Comorbidity Clusters in Autism Spectrum Disorders: An Electronic Health Record Time-Series Analysis

WHAT'S KNOWN ON THIS SUBJECT: Individuals with autism spectrum disorders have a higher comorbidity burden than the general pediatric population, including higher rates of seizures, psychiatric illness, and gastrointestinal disorders.

WHAT THIS STUDY ADDS: Comorbidities do not occur evenly. Our clustering analysis reveals subgroups characterized by seizure, psychiatric disorders, and complex multisystem disorders including auditory and gastrointestinal disorders. Correlations between seizure, psychiatric disorders, and gastrointestinal disorders are validated on a sample from a second hospital.

AUTHORS: Finale Doshi-Velez, PhD,[a] Yaorong Ge, PhD,[b] and Isaac Kohane, MD, PhD[a]

[a]Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts; and [b]Center for Biomedical Informatics, Wake Forest University, Winston-Salem, North Carolina

abstract

OBJECTIVE: The distinct trajectories of patients with autism spectrum disorders (ASDs) have not been extensively studied, particularly regarding clinical manifestations beyond the neurobehavioral criteria from the Diagnostic and Statistical Manual of Mental Disorders. The objective of this study was to investigate the patterns of co-occurrence of medical comorbidities in ASDs.

METHODS: *International Classification of Diseases, Ninth Revision* codes from patients aged at least 15 years and a diagnosis of ASD were obtained from electronic medical records. These codes were aggregated by using phenotype-wide association studies categories and processed into 1350-dimensional vectors describing the counts of the most common categories in 6-month blocks between the ages of 0 to 15. Hierarchical clustering was used to identify subgroups with distinct courses.

RESULTS: Four subgroups were identified. The first was characterized by seizures ($n = 120$, subgroup prevalence 77.5%). The second ($n = 197$) was characterized by multisystem disorders including gastrointestinal disorders (prevalence 24.3%) and auditory disorders and infections (prevalence 87.8%), and the third was characterized by psychiatric disorders ($n = 212$, prevalence 33.0%). The last group ($n = 4316$) could not be further resolved. The prevalence of psychiatric disorders was uncorrelated with seizure activity ($P = .17$), but a significant correlation existed between gastrointestinal disorders and seizures ($P < .001$). The correlation results were replicated by using a second sample of 496 individuals from a different geographic region.

CONCLUSIONS: Three distinct patterns of medical trajectories were identified by unsupervised clustering of electronic health record diagnoses. These may point to distinct etiologies with different genetic and environmental contributions. Additional clinical and molecular characterizations will be required to further delineate these subgroups. *Pediatrics* 2014;133:e54–e63

Clinical manifestations of autism spectrum disorders (ASDs) beyond the core *Diagnostic and Statistical Manual of Mental Disorders* criteria have been gaining increasing attention.[1–5] With the prevalence of autism near 1%,[6] understanding this comorbidity burden is especially important (not least because of the clinical resources this burden entails). Understanding the co-occurrence patterns among comorbidities in ASD is the first step for uncovering the underlying etiologies associated with ASD and stratifying the risk of various conditions across individuals with ASD.

Previous research has quantified the prevalence of various comorbidities in ASD, including the rates of gastrointestinal disorders and complaints,[7,8] epilepsy,[9,10] sleep disorders,[11] muscular dystrophy,[12–14] and psychiatric illness.[15] However, with the exception of Kohane et al,[5] these studies have involved small samples (under 200 individuals) and have focused on the prevalence of a single disorder. The distinct clinical trajectories of patients with ASD have not been extensively studied, particularly as regards to these comorbidities.

The objective of this study was to better understand the co-occurrence patterns of comorbidities in ASD. ASD is known to be a heterogeneous disorder with complex genetic underpinnings.[16] Subsets of these variants are common to other diseases. For example, there exist mutations common to ASD, attention-deficit/hyperactivity disorder, and schizophrenia.[17–20] Mutations responsible for Rett syndrome[21] and Fragile X[22] carry a much higher risk of ASD and epilepsy. The association between peripartum infections and ASD is also well-documented.[23–25] Subgroups clustered on clinical criteria may be enriched for different etiologies; if so, these criteria may then be used to specifically target and evaluate therapies or preventive measures.

We use electronic health records (EHRs) to cluster co-occurrence patterns of clinical conditions. EHR and claims data have large research potential to discover both disease–disease and disease–gene correlations.[26–28] Studies have described procedures for assessing patient similarities between patients[29,30]; applications include risk stratification strategies for diabetes and cardiovascular disease.[31] Carney and Jones[32] use claims data from a large provider to identify comorbidities associated with bipolar disorders. Particularly valuable is that these data already document the date of the clinic visit during which a disorder was reported.

Previous work in clustering phenotypes in ASD has relied on surveys and diagnostic tests, limiting the sample size. For example, Miles et al[33] divide ASD into 2 clusters, "essential" and "complex" based on the manifestation of significant dysmorphology or microcephaly. They find that patients with complex ASDs have poorer outcomes, including lower IQ and more seizures. Other studies have focused on the core neurobehavioral criteria. Wiggins et al[34] find clusters along disease severity, whereas Lane et al[35] discover sensory processing subtypes. Other studies[36–39] find clusters along cognitive, language, and behavioral criteria. Sacco et al[40] find patterns among both neurodevelopmental factors as well as immune and circadian dysfunction.

Through the use of EHR data, we find distinct clinical trajectories of comorbidities outside of the core neurobehavioral ASD *Diagnostic and Statistical Manual of Mental Disorders* criteria. From a sample of 4927 patients aged 15 years from a tertiary-care pediatric hospital (mean follow-up 11 years, SD 4.8 years), our clustering analyses revealed 3 high-morbidity subgroups: 1 characterized by seizures, 1 characterized by psychiatric disorders, and 1 characterized by more complex multi-system disorders. These phenotypic distinctions may point to distinct etiologies with different genetic and environmental contributions.

METHODS

Patients

We identified 13 740 individuals with at least 1 *International Classification of Diseases, Ninth Revision* (ICD-9) code for ASD (299.00, 299.01, 299.80, 299.81, 299.90, 299.91) by using infrastructure from the i2b2 National Center for Biomedical Computing[41] at Boston Children's Hospital. Of these, 4934 individuals (78% boys) were at least 15 years old. Key patterns found among these individuals were examined for in a sample of 496 (80% boys) individuals from Wake Forest University Health Sciences (the full study could not be replicated because of the small sample size). The institutional review boards of Harvard Medical School, Boston Children's Hospital, and Wake Forest University reviewed and approved the research protocol.

Methods

The 6905 ICD-9 codes present in our data were aggregated into 802 categories used in phenotype-wide association studies (PheWAS).[42] Procedure codes were ignored. The ICD-9 codes for key categories are provided in the Supplemental Information. Individuals with more than 50 instances of the same category code in a 6-month period were excluded because their records were dominated by conditions unrelated to ASD (eg, renal failure, oncology visits). Finally, we only considered categories that had at least 5% prevalence in the sample. This pre-processing resulted in 45 common category codes and 4927 individuals.

For each patient, we constructed a time-series with 30 6-month windows from

birth to age 15. For each time window, we counted the number of occurrences of each of the 45 categories for that patient in that window. This processing step resulted in a 30 × 45 = 1350-dimensional vector of counts per patient. Hierarchical clustering by using a Euclidean distance and Ward's method resulted in 4 clusters, where clusters were constrained to a minimum size of 2% of the overall sample.

Differences between the clusters were assessed by using permutation tests. We computed a χ^2 statistic comparing the expected number of code counts in each time window in each cluster to the observed numbers for each patient. Next, the same statistic was recomputed on 15 000 permutations of the patients. Thus, the permutations preserved the cluster sizes and the intrapatient comorbidity statistics. The empirical P value for each cluster was computed by comparing the χ^2 statistic for our clustering to the empirical distribution created from the permutations. The number of permutations was chosen to estimate small P values to sufficient precision to then apply a Bonferroni multiple hypothesis correction to these P values.

Analyses were performed by using Matlab 7.14.0.739 (Mathworks; Natick, MA) (R2012a) and R 2.15.2 (R Foundation; Vienna, Austria).

RESULTS

The hierarchical clustering resulted in 4 clusters (Table 1). Three clusters were small (n = 120; 197; 212), and the final cluster (n = 4316) could not be further resolved into subgroups. Eighty-two outliers were excluded because they were far from all the other clusters and each other. Individuals in the smaller subgroups averaged over 5 times more codes than the larger subgroup, suggesting that the smaller subgroups had higher morbidity than the overall sample. Subgroup 1 had slightly fewer

boys and the oldest age of first diagnosis. Individuals in subgroups 1 and 2 had more diagnoses of autism than Asperger syndrome, whereas subgroup 3 had more individuals with Asperger syndrome.

Subgroup characteristics are summarized in Table 2 and Fig 1 (the ICD-9 codes used to define these comorbidities are in the Supplemental Information). Subgroup 1 was characterized by seizures (prevalence 77.5%). Subgroup 2 was characterized by multisystem disorders including gastrointestinal disorders (24.4%) and early ear infections and auditory disorders (87.8%). Subgroup 3 was characterized by psychiatric disorders (prevalence 33.0%). All of these subgroups had higher levels of cardiac disorders (30.8%, 33.0%, and 24.0%, compared with 6.9%) and intellectual disability (60.0%, 48.7%, and 27.8%, compared with 12.7%). The final subgroup had no significantly elevated comorbidities, so we now focus on the 3 high-morbidity subgroups.

The temporal patterns of developmental delays varied between the subgroups. Individuals in subgroup 2, characterized by multisystem disorders,

had a spike in diagnoses for developmental delays (ICD-9 codes starting with 315) at age 2.5 (Fig 2). Specific developmental delays for individuals in subgroup 1 rose steadily through age 5. In contrast, subgroup 3 had steady and relatively low prevalence of specific developmental delays through age 15. Table 3 shows the specific ICD-9 codes contributing to the developmental delays for each subgroup. Codes for expressive language disorder (315.31) were more common in subgroups 2 and 3, and codes for mixed developmental disorder (315.5) were more common in subgroups 1 and 2. These patterns are consistent with higher proportion of autism in subgroups 1 and 2 and Asperger syndrome in subgroup 3.

Finally, the rates of visits where ASD diagnoses were recorded varied widely between the different subgroups (Fig 3). Individuals in subgroup 3, the group characterized by psychiatric disorders, had the earliest ASD diagnoses; whether that was the case because their ASDs were more clear and thus diagnosed earlier, because their psychiatric disorder was initially

TABLE 1 Characteristics of Each Subgroup (With 95% Confidence Intervals)

Subgroup	Subgroup 1	Subgroup 2	Subgroup 3	Subgroup 4
Size	120	197	212	4316
Average ICD-9 codes	150.0 (131.1–168.9)	178.6 (158.0–199.3)	103.6 (94.5–112.6)	20.8 (19.9–21.6)
Proportion of boys	0.6 (0.5–0.7)	0.7 (0.6–0.8)	0.8 (0.8–0.9)	0.8 (0.8–0.8)
Mean age of diagnosis in years	9.5 (8.8–10.1)	7.7 (7.1–8.3)	7.6 (7.1–8.0)	8.0 (7.9–8.1)
Autism proportion	0.8 (0.7–0.8)	0.7 (0.6–0.8)	0.6 (0.6–0.7)	0.7 (0.7–0.7)
Asperger proportion	0.4 (0.3–0.5)	0.5 (0.4–0.6)	0.8 (0.7–0.9)	0.5 (0.5–0.5)
PDD-NOS proportion	0.2 (0.1–0.3)	0.1 (0.1–0.2)	0.2 (0.1–0.2)	0.1 (0.1–0.1)

PDD-NOS, Pervasive Developmental Disorders, Not Otherwise Specified.

TABLE 2 Rate of Disorders in Each Subgroup (With 95% Confidence Intervals)

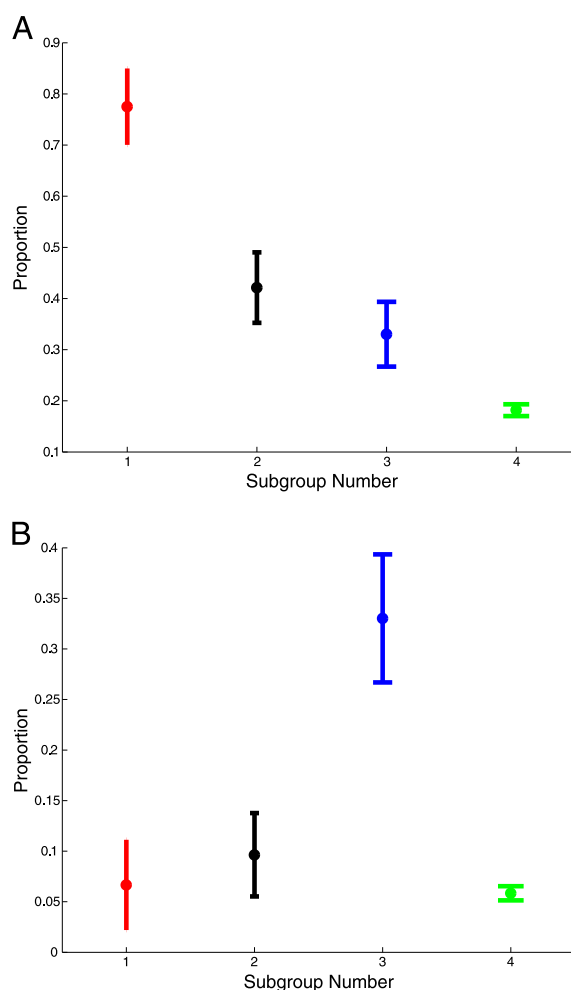| Outcome	Subgroup 1	Subgroup 2	Subgroup 3	Subgroup 4
Seizure	77.50 (70.03–84.97)	42.13 (35.24–49.03)	33.02 (26.69–39.35)	18.16 (17.01–19.32)
Psychiatric disorders	6.67 (2.20–11.13)	9.64 (5.52–13.77)	33.02 (26.69–39.35)	5.84 (5.14–6.54)
GI disorders	14.17 (7.93–20.41)	24.37 (18.37–30.36)	10.85 (6.66–15.04)	3.43 (2.89–3.97)
Intellectual disability	60.00 (51.23–68.77)	48.73 (41.75–55.71)	27.83 (21.80–33.86)	12.70 (11.70–13.69)
Auditory disorders and infections	55.83 (46.95–64.72)	87.82 (83.25–92.38)	47.17 (40.45–53.89)	23.12 (21.87–24.38)
Cardiac disorders	30.83 (22.57–39.10)	32.99 (26.43–39.56)	24.06 (18.30–29.81)	6.93 (6.17–7.69)

FIGURE 1
Prevalence of seizures and psychiatric disorders in each subgroup (with 95% confidence intervals). Subgroup 1 is characterized by seizures, whereas subgroup 3 is characterized by psychiatric disorders. Prevalence is with respect to the subgroup sizes recorded in Table 1. A, Prevalence of seizures. B, Prevalence of psychiatric disorders.

rate of convulsions rises at approximately age 3 then stays steady; increasing diagnoses of epilepsy are seen starting at age 3 and continuing through age 15, likely because the convulsions are now being diagnosed as epilepsy.

Figure 5 shows the prevalence trajectory of comorbidities associated with subgroup 2. This subgroup contained multisystem disorders including gastrointestinal disorders, disorders and infections of the ear, cardiac disorders, and congenital anomalies. Individuals in this clinical phenotype had an elevated prevalence of seizures, psychiatric disorders, and cerebral palsy. However, they appeared to be distinct individuals with just one of those conditions as they also had the most diagnostic codes overall (Table 1).

Subgroup 3 was characterized by psychiatric disorders, including episodic mood disorders, bipolar disorder, depression, anxiety dissociative and somatoform disorders, conduct disorders, and hyperkinetic syndrome of childhood (Fig 6). Except for hyperkinetic syndrome of childhood, which rose in prevalence starting at age 5, these disorders were diagnosed later, rising gradually between the ages of 5 and 15. Diagnoses of anxiety-related disorders spiked sharply after age 11.

Finally, diagnosing ASD in an individual with cerebral palsy or blindness can be challenging. We recomputed the clustering with individuals with those diagnoses removed. The first and third subgroups (characterized by seizures and psychiatric disorders) stayed similar. The second subgroup, with individuals characterized by multisystem disorders, absorbed some of the higher-morbidity individuals from subgroup 3 when many of previous higher-morbidity individuals were removed. Thus, the clustering patterns described above appear to be legitimate patterns in ASD, not just artifacts of other conditions.

misdiagnosed as an ASD, or because of the specialties present at the hospital cannot be determined from these data.

Characteristics of Each Subgroup

We now summarize the comorbidities that were statistically significant at the $\alpha = .05$ level for each subgroup after Bonferroni correction. All 802 comorbidities were included as part of the multiple hypothesis correction, and complete P value tables are included in the Supplemental Information.

To summarize the most relevant information, the plots do not reveal all of the statistically significant comorbidities for each cluster, as even after

Bonferroni correction some subgroups had tens of associated conditions. We only plot categories associated with specific conditions (excluding categories with names containing "other," "nonspecific," "unspecified," or "symptom"). We also only plot comorbidities with elevated rates in the subgroup, have a subgroup prevalence >5%, and an interquartile range >3% (reveal changes over time).

The defining characteristic of the first subgroup was seizures (Fig 4), although several comorbidities (including gastrointestinal conditions, cerebral palsy, and disorders of the visual pathways) survived the Bonferroni correction. The
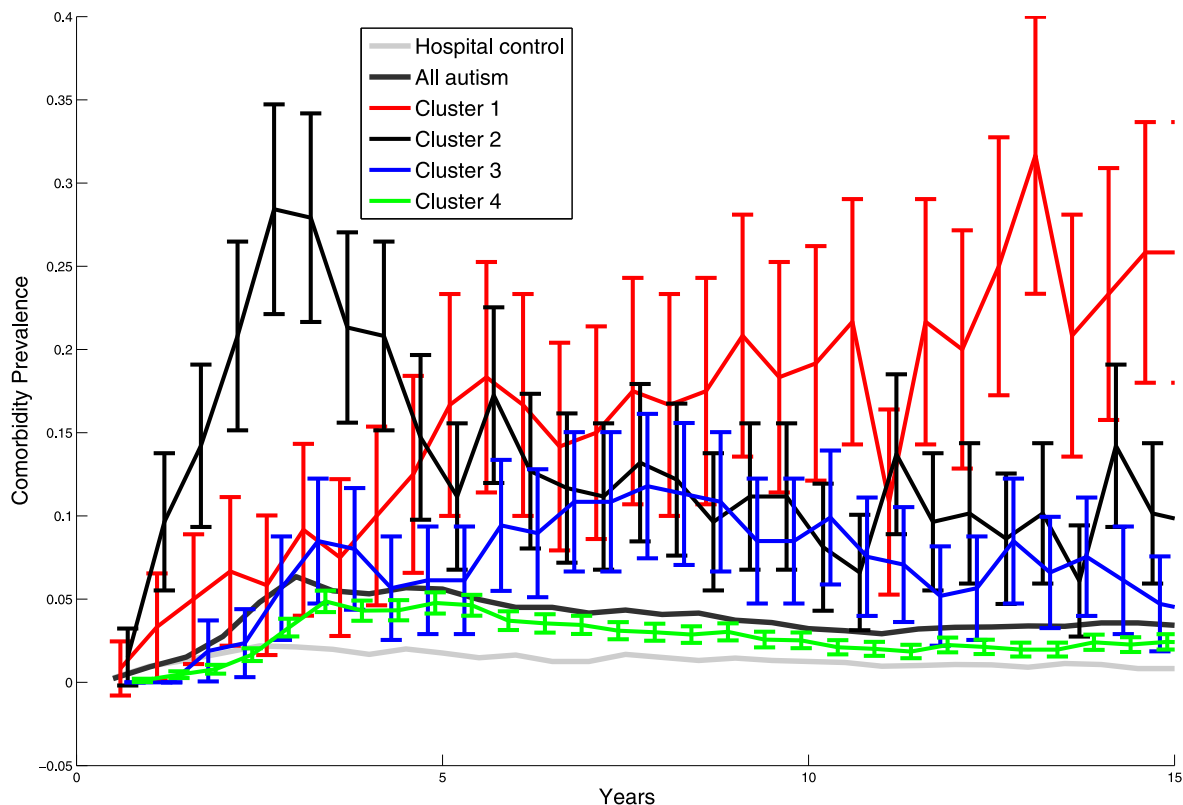
FIGURE 2
Prevalence of various specific delays in development (eg, reading and coordination delays) for each of the 4 subgroups.

TABLE 3 ICD-9 Codes Contributing to at Least 10% of the Specific Delays in Development

Subgroup	Specific Delays in Development (%)
Subgroup 1	315.4, Developmental coordination disorder (10.0)
	315.5, Mixed developmental disorder (18.1)
	315.8, Other specified delays in development (29.4)
	315.9, Unspecified delay in development (26.2)
Subgroup 2	315.31, Expressive language disorder (22.1)
	315.5, Mixed developmental disorder (11.3)
	315.8, Other specified delays in development (15.4)
	315.9, Unspecified delay in development (29.5)
Subgroup 3	315, Specific delays in development (11.7)
	315.2, Other specific developmental learning difficulties (10.4)
	315.31, Expressive language disorder (10.7)
	315.9, Unspecified delay in development (33.9)
Subgroup 4	315.31, Expressive language disorder (16.0)
	315.39, Other developmental speech or language disorder (10.5)
	315.5, Mixed developmental disorder (12.4)
	315.9, Unspecified delay in development (26.1)

Validation: Separation Between Psychiatric and Multisystem Disorders

The 3 subgroups from our original clustering analysis consisted of <10% of the overall sample. Thus, we could not validate the subgroups in a smaller sample of only 496 individuals from a different hospital (a comparison of the 2 hospital samples are in the Supplemental Information). However, we could validate the apparent distinction between individuals with only psychiatric disorders (subgroup 3) and individuals with seizures and multisystem disorders (subgroups 1 and 2).

We hypothesized that psychiatric disorders would not be significantly correlated with seizures or gastrointestinal disorders, whereas seizures and gastrointestinal disorders would be significantly correlated. In the original Boston Children's Hospital sample, there was no evidence for correlation between psychiatric disorders and seizures (Fisher's exact uncorrected $P = .17$) or psychiatric disorders and gastrointestinal disorders (Fisher's exact uncorrected $P = .04$) and strong evidence for a correlation between seizures and gastrointestinal disorders (Fisher's exact uncorrected $P < .001$).

These hypotheses were supported in the Wake Forest individuals: after a Bonferroni correction, we found no evidence for a correlation between psychiatric disorders and seizures (Fisher's exact uncorrected $P = .13$) or psychiatric disorders and gastrointestinal
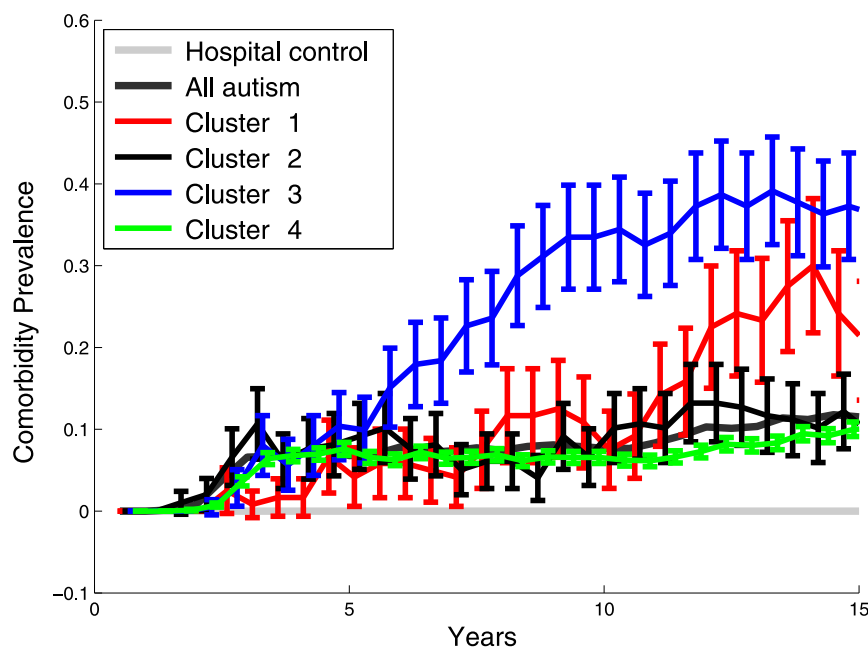
FIGURE 3
Prevalence of ASD diagnoses over time. ASD diagnoses spike dramatically starting at age 10 for individuals for subgroup characterized by seizures (subgroup 1). In contrast, individuals in the subgroup characterized by psychiatric disorders (subgroup 3) have higher rates of ASD-related diagnoses starting at age 5.
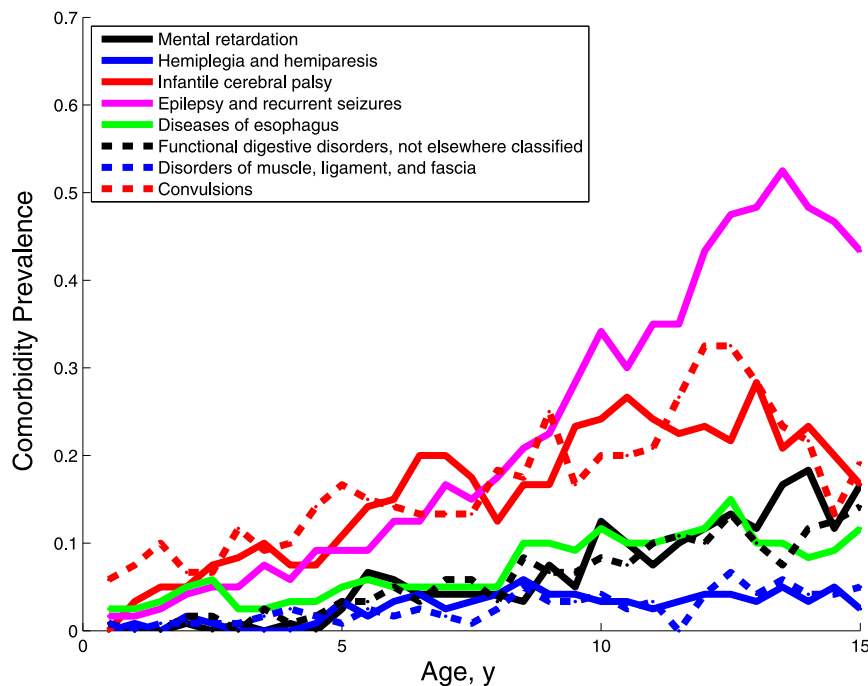

FIGURE 4
Prevalence of the diagnoses characterizing subgroup 1.

$P = .22$). Our findings are consistent with the lack of correlation between medical and psychiatric comorbidities in Ming et al.[1]

DISCUSSION

The prevalence of the comorbidities described in here are higher in the ASD population than in the general pediatric population, even in tertiary care centers,[5] and echo the differences between essential and complex ASDs described by Miles et al.[33] Thus, although these co-occurrence patterns may occur for many reasons (a disease, or its treatment, may make the patient more vulnerable for another; genetic variants may have pleiotropic effects that include autism and other comorbidities; and environmental insults may also have pleiotropic effects), we do not believe these findings are simply incidental.

Subgroup 1

Seizures, present in 77.5% of individuals in subgroup 1, are 1 of the best-known comorbidities of autism. The mutations responsible for Rett syndrome[21] and Fragile X[22] carry a much higher risk of ASD and epilepsy, but the mechanism remains poorly understood.[43] The high rate of intellectual disability (60.0%) in this subgroup is consistent with connections between intellectual disability, epilepsy, and ASD to the ARX (Aristaless related homeobox) gene.[44] More generally, high rates of intellectual disability can act as a proxy for more severe ASDs, and epilepsy is associated with more severe ASDs.[9,10] Subgroup 1 had the lowest proportion of boys, and past studies have revealed the association between epilepsy, intellectual disability, and ASD to be stronger in female patients.[45] Overlap between ASD, epilepsy, and cerebral palsy has also been documented,[46] with recent work pointing to common genetic etiologies.[47]

disorders (Fisher's exact uncorrected $P = .64$) but strong evidence for a correlation between seizures and gastrointestinal disorders (Fisher's exact uncorrected $P < .001$). There was no evidence for a difference between the statistics of the 2 samples (χ^2 simulated P value with Yates correction
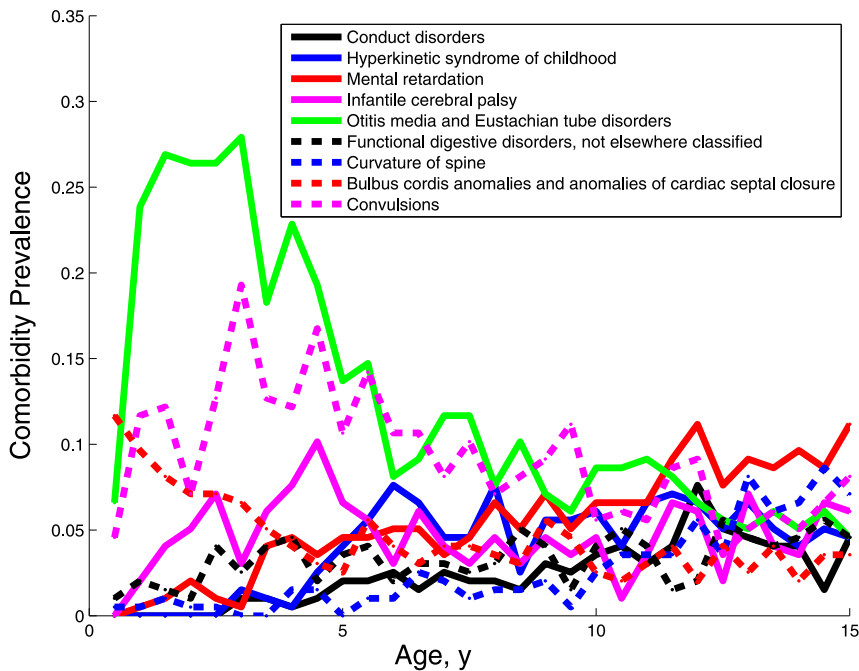
FIGURE 5
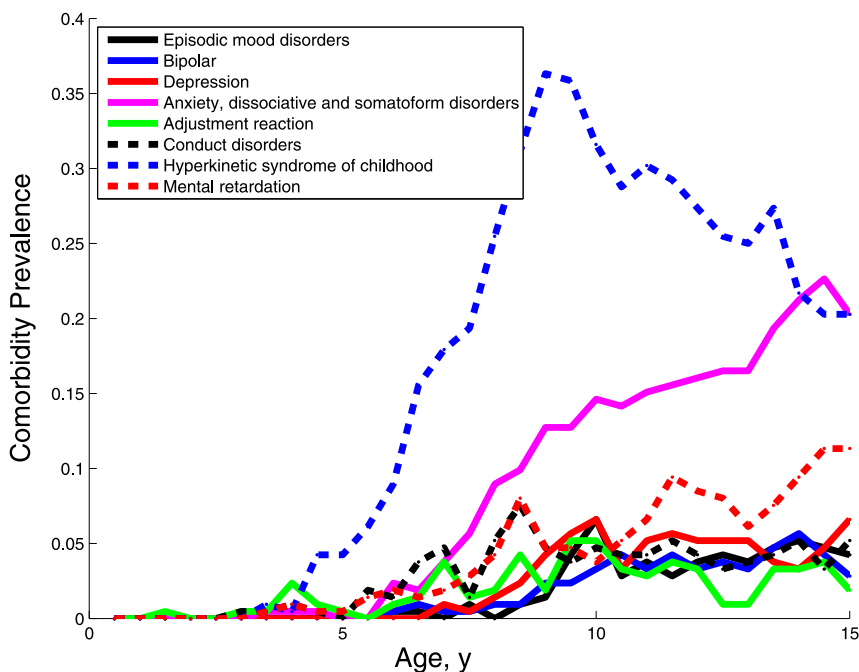Prevalence of the diagnoses characterizing subgroup 2.

FIGURE 6
Prevalence of the diagnoses characterizing subgroup 3 over the first 15 years of life._

Subgroup 2

With an intellectual disability rate of 48.7%, subgroup 2 was also characterized by relatively severe ASDs. Several correlations found in subgroup 2 have been previously reported in autism, including cardiac and auditory disorders,[48] asthma and other autoimmune disorders,[49] and congenital anomalies involving the ear, eye, and cranial nerve.[50] Konstantareas and Homatidis[51] noted that the severity of autistic features was correlated with ear infections; higher rates of hearing loss have also been observed in the overall sample.[52]

The complex multisystem interactions in this group suggest a different etiology than in subgroup 1. Multiple studies of autism have demonstrated abnormal chemokine responses to toll-like receptor ligands[53,54] and abnormal natural killer cell response to stimulation.[55] These might contribute to an abnormal immune response to infection and manifest itself as increased ear infections. The association between peripartum infections and ASD may be a manifestation of the same pleiotropy.[23–25] Mutations in the CaV1.2 channel[56] have been associated with a combination of autism, bipolar disorder, cardiac disorders, and immunologic disorders. However, more refinement will be needed to understand these interactions.

Subgroup 3

Subgroup 3 had the highest rate of individuals with Asperger syndrome and the lowest rate of intellectual disability (27.8%) among the 3 high-morbidity subgroups. The comorbidities in this group were largely psychiatric disorders (especially anxiety). High functioning children with autism often suffer more from anxiety than their normally developing counterparts,[57] and correlations have been found between higher-functioning individuals with ASD and bipolar disorder.[58] A growing number of studies have revealed mutations common to autism, attention-deficit/hyperactivity disorder, and schizophrenia.[17–20]

The only nonpsychiatric comorbidities associated with subgroup 3 were asthma and cardiac dysrhythmia.* Known

---

*In the PheWAS[42] clustering, cardiac dysrhythmias include ventricular flutter, fibrillation, and premature beats; atrial fibrillation and flutter and tachycardias are not included.

associations with cardiac dysrhythmias and autism include velocardiofacial syndrome and chromosomal disorders of 22q11. Velocardiofacial syndrome is also associated with a variety of affective and neuropsychiatric disorders besides ASD.[59] There also exist associations between heart disease and both anxiety[60,61] and major depression,[62] as well as the drugs used to treat them.[63,64] The connections between the nonpsychiatric and the psychiatric disorders suggest that psychiatric disorders are the main characteristic of this subgroup.

The 3 high-morbidity subgroups had distinct pathophysiologies; the heterogeneity of the ASD comorbidity spectrum mirrors the heterogeneity observed in genome-wide studies of variants associated with ASD[65–67] and gene expression.[68] How our pathophysiological subgroups map to genome-scale heterogeneity remains unknown, but our subgroups suggest a group structure worth investigating with these molecular measures. Specifically, analyzing individuals with ASD as a single group may have blurred the different etiologies responsible for this heterogeneous disease. The study presented here may support the exploration of the underlying distinct pathobiologies of children with ASD by providing subpopulations, which are enriched for these distinct mechanisms. If these subpopulations have homogeneous etiological mechanisms, then specifically targeted therapies or preventive measures can be evaluated.

Limitations

Although many of the observed co-occurrence patterns are supported in the literature, our analysis can only be considered preliminary because of our reliance on ICD-9 codes and intermittent hospital visits. Originally developed for billing, one cannot distinguish between diseases and symptom-complexes just from these codes. In particular, 364 individuals in our sample had diagnoses for ASD and cerebral palsy or blindness, which makes the diagnosis of ASD challenging. Without inspecting the clinical notes for these patients, we cannot be sure if these individuals had one, both, or none of these conditions.

We also have no information about conditions that were not diagnosed at the particular hospital. The mean ages of the first ASD code in our study is high (between 7.6 and 9.5 years) for a disorder that is usually diagnosed in early childhood. Thus, these individuals were likely diagnosed, and treated, elsewhere before coming to the tertiary-care hospital. Finally, the reported conditions are biased by the specialists available at the hospital; the absence of a condition may only indicate that the patient sought treatment of that condition elsewhere.

Finally, from a technological perspective, our analyses are easy to replicate in other hospital samples (as we did) because we rely only on ICD-9 codes recorded in EHRs. However, extracting data from health care systems does require that the institution invest in extracting and standardizing the data from their EHR system. Although the i2b2 infrastructure[41] is a free and open source platform used by over 100 academic health centers, each of these centers has invested in managing the data extraction process.[69] This investment is typically $< 1\%$ of the cost of implementing an EHR, but that portion remains a large number in absolute terms.

CONCLUSIONS

Three distinct medical trajectories were identified by unsupervised clustering of EHR diagnoses of individuals with ASD. Our analysis confirms the heterogeneity of the ASD, now in the landscape of comorbidities. Each of these subgroups averaged more diagnostic codes in the first 15 years of life than the remaining overall sample. The first subgroup was characterized by seizures, the second by multisystem disorders, and the third by psychiatric disorders. Each of these groups may point to distinct etiologies with different genetic and environmental contributions.

Although preliminary, our study provides guidance for the expensive prospective, longitudinal studies that would be needed to validate these findings. By providing hypotheses of groups to follow, our work may help target recruitment efforts and focus analysis objectives. Meanwhile, further refinement of these categories by using additional clinical and molecular characterizations, as well as more sophisticated time-series analysis techniques, will undoubtedly recover finer patterns in the clinical trajectories of ASD.

REFERENCES

1. Xue Ming , Brimacombe M, Chaaban J, Zimmerman-Bier B, Wagner GC. Autism spectrum disorders: concurrent clinical disorders. *J Child Neurol*. 2008;23(1):6–13

2. Bauman ML. Medical comorbidities in autism: challenges to diagnosis and treatment. *Neurotherapeutics*. 2010;7(3):320–327

3. Coury D. Medical treatment of autism spectrum disorders. *Curr Opin Neurol*. 2010;23(2):131–136

4. Smith RD. Abnormal head circumference in learning-disabled children. *Dev Med Child Neurol*. 1981;23(5):626–632

5. Kohane IS, McMurry A, Weber G, et al. The co-morbidity burden of children and young adults with autism spectrum disorders. *PLoS ONE*. 2012;7(4):e33224

6. Baio J; Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators; Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveill Summ*. 2012;61(3):1–19

7. Horvath K, Papadimitriou J, Rabsztyn A, Drachenberg C, Tildon J. Gastrointestinal abnormalities in children with autistic disorder. *J Pediatr*. 1999;135(5):559–563

8. Horvath K, Perman JA. Autistic disorder and gastrointestinal disease. *Curr Opin Pediatr*. 2002;14(5):583–587

9. Mouridsen SE, Rich B, Isager T. Epilepsy in disintegrative psychosis and infantile autism: a long-term validation study. *Dev Med Child Neurol*. 1999;41(2):110–114

10. Tuchman R, Rapin I. Epilepsy in autism. *Lancet Neurol*. 2002;1:352–358

11. Richdale AL, Schreck KA. Sleep problems in autism spectrum disorders: Prevalence, nature, and possible biopsychosocial etiologies. *Sleep Med Rev*. 2009;13(6):403–411

12. Wu JY, Kuban KCK, Allred E, Shapiro F, Darras BT. Association of Duchenne muscular dystrophy with autism spectrum disorder. *J Child Neurol*. 2005;20(10):790–795

13. Hendriksen JGM, Vles JSH. Neuropsychiatric disorders in males with Duchenne muscular dystrophy: frequency rate of attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder, and obsessive-compulsive disorder. *J Child Neurol*. 2008;23(5):477–481

14. Hinton VJ, Cyrulnik SE, Fee RJ, Batchelder A, Kiefel JM. Association of autistic spectrum disorders with dystrophinopathies. *Pediatr Neurol*. 2009;41(5):339–346

15. Morgan CN, Roy M, Chance P. Psychiatric comorbidity and medication use in autism: a community survey. *Psychiatr Bulletin*. 2003;27:378–381

16. State MW, Šestan N. Neuroscience. The emerging biology of autism spectrum disorders. *Science*. 2012;337(6100):1301–1303

17. Ingason A, Rujescu D, Cichon S, et al; GROUP Investigators. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry*. 2011;16(1):17–25

18. Smoller JW, Craddock N, Kendler K, et al; Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 2013;381(9875):1371–1379

19. Tabarés-Seisdedos R, Rubenstein JL. Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: implications for schizophrenia, autism and cancer. *Mol Psychiatry*. 2009;14(6):563–589

20. Murdoch JD, State MW. Recent developments in the genetics of autism spectrum disorders. *Curr Opin Genet Dev*. 2013;23(3):310–315

21. Glaze DG, Percy AK, Skinner S, et al. Epilepsy and the natural history of Rett syndrome. *Neurology*. 2010;74(11):909–912

22. Bear MF, Huber KM, Warren ST. The mGluR theory of fragile X mental retardation. *Trends Neurosci*. 2004;27(7):370–377

23. Garbett KA, Hsiao EY, Kalman S, Patterson PH, Mirnics K. Effects of maternal immune activation on gene expression patterns in the fetal brain. *Transl Psychiatry*. 2012;2(4):e98

24. Hagberg H, Gressens P, Mallard C. Inflammation during fetal and neonatal life: implications for neurologic and neuropsychiatric disease in children and adults. *Ann Neurol*. 2012;71(4):444–457

25. Atladóttir HO, Thorsen P, Østergaard L, et al. Maternal infection requiring hospitalization during pregnancy and autism spectrum disorders. *J Autism Dev Disord*. 2010;40(12):1423–1430

26. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395–405

27. Roque FS, Jensen PB, Schmock H, et al. Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLOS Comput Biol*. 2011;7(8):e1002141

28. Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet*. 2011;12(6):417–428

29. Cao H, Melton GB, Markatou M, Hripcsak G. Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases. *J Biomed Inform*. 2008;41(6):882–888

30. Melton GB, Parsons S, Morrison FP, Rothschild AS, Markatou M, Hripcsak G. Inter-patient distance metrics using SNOMED CT defining relationships. *J Biomed Inform*. 2006;39(6):697–705

31. Hivert MF, Grant RW, Shrader P, Meigs JB. Identifying primary care patients at risk for future diabetes and cardiovascular disease using electronic health records. *BMC Health Serv Res*. 2009;9:170

32. Carney CP, Jones LE. Medical comorbidity in women and men with bipolar disorders: a population-based controlled study. *Psychosom Med*. 2006;68(5):684–691

33. Miles JH, Takahashi TN, Bagby S, et al. Essential versus complex autism: definition of fundamental prognostic subtypes. *Am J Med Genet A*. 2005;135(2):171–180

34. Wiggins LD, Robins DL, Adamson LB, Bakeman R, Henrich CC. Support for a dimensional view of autism spectrum disorders in toddlers. *J Autism Dev Disord*. 2012;42(2):191–200

35. Lane AE, Young RL, Baker AE, Angley MT. Sensory processing subtypes in autism: association with adaptive behavior. *J Autism Dev Disord*. 2010;40(1):112–122

36. Wing L, Gould J. Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *J Autism Dev Disord*. 1979;9(1):11–29

37. Ben-Sasson A, Cermak SA, Orsmond GI, Tager-Flusberg H, Kadlec MB, Carter AS. Sensory clusters of toddlers with autism spectrum disorders: differences in affective symptoms. *J Child Psychol Psychiatry*. 2008;49(8):817–825

38. Bitsika V, Sharpley CF, Orapeleng S. An exploratory analysis of the use of cognitive, adaptive and behavioural indices for cluster analysis of ASD subgroups. *J Intellect Disabil Res*. 2008;52(11):973–985

39. Hu VW, Steinberg ME. Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. *Autism Res*. 2009;2(2):67–77

40. Sacco R, Lenti C, Saccani M, et al. Cluster analysis of autistic patients based on principal pathogenetic components. *Autism Res*. 2012;5(2):137–147

41. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc*. 2010;17(2):124–130

42. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010; 26(9):1205–1210

43. Tuchman R, Moshé SL, Rapin I. Convulsing toward the pathophysiology of autism. *Brain Dev*. 2009;31(2):95–103

44. Sherr EH. The ARX story (epilepsy, mental retardation, autism, and cerebral malformations): one gene leads to many phenotypes. *Curr Opin Pediatr*. 2003;15(6):567–571

45. Danielsson S, Gillberg IC, Billstedt E, Gillberg C, Olsson I. Epilepsy in young adults with autism: a prospective population-based follow-up study of 120 individuals diagnosed in childhood. *Epilepsia*. 2005;46(6):918–923

46. Surén P, Bakken IJ, Aase H, et al. Autism spectrum disorder, ADHD, epilepsy, and cerebral palsy in Norwegian children. *Pediatrics*. 2012;130(1). Available at: www.pediatrics.org/cgi/content/full/130/1/e152

47. Talkowski ME, Maussion G, Crapper L, et al. Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities. *Am J Hum Genet*. 2012;91(6): 1128–1134

48. Porges SW, Macellaio M, Stanfill SD, et al Respiratory sinus arrhythmia and auditory processing in autism: modifiable deficits of an integrated social engagement system? *Int J Psychophysiol*. 2013;88(3):261–270

49. Becker KG. Autism, asthma, inflammation, and the hygiene hypothesis. *Med Hypotheses*. 2007;69(4):731–740

50. Miller MT, Strömland K, Ventura L, Johansson M, Bandim JM, Gillberg C. Autism associated with conditions characterized by developmental errors in early embryogenesis:

a mini review. *Int J Dev Neurosci*. 2005;23 (2–3):201–219

51. Konstantareas MM, Homatidis S. Ear infections in autistic and normal children. *J Autism Dev Disord*. 1987;17(4):585–594

52. Rosenhall U, Nordin V, Sandström M, Ahlsén G, Gillberg C. Autism and hearing loss. *J Autism Dev Disord*. 1999;29(5):349–357

53. Ashwood P, Krakowiak P, Hertz-Picciotto I, Hansen R, Pessah I, Van de Water J. Elevated plasma cytokines in autism spectrum disorders provide evidence of immune dysfunction and are associated with impaired behavioral outcome. *Brain Behav Immun*. 2011;25(1):40–45

54. Enstrom AM, Onore CE, Van de Water JA, Ashwood P. Differential monocyte responses to TLR ligands in children with autism spectrum disorders. *Brain Behav Immun*. 2010;24(1):64–71

55. Enstrom AM, Lit L, Onore CE, et al. Altered gene expression and function of peripheral blood natural killer cells in children with autism. *Brain Behav Immun*. 2009;23(1): 124–133

56. Liao P, Soong TW. CaV1.2 channelopathies: from arrhythmias to autism, bipolar disorder, and immunodeficiency. *Pflugers Arch*. 2010;460(2):353–359

57. Gillott A, Furniss F, Walter A. Anxiety in high-functioning children with autism. *Autism*. 2001;5(3):277–286

58. DeLong GR, Dwyer JT. Correlation of family history with specific autistic subgroups: Asperger's syndrome and bipolar affective disease. *J Autism Dev Disord*. 1988;18(4): 593–600

59. Fabbro A, Rizzi E, Schneider M, Debbane M, Eliez S. Depression and anxiety disorders in children and adolescents with velo-cardio-facial syndrome (VCFS). *Eur Child Adolesc Psychiatry*. 2012;21(7):379–385

60. Kubzansky LD, Kawachi I, Weiss ST, Sparrow D. Anxiety and coronary heart disease: a synthesis of epidemiological, psychological, and experimental evidence. *Ann Behav Med*. 1998;20(2):47–58

61. Watkins LL, Grossman P, Krishnan R, Sherwood A. Anxiety and vagal control of heart rate. *Psychosom Med*. 1998;60(4):498–502

62. Stapelberg NJ, Hamilton-Craig I, Neumann DL, Shum DH, McConnell H. Mind and heart: heart rate variability in major depressive disorder and coronary heart disease—a review and recommendations. *Aust N Z J Psychiatry*. 2012;46(10):946–957

63. Witchel HJ, Hancox JC, Nutt DJ. Psychotropic drugs, cardiac arrhythmia, and sudden death. *J Clin Psychopharmacol*. 2003;23(1): 58–77

64. Taylor D. Antidepressant drugs and cardiovascular pathology: a clinical overview of effectiveness and safety. *Acta Psychiatr Scand*. 2008;118(6):434–442

65. Liu L, Sabo A, Neale BM, et al Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genet*. 2013;9(4): e1003443

66. Yu TW, Chahrour MH, Coulter ME, et al. Using whole-exome sequencing to identify inherited causes of autism. *Neuron*. 2013;77 (2):259–273

67. Michaelson JJ, Shi Y, Gujral M, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*. 2012;151(7):1431–1442

68. Kong SW, Collins CD, Shimizu-Motohashi Y, et al Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS ONE*. 2012;7(12):e49475

69. Masys DR, Harris PA, Fearn PA, Kohane IS. Designing a public square for research computing. *Sci Transl Med*. 2012;4(149): 49fs32