Linkage of whole genome sequencing with administrative health, and electronic medical record data for the study of autism spectrum disorder: Feasibility, Opportunities and Challenges

Brooks, J[1], Anagnostou, E[2], Rahman, F[3], Tu, K[1], Uruthiramoorthy, L[1], Nylen, N[4], McLaughlin, J[5], Schull, M[3], and Bronskill, S[3]

[1]University of Toronto

[2]Holland Bloorview Kids Rehabilitation Hospital

[3]Institute of Clinical Evaluative Sciences

[4]Ontario Brain Institute

[5]Public Health Ontario

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder (NDD) that presents with a high degree of heterogeneity (e.g., co-occurrence of other NDDs and other co-morbid conditions), contributing to differential health system needs. Genetics are known to play an important role in ASD and may be associated with different disease trajectories.

Objectives and Approach

In this proof of principle project, our objective is to link >2,200 children with a confirmed diagnosis of a NDD from the Province of Ontario Neurodevelopmental (POND) Study to administrative health data and electronic medical record (EMR) data in order to identify subgroups of ASD with unique health system trajectories. POND includes detailed phenotype and whole genome sequencing (WGS) data. Identified subgroups will be characterized based on clinical phenotype and genetics. To meet this goal, consideration of WGS-specific privacy and data issues is needed to implement processes which are above and beyond traditional requirements for analyzing individual-level administrative health data.

Results

Linkage of WGS data with administrative health data is an emerging area of research. As such it has presented a number of initial challenges for our study of ASD. Privacy concerns surrounding the use of WGS data and rare-variant analysis are of particular importance. Practical issues required the need for analysts with expertise in administrative data, EMR data and genetic analyses, and specialized software and sufficient processing power to analyze WGS data. Transdisciplinary discussions of the scope and significance of research questions addressed through this linkage were crucial. The identification of genetic determinants of phenotypes and trajectories in ASD could support targeted early interventions; EMR linkage may inform algorithms to identify ASD in broader populations. These approaches could improve both patient outcome and family experience.

Conclusion/Implications

As the cost of genetic sequencing decreases, WGS data will become part of the routine clinical management of patients. Linkage of WGS, EMR and administrative data has tremendous potential that has largely not been realized; including population-level ASD research to improve our ability to predict long-term outcomes associated with ASD.