Analysis of electronic health record data identifies features associated with future genetic evaluation

Elmer Rajah, Nhat Duong, Andrea Cohen, Eric Villain, Andrea Gropman, Seth I. Berger

Children's National

INTRODUCTION

Neurodevelopmental delay is a feature of most rare genetic diseases and is often the first presenting sign. Pediatricians are the first care providers to screen for developmental delays using standardized screening batteries in conjunction with observation and parent/caregiver histories. Automated evaluation of the data generated by these primary care providers can suggest earlier referral for children with first signs of neurodevelopmental disorders. Our analysis of primary care electronic medical record data identified predictive features that are correlated with risk for neurodevelopmental disorders potentially due to genetic etiologies.

METHODS

- Retrospective patient data was obtained from Cerner and eClinicalWorks, representing 25,281 individual visits for well-child checks for 14,808 patients at three separate time points
- ICD-10 codes, encounters, newborn screening results, biometrics (height, weight, and head circumference) and medications were extracted from 5 years of data from the Goldberg Center's primary care records cross referenced with genetics and neurogenetics records
- Point biserial correlation coefficients were calculated for each feature and the outcome of genetics evaluation at 9 months, 18 months, and 30 months old.
- Kolmogorov-Smirnov Goodness of Fit was calculated between each feature and the identified outcome variable.
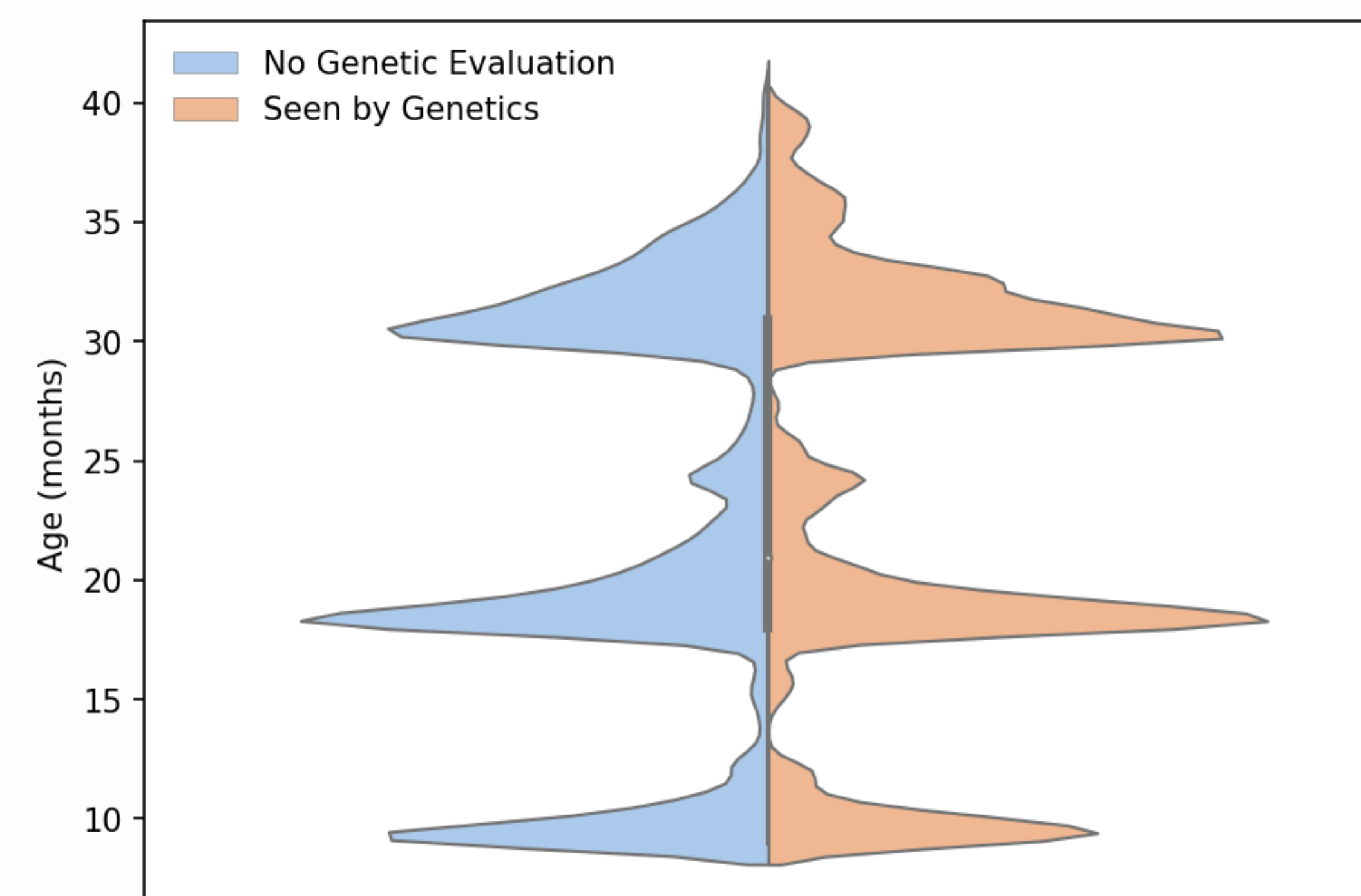
RESULTS

Figure 1: Age distribution
Distribution of age in months, demonstrating three distinct peaks that correlate with developmental screening timelines.
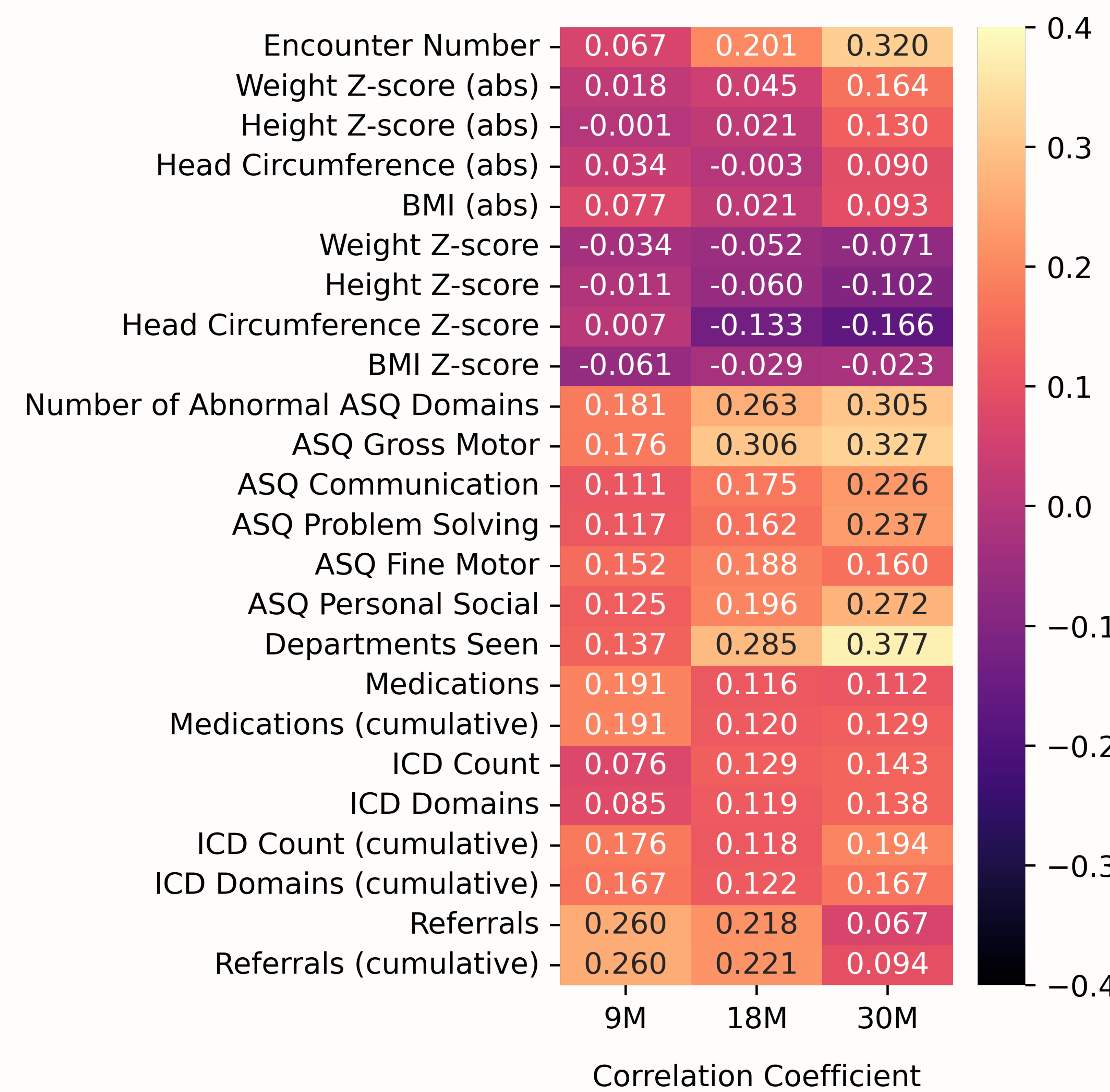

Figure 2: Point biserial correlation of features with genetic evaluation
Correlation coefficients between features extracted from electronic medical record and genetic evaluation.
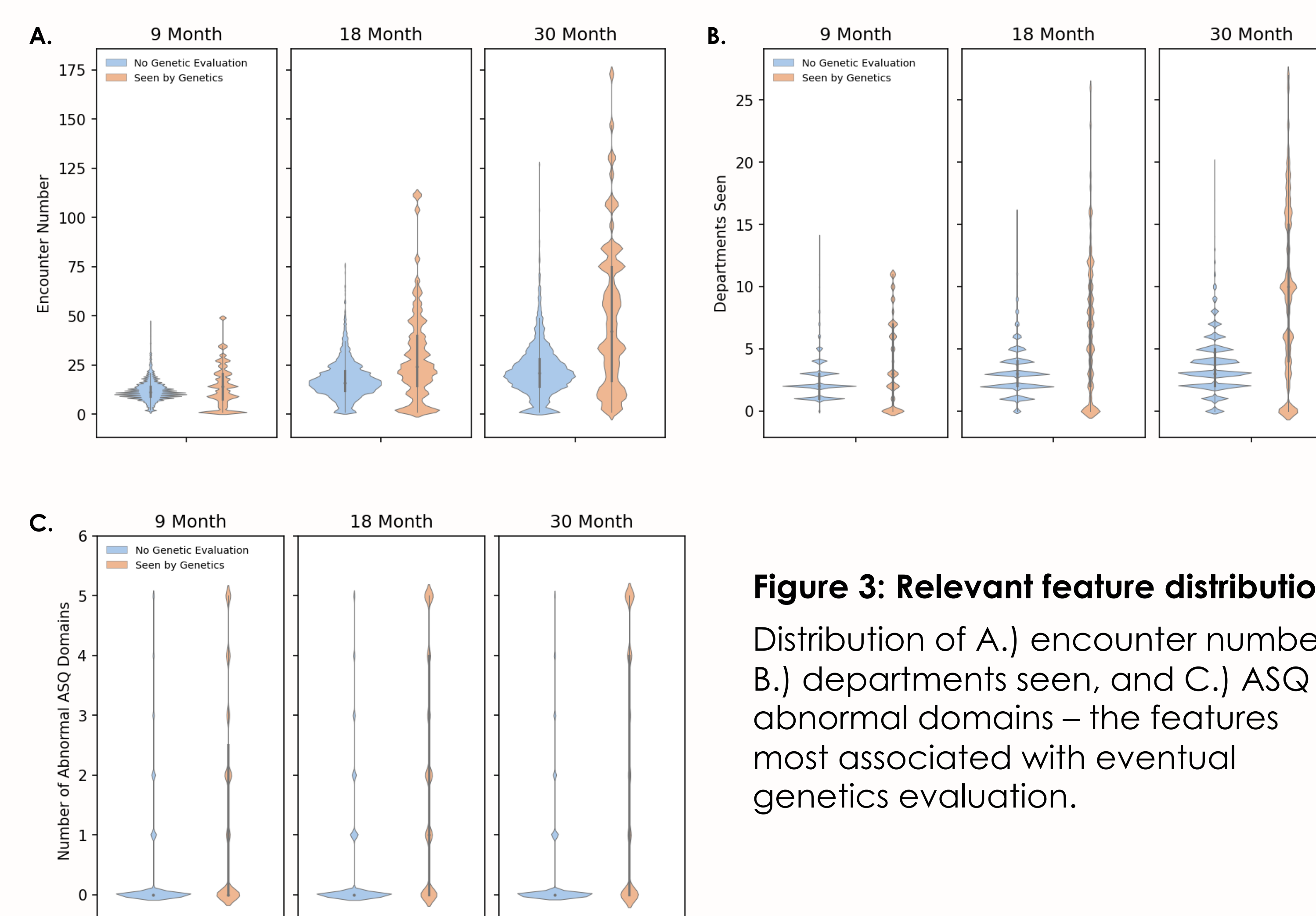

Figure 3: Relevant feature distribution
Distribution of A.) encounter number, B.) departments seen, and C.) ASQ abnormal domains – the features most associated with eventual genetics evaluation.

DISCUSSION

- 25,281 records were accessed, with 11,118 meeting the inclusion criteria for analysis
- Relevant features
  - 9 Month: referrals, medications, number of abnormal ASQ domains, ICD codes
  - 18 Month: departments seen, ASQ gross motor score, number of abnormal ASQ domains
  - 30 Month: departments seen, ASQ gross motor score, encounter number, number of abnormal ASQ domains
- Number of encounters, number of abnormal ASQ domains, number of departments seen, and ICD codes/domains demonstrated stronger correlations as children got older.
- Persistence of individual abnormal ASQ domains was also associated with eventual genetics evaluation.

CONCLUSION

Features extracted from primary care records associated with abnormal screening results and frequent encounters/intervention with medical services are predictive of future genetic evaluation. The relationship between these variables and their temporal predictive value may be used to develop and train a machine learning assisted algorithm that can identify patients in need of genetic testing and evaluation.

REFERENCES

1. Boyle, C.A., et al., *Trends in the prevalence of developmental disabilities in US children, 1997-2008.* Pediatrics, 2011. 127(6): p. 1034-42.
2. Office of Special Education and Rehabilitative Services, D.p.o.E., *Early intervention program for infants and toddlers with disabilities. Final regulations.* Fed Regist, 2011. 76(188): p. 60140-309.
3. Moeschler, J.B., M. Shevell, and G. Committee on, *Comprehensive evaluation of the child with intellectual disability or global developmental delays.* Pediatrics, 2014. 134(3): p. e903-18.