Examining differential measurement error due to race, age, and sex in mental health disorders using PheValuator.

Joel N. Swerdel[1,2] and Dmytro Dymshyts[1,2]

[1]Observational Health Data Analytics, Global Epidemiology, Janssen Research and Development, Titusville, NJ, USA

[2]Observational Health Data Sciences and Informatics, New York, NY, USA

Background

Misclassification of health condition is a serious threat to validity in research involving observational data. The problem would be exacerbated if there was differential misclassification between population subgroups. As the importance of understanding differential effects of medical treatments between subpopulations increases, so does the importance of understanding differential measurement error of the phenotype algorithms used within observational studies examining these treatment effects.

PheValuator is a methodology within the OHDSI toolstack that uses diagnostic predictive modeling to determine the probability that a subject has a specific health outcome during a specified period of time.[1, 2] It was designed to evaluate the performance characteristics, i.e., sensitivity, specificity, and positive and negative predictive value, of phenotype algorithms in observational data.

The objective of this study was to use the results from PheValuator to estimate subpopulation differences between phenotype algorithm sensitivity and positive predictive value (PPV) across a set of mental health disorders. Populations were subgrouped by race, sex, and age.

Methods

We developed phenotype algorithms for eight mental health disorders: anxiety disorder, attention deficit hyperactivity disorder (ADHD), autism, bipolar disorder, depression, post-traumatic stress disorder (PTSD), schizoaffective disorder, and schizophrenia. We examined these conditions in three databases which include subjects of all ages: IBM® MarketScan® Multi-State Medicaid Database (MDCD), Optum's Clinformatics® Data Mart (SES), and Optum's Longitudinal EHR repository (EHR). We stratified the subjects in the analysis by sex; race, Black and White; and age, 65 years old (YO) and younger and 66 YO and older. We used PheValuator (V2.2.6) for the analyses. We developed algorithms for each condition using an empirical process previously documented involving the use of the standard OHDSI tools ATLAS, CohortDiagnostics, PHOEBE, and PheValuator. We compared sensitivity, estimated as true positives/(true positives + false negatives), and PPV, estimated as true positives/(true positives + false positives), estimated for each condition across the three databases.

Results

Figure 1 shows the results for differences in sensitivity by sex (sensitivity females – sensitivity males) across the eight health conditions. We found higher estimates for sensitivity for female subjects compared to male subjects for anxiety, bipolar, depression, and PTSD as shown by the positive values in each graph. We found lower estimates for sensitivity for female subjects compared to male subjects for ADHD, autism, schizoaffective disorder, and schizophrenia as shown by the negative values in each graph. For example, in SES, the estimate for sensitivity for PTSD in females was 73% (95% confidence interval (CI) 72 - 74%); for males the sensitivity for PTSD was 65% (95% CI 65 – 66%) resulting in a +8% difference in sensitivity.

Figure 1: Differences in phenotype algorithm sensitivity between Females and Males in mental health disorders

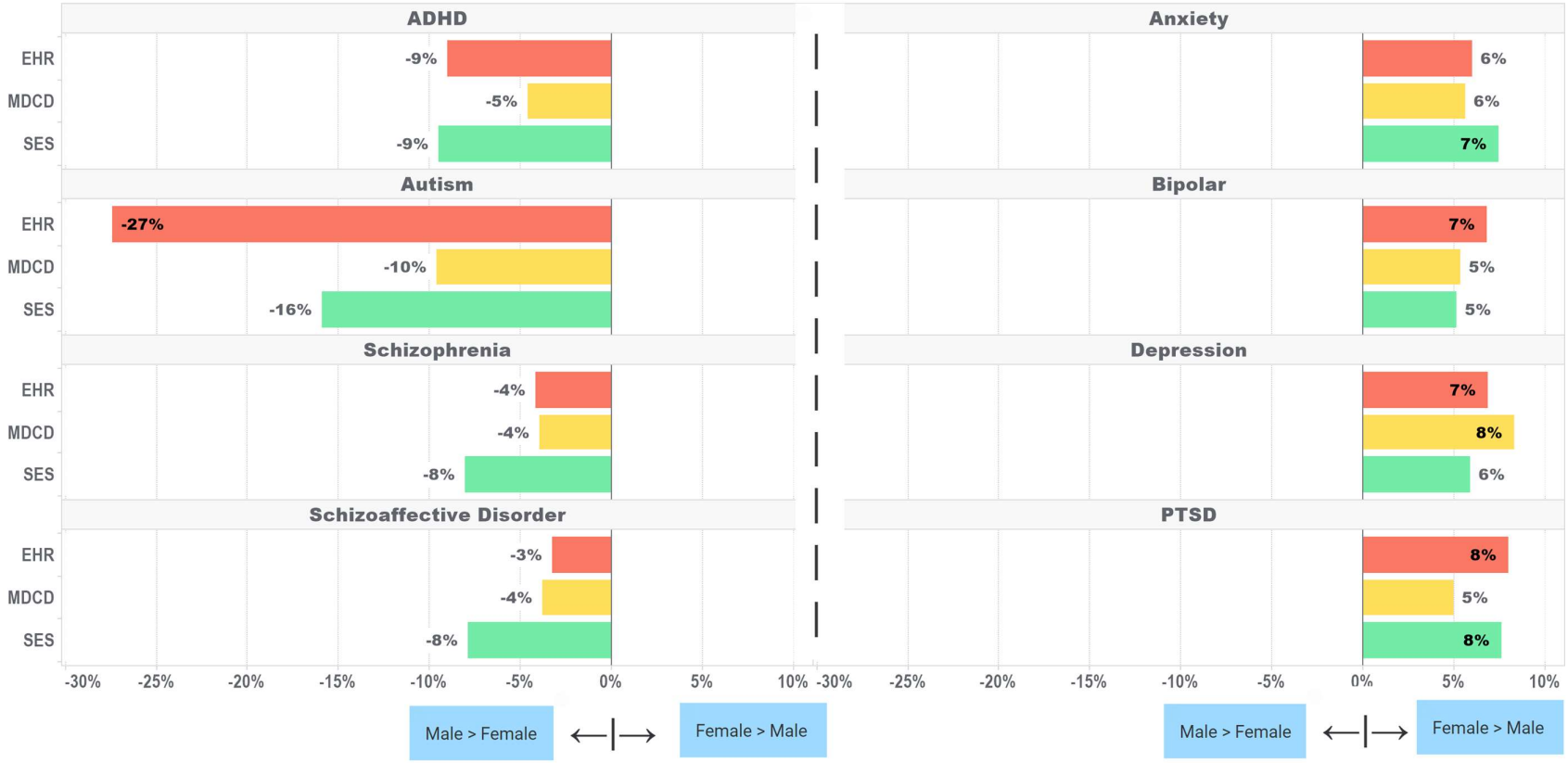


Figure 2 shows the results for differences in sensitivity by race (Sensitivity Black subjects - Sensitivity White subjects). We found large differences in sensitivity for schizoaffective disorder and schizophrenia between Blacks and Whites where the sensitivity for Blacks was higher than that for Whites. We found consistently lower sensitivities for Blacks compared to Whites for anxiety, bipolar disorder, and depression. For example, in SES, the sensitivity for schizoaffective disorder was 71% (95% CI 68 – 74%) for Blacks whereas for Whites it was 58% (95% CI 56 – 61%) resulting in a +13% difference in sensitivity. The differences in sensitivities for the other three conditions were smaller and varied between the databases tested.

Figure 2: Differences in phenotype algorithm sensitvity between Blacks and Whites in mental health disorders

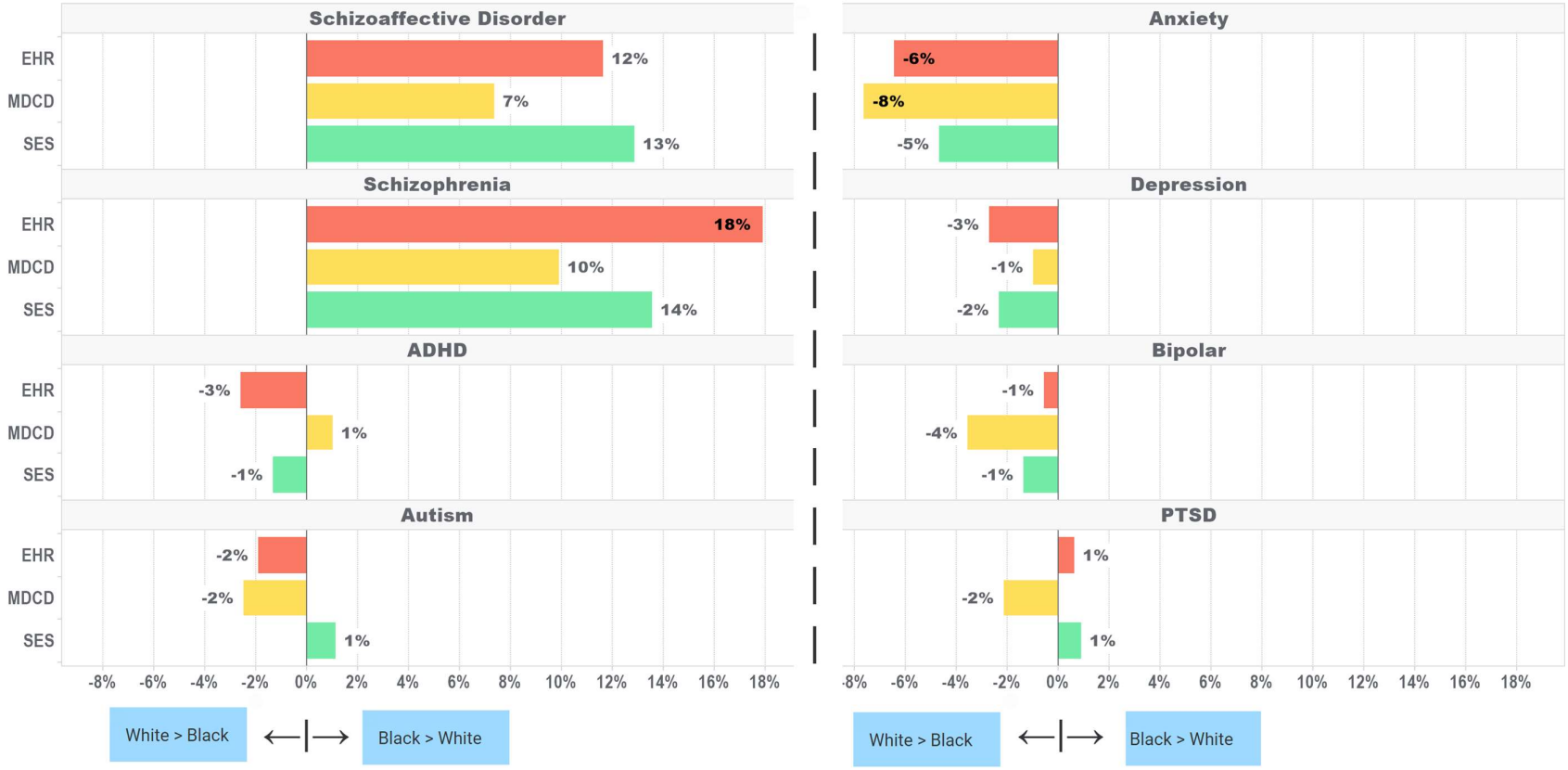
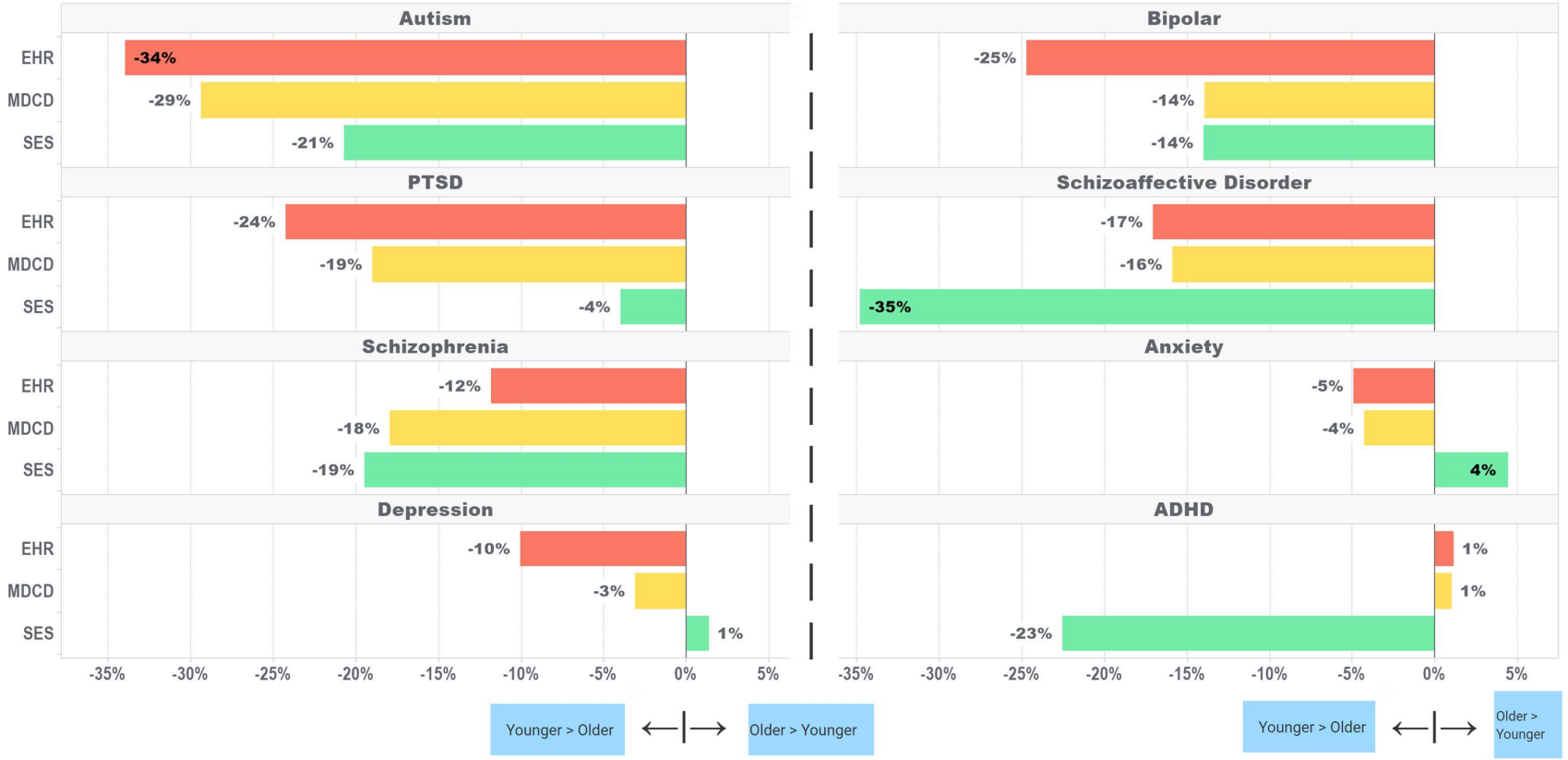


Figure 3 shows the results comparing estimates for sensitivity by age groups (older age group, 66YO+ - younger age group, 65YO and younger). We found that in five of the disorders, autism, bipolar disorder, PTSD, schizoaffective disorder, and schizophrenia, the estimates for sensitivity were much lower in the older age group than the younger age group. The largest difference was in schizoaffective disorder in SES. The sensitivity estimate for the older subjects was 39% (95% CI 36 – 41%) and in the younger subjects the estimate for sensitivity was 74% (95% CI 72 – 76%) resulting in a -35% difference in sensitivity.

Figure 3: Differences in phenotype algorithm sensitivity between those ≤ 65 years and those > 65 years in mental health disorders



The directionality of differences in sensitivity between sex, race, and age groups in mental health disorders is summarized in Table 1.

Table 1. Directionality of differences in sensitivity between sex, race, and age groups in mental health disorders

	Sex	Race	Age
Anxiety	F > M	W > B	
Attention deficit hyperactivity disorder	M > F		
Autism	M > F		Y > O
Bipolar disorder	F > M		Y > O
Depression	F > M		
Post-traumatic stress disorder	F > M		Y > O
Schizoaffective disorder	M > F	B > W	Y > O
Schizophrenia	M > F	B > W	Y > O

F – Female; M – Male; W – White; B- Black; Y – Age 65 and younger; O – Age 66 and older

We examined differences in PPV between the groups. The differences were much smaller for PPV between the groups compared to the sensitivity differences for race and sex. The differences were larger when comparing age differences. The results were similar in direction though smaller in magnitude as those for sensitivity where the PPV's for the older subjects were lower than the estimates for the younger subjects.

Conclusion

In this study we examined differences in the performance characteristics, sensitivity and PPV, for phenotype algorithms for eight mental health disorders for subgroup populations divided by race, sex, and age. We found large differences in sensitivity for many of the conditions in each of the subgroups.

The results from this study parallel findings in previous research examining sex, race, and age disparities in diagnosis and treatment of different mental health disorders. Hull et al suggest that females are underdiagnosed for autism compared to males possibly due to the expression of autism in females that do not meet diagnostic criteria.[3] In our estimates the sensitivity of the autism algorithm was significantly lower for females indicating that the number of false negatives, i.e., missing diagnosis codes for autism, was higher in females than males. van Niekerk and colleagues report that autism disorder is underdiagnosed in the older population especially those presenting with comorbid psychiatric disorders.[5] In our current study, we find lower sensitivity for autism in those over age 65. Vanderminden and Esala found that females were more likely diagnosed with anxiety disorder compared to males as were Whites compared to Blacks.[4] This is similar to our findings of higher sensitivity, i.e., fewer missed diagnoses, for females compared to males as well as lower sensitivity in Blacks compared to Whites.

The differences we found in sensitivity of phenotype algorithms for mental disorders across sex, race, and age suggest that researchers may be introducing bias into their research if they assume phenotype algorithm performance characteristics are the same across subgroups. Future research should be conducted to determine how these differences may affect study results such as those from drug comparative effectiveness analyses.

References

1. Swerdel JN, Hripcsak G, Ryan PB. PheValuator: Development and evaluation of a phenotype algorithm evaluator. Journal of Biomedical Informatics. 2019;97:103258.
2. Swerdel JN, Schuemie M, Murray G, Ryan PB. PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation. J Biomed Inform. 2022:104177.
3. Hull L, Petrides KV, Mandy W. The Female Autism Phenotype and Camouflaging: a Narrative Review. Review Journal of Autism and Developmental Disorders. 2020;7(4):306-17.
4. Vanderminden J, Esala JJ. Beyond Symptoms: Race and Gender Predict Anxiety Disorder Diagnosis. Society and Mental Health. 2018;9(1):111-25.
5. van Niekerk MEH, Groen W, Vissers CTWM, van Driel-de Jong D, Kan CC, Oude Voshaar RC. Diagnosing autism spectrum disorders in elderly people. Int Psychogeriatr. 2010;23(5):700-10.