Symposium Title: Using Electronic Health Records in Research on Neurodevelopmental Disorders: Phenotypic Discovery and Implications for Privacy

Chair: Marsha Mailick, Waisman Center, University of Wisconsin-Madison

Discussant: Matthew Maenner, Centers for Disease Control and Prevention

Overview: Electronic health records (EHRs) offer data on the health of individuals with neurodevelopmental disorders. Unlike most sources of data used in IDD research (e.g., direct observation, testing, informant report), EHR data are drawn from billing codes in the health care system and are not tailored to address phenotypic characteristics associated with neurodevelopmental disorders. Further, EHR data are highly personal, reflecting the individual’s health characteristics and health care utilization, and as such there are privacy issues that must be considered in utilizing such data in research.

This symposium brings together three studies that differ in their focus but together have the potential to shed light on the use of EHR data in research on neurodevelopmental disorders. The first presentation focuses on potential sex differences in the health of adults with autism spectrum disorders. Drawing from a large data set of ICD-9 and ICD-10 codes assigned to men and women with and without autism during their medical encounters, the heavy burden of disease experienced particularly by women with autism is revealed. The second presentation focuses on health characteristics that differentiate individuals with FXS from controls, based on their EHR data. Even after eliminating EHR codes for mental disorders and congenital anomalies, those with FXS have a greater profile of health conditions, many of which have never been reported before as associated with FXS. This discovery was made possible by the application of machine learning to the corpus of EHR data in individuals with FXS and controls. The third presentation takes a different perspective, focusing on the attitudes of parents of individuals with autism, FXS, and unaffected children toward allowing their child’s EHR data to be used in research. Although there are significant privacy concerns associated with the EHRs of individuals with neurodevelopmental conditions, it was their parents who were most supportive of the use of the record data in research.

Discussion of these presentations will be offered by Matthew Maenner, who serves as the Surveillance Team Lead, Developmental Disabilities Branch, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention in the Centers for Disease Control and Prevention. He has extensively analyzed medical records data in autism surveillance, and used machine learning approaches in identifying individuals with autism spectrum disorders. As such, he is uniquely qualified to comment on these three presentations and point to a future research agenda.

Paper 1 of 3

Paper Title: Discovery of Sex Differences in Health Profiles for Adults with ASD using Electronic Health Records

Authors: Leann Smith DaWalt¹, Julie Lounds Taylor², Arezoo Movaghar¹, Jinkuk Hong¹, Bryan Kim¹, Murray Brilliant³¹, Marsha Mailick¹

¹Waisman Center, University of Wisconsin-Madison; ²Vanderbilt University Medical Center; ³Marshfield Clinic Research Institute

Introduction: Individuals with autism spectrum disorder (ASD) face many challenges during adulthood, including elevated health problems relative to the general population. An emerging body of evidence suggests that females with ASD may be doubly-vulnerable for poor outcomes across a range of domains because of both their sex and their autism status. However, few studies have investigated sex differences in health outcomes for adults with ASD, and how such risk compares to the general population. The present study addresses this gap. We examined if sex differences in diagnostic codes in electronic health records (EHR) for adults with ASD were similar to or different from sex differences in the general population.
Methods: We utilized EHRs from the Marshfield Clinic population (a large, private, multispecialty group practice with records for over 1 million patients) to characterize health profiles. We selected ASD cases based on a patient having an ICD-9 code of pervasive developmental disorder (299) on at least two occasions (“rule of two”), to reduce risk of misclassification due to miscoding or misdiagnosis. There were 2119 patients with ASD (458 females, 1661 males) in the Marshfield EHR data who met these criteria and were therefore included in this study. The comparison group of community controls included 21,228 age- and sex-matched individuals without disabilities who were patients at the Marshfield Clinic, selected randomly with a ratio of 10 controls for every one patient with ASD. Following the methods of Croen et al. (2015), data were extracted from the EHR to reflect seven specific domains of health conditions: cardiovascular disease, endocrine disorders, gastrointestinal disorders, neurologic diseases, nutrition conditions, psychiatric conditions, and sleep disorders. Two variables were computed for each patient with respect to each of the seven health domains: the prevalence of disease for each domain and the burden of disease within each domain. To measure prevalence of disease, the presence (coded as 1) or absence (coded as 0) of any disease code for each domain was determined, following the rule of two. To measure burden of disease, the number of encounters in the EHR across disease codes within each domain was determined.

Results: Data were analyzed by two (ASD status) by two (sex) analyses of variance. There were significant Sex X ASD status interactions for prevalence of disease for three domains: endocrine diseases, neurological conditions, and sleep disorders. For these three domains, patients with ASD had a greater likelihood of having a condition than community controls, with a statistically significant divergent pattern of sex effects within each group. Specifically, whereas women with ASD were much more likely to have endocrine, neurological, and sleep disorders than men with ASD, for the community controls the sex differences were less pronounced (endocrine disorders, neurological diseases), or in the reverse direction (sleep disorders). There also were significant Sex X ASD status interactions for burden of disease within four domains: nutrition conditions, neurologic diseases, psychiatric conditions, and sleep disorders. For these four domains, women had a greater number of encounters within each of these domains than men, with significantly greater male-female differences within the ASD group than within the community control group. Particularly striking, for psychiatric conditions, women with ASD had nearly 4 times more encounters than control women, whereas men with ASD had 2.5 times more encounters than control men.

Discussion: The present analyses suggest newly-identified ways in which women with ASD might be at greater risk than men with ASD (relative to their counterparts in the general population). Though both men and women with ASD were at greater risk for a number of conditions (consistent with extant research), our analyses suggest that there is an increased burden of disease for women with ASD well beyond what would be expected for being a woman and for having ASD.


Support: This research was supported by the National Institute of Mental Health (R03 MH112783, PIs: DaWalt & Taylor) with core support from the National Institute of Child Health and Human Development (U54 HD090256, PI: Chang; U54 HD083211, PI: Neul).
**Paper Title:** Using Machine Learning to Discover Clinical Phenotypes Associated with Fragile X Syndrome

**Authors:** Arezoo Movaghar\(^1,2\), David Page\(^3\), Jan Greenberg\(^1\), Jinkuk Hong\(^2\), Leann Smith Dawalt\(^1\), Danielle Scholze\(^1,5\), Murray Brilliant\(^4\), Marsha Mailick\(^1\)

\(^1\)Waisman Center, University of Wisconsin-Madison  
\(^2\)Department of Biomedical Engineering, University of Wisconsin-Madison  
\(^3\)Department of Biostatistics and Bioinformatics, Duke University  
\(^4\)Marshfield Clinic Research Institute  
\(^5\)Department of Pediatrics, University of Wisconsin School of Medicine and Public Health

**Introduction:** Fragile X syndrome (FXS) is the most common inherited cause of developmental disability and autism. FXS is associated with a range of clinical phenotypes including social anxiety, intellectual disability, behavioral problems, language deficits, motor problems and sensory integration problems. Several studies have reported that medical issues such as recurrent otitis media and recurrent sinusitis, joint laxity with hyperextensible finger joints and pes planus (flat feet), gastroesophageal reflux diseases and seizures are common in individuals with FXS during childhood. However, the effect of FXS on the physical health, especially in adulthood is not fully explored. We have used a novel approach based on computational phenotyping to discover the pattern of health problems seen in individuals with FXS from electronic health records (EHRs).

**Method:** Using a machine learning method, we examined de-identified diagnostic codes from 51 individuals with FXS and 5100 controls. FXS cases where identified based on the presence of diagnostic code for fragile X syndrome (759.83 or Q99.2) in their medical records on at least two occasions. Cases and controls were matched on sex and year of birth with 1 to 100 ratio. Focusing on physical health, we excluded all of the diagnostic codes related to mental health, neurological disorders and congenital anomalies. In order to filter possible noise and errors in the data, we only included the codes that were recorded at least twice (rule of 2) for an individual and were observed in at least 5 participants. We used a supervised classification method called random forest to identify FXS cases from controls. To measure the success of classification, the area under receiver operating characteristic curve (AUROC) was reported. We also used Phenome-Wide Association Study (PheWAS) in which we mapped diagnostic codes to phecodes and investigated the association of FXS with clinical phenotypes. We reported the conditions that survived adjustment for multiple comparisons using Bonferroni corrections.

**Results:** We were able to successfully differentiate individuals with FXS from general population. Our random forest classifier could predict the FXS status with AUROC of 0.83, by utilizing only diagnostic codes related to physical health. PheWAS analysis suggests that individuals with FXS are at greater risk of having **circulatory problems** (i.e. ventricular fibrillation and flutter, mitral valve stenosis and aortic valve stenosis, rheumatic disease of the heart valves, cardiac arrest and ventricular fibrillation); **respiratory problems** (i.e. empyema and pneumothorax, pleurisy/pleural effusion, other pulmonary inflammation or edema and other diseases of respiratory system); and **digestive problems** (i.e. intestinal obstruction, ulceration of intestine, irritable bowel syndrome, dyspepsia and other specified disorders of function of stomach). They also had higher rates of **endocrine/metabolic problems** (i.e. abnormal weight gain, lack of normal physiological development, disturbances of amino-acid transport, acquired hypothyroidism and electrolyte imbalance); **genitourinary problems** (i.e. retention of urine, oliguria and anuria, noninflammatory disorders of cervix, and other symptoms/disorders of the urinary system); conditions related to **sense organs** (i.e. macular degeneration, cholesteatoma, impacted cerumen, dizziness and giddiness [Light-headedness and vertigo], conductive hearing loss, otitis media and disorders of acoustic nerve); and **injuries and poisonings** (i.e. urticarial, complication due to other implant and internal device, poisoning by anticonvulsants and anti-Parkinsonism drugs). These results provide strong evidence that individuals are at greater risk for a wide range of health conditions beyond current clinical reports.
Discussion: Although there have been multiple studies investigating the health and well-being of individuals with FXS, most studies are focused on children and they are usually limited to a small range of targeted phenotypes. Our discovery-oriented approach investigates the health characteristics of adults with FXS based on population-based data incorporating the entire spectrum of available health diagnoses. While our computational phenotyping showed that individuals with FXS suffer from more health difficulties in their life span, additional studies on larger populations are required to replicate the results.

Acknowledgement: This research was supported by grants from NICHD (R01 HD082110) to M. Mailick and by the core grant of the Waisman Center IDDRC (U54 HD090256) to Q. Chang. We also received support from NHGRI U01HG8701 (eMERGE) and NCATS UL1TR000427 (ICTR). We are very grateful to the Marshfield Clinic Research Institute and all individuals who have participated in our research study.

Paper Title: Research Use of Electronic Health Records: Preferences of Parents of Children with Fragile X Syndrome or An Autism Spectrum Disorder

Authors: Melissa Raspa1, Sara Andrews1, Anne Edwards1, Anne Wheeler1

1 RTI International

Contact info: mraspa@rti.org

Introduction

The advent of electronic health records (EHRs) has enabled researchers to mine vast amounts of clinical data in order to better understand an individual’s health, service utilization, and treatment effectiveness.1,2 Although the general public has supported EHR research, concerns related to privacy and security are pervasive.3,4 Not much is known about the views of individuals with known or suspected genetic conditions, such as fragile X syndrome (FXS) and autism spectrum disorders (ASD). These individuals may be at increased risk related to data breaches, but also have much to gain from EHR research. Therefore, it is important to understand the preferences of parents of children with FXS or ASD regarding the research use of their child’s EHRs.

Methods

After extensive formative work, a discrete-choice experiment (DCE) was designed consisting of 5 attributes, each with 2 or 3 levels, including (1) type of researcher accessing the EHR (for-profit, non-profit, government), (2) personally identifiable information included in the research (identifiable, non-identifiable), (3) sensitive information included in the research (sensitive, non-sensitive), (4) personal importance of research (important, not important), and (5) whether results of the research were shared (individual, summary, or no results). Three groups of parents were included: (a) those with FXS (b) those with an autism spectrum disorder (ASD), and (c) those who did not have a known or suspected genetic condition. Parents of children with FXS or ASD were recruited through research registries and control parents were recruited through a panel management company. The DCE was conducted online in English.

Results

In total, 1,543 parents completed the survey. Conditional logit models were conducted to predict parents’ choices to give permission for use of their child’s EHR in research by genetic status. When compared with control group parents (β=0.15), parents of children with known (β=0.41) or suspected (β=0.33) genetic conditions had stronger preferences for EHR studies conducted by non-profit researchers. Parents of children with known (β=0.21) or suspected (β=0.24) genetic conditions also had stronger preferences for return of individual research results when compared with control parents (β=0.09). Control group parents valued EHR studies on a research topic that was personally important to them (β=0.27) less than parents of children with
known (β=0.60) or suspected (β=0.51) genetic conditions. Parents of children with known (β=-0.10) genetic conditions also considered studies involving sensitive information about their child more desirable than control parents (β=-0.20), but on balance all parents preferred studies using non-sensitive information. Similarly, all parents preferred EHR studies that used non-identifiable information to studies that used identifiable information.

Discussion

Overall, parents who have a child with FXS or ASD were fairly similar in their preferences but they differed from those of parents who had a child that did not have a known or suspected genetic condition. The results indicate that trust and transparency of the research, importance and benefits of research, and return of results were the most important motivators for participation.

Citations


Acknowledgement: Research reported in this presentation was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number 5R01HD086702, PI: Raspa. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.