Machine Learning Prediction of Autism Spectrum Disorder Through Linking Mothers’ and Children’s Electronic Health Record Data

Yongqiu Li, BS¹, Yu Huang, PhD¹, Shuang Yang, MS¹, Elahe M. Shychuk, MD¹, Elizabeth A. Shenkman, PhD¹, Jiang Bian, PhD¹, Amber M. Angell, PhD, OTR/L², Yi Guo, PhD¹*¹

¹Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, Florida, USA; ²Division of Occupational Science and Occupational Therapy, University of Southern California, Los Angeles, California, USA

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder typically diagnosed in children. Early detection of ASD, particularly in girls who are often diagnosed late, can aid long-term development for children. We aimed to develop machine learning models for predicting ASD diagnosis in children, both boys and girls, using child-mother linked electronic health records (EHRs) data from a large clinical research network. Model features were children and mothers’ risk factors in EHRs, including maternal health factors. We tested XGBoost and logistic regression with Random Oversampling (ROS) and Random Undersampling (RUS) to address imbalanced data. Logistic regression with RUS considering a three-year observation window for children’s risk factors achieved the best performance for predicting ASD among the overall study population (AUROC = 0.798), boys (AUROC = 0.786), and girls (AUROC = 0.791). We calculated SHAP values to quantify the impacts of important clinical and sociodemographic risk factors.

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder typically diagnosed in children. ASD presents challenges in social interaction, and is often accompanied by repetitive behaviors and limited interests¹. Children with ASD may also learn, move, or focus in different ways. It is reported that the prevalence of ASD is 2.76% among children in the United States², and the lifetime costs for supporting an individual with an ASD range from $1.4 to $2.4 million per individual³⁴. It is crucial to identify ASD as early as possible since early intervention for ASD is associated with improved social communication, cognition, and adaptive functioning¹. However, detecting ASD can be difficult because no medical tests, such as a blood test, exist for ASD. To screen and diagnose ASD, children are usually assessed for early signs, and if of high risk, referred for clinical diagnostic evaluation. Yet, no single screening or diagnostic tool is appropriate for all clinical setting, clinicians need to look at child’s developmental history and behavior to make a final diagnosis⁶⁷.

Although the early reliable signs of ASD may be observed as young as two years of age, the average age of ASD diagnosis is around four years. This large time gap, or delay in ASD diagnosis, may sabotage the efficacy of early interventions. Furthermore, under-diagnosis of ASD is a severe problem in the United States. In the Autism and Developmental Disabilities Monitoring (ADDM) Network, the largest tracking system of ASD among children aged eight years in the United States, only 74% of the ASD children had a recorded clinical diagnosis. In other words, one in four 8-year-old American children with ASD were not diagnosed, suggesting a significant under-diagnosis and highlighting the potential for late diagnosis to exacerbate psychological distress and functional challenges in daily life. The issue of delayed and under-identification of ASD is most acute among girls. Despite similar ages of parental initial concerns, the most recent surveillance data showed that the prevalence of ASD remains significantly lower among girls than among males, and research consistently finds that a critical reason for the low prevalence is under-ascertainment rather than biological or genetic causes. Girls are more likely to be diagnosed later than boys, even though they do not differ from boys in the age at their parents’ “first concerns” about their development⁹.

Currently, there exist several ASD screening tools, such as the Modified Checklist for Autism in Toddlers (MCHAT). But these tools are survey instruments that require extra efforts and resources to use, and are of sub-optimal accuracy¹⁰. For example, despite MCHAT’s popularity as a screening tool for ASD, using it alone may not yield sufficient accuracy in detecting ASD cases. A recent analysis has demonstrated that MCHAT has a sensitivity of just 39% and a positive predictive value (PPV) of 15% for ASD detection. Moreover, these instruments often fail to perform equally across population subgroups, especially in girls¹¹, leading to diagnostic disparities. On the other hand, in seeking more affordable and reliable methods, electronic health record (EHR) data may be an alternative approach for early ASD

*Correspondence: Yi Guo, PhD (viguo@ufl.edu)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
detection. Many risk factors for ASD can be found in EHRs, including low birth weight, preterm birth, low Apgar scores, other perinatal complications, and related conditions such as attention deficit hyperactivity disorders (ADHD). While EHR data have been instrumental in the early identification of phenotypes for chronic diseases such as heart failure, diabetes, and Alzheimer’s disease before symptoms appear, using EHRs for early ASD detection is not yet fully realized. To date, EHRs have been utilized to distinguish ASD subtypes and assess suicide risk among adolescents with the condition. However, there is a scarcity of studies that have used EHR data to predict ASD in children for early detection. Moreover, there is a notable lack of research that considers risk factors from both the child and the mother using EHRs, which is crucial for more accurate ASD detection.

Therefore, to avoid delayed diagnosis and under-identification of ASD in children, it is imperative to build prediction models that can be easily adopted in EHRs for early and accurate ASD detection. In this study, we aimed to (1) develop gender-specific machine learning models for ASD risk prediction utilizing EHRs in the OneFlorida+ Clinical Research Consortium, part of the national Patient-Centered Clinical Research Network (PCORnet), and (2) assess the gender disparities in ASD risk prediction and risk factors by exploring feature importance in each population group.

Methods

Data Source and Study Population

We obtained 2012-2023 EHR data from the OneFlorida+ Clinical Research Consortium which contains patient-level information including demographics, diagnoses, medications, procedures, vital signs, lab tests, and more from 17 million residents in Florida, 2.1 million in Georgia, and 1.1 million in Alabama. Our study population included a cohort of ASD children and a matching cohort of non-ASD children identified in OneFlorida+ EHRs as outlined in Figure 1.

![Figure 1. Overview of study population extraction from the OneFlorida+ clinical research consortium.](image_url)
before two years old only. For the non-ASD cohort, we additionally removed children with any intellectual disability (ID) or developmental delay (DD) diagnosis. We defined the index date as the first encounter with an ASD diagnosis at or after two years old for the cases, and as a random encounter date within the index year of the case for the controls. We then matched each ASD child (i.e., cases) with two non-ASD children (i.e., controls) based on age and year.

**Overview of Data Analysis Plan**

The main goal of our data analysis is to predict ASD diagnosis (ASD vs. non-ASD) using children and mothers’ risk factors in EHRs, including maternal health factors. We summarized the steps in our data analysis plan in Figure 2. In Step 1, we identified potential ASD risk factors based on a literature review and expert input. In Step 2, we extracted features associated with these factors from EHRs and balanced the dataset through preprocessing techniques. In Step 3, a set of machine learning models were trained using grid search cross-validation to optimize hyperparameters. In Step 4, we evaluated model performance using various metrics, and chose the best performing models. In Step 5, we employed SHAP\(^{20}\) (SHapley Additive exPlanations), a commonly used XAI technique, to discern important features for predicting ASD risk. Python version 3.7 with the Python libraries Sciki-learn\(^{21}\), Imbalanced-learn\(^{22}\), and statsmodels\(^{23}\) was used for data processing, machine learning modeling, and SHAP analysis.

**Figure 2.** Data analysis plan overview.

**Step 1: Identify Potential ASD Risk Factors**

We conducted a review of systematic reviews from 2017–2023 by searching ASD-related terms (e.g., Autism) in PubMed to identify potential ASD risk factors associated with children and mothers (during and before pregnancy). The initial list of ASD risk factors was also reviewed and expanded by domain experts. We displayed the observation windows for risk factors associated with children and mothers in Figure 3. We adopted: (1) a varying observation window (1-, 2-, or 3-year) prior to the index date for observing risk factors from children, and (2) a 22-month observation window (10 months during pregnancy and 12 months before pregnancy) prior to childbirth for observing risk factors from mothers.

**Figure 3.** Patient timeline with observation windows for identifying risk factors.

All risk factors were identified in the children’s and their mothers’ EHRs using ICD codes, Current Procedural Terminology (CPT) codes, Healthcare Common Procedure Coding System (HCPCS) codes, RxNorm codes, and National Drug Codes (NDC). We excluded risk factors with lower than 1% prevalence within each observation window, and ultimately included 27 factors in subsequent analysis (Table 1).

**Table 1.** List of risk factors used for machine learning.
### Characteristics of Study Population

We summarized the basic demographics of the ASD (n = 2,143) and non-ASD (n = 4,286) cohorts in Table 2. The ASD and non-ASD children had the same age distribution (mean = 3.9 years, standard deviation = 1.8 years). There was a significantly lower percentage of girls in the ASD cohort compared to the non-ASD cohort (24.2% vs. 31.2%; \( p < 0.001 \)). We observed a significant difference in the racial-ethnic distribution between the ASD and non-ASD cohorts (\( p < 0.001 \)). Compared to the non-ASD cohort, the ASD cohort had lower percentages of NHW (32.0% vs. 37.4%) and NHB (24.5% vs. 25.2%) yet higher percentages of NHO (11.5% vs. 9.9%) and Hispanics (24.5% vs. 20.8%). The mothers of the ASD and non-ASD children had a similar age distribution (32.5 vs. 32.0 years, \( p = 0.165 \)).
Compared to the mothers of the non-ASD cohort, mothers of the ASD cohort had lower percentages of NHW (37.8% vs. 42.6%), NHB (27.4% vs. 28.1%), and NHO (6.9% vs. 7.6%), yet higher percentages of Hispanics (24.9% vs. 16.7%).

Table 2. Demographics of study population.

<table>
<thead>
<tr>
<th>Children</th>
<th>ASD (n = 2,143)</th>
<th>Non-ASD (n = 4,286)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at index date (mean, SD)</strong></td>
<td>3.9, 1.8</td>
<td>3.9, 1.8</td>
<td>= 1.000</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,624 (75.8%)</td>
<td>2,947 (68.8%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>519 (24.2%)</td>
<td>1,339 (31.2%)</td>
<td></td>
</tr>
<tr>
<td><strong>Race-ethnicity</strong></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHW</td>
<td>686 (32.0%)</td>
<td>1,604 (37.4%)</td>
<td></td>
</tr>
<tr>
<td>NHB</td>
<td>524 (24.5%)</td>
<td>1,081 (25.2%)</td>
<td></td>
</tr>
<tr>
<td>NHO</td>
<td>247 (11.5%)</td>
<td>422 (9.9%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>525 (24.5%)</td>
<td>891 (20.8%)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>161 (7.5%)</td>
<td>288 (6.7%)</td>
<td></td>
</tr>
<tr>
<td><strong>Mothers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age at childbirth (mean, SD)</strong></td>
<td>32.5, 13.1</td>
<td>32.0, 12.6</td>
<td>= 0.165</td>
</tr>
<tr>
<td><strong>Race-ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHW</td>
<td>811 (37.8%)</td>
<td>1,826 (42.6%)</td>
<td></td>
</tr>
<tr>
<td>NHB</td>
<td>587 (27.4%)</td>
<td>1,205 (28.1%)</td>
<td></td>
</tr>
<tr>
<td>NHO</td>
<td>147 (6.9%)</td>
<td>327 (7.6%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>534 (24.9%)</td>
<td>717 (16.7%)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>64 (3.0%)</td>
<td>211 (4.9%)</td>
<td></td>
</tr>
</tbody>
</table>

ASD: Autism spectrum disorder; SD: Autism spectrum disorder; NHW: non-Hispanic White; NHB: non-Hispanic Black; NHO: non-Hispanic Other.

Performance of Machine Learning Models

We summarized the performance (i.e., AUROC, sensitivity, and specificity) of machine learning models by study population group in Table 3. Based on the AUROC, the results illustrated that the three-year observation window was the optimal period for ASD prediction across all study population groups as compared to the one- or two-year observation window in most cases. We also found that RUS consistently produced satisfactory results across different study population groups and observation windows, and logistic regression consistently outperformed XGboost, although slightly. Under RUS, logistic regression with a three-year observation window achieved the best AUROC of 0.786, 0.791, 0.798 for predicting ASD in the overall, boys, and girls study population, respectively.

Table 3. Overall performance of machine models.

<table>
<thead>
<tr>
<th></th>
<th>AUROC</th>
<th>1-year</th>
<th>2-year</th>
<th>3-year</th>
<th>Sensitivity</th>
<th>1-year</th>
<th>2-year</th>
<th>3-year</th>
<th>Specificity</th>
<th>1-year</th>
<th>2-year</th>
<th>3-year</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall (Boys + Girls)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROS LR</td>
<td>0.789</td>
<td>0.795</td>
<td>0.798</td>
<td>0.576</td>
<td>0.584</td>
<td>0.588</td>
<td>0.886</td>
<td>0.887</td>
<td>0.891</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XGboost</td>
<td>0.785</td>
<td>0.793</td>
<td>0.796</td>
<td>0.526</td>
<td>0.558</td>
<td>0.569</td>
<td>0.921</td>
<td>0.899</td>
<td>0.899</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUS LR</td>
<td>0.789</td>
<td>0.795</td>
<td>0.798</td>
<td>0.578</td>
<td>0.581</td>
<td>0.590</td>
<td>0.883</td>
<td>0.891</td>
<td>0.888</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XGboost</td>
<td>0.782</td>
<td>0.790</td>
<td>0.794</td>
<td>0.525</td>
<td>0.565</td>
<td>0.544</td>
<td>0.925</td>
<td>0.895</td>
<td>0.921</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Boys</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROS LR</td>
<td>0.774</td>
<td>0.781</td>
<td>0.786</td>
<td>0.561</td>
<td>0.571</td>
<td>0.578</td>
<td>0.866</td>
<td>0.874</td>
<td>0.874</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XGboost</td>
<td>0.765</td>
<td>0.775</td>
<td>0.781</td>
<td>0.561</td>
<td>0.514</td>
<td>0.540</td>
<td>0.866</td>
<td>0.922</td>
<td>0.897</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUS LR</td>
<td>0.774</td>
<td>0.781</td>
<td>0.786</td>
<td>0.531</td>
<td>0.567</td>
<td>0.584</td>
<td>0.884</td>
<td>0.878</td>
<td>0.872</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XGboost</td>
<td>0.767</td>
<td>0.769</td>
<td>0.780</td>
<td>0.566</td>
<td>0.553</td>
<td>0.520</td>
<td>0.863</td>
<td>0.871</td>
<td>0.925</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Girls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROS LR</td>
<td>0.778</td>
<td>0.789</td>
<td>0.784</td>
<td>0.566</td>
<td>0.572</td>
<td>0.530</td>
<td>0.877</td>
<td>0.905</td>
<td>0.937</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XGboost</td>
<td>0.776</td>
<td>0.786</td>
<td>0.784</td>
<td>0.483</td>
<td>0.514</td>
<td>0.527</td>
<td>0.971</td>
<td>0.957</td>
<td>0.952</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ROS: Random Oversampling; RUS: Random Undersampling, LR: Logistic Regression.

1-year, 2-year, 3-year are varying observation windows before index for children’s risk factors.

**SHAP Values Measuring Feature Importance**

We calculated SHAP values based on the best performing model, logistic regression with RUS considering a three-year observation window for children’s risk factors. We displayed the SHAP summary plots for the top 15 features in each population group in Figure 4.

Figure 4A shows the important features for the overall study population of children. DD and ADHD were the two most impactful predictors for increased risk of ASD. Being NHO or having diagnoses such as congenital malformations, ID, and anxiety was associated with elevated ASD risk, whereas being girl or NHW was associated with lower risk of ASD. Additionally, risk factors from the mothers also played a significant role in ASD risk prediction. Significant maternal health risk factors for ASD included having autoimmune diseases and obesity. Mother being Hispanic or NH as well as pre-pregnancy obesity and use of antidepressant medications were all associated with increased risk of ASD.

Figure 4B highlights the influential factors for predicting ASD in boys. ADHD and DD were marked as the two most significant predictors of ASD. Other predictors contributing to increased risk of ASD included congenital malformations and genetic abnormalities. Being Hispanic or NHW was associated with a decreased ASD risk, whereas being NHO was associated with an increased ASD risk. Maternal health risk factors included the use of antidepressant medications and having fever, obesity, and autoimmune diseases. Other risk factors from mothers included being Hispanic as well as pre-pregnancy obesity and use of antidepressant medications.

Figure 4C highlights the important factors for predicting ASD risk in girls. The plots revealed that the diagnosis of ADHD and DD were the two primary predictors of ASD risk. Other predictors such as congenital malformations also contributed to an increased risk of ASD. Being Hispanic was associated with an increased ASD risk, while being NHW was associated with a decreased ASD risk. Maternal health risk factors for ASD such as having autoimmune diseases and obesity were associated an increased ASD risk, whereas fever was associated with a decreased ASD risk. Regarding other risk factors from mothers, being Hispanic as well as having pre-pregnancy hypertension, obesity and use of antidepressant medications were associated with an increased risk of ASD, while being NHB or NHW was associated with a decreased risk of ASD.

<table>
<thead>
<tr>
<th>Model</th>
<th>RUS</th>
<th>LR</th>
<th>0.782</th>
<th>0.788</th>
<th>0.791</th>
<th>0.570</th>
<th>0.577</th>
<th>0.577</th>
<th>0.883</th>
<th>0.891</th>
<th>0.888</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGboos</td>
<td>0.768</td>
<td>0.773</td>
<td>0.784</td>
<td>0.494</td>
<td>0.512</td>
<td>0.544</td>
<td>0.959</td>
<td>0.956</td>
<td>0.935</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4. SHAP plots of the top 15 features from logistic regression using the RUS resampling technique based on a three-year observation window for children’s ASD risk factor. (A) Overall (i.e., both boys and girls), (b) boys, and (c) girls.

Discussion and Conclusions

In this study, we developed machine learning models for identifying children with ASD within three study population groups, including overall (i.e., both boys and girls), boys, and girls using OneFlorida+ EHR data. We tested both XGBoost and logistic regression using the ROS and RUS resampling techniques to create the prediction models considering a varying observation window (i.e., one-year, two-year, or three-year before index date) for children’s ASD risk factors. Logistic regression using the RUS resampling method with a three-year observation window achieved the best performance for predicting ASD among the overall population (AUROC = 0.798), boys (AUROC = 0.786), and girls (AUROC = 0.791), which indicates that our final prediction models are accurate tools for predicting ASD among children. Particularly, the model for girls achieved an AUROC of 0.791, which is as good as the model for boys and the overall population, indicating our model can be used to detect ASD in girls who are usually face delays in diagnosis.

In each population group (i.e., overall, boys, and girls), we identified important features for ASD risk prediction by calculating SHAP values using our three best performing models. We observed some of the same predictors in the top features in all population groups. In particular, DD and ADHD were consistently underscored as the most important predictors of ASD in all groups, which aligns with existing literature. For instance, Riaz et al reported that the majority of children with ASD experience did not achieve developmental milestones and suffered from developmental delays in a few skills such as language and social skills. Similarly, ADHD’s co-occurrence with ASD was supported by...
several studies, which suggested that 20-50% of children with ADHD also fulfilled the criteria for ASD, and 30-80% of children with ASD might exhibit symptoms of ADHD\(^{4-8}\). These prevalence rates serve to rationalize the prominence of these features in our models across all population groups. In the model for the overall study population (i.e., both boys and girls) (Figure 4A), we observed a reduced ASD risk in girls, which aligns with previous findings that girls often face delays and underdiagnosis for ASD\(^9\). Examination of the SHAP plots also highlighted that congenital abnormalities had a more pronounced impact in boys. This could be attributed to the higher prevalence of most congenital anomalies in males at birth, as supported by research from Lary et al\(^{41}\).

Maternal health factors, such as having autoimmune diseases and obesity were consistently shown as predictors of ASD in all population groups. This finding of ours aligns with existing literature. For instance, Chen et al found that maternal autoimmune diseases significantly increased the risk of ASD in children\(^5\). Similarly, Gardner et al revealed that maternal obesity was associated with an increased ASD risk\(^6\). Other factors from mothers such as being Hispanic as well as pre-pregnancy obesity and use of antidepressant medications were also showed to consistently correlate with increased ASD risk in all population groups. Andrade et al reported that antidepressant use 3-12 months prior to pregnancy was correlated with a rise in ASD risk in children\(^7\). Krakowlak et al observed that children whose mother was obese before pregnancy had 67% higher risk of ASD\(^8\). The identification of ASD risk factors from mothers, independent of those from children, underscores the value of linking children’s and mothers’ EHRs, which offers a more comprehensive view of factors contributing to ASD. This unique approach of ours indicates the importance of considering risk factors from mothers, both pre-pregnancy factors and maternal health factors, a perspective not extensively explored in prior EHR studies.

Our study has several strengths, primarily the innovative integration and analysis of linked mother-child EHR data, which provided a more comprehensive perspective on potential ASD risk factors from both maternal and child health. Furthermore, our gender-specific models, which showed good performance in both boys and girls, may be leveraged to accurately predict ASD in girls, who are typically under-diagnosed. Our study is not without limitations. Although we utilized EHR data from a large clinical research network, the models we built were not externally validated across diverse datasets, which may limit the generalizability of our findings. Additionally, we were unable to examine certain risk factors potentially important for ASD prediction, such as family history (e.g., risk factors related to father) and environmental exposures, due to the lack of data on these factors in EHRs. Future studies should aim to include these important yet omitted factors for ASD prediction, and validate the resulted models in multi-site settings to ensure the models’ applicability on a broader scale.

In summary, our study developed machine learning prediction models for ASD in the overall children population, and boys and girls separately. Our models used child-mother linked EHR data and can be used for accurate ASD prediction in both boys as well as girls who are usually under-diagnosed. Future research should prioritize external model validation and investigate the models’ clinical applications as clinical decision support tools. Incorporating additional data types, such as clinical notes, genomic data, and exposome data, could further be considered to enhance the predictive ability of ASD risk models.

**Acknowledgement**

This study was supported by grant R21MH129682 from the National Institutes of Health (NIH). This study was also partially supported by NIH grants R01CA246418, R01CA246418-02S1, R21CA245858, R21CA245858-01A1S1, R21CA253394-01A1, R01AG080624, and R21AG068717. The authors wish to thank the Cancer Informatics Shared Resource in the UF Health Cancer Center for data analytics support.

**References**


33. Patterns of Developmental Delay in Children with Autism Spectrum Disorder: A Perspective from a Developing Country.