

Prediction of Autism Risk From Family Medical History Data Using Machine Learning: A National Cohort Study From Denmark

Linda Ejlskov, Jesper N. Wulff, Amy Kalkbrenner, Christine Ladd-Acosta, M. Danielle Fallin, Esben Agerbo, Preben Bo Mortensen, Brian K. Lee, and Diana Schendel

ABSTRACT

BACKGROUND: A family history of specific disorders (e.g., autism, depression, epilepsy) has been linked to risk for autism spectrum disorder (ASD). This study examines whether family history data could be used for ASD risk prediction.

METHODS: We followed all Danish live births, from 1980 to 2012, of Denmark-born parents for an ASD diagnosis through April 10, 2017 ($N = 1,697,231$ births; 26,840 ASD cases). Linking each birth to three-generation family members, we identified 438 morbidity indicators, comprising 73 disorders reported prospectively for each family member. We tested various models using a machine learning approach. From the best-performing model, we calculated a family history risk score and estimated odds ratios and 95% confidence intervals for the risk of ASD.

RESULTS: The best-performing model comprised 41 indicators: eight mental conditions (e.g., ASD, attention-deficit/hyperactivity disorder, neurotic/stress disorders) and nine nonmental conditions (e.g., obesity, hypertension, asthma) across six family member types; model performance was similar in training and test subsamples. The highest risk score group had 17.0% ASD prevalence and a 15.3-fold (95% confidence interval, 14.0–17.1) increased ASD risk compared with the lowest score group, which had 0.6% ASD prevalence. In contrast, individuals with a full sibling with ASD had 9.5% ASD prevalence and a 6.1-fold (95% confidence interval, 5.9–6.4) higher risk than individuals without an affected sibling.

CONCLUSIONS: Family history of multiple mental and nonmental conditions can identify more individuals at highest risk for ASD than only considering the immediate family history of ASD. A comprehensive family history may be critical for a clinically relevant ASD risk prediction framework in the future.

<https://doi.org/10.1016/j.bpsgos.2021.04.007>

Autism spectrum disorder (ASD) is a severe neurodevelopmental condition affecting 1%–3% of the population (1). It has lifelong impacts and is associated with considerable personal and societal costs (2–4). With increasing incidence of ASD (5) and accompanying societal expense (6), better understanding of ASD etiology and better capability to identify groups at a high risk are of critical importance.

While the causes of ASD are not completely understood, it is suspected to be multifactorial, involving polygenic inheritance as well as environmental and behavioral risk factors (1,7). One of the most well-established ASD risk factors is a family history of autism (8,9). Such a history is believed to reflect especially genetic factors but also shared environmental, social, nutritional, and other potentially modifiable risk factors among relatives (10). Traditionally, research into family morbidity in ASD has focused on mental disorders within the immediate family (11,12). Other studies have shown elevated risks associated with immediate family members diagnosed with autoimmune disorders (13,14), congenital defects (15,16), neurologic disorders (8,17,18), cardiometabolic disorders

(19,20), and asthma and allergies (21) as well as with relatives of different degrees of relatedness (8). It is hypothesized that these associations are based on shared pathogenic mechanisms, especially genetic factors, between ASD and these other conditions (22). Although some previous studies demonstrate that extended family history is associated with ASD, they typically consider single or few disorders or limited types of family members such as parents, siblings, or cousins in 1-by-1 analyses (13,23). For this reason, the structure of family morbidity that underlies ASD occurrence is not well understood, and furthermore, it is unclear as to which components of family morbidity are the most important in predicting autism occurrence.

This study was designed to rigorously and systematically identify whether family history of mental and nonmental conditions could be used to predict ASD risk. To achieve this aim, we gathered family history data on a nationwide Danish cohort and their three-generation family members. We used state-of-the-art machine learning techniques to explore competing models. From the optimal model, we calculated a continuous

risk score and tested its performance, including comparison with common measures of psychiatric family history in ASD.

METHODS AND MATERIALS

Study Cohort

The study cohort consisted of all people born in Denmark from January 1, 1980, through December 2012 from Denmark-born parents who could be identified in the Danish Civil Registration System (24) and the Danish Medical Birth Register (25). Using the unique personal identification number given to all Danish residents, we linked cohort members to their mother, father, full siblings, grandparents, aunts, uncles, and cousins via the Danish Civil Registration System (Figure S1).

Outcome

The study cohort was followed from birth through April 10, 2017, in the Psychiatric Central Research Register (26). We defined the ASD outcome as the cohort member receiving one of the following ICD codes: ICD-8 299.00, 299.01, 299.02, 299.03; ICD-10 F84.0, F84.1, F84.5, F84.8, F84.9 (27). In Denmark, persons suspected of having ASD or other mental or behavioral disorders are referred, e.g., by general practitioners or school psychologists, to a psychiatric department for evaluation and are assigned a diagnosis by a psychiatrist; Danish health care is universal and free of charge. All diagnoses are reported to the Psychiatric Central Research Register without regard to the need for treatment or educational provisions.

Family History

We followed all cohort family members for mental disorder diagnoses reported to the Psychiatric Central Research Register (beginning in 1969, when it was established) and for nonmental disorder diagnoses reported to the National Patient Register (28) (beginning in 1977, when it was established) through April 10, 2017. Reporting of diagnoses to these registers is done only by medical specialists. Based on literature and our previous work, we selected 73 candidate disorders spanning mental, cardiometabolic, neurologic, congenital defects, autoimmune, asthma, and allergy conditions (Table S1). The first time a family member received one of the 73 diagnoses during follow-up (1969 to April 10, 2017, for mental diagnoses; 1977 to April 10, 2017, for nonmental diagnoses), we registered the family member type (6 possible: mother, father, full sibling, grandparent, aunt/uncle, cousin) (Figure S1) and diagnosis to create a morbidity indicator (1 = occurred, 0 = not occurred), for a total of 438 family member type–disorder indicators (Figures S2 and S3 for the prevalence of each morbidity indicator).

Covariates

We obtained information from the Danish Medical Birth Register, Danish Civil Registration System, and Statistics Denmark on each cohort member's birth weight, gestational age, birth year, maternal and paternal age, and parental educational attainment at birth of the child (Table 1).

Statistical Approach

Our goal was to use machine learning to find an optimal model predicting ASD diagnosis. The phases of analysis are detailed in Figure 1 and Table S2. A brief description follows.

Model Development Phase. Stage 1: To avoid overfitting, we randomly split the cohort into test (20%) and training (80%) subsamples.

Stage 2: We developed and tuned multiple machine learning algorithms on the training subsample using 10-fold cross-validation. Algorithms included (29) random forest (RF), extreme gradient boosting (EGB), the traditional generalized linear model, elastic net, neural networks, support vector machines, K-nearest neighbors, and three different ensemble learning methods (30)—averaging, stacking, and hill climbing. Because of class imbalance (a ratio of 1 ASD case to 63 noncases) (31), we used undersampling to balance the class distribution at a ratio of 1:1. We optimized each model using the *F* measure, the harmonic mean of positive predictive value, and sensitivity (32,33).

Stage 3: We then used the *F* score, area under the curve, sensitivity (true positive rate), positive predictive value, deviance, kappa, and specificity (negative predictive value) to compare performance across different algorithms (34,35).

Stage 4: We repeated stages 2 and 3 for different candidate morbidity indicators. First, we trained a model using all indicators. Then, we restricted them to the best indicators as defined by variable importance (Figure S4) using the RF and EGB algorithms: the top 41 most important (i.e., present among the top 30 in the RF or the EGB), the top 21 (i.e., among the top 15 in the RF or the EGB), and the top 3. We compared each model against a basic model with one indicator for any parental psychiatric history before birth of the index person (Yes/No). The latter representation of family psychiatric history is widely used in ASD research (11,12).

To detect effects of sex or family size on model, we conducted several sensitivity analyses replicating stages 2 and 3. First, we trained models for males and females, separately. Second, we trained a model with the candidate morbidity indicators as counts, denoting how many times a particular diagnosis had occurred within each family member type, rather than the bivariate Yes/No indicator.

Stage 5: The resulting best algorithm was evaluated on the test subsample to obtain an out-of-sample estimate of performance (36).

Model Deployment Phase. Stage 6: We calculated the family morbidity risk score (FMRS) using predicted probabilities based on the best-performing model and divided the FMRS range into 10 segments of equal length. To evaluate the association between the FMRS and ASD, we modeled a logistic regression with ASD occurrence as a function of the FMRS adjusted for sex, birth weight, gestational age, birth year, maternal, paternal age, and highest parental educational attainment at child's birth, estimating odds ratios (ORs) with 95% confidence intervals (CIs). We evaluated potential interactions between the FMRS and sex, birth weight, and parental socioeconomic position, respectively. We also

estimated associations considering parental psychiatric history (Yes/No) or full sibling ASD history (Yes/No) as the family morbidity measure in each model. There were little missing

Table 1. Descriptive Characteristics of the Study Cohort

Characteristics	No ASD, <i>n</i> = 1,670,391, <i>n</i> (%)	ASD, <i>n</i> = 26,840, <i>n</i> (%)
Sex, Male	851,349 (51.0%)	19,900 (74.1%)
Intellectual Disability	7697 (0.5%)	3342 (12.5%)
Gestational Weeks		
≤36	96,874 (5.9%)	2040 (7.7%)
37–40	1,111,458 (67.6%)	17,467 (65.6%)
≥41	436,989 (26.6%)	7112 (26.7%)
Maternal Age >35 Years	211,979 (12.7%)	3797 (14.1%)
Paternal Age >35 Years	430,612 (25.8%)	7654 (28.5%)
Birth Weight, g		
<2501	88,542 (5.3%)	1810 (6.8%)
2501–3000	210,534 (12.7%)	3448 (12.9%)
3001–4000	1,089,957 (65.6%)	16,506 (61.8%)
≥4001	273,317 (16.4%)	4929 (18.5%)
Birth Year		
1980	51,048 (3.1%)	153 (0.6%)
1981	47,172 (2.8%)	172 (0.6%)
1982	46,860 (2.8%)	236 (0.9%)
1983	45,259 (2.7%)	248 (0.9%)
1984	45,873 (2.7%)	292 (1.1%)
1985	47,435 (2.8%)	342 (1.3%)
1986	48,334 (2.9%)	353 (1.3%)
1987	48,578 (2.9%)	399 (1.5%)
1988	50,616 (3.0%)	500 (1.9%)
1989	52,444 (3.1%)	583 (2.2%)
1990	53,925 (3.2%)	692 (2.6%)
1991	53,912 (3.2%)	788 (2.9%)
1992	56,527 (3.4%)	877 (3.3%)
1993	55,769 (3.3%)	970 (3.6%)
1994	57,302 (3.4%)	1213 (4.5%)
1995	57,101 (3.4%)	1189 (4.4%)
1996	54,272 (3.2%)	1279 (4.8%)
1997	53,666 (3.2%)	1416 (5.3%)
1998	52,138 (3.1%)	1470 (5.5%)
1999	51,757 (3.1%)	1546 (5.8%)
2000	51,834 (3.1%)	1668 (6.2%)
2001	50,275 (3.0%)	1496 (5.6%)
2002	49,391 (3.0%)	1361 (5.1%)
2003	50,024 (3.0%)	1277 (4.8%)
2004	50,572 (3.0%)	1158 (4.3%)
2005	50,565 (3.0%)	1055 (3.9%)
2006	51,207 (3.1%)	985 (3.7%)
2007	50,321 (3.0%)	857 (3.2%)
2008	50,788 (3.0%)	656 (2.4%)
2009	48,470 (2.9%)	561 (2.1%)
2010	48,593 (2.9%)	503 (1.9%)
2011	44,783 (2.7%)	345 (1.3%)
2012	43,579 (2.6%)	200 (0.7%)

Table 1. Continued

Characteristics	No ASD, <i>n</i> = 1,670,391, <i>n</i> (%)	ASD, <i>n</i> = 26,840, <i>n</i> (%)
Highest Parental Educational Attainment at Child's Birth		
Early childhood education, ISCED 0–2	212,329 (12.7%)	4183 (15.6%)
Upper secondary education, ISCED 3	794,704 (47.7%)	12,859 (48.0%)
Short, bachelor's or equivalent, ISCED 5–6	469,097 (28.1%)	6940 (25.9%)
A Part of the Test Subsample	334,055 (20.0%)	5392 (20.1%)
Have Information on at Least One of Each Family Member Type	1,178,351 (70.5%)	18,115 (67.5%)
Only Child	273,021 (16.3%)	6202 (23.1%)
	Mean (SD)	Mean (SD)
Number of Siblings	1.3 (0.9)	1.1 (0.9)
Number of Grandparents	3.6 (1.1)	3.8 (0.8)
Number of Aunts/Uncles	2.6 (1.8)	2.7 (1.7)
Number of Cousins	4.6 (4.1)	4.6 (4.0)

ASD, autism spectrum disorder; ISCED, International Standard Classification of Education.

data on covariates (a total of 2%), and for this reason, we choose not to perform multiple imputation (37).

Effects of Family Size and Missing Family Linkages.

We adjusted the association between ASD and FMRS for family size, including the number of aunts and uncles, cousins, siblings, and grandparents in the same model. We further assessed the association in different subpopulations: cohort members who were an only child, cohort members who had at least one missing link to a family member type, and cohort members who had at least one of each family member type (including at least 1 full sibling). Finally, we assessed whether the FMRS performed similarly in those who had been a part of the test subsample and not included in model development. All analyses were done using the mlr package (38) with R version 3.5.2 (39).

Ethical Approval

This study was approved by the Danish Data Protection Agency. Informed consent is not required for purely register-based research of preexisting personal data in Denmark.

RESULTS

Of the 1,697,231 cohort members, 26,840 (1.6%) received an ASD diagnosis during follow-up. Compared with persons without ASD, cohort members with ASD were more likely to be either below or above normal birth weight, be born prematurely, be an only child, have parents older than 35 years and with lower educational attainment, and have a slightly lower mean number of siblings but similar mean numbers of linked grandparents, aunts and uncles, and cousins (Table 1).

Model Development

After initial assessment, we excluded 85 indicators with fewer than 40 exposed ASD cases, because these rare indicators increased computation time without improving model

Risk Prediction of Autism

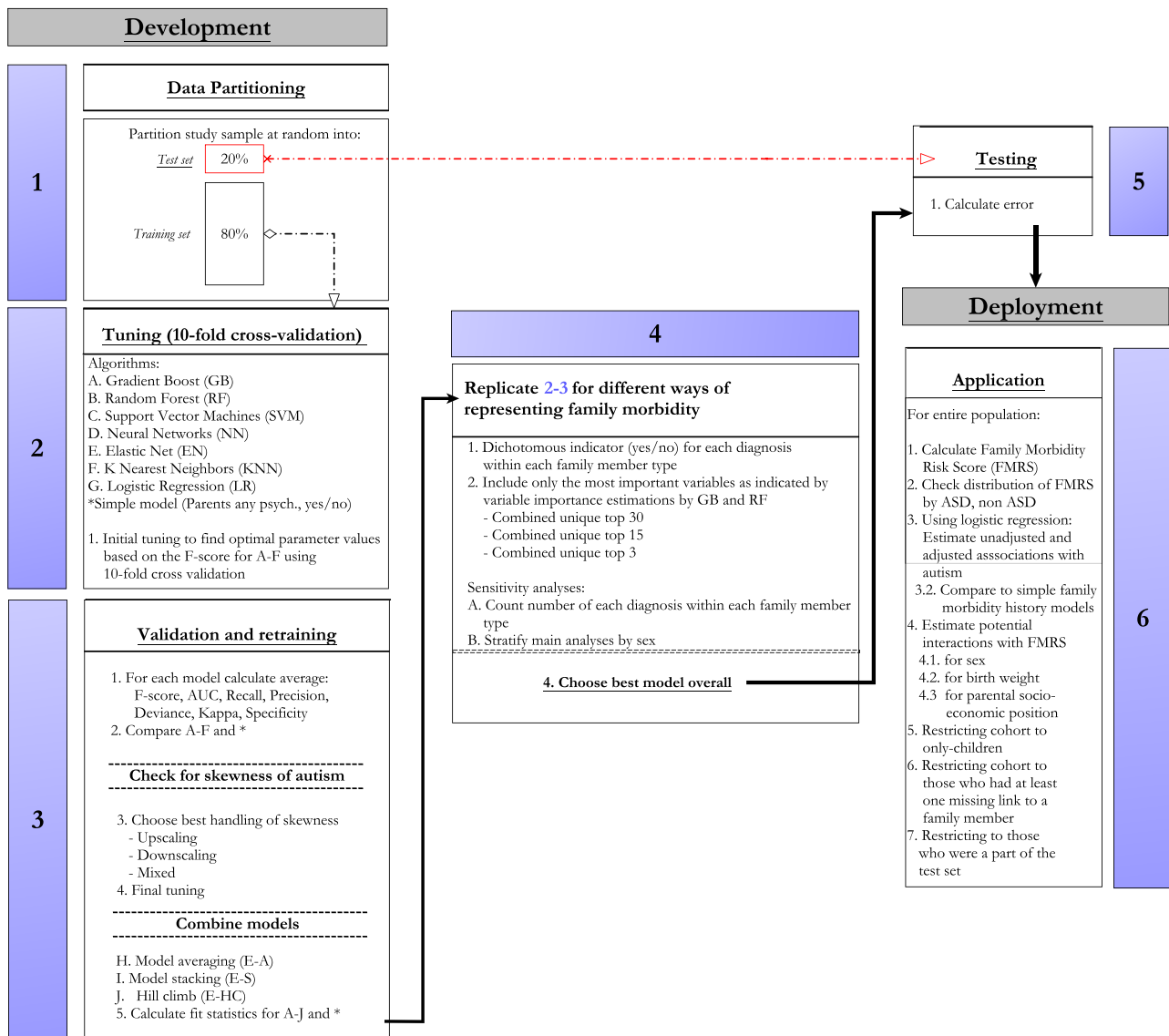


Figure 1. Overview of prediction model development. ASD, autism spectrum disorder; AUC, area under the curve.

performance, leaving 353 candidate morbidity indicators for further analysis (Table S3). We chose the EGB algorithm to estimate the FMRS because EGB has the best balance of a high relative performance on all measures compared with the other algorithms combined with a low standard deviation (see Table S4 and Figure S5). As can be seen in Figure 2, the four EGB competing models (using all, top 41, top 21, or top 3 indicators) performed better on all investigated performance measures compared with parental psychiatric history alone. However, the models were still far from being clinically relevant for ASD prediction (e.g., test subsample *F* score = 0.054, area under the curve = 0.64, true positive rate = 0.44, positive predictive value = 0.029; virtually identical performance measures of optimal model in training subsample and test subsample). For the algorithm based on the top 41 indicators, we observed a 12% increase in the *F* score, a 15% increase in

area under the curve and *F* score, a 50% increase in true positive rate, and a 9% increase in positive predictive value (see Table S5). Similar performance measures were observed for the model developed when using the count indicators (Table S6). There was little indication of a difference in the optimal model depending on the sex of the cohort member (Figures S6 and S7).

Deployment Phase: FMRS

The distribution of the FMRS based on the EGB algorithm can be seen in Figure S8. Regardless of the number of morbidity indicators included, most cohort members had an FMRS on the lower end of the spectrum. Cohort members with an ASD diagnosis had a slightly higher average FMRS (0.53 vs. 0.48, $p < .0001$) and were much more likely to be in the uppermost

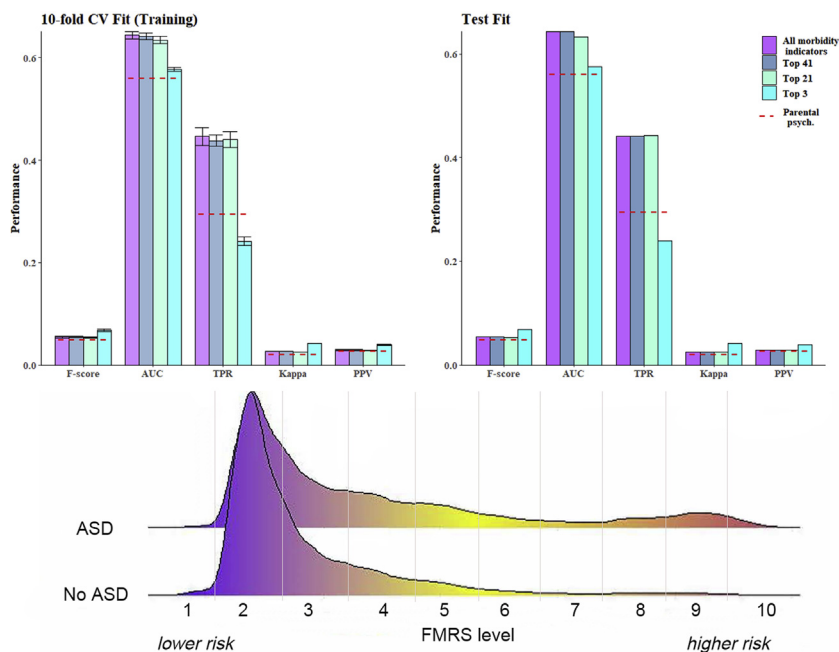


Figure 2. Fit statistics for the extreme gradient boosting algorithm when including all 353 indicators, the top 41, the top 21, and the top 3 estimated on the training (top left) and the test (top right) set. Smoothed density plot (bottom) of family morbidity risk score (FMRS) based on the top 41 indicators stratified by autism spectrum disorder (ASD)/no ASD. AUC, area under the curve; CV, cross-validation; PPV, positive predictive value; TPR, true positive rate.

end of the FMRS spectrum compared with non-ASD cohort members. Because of the similar model performance measures and distributions of the top 41 indicators compared with the FMRS calculated using all 353 indicators (Figure 2 and Tables S4 and S5) and somewhat improved FMRS distribution compared with the model using the top 21 indicators

(Figure S8), we view the FMRS based on the top 41 indicators as the overall best and simplest representation of family risk.

Table 2 shows an overview of the 41 morbidity indicators, comprising all family member types and both mental (8) and nonmental (9) disorders, used to calculate the FMRS, and Figure 3 shows the importance of each indicator. Among the

Table 2. An Overview of the 41 Morbidity Indicators Included in the Family Morbidity Risk Score by Diagnosis and Family Member Type

Diagnosis ^a	Family Member Type					
	Mother	Father	Full Sibling	Grandparent	Aunt/Uncle	Cousin
Neurotic/Stress/Somatoform Disorder (Not OCD)	x	x	x	x	x	x
Obesity	x		x	x	x	x
Hypertension	x		x	x	x	
Personality Disorder	x	x		x	x	
ASD	x		x			x
ADHD	x		x			x
Asthma	x			x		x
Depression	x	x				
Psychological Development Disorder (Not ASD)			x			x
Psychoactive Substance Abuse			x	x		
Diabetes				x		
Episodic and Paroxysmal Disorder (Not Epilepsy)				x		
Nerve, Nerve Root, and Plexus Disorders				x	x	
Other Neurologic Disorders				x		
Musculoskeletal Defects						x
Allergies						x
Behavioral/Emotional Disorder of Juvenile Onset (Not ADHD or Tic)			x			

See Table S1 for more information on the specific ICD codes.

ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; OCD, obsessive-compulsive disorder.

^aSorted by number of family member types.

Risk Prediction of Autism

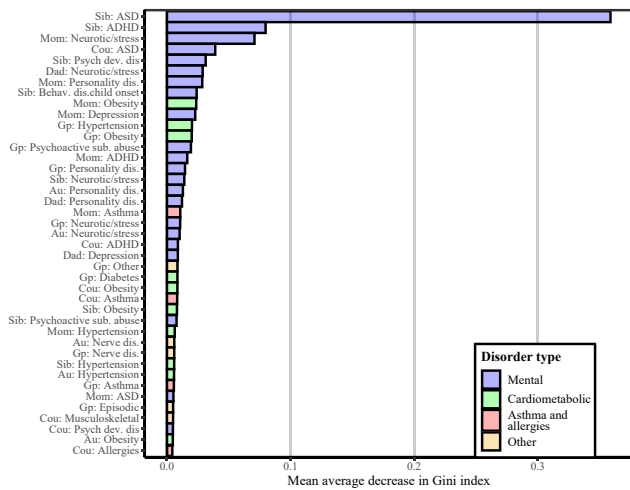


Figure 3. Importance ranking of 30 most important predictors in the family morbidity risk score. ADHD, attention-deficit/hyperactivity disorder; AS, autism spectrum disorder; Au, aunt; Behav. dis child onset, behavioral disorder childhood onset; Cou, cousin; dis., disorder; Gp, grandparent; Psych dev. dis, psychological developmental disorder; Sib, sibling; sub. abuse, substance abuse.

psychiatric disorders were stress/neurotic disorders for all family member types; ASD or attention-deficit/hyperactivity disorder for mother, sibling, and cousins; other psychological development disorder for sibling and cousins; personality disorder for mother, father, grandparents, and aunts/uncles; and parental depression. Sibling ASD was the most important of all indicators. Among the 57 nonmental disorders, obesity was most common and present for 5 of 6 family member types (save fathers); hypertension for mother, sibling, grandparents, and aunts/uncles; and asthma for mother, grandparents, and cousins.

There was a dose-response relationship, with higher levels of family morbidity corresponding to higher ASD risk, reaching a 15.4-fold increase in the highest versus the lowest FMRS group (OR, 15.3; 95% CI, 14.0–17.1) (Figure 4 and Table 3; Table S7). The risk for ASD was increased twofold when considering parental psychiatric history alone (OR, 1.98; 95% CI, 1.97–2.08) and 6.1-fold (OR, 6.1; 95% CI, 5.86–6.38) considering full sibling ASD history alone. The ratio of cases

to noncases decreased steadily with increasing family morbidity, reaching a prevalence of 17% (corresponding to 1 ASD case per 6 people) in the highest FMRS group. Table S8 provides the adjusted ORs for the three other family morbidity models (all indicators, top 21, top 3).

The FMRS performed similarly for only-children, for cohort members with complete linkages, and when restricted to the test subsample (Table S9). For the cohort members with missing information on at least one family member type ($n = 485,814$), the highest adjusted ORs for each FMRS level were attenuated slightly, but the dose-response relationship remained. Adjusting for the number of aunts and uncles, cousins, siblings, and grandparents also attenuated the associations slightly (Table S10).

DISCUSSION

We applied state-of-the-art machine learning techniques to a total population sample to derive an optimal ASD risk prediction model based on disorders diagnosed in family members across three generations. The FMRS derived from the optimal model demonstrated better predictive performance and provided a more fine-grained differentiation between levels of family risk for ASD compared with dichotomous measures of any parental psychiatric history or full sibling ASD. For example, the FMRS levels 4–10 were associated with a gradient of elevated ASD risk (80% to 15-fold increased risk) and comprised 20% of the entire cohort and 40% of all ASD cases, while having a sibling with ASD (Yes/No) is associated with a quite high risk (sixfold) but comprises only 1.7% of the cohort and only 9.9% of ASD cases. Furthermore, the FMRS had stronger associations with ASD than many other well-studied population-based risk factors (1) such as fetal growth, which has about a 50% increased ASD risk (40).

Notably, the most influential indicators spanned all family member types and both mental and nonmental disorders. While previous studies have shown 1-by-1 associations between a wide range of different disorders and family member types and ASD (8), to our knowledge, this is the first study to demonstrate which combination of morbidities and family member types may be the most important to consider and further elucidates the complex underlying family morbidity structure in ASD risk. For example, across all models, sibling ASD and attention-deficit/hyperactivity disorder were the most

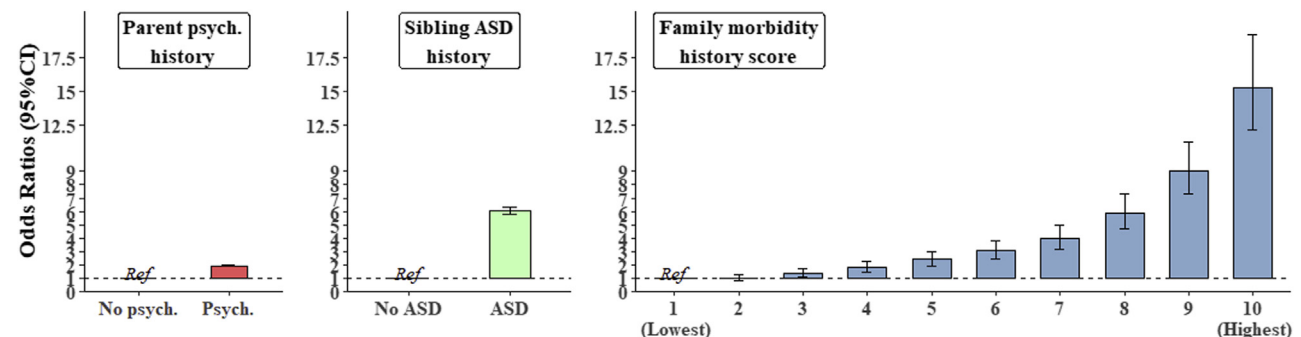


Figure 4. Adjusted odds ratios with 95% confidence interval (CI) for the association between family morbidity history and autism spectrum disorder (ASD), adjusted for sex, birth weight, gestational age, birth year, maternal, paternal age, and highest parental educational attainment. psych., psychiatric.

Table 3. Number of ASD Cases, Prevalence, OR, and Expected Frequency of ASD for Different Levels of the Three Different Family History Models

No ASD	ASD	% ASD	OR ^a	1 ASD Case Per No. of Cohort Members	Levels	Family History Model
1,353,858	18,686	1.4%	Ref	1/73	No psych.	Parental psych.
285,964	7832	2.7%	1.9	1/38	Psych.	
1,614,787	23,896	1.5%	Ref	1/69	No ASD	Sibling ASD
25,035	2622	9.5%	6.1	1/11	ASD	
13,483	88	0.6%	Ref	1/154	1 (lowest)	FMRS
961,695	10,409	1.1%	1.1	1/93	2	
346,825	5447	1.5%	1.4	1/65	3	
163,238	3474	2.1%	1.8	1/48	4	
84,439	2410	2.8%	2.4	1/36	5	
31,178	1185	3.7%	3.0	1/27	6	
12,303	649	5.0%	3.9	1/20	7	
11,787	893	7.0%	5.9	1/14	8	
12,443	1465	10.5%	9.0	1/9	9	
2431	498	17.0%	15.3	1/6	10 (highest)	

ASD, autism spectrum disorder; FMRS, family morbidity risk score; OR, odds ratio; Ref, reference; Psych., psychiatric.

^aAdjusted for sex, birth weight, gestational age, birth year, maternal, paternal age, and highest parental educational attainment.

important predictors, which is consistent with the literature (41); however, stress-based and personality disorders were also influential psychiatric indicators across family member types. These results are consistent with recent findings showing significant genetic (22) and nongenetic (42) correlations between ASD and these diagnoses.

Cardiometabolic disorders (specifically obesity and hypertension) emerged as the most influential indicators among 57 nonmental disorders. Notably, 5 of 6 obesity indicators across family member types were present among the top 41 morbidity indicators. This finding is supported by several reports of maternal obesity risk for ASD and other neurodevelopmental disorders (20, 43–45), to which our study adds obesity in the extended family to the list of potentially relevant risk factors. The predictive model provides no information regarding any potential causal link between familial obesity and ASD, such as whether the extended family obesity risk reflects shared detrimental lifestyles and other social and environmental factors that may influence epigenetic and nongenetic prenatal ASD risks (instead of, or in addition to, genetic connections). Whether both psychiatric and cardiometabolic disorders are the best indicators of the underlying risk of ASD or, as predictors, provide clues as to ASD pathogenesis would need to be further investigated.

Strengths and Limitations

The FMRS was developed on an entire population, using a machine learning approach, and is estimable for all members of a population, regardless of missing information on one more of the family member types, similar to a polygenic score. We performed a wide variety of sensitivity analyses, which indicate that the FMRS is robust to size and structure of family, showing similar results in only-child, small families, and big families, as well as by sex of the cohort member. The optimal model performance measures and FMRS results were virtually the same in both the training and test subsamples. However, this study also has limitations. While we have

adjusted for birth year in the regressions, the presence of certain morbidities in certain family member types included in the risk score may be influenced by changes in diagnostic criteria and practice over time, leading to possible diagnosis misclassification (measurement error) and reduced model predictive accuracy. We included diagnoses that had been reported in the literature regarding conditions in the family and co-occurring conditions that were associated with ASD, but future work could consider a more comprehensive array of conditions. While we considered possible model differences by sex, future work could also consider whether predictive models might also differ among ASD subgroups, such as ASD with or without intellectual disability. Because we are using the models for prediction and not causal inference, confounding is not a concern. There is some tendency for more common diagnoses to be included in the FMRS, which may partly reflect structural tendencies in the data or specific underlying morbidity structures. The timing of diagnoses has not been taken into account; therefore, future studies investigating the nature of the association between family history and ASD risk should consider the potential for reverse causation. The extent to which the results from this study can be replicated in other study populations is unknown; however, identification of the most influential indicators being consistent with existing literature supports generalizability of the FMRS. From a clinical feasibility perspective, future work should also consider the predictive ability of models limited to mental and nonmental disorders in the immediate family. We will make our FMRS model available to investigators upon request to promote replication and further study.

Conclusions

In conclusion, family history of multiple mental and nonmental conditions can identify more individuals at highest risk for ASD than only considering the immediate family history of ASD. Results, which yielded a family history risk score estimable for every person, underscore the complexity of family history

Risk Prediction of Autism

underlying ASD occurrence and the potential importance of conditions beyond immediate family history of mental disorders in identification of at-risk groups. A comprehensive family history may be a critical component in a clinically relevant ASD risk prediction framework in the future.

ACKNOWLEDGMENTS AND DISCLOSURES

The work was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health (NIH) (Grant No. R01ES026993 [to DS, AK]) and by the Lundbeck Foundation (IPSYCH; Grant Nos. R102-A9118 and R155-2014-1724 [to PBM]).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Economics and Business (LE, EA, PBM, DS), National Center for Register-based Research; Department of Econometrics and Business Analytics (JNW); and the Centre for Integrated Register-based Research (EA, PBM), Aarhus University; Lundbeck Foundation Initiative for Integrative Psychiatric Research (IPSYCH) (LE, EA, PBM, DS), Aarhus, Denmark; Joseph J Zilber School of Public Health (AK), University of Wisconsin-Milwaukee, Milwaukee, Wisconsin; Department of Epidemiology (CL-A), Department of Mental Health (MDF), and Wendy Klag Center for Autism and Developmental Disabilities (MDF), Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; and Department of Epidemiology and Biostatistics (BKL) and A.J. Drexel Autism Institute (BKL, DS), Drexel University, Philadelphia, Pennsylvania.

Address correspondence to Diana Schendel, Ph.D., at des348@drexel.edu.

Received Feb 3, 2021; revised Mar 9, 2021; accepted Apr 18, 2021.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.bpsgos.2021.04.007>.

REFERENCES

- Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, *et al.* (2017): The changing epidemiology of autism spectrum disorders. *Annu Rev Public Health* 38:81–102.
- Xu G, Strathearn L, Liu B, Bao W (2018): Prevalence of autism spectrum disorder among US children and adolescents, 2014–2016. *JAMA* 319:81–82.
- Buescher AVS, Cidav Z, Knapp M, Mandell DS (2014): Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatr* 168:721–728.
- Van Heijst BFC, Geurts HM (2015): Quality of life in autism across the lifespan: A meta-analysis. *Autism* 19:158–167.
- Schendel DE, Thorsteinsson E (2018): Cumulative incidence of autism into adulthood for birth cohorts in Denmark, 1980–2012. *JAMA* 320:1811–1813.
- Leigh JP, Du J (2015): Brief report: Forecasting the economic burden of autism in 2015 and 2025 in the United States. *J Autism Dev Disord* 45:4135–4139.
- Schmidt RJ, Iosif AM, Guerrero Angel E, Ozonoff S (2019): Association of maternal prenatal vitamin use with risk for autism spectrum disorder recurrence in young siblings. *JAMA Psychiatry* 76:391–398.
- Xie S, Karlsson H, Dalman C, Widman L, Rai D, Gardner RM, *et al.* (2019): Family history of mental and neurological disorders and risk of autism. *JAMA Netw Open* 2:e190154.
- Hansen SN, Schendel DE, Francis RW, Windham GC, Bresnahan M, Levine SZ, *et al.* (2019): Recurrence risk of autism in siblings and cousins: A multinational, population-based study. *J Am Acad Child Adolesc Psychiatry* 58:866–875.
- Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, *et al.* (2019): Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA Psychiatry* 76:1035–1043.
- Daniels JL, Forssen U, Hultman CM, Cnattingius S, Savitz DA, Feychting M, Sparen P (2008): Parental psychiatric disorders associated with autism spectrum disorders in the offspring. *Pediatrics* 121:e1357–e1362.
- Jokiranta E, Brown AS, Heinimaa M, Cheslack-Postava K, Suominen A, Sourander A (2013): Parental psychiatric disorders and autism spectrum disorders. *Psychiatry Res* 207:203–211.
- Wu S, Ding Y, Wu F, Li R, Xie G, Hou J, Mao P (2015): Family history of autoimmune diseases is associated with an increased risk of autism in children: A systematic review and meta-analysis. *Neurosci Biobehav Rev* 55:322–332.
- Atladóttir HO, Pedersen MG, Thorsen P, Mortensen PB, Deleuran B, Eaton WW, Parner ET (2009): Association of family history of autoimmune diseases and autism spectrum disorders. *Pediatrics* 124:687–694.
- Wier ML, Yoshida CK, Odouli R, Grether JK, Croen LA (2006): Congenital anomalies associated with autism spectrum disorders. *Dev Med Child Neurol* 48:500–507.
- Schendel DE, Autry A, Wines R, Moore C (2009): The co-occurrence of autism and birth defects: Prevalence and risk in a population-based cohort. *Dev Med Child Neurol* 51:779–786.
- Christensen J, Overgaard M, Parner ET, Vestergaard M, Schendel D (2016): Risk of epilepsy and autism in full and half siblings—A population-based cohort study. *Epilepsia* 57:2011–2018.
- Christensen D, Van Naarden Braun K, Doernberg NS, Maenner MJ, Arneson CL, Durkin MS, *et al.* (2014): Prevalence of cerebral palsy, co-occurring autism spectrum disorders, and motor functioning—Autism and Developmental Disabilities Monitoring Network, USA, 2008. *Dev Med Child Neurol* 56:59–65.
- Xiang AH, Wang X, Martinez MP, Walthall JC, Curry ES, Page K, *et al.* (2015): Association of maternal diabetes with autism in offspring. *JAMA* 313:1425–1434.
- Li M, Fallin MD, Riley A, Landa R, Walker SO, Silverstein M, *et al.* (2016): The association of maternal obesity and diabetes with autism and other developmental disabilities. *Pediatrics* 137:e20152206.
- Croen LA, Grether JK, Yoshida CK, Odouli R, Van de Water J (2005): Maternal autoimmune diseases, asthma and allergies, and childhood autism spectrum disorders: A case-control study. *Arch Pediatr Adolesc Med* 159:151–157.
- Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, *et al.* (2019): Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* 51:431–444.
- Sullivan PF, Magnusson C, Reichenberg A, Boman M, Dalman C, Davidson M, *et al.* (2012): Family history of schizophrenia and bipolar disorder as risk factors for autism. *Arch Gen Psychiatry* 69:1099–1103.
- Pedersen CB (2011): The Danish Civil Registration System. *Scand J Public Health* 39(suppl 7):22–25.
- Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J (2018): The Danish Medical Birth Register. *Eur J Epidemiol* 33:27–36.
- Mors O, Perto GP, Mortensen PB (2011): The Danish Psychiatric Central Research Register. *Scand J Public Health* 39(suppl 7):54–57.
- Pedersen CB, Mors O, Bertelsen A, Waltoft BL, Agerbo E, McGrath JJ, *et al.* (2014): A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. *JAMA Psychiatry* 71:573–581.
- Lynge E, Sandegaard JL, Rebolj M (2011): The Danish National Patient Register. *Scand J Public Health* 39(suppl 7):30–33.
- Hastie T, Tibshirani R, Friedman J (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Switzerland: Springer.
- Dietterich TG (2000): *Ensemble Methods in Machine Learning Available at: <https://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf>. Accessed April 20, 2021.*
- Krawczyk B (2016): Learning from imbalanced data: Open challenges and future directions. *Prog Artif Intell* 5:221–232.
- Poldrack RA, Huckins G, Varoquaux G (2020): Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry* 77:534–540.

33. Sasaki Y (2007): The truth of the *F*-measure Available at <https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf>. Accessed May 15, 2021.
34. Sun Y, Wong AKC, Kamel MS (2009): Classification of imbalanced data: A review. *Int J Patt Recognit Artif Intell* 23:687–719.
35. Phung S, Bouzerdoum A, Nguyen G (2009): Learning pattern classification tasks with imbalanced data sets. In: Yin P, editor. *Pattern Recognition*. Vukovar, Croatia: InTechOpen, 193–208.
36. James G, Witten D, Hastie T, Tibshirani R (2014): *An Introduction to Statistical Learning: With Applications in R*, 6th ed. New York: Springer-Verlag New York.
37. Wulff J, Ejlskov L (2017): Multiple imputation by chained equations in praxis: Guidelines and review. *Electron J Bus Res Methods* 15:41–56.
38. Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, *et al.* (2016): *Mr: Machine learning in R*. *J Mach Learn Res* 17:1–5.
39. R Core Team (2015): *R: A language and environment for statistical computing* Available at: <https://www.r-project.org/>. Accessed January 6, 2021.
40. Abel KM, Dalman C, Svensson AC, Susser E, Dal H, Idring S, *et al.* (2013): Deviance in fetal growth and risk of autism spectrum disorder. *Am J Psychiatry* 170:391–398.
41. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A (2014): The familial risk of autism. *JAMA* 311:1770–1777.
42. Schnabel A, Youssef GJ, Hallford DJ, Hartley EJ, McGillivray JA, Stewart M, *et al.* (2020): Psychopathology in parents of children with autism spectrum disorder: A systematic review and meta-analysis of prevalence. *Autism* 24:26–40.
43. Kong L, Norstedt G, Schalling M, Gissler M, Lavebratt C (2018): The risk of offspring psychiatric disorders in the setting of maternal obesity and diabetes. *Pediatrics* 142:e20180776.
44. Sanchez CE, Barry C, Sabhlok A, Russell K, Majors A, Kollins SH, Fuemmeler BF (2018): Maternal pre-pregnancy obesity and child neurodevelopmental outcomes: A meta-analysis. *Obes Rev* 19:464–484.
45. Rahman R, Kodesh A, Levine SZ, Sandin S, Reichenberg A, Schlessinger A (2020): Identification of newborns at risk for autism using electronic medical records and machine learning. *Eur Psychiatry* 63:e22.