Genetic and Clinical Characterization of Multimorbidity Trajectories in Major Depressive Disorder

Juan F. De la Hoz; Justin Tubbs; Yu Zhou; Evan Giangrande; Na Cai; Jordan W Smoller; Tian Ge

Background

Major Depressive Disorder (MDD) is often studied as a discrete diagnosis despite its heterogeneity in clinical presentation, treatment response, and course of illness. Using electronic medical records (EMR), we identify distinct disease trajectories that delineate the longitudinal heterogeneity of MDD.

By characterizing subgroups of individuals with shared comorbidities before and after the onset of MDD, we aim to identify pathways that link pre-existing conditions to the onset of MDD and to subsequent somatic problems. Our analyses suggest the existence of distinct subtypes with shared etiologies and outcomes that can inform personalized interventions for this complex disorder.

Materials and Methods

We used diagnostic codes from longitudinal EMR in the UKBB, including Hospital Episode Statistics (HES) and General Practice (GP). Read 2, Read 3, and ICD-9 codes were mapped onto 3-character ICD-10 codes.

Our study includes individuals with an MDD code (F32 or F33) and at least 20 years of EMR data, with 10 years flanking the first MDD code on each side. We split these 20 years into three windows: before (-10 to -1), during (-1 to +1), and after (+1 to +10), with t=0 at the time of the first MDD code. For each individual, we collected diagnostics codes from these three periods along with the type of visit (Outpatient, Inpatient, or ER). Codes present in at least 1% of individuals were retained.

We applied topic modeling to summarize the sparse, high-dimensional space of co-occurring diseases into distinct comorbidity profiles and used these to group individuals into discrete clinical trajectories using agglomerative clustering with Ward's linkage criterion. We characterized the resulting MDD trajectories using demographic information, healthcare usage, antidepressant (AD) treatment, and genetic risk for a set of common diseases.

Results

Our study included 14,010 individuals with MDD and comprehensive EMR data (68% female; mean age at MDD onset: 50.8 years, range: 17-74 years). The average EMR length was 42 years, with 47 different ICD-10 codes per individual. The average duration of MDD was 6 years, with 4 MDD codes per individual.

We repeated our clustering algorithm for 12 different number of topics (between 20-75) and combined them to obtain a set of 5 “consensus clusters”, representing distinct multimorbidity trajectories of MDD. These trajectories are associated with unique demographic patterns, medication use, and polygenic risk. For example, two trajectories (N=2,863 and 3,865) were associated with immune system dysregulation, but at different times relative to the onset of MDD. The one with earlier immune comorbidities was more

likely to be resistant to AD treatment, while the one with later immune comorbidities had increased polygenic risk for rheumatoid arthritis and asthma. The trajectory characterized by cardiometabolic conditions (*N=2,049*) had the oldest average age of MDD onset and the highest proportion of males *(45%,)*, had significantly increased polygenic risk for externalizing conditions (ADHD and alcohol use disorder), and was more likely to be treated with atypical ADs (Wellbutrin).

Discussion

Our study leverages EHRs linked to the UKBB to identify 5 distinct clinical trajectories of MDD based on comorbidity patterns and their temporal relationship to the onset of MDD. These trajectories, characterized by specific treatment regimens and polygenic risk scores, represent potential subtypes to be used in the identification of risk factors, the development of early intervention strategies and personalized treatments, and the long-term management of MDD. Ongoing efforts focus on the genetic validation and external replication of these trajectories.

References

Comorbidity in MDD

Berk, M., … Marx, W. (2023). Comorbidity between major depressive disorder and physical diseases: a comprehensive review of epidemiology, mechanisms and management. World Psychiatry, 22(3), 366–387.

Trajectories in MDD

Han, X., … Song, H. (2021). Disease trajectories and mortality among individuals diagnosed with depression: a community-based cohort study in UK Biobank. Molecular Psychiatry, 26(11), 6736–6746.

Plana-Ripoll, O., … McGrath, J. J. (2019). Exploring Comorbidity Within Mental Disorders Among a Danish National Population. JAMA Psychiatry, 76(3), 259–270.

Susan K. Service, De La Hoz, J., … Olde Loohuis, L. M. (2023). Predicting diagnostic conversion from major depressive disorder to bipolar disorder: an EHR based study from Colombia. In medRxiv (p. 2023.09.28.23296092). https://doi.org/10.1101/2023.09.28.23296092

Examples in other psychiatric disorders

Doshi-Velez, F., Ge, Y., & Kohane, I. (2014). Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. Pediatrics, 133(1), e54–e63.

Krebs, M. D., … Thompson, W. K. (2021). Associations between patterns in comorbid diagnostic trajectories of individuals with schizophrenia and etiological factors. Nature Communications, 12(1), 1–12.

Lu, C., … Yu, K.-H. (2022). Large-scale real-world data analysis identifies comorbidity patterns in schizophrenia. Translational Psychiatry, 12(1), 1–8.

Methods

Ge, T., … Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nature Communications, 10(1), 1–10.

Gu, Z., Schlesner, M., & Hübschmann, D. (2020). cola: an R/Bioconductor package for consensus partitioning through a general framework. Nucleic Acids Research, 49(3), e15–e15.

Pivovarov, R., … Elhadad, N. (2015). Learning probabilistic phenotypes from heterogeneous EHR data. Journal of Biomedical Informatics, 58, 156–165.

Sharma, A., … Doshi-Velez, F. (2024). Identifying data-driven subtypes of major depressive disorder with electronic health records. Journal of Affective Disorders, 356, 64–70.

Stroganov, O., …Bryant, J. (2022). Mapping of UK Biobank clinical codes: Challenges and possible solutions. PLOS ONE, 17(12), e0275816.

Thompson, D. J., … Weale, M. E. (2022). UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. In medRxiv (p. 2022.06.16.22276246). https://doi.org/10.1101/2022.06.16.22276246