental and epilepsy disorders. Speech conditions show both phenotypic and genetic overlap with these broad phenotypes.
- Rigorous prospective studies and genetic testing of individuals affected by speech disorders can provide further insights into how mutations in specific genes contribute to distinct speech clinical presentations.

ACKNOWLEDGEMENTS

This study was supported by the College Alumni Society Undergraduate Research Grant.