Clustering of Co-occurring Conditions in Autism Spectrum Disorder during Early Childhood: A Retrospective Analysis of Medical Claims Data

Running title: Co-occurring condition clusters in ASD

Troy Vargason^{1,2,3}, Richard E. Frye^{4,5}, Deborah L. McGuinness^{6,7}, Juergen Hahn^{1,2,8,*}

¹Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

²Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

³OptumLabs Visiting Fellow, OptumLabs, Cambridge, Massachusetts 02142, United States

⁴Department of Child Health, University of Arizona College of Medicine, Phoenix, Arizona 85004, United States

⁵Phoenix Children's Hospital, Phoenix, Arizona 85016

⁶Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

⁷Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

⁸Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

*Corresponding author. Email: hahnj@rpi.edu

Acknowledgements

The authors gratefully acknowledge partial financial support from the National Institutes of Health (grant R01AI110642). Support for this research was also received from the Rensselaer Institute for Data Exploration and Applications. The authors express their gratitude to John Erickson at Rensselaer Polytechnic Institute for supporting the interactions with OptumLabs, and to the staff at OptumLabs for supporting the study design.

Lay Summary

Medical conditions that co-occur with autism spectrum disorder (ASD) vary significantly from person to person. This study analyzed patterns in diagnosis of co-occurring conditions from medical claims data and observed three subtypes of children with ASD. These results may aid with screening for co-occurring conditions in children with ASD and with understanding ASD subtypes.

Abstract

Individuals with autism spectrum disorder (ASD) are frequently affected by co-occurring medical conditions (COCs), which vary in severity, age of onset, and pathophysiological characteristics. The presence of COCs contributes to significant heterogeneity in the clinical presentation of ASD between individuals and a better understanding of COCs may offer greater insight into the etiology of ASD in specific subgroups while also providing guidance for diagnostic and treatment protocols. This study retrospectively analyzed medical claims data from a private United States health plan between years 2000–2015 to investigate patterns of COC diagnoses in a cohort of 3,278 children with ASD throughout their first five years of enrollment compared to 279,693 children

from the general population without ASD diagnoses (POP cohort). Three subgroups of children with ASD were identified by *k*-means clustering using these COC patterns. The first cluster was characterized by generally high rates of COC diagnosis and comprised 23.7% ($n = 776$) of the cohort. Diagnoses of developmental delays were dominant in the second cluster containing 26.5% ($n = 870$) of the cohort. Children in the third cluster, making up 49.8% ($n = 1,632$) of the cohort, had the lowest rates of COC diagnosis, which were slightly higher than rates observed in the POP cohort. A secondary analysis using these data found that GI and immune disorders showed similar longitudinal patterns of prevalence, as did seizure and sleep disorders. These findings may help to better inform the development of diagnostic workup and treatment protocols for COCs in children with ASD.

Keywords

autism spectrum disorder, co-occurring condition, comorbidity, *k*-means clustering, medical claims, retrospective analysis

Introduction

The diagnostic criterion for autism spectrum disorder (ASD) is based upon behavioral symptoms that include a combination of social-communication deficits occurring along with repetitive and restrictive behaviors and interests (American Psychiatric Association, 2013). Despite the fact that ASD is defined by behavior, it is now recognized as being associated with a large number of co-occurring medical conditions (COCs) beyond the core behavioral symptoms (Tye, Runicles, Whitehouse, & Alvares, 2019). A recent study from the Autism and Developmental Disabilities Monitoring Network found over 95% of children with ASD to be affected by at least one COC (Soke, Maenner, Christensen, Kurzius-Spencer, & Schieve, 2018). While the economic costs of caring for a child with ASD are already elevated compared to those for a typically developing child, these costs are further increased by the presence of COCs (Buescher, Cidav, Knapp, & Mandell, 2014; Horlin, Falkmer, Parsons, Albrecht, & Falkmer, 2014; Lavelle et al., 2014). Rising numbers of diagnosed COCs in children with ASD (Rubenstein et al., 2018) necessitate further understanding of COCs' etiologies and how they relate to ASD pathophysiology.

Conditions that co-occur with ASD have diverse pathophysiological characteristics. Studies have found gastrointestinal (GI) symptoms (Holingue, Newill, Lee, Pasricha, & Daniele Fallin, 2018), sleep disorders (Díaz-Román, Zhang, Delorme, Beggiato, & Cortese, 2018), immune-mediated disorders (H. Hughes, Mills Ko, Rose, & Ashwood, 2018), epilepsy (Woolfenden, Sarkozy, Ridley, Coory, & Williams, 2012), and many other conditions (Kohane et al., 2012; Mannion, Leader, & Healy, 2013; Muskens, Velders, & Staal, 2017; Rydzewska et al., 2018) to affect substantial numbers of individuals with ASD. Although seemingly distinct, some studies have suggested that certain COCs tend to co-occur together (Aldinger, Lane, Veenstra-VanderWeele, & Levitt, 2015; Fulceri et al., 2016; Hirata et al., 2016). For example, sleep and seizure disorders may be closely linked (Accardo & Malow, 2015), while GI and immune-mediated symptoms may also be physiologically connected (Rose et al., 2018). Increased liability for certain conditions may arise from genetic variants shared with ASD (Ingason et al., 2011; Murdoch & State, 2013; Smoller et al., 2013; Diaz-Beltran et al., 2017) or from perturbations of metabolic pathways held in common (Frye, 2015; Cheng, Rho, & Masino, 2017).

Addressing COCs can be just as important as the underlying diagnosis of ASD. For example, sleep disruption can decrease the quality of life for both the child and the parent (Tilford et al., 2015). Most importantly, early morbidity and mortality in individuals with ASD may be related to COCs (Hirvikoski et al., 2016), particularly epilepsy (Pickett, Xiu, Tuchman, Dawson, & Lajonchere, 2011) and psychiatric conditions (Hirvikoski et al., 2016). Certain COCs including psychiatric conditions (Spain, Sin, Chalder, Murphy, & Happé, 2015), epilepsy (Hirota, Veenstra-VanderWeele, Hollander, & Kishi, 2014), and sleep disorders (Souders et al., 2017) may be treatable, however, especially if they are recognized early. This highlights the need for timely diagnosis of COCs, which is made difficult by substantial variation in their occurrence with age. One study reported a higher prevalence of most evaluated COCs in 8-year-old children with ASD compared to 4-year-olds (Soke et al., 2018). Changes in prevalence of GI symptoms and sleep problems over two years of follow-up have also been observed in individuals with ASD (Mannion & Leader, 2016), although it is unclear whether this difference in prevalence simply results from difficulty diagnosing COCs early in life. For example, many GI disturbances such as gastroesophageal reflux and eosinophilic esophagitis are most likely under-recognized because they present with behavioral symptoms that can be confused with core symptoms of ASD (Buie et al., 2010).

Since many COCs can present with atypical symptoms in children with ASD and may even present with ASD-related symptoms that are considered to be core symptoms, it is important that medical practitioners who manage individuals with ASD have a high index-of-suspicion for these conditions. Thus, developing standards for COC screening can aid with detection, with one study finding screening to be effective for identifying psychiatric disorders in adults with ASD (Findon et al., 2016) and a more recent study developing a brief screening tool for common GI disorders in individuals with ASD (Margolis et al., 2019). Detecting and treating COCs can have a significant impact on the clinical course of the child; for example, improving sleep can improve the quality of life of the whole family (Tilford et al., 2015) and improve behavioral problems during the day (Rossignol & Frye, 2011). It is essential to better understand COCs so they can be detected early using targeted screening protocols based on specific patient history to provide optimal treatment for individuals with ASD.

Differences in incidence and severity of COCs may contribute to significant heterogeneity in ASD's clinical manifestations. There is thus considerable variation in the age that ASD symptoms are first recognized and, consequently, great non-uniformity in the timing of ASD diagnosis (Zwaigenbaum et al., 2015). A study of children from the Autism and Developmental Disabilities Monitoring Network found the median age of ASD diagnosis in the United States to be 52 months (Baio et al., 2018), but the fact that the presence of specific COCs can inadvertently increase or decrease the age at which ASD is first diagnosed (Soke et al., 2018) suggests that the relationships between COCs and ASD are still poorly understood. Demographic and socioeconomic factors can also influence the timing of diagnosis, although these account for a relatively small portion of the variation (Herlihy et al., 2014; Emerson, Morrell, & Neece, 2016). An improved understanding of COC trajectories in young children may hold potential for more informed and consistently-timed ASD diagnoses.

Understanding how COCs relate to ASD diagnosis and symptoms can also help with understanding the heterogeneity of ASD by potentially categorizing affected individuals into relatively

homogeneous subtypes. While most studies identify subtypes of participants based upon criteria other than COCs, a few incorporate COCs as the primary criteria for clustering individuals with ASD. For example, Sacco et al. (2012) performed a cluster analysis of questionnaire data describing circadian, immune, and developmental abnormalities, along with certain core symptoms, in a clinical sample of individuals with ASD aged 2–30 years old. Electronic health records from a tertiary-care pediatric hospital were analyzed with hierarchical clustering by Doshi-Velez, Ge, and Kohane (2014) to group individuals based on temporal patterns in COC diagnoses from birth through 15 years of age. Lingren et al. (2016) examined COC diagnoses in electronic health records from three clinical sites and used cluster analysis to identify subtypes of ASD in a pediatric population. Clustering based on medical conditions diagnosed prior to ASD diagnosis, including COCs, was also performed by Alexeeff et al. (2017) to predict future ASD diagnosis risk from electronic health records and medical claims of health plan members aged 2–12 years old. There is thus great potential for using recorded healthcare data to study COC diagnosis patterns and characterize possible subtypes of ASD from these diagnoses. It would be of significant interest to focus these analyses on young children for whom complete medical histories are available (i.e., not just data from a few sites), allowing for more comprehensive understanding of COC diagnosis patterns during early childhood.

We perform here a retrospective analysis of medical claims data to study COC diagnoses in individuals with ASD enrolled in a private United States health plan. Specifically, diagnosis codes for young children with ASD are collected throughout the first five years of health plan enrollment and temporal patterns of COC diagnosis are used to identify distinct clusters of these children. In addition to including a large and diverse cohort, the strengths of this study come from its focus on early childhood and having comprehensive medical diagnosis histories for this critical period of development. This analysis aims to provide a greater understanding of COCs contributing to ASD's various clinical manifestations, an understanding that could aid development of better-informed screening for COCs in children with ASD.

Methods

Claims data

This study involves a retrospective analysis of claims data from the OptumLabs® Data Warehouse (OLDW), which includes de-identified claims data for privately insured and Medicare Advantage enrollees in a large, private, U.S. health plan. The database contains longitudinal health information on enrollees, representing a diverse mixture of ages, ethnicities, and geographical regions across the United States. The health plan provides comprehensive full insurance coverage for physician, hospital, and prescription drug services (OptumLabs, 2018). Approval from the Institutional Review Board at Rensselaer Polytechnic Institute was not required for this study since all data were pre-existing and de-identified. Formal consent was also not required for this type of study.

Cohort definitions

Inclusion and exclusion criteria described in a previous study by the authors (Vargason, McGuinness, & Hahn, 2019) were used to identify a cohort of 3,278 privately insured children with ASD and a cohort of 279,693 privately insured children from the general population without an ASD diagnosis (POP cohort). The total study period spanned from January 1, 2000 to September 30, 2015, and all included children were born between years 2000–2010 with their first

day of health plan enrollment (index date) being within one year of their birth year. At least five years of continuous medical, pharmacy, and mental health coverage in the health plan beyond the index date were required, with gaps in coverage of 45 days or fewer counting towards continuous enrollment (Figure 1). To ensure that all diagnosis codes were assigned according to *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) coding, continuous enrollment criteria had to be satisfied before September 30, 2015.

Children were assigned to the ASD cohort if they had at least two medical claims, on separate dates, with an ICD-9-CM code for autistic disorder (code 299.0x), Asperger syndrome (299.8x), or an unspecified pervasive developmental disorder (299.9x), and no claims containing diagnoses for childhood disintegrative disorder (299.1x) or Rett syndrome (330.8x); these criteria have previously yielded high positive predictive value for identifying ASD from medical records (Burke et al., 2013; Coleman et al., 2015). Only ASD diagnoses made at least two years after the birth year were considered, as perceived developmental deficits before this time may not necessarily be ASD-related. Furthermore, to ensure that all ASD diagnoses were made according to criteria defined in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (American Psychiatric Association, 2000), any diagnoses made after the year 2012 were ignored. The POP cohort consisted of children with no claims for any aforementioned ASD diagnoses made at least two years after the birth year.

Clustering of individuals with ASD

Seven categories of conditions potentially co-occurring with ASD were defined (Table 1). COCs were categorized by auditory disorders, developmental delays (DDs), GI symptoms, immune-related conditions, psychiatric disorders, seizure disorders, and sleep disorders. For each individual, a time-series consisting of ten six-month windows was constructed to span for five years beyond the index date. The diagnosis of a COC category in a particular window for an individual was indicated by a binary 0/1 flag, yielding a vector of length 70 per individual (seven categories across ten windows). Principal component analysis was then applied to the mean-centered ASD data set to transform it from a series of binary variables to continuous variables (70 principal components) prior to cluster analysis.

To identify subgroups of children with ASD, *k*-means clustering with the Hartigan-Wong algorithm (Hartigan & Wong, 1979) was used. This algorithm inputs a data matrix containing one row for each member/sample of the ASD cohort and one column for each of the 70 principal components. Initial centers for a pre-specified number of clusters *k* must also be provided and each sample is first assigned to the cluster whose center is nearest. Cluster centers are then updated and the algorithm proceeds by iteratively reassigning points to the nearest cluster and updating cluster centers until the within-cluster sum of squares is minimized. Since the most meaningful number of ASD clusters is unknown, *k* was varied from 1 to 5; for each number of clusters, the initial cluster centers were randomly chosen for 100 trials and the result yielding the minimum within-cluster sum of squares was accepted as the clustering result for that value of *k*. The average values of the silhouette coefficient (Rousseeuw, 1987) over all data points for each value of *k* were used to select an appropriate final number of clusters.

Clustering of co-occurring conditions via temporal trajectories

Our secondary analysis aims to study apparent temporal differences in COC prevalence profiles between the ASD and POP cohorts, and how these differences might be similar between COCs in ASD. For this analysis, a time-series of ten six-month windows (from the index date to five years beyond) was created for each COC category to describe the percentage of individuals in the ASD cohort with diagnoses in that category during the specified period; the same procedure was then repeated for the POP cohort. The difference in prevalence between the cohorts for each COC category during each window was taken to represent the increased COC prevalence for the ASD cohort. *k*-means clustering was used to cluster the seven COC categories based on these temporal trajectories of increased prevalence in the ASD cohort, with *k* varying between 1 and 6. Since the silhouette coefficient is not meaningful when only seven objects are to be clustered, this analysis relied on qualitative comparison of COC category patterns for determining the final clusters.

Results

Cluster analysis yielded three ASD clusters. These clusters' demographic characteristics are compared with those of the ASD and POP cohorts in Table 2. The first and second clusters were of comparable sizes, constituting 23.7% and 26.5% of the ASD cohort, respectively. The third and largest cluster comprised 49.8% of the cohort. The proportion of males was greatest in the first cluster, but gender ratios were overall similar to that in the total ASD cohort (4.41 males per female). The distributions of race/ethnicity and census divisions only modestly shifted between clusters. Members of the ASD cohort were affected, on average, by 65% more COC categories than members of the POP cohort; this number was also considerably higher in the first cluster (the high-prevalence cluster) and lower in the third cluster (the low-prevalence cluster), with the second cluster (the mid-prevalence cluster) having a number most similar to the total ASD cohort. The POP cohort's mean number of diagnosed COC categories was most similar to, although still less than, that of the low-prevalence cluster.

All COC categories were more prevalent in the ASD cohort compared to the POP cohort across the five-year enrollment period (Figure 2A). Prevalence was considerably higher in the ASD cohort for DDs (74.1% to 8.4%), psychiatric disorders (33.2% to 9.9%), seizures (17.4% to 2.9%), and sleep disorders (20.0% to 8.6%), while differences in prevalence between the two groups were smaller, but still significant, for auditory (>93.2% to 86.3%), GI (67.8% to 50.7%), and immune-related (>73.3% to 62.8%) disorders. Among the clusters, the high-prevalence cluster had the highest prevalence in every category except DDs and seizures, which were most prevalent in the mid-prevalence cluster (Figure 2B); all individuals in the mid-prevalence cluster had at least one DD diagnosis. Other than these two categories, the prevalence patterns of the mid-prevalence cluster closely resembled those of the total ASD cohort. The low-prevalence cluster had the lowest prevalence in every category. The prevalence of auditory, GI, and immune-related disorders in this cluster were similar to those in the POP cohort, but the remaining categories were notably more prevalent relative to the POP cohort. No notable differences in prevalence between genders were observed in either cohort.

Temporal trends of COC diagnosis in the ASD and POP cohorts were clearly distinguishable (Figure 3). Prevalence in the ASD cohort was higher than in the POP cohort for every COC category during almost every six-month window, with the exception of auditory disorders where POP prevalence was initially higher. The trajectories' shapes were most similar between the two cohorts (albeit at higher levels in the ASD cohort) for auditory, GI, and immune-related conditions,

but trended towards much higher prevalence with time in the ASD cohort for the remaining categories. Each ASD cluster also had distinct temporal trends. The high-prevalence cluster had consistently higher rates of diagnosis during each six-month window for all categories except DDs and seizure disorders. The trajectories of the mid-prevalence cluster were generally most similar to those of the total ASD cohort for all categories except DDs and seizure disorders, for which the mid-prevalence cluster had significantly higher diagnosis rates. The low-prevalence cluster was characterized by the lowest rates of COC diagnosis throughout the enrollment period; trajectories in this cluster still indicated generally higher prevalence over time compared to the POP cohort for most COC categories (DD, GI, psychiatric, seizure, sleep), although prevalence was also lower than the POP cohort for auditory and immune-related conditions for most six-month windows.

Members of the ASD cohort more frequently had diagnoses from a greater number of COC categories than members of the POP cohort (Figure 4). Although the percentages of individuals with at least one COC category diagnosis were similar between the ASD (99.1%) and POP (94.6%) cohorts, the percentage of children with a diagnosis in at least four COC categories was nearly 4.5-fold greater in the ASD cohort (58.2%) than in the POP cohort (13.0%). All members of the high- and mid-prevalence clusters had at least one diagnosed COC, and only a small percentage of the low-prevalence cluster (1.8%) did not. Children in the high-prevalence cluster tended to have diagnoses from slightly more COC categories than children in the mid-prevalence cluster, but children in both of these clusters were diagnosed with a greater number of COC categories much more frequently than children in the low-prevalence cluster. Despite having the lowest overall rates of COC diagnosis among the ASD clusters, the low-prevalence cluster featured a significantly greater number of diagnosed COC categories per individual than the POP cohort.

Patterns in timing of first ASD diagnosis varied by cluster (Figure 5). The median time of diagnosis in the ASD cohort was 1,256 days after the index date (25th/75th percentile 902/1,800 days). This was similar to the distribution for the high-prevalence cluster (median 1,210 days; 25th/75th percentile 871/1,716 days). The mid-prevalence cluster had the earliest median diagnosis at 1,054 days (25th/75th percentile 826/1,405 days), while the low-prevalence cluster had the latest diagnoses on average (median 1,432 days, 25th/75th percentile 996/2,046 days). It should be noted that these distributions are not limited to diagnoses made at least two years after individuals' birth years, although this was a requirement for initially identifying individuals in the ASD cohort.

Clustering of the seven COC categories yielded two pairs of categories (i.e., five clusters, two containing multiple categories) following similar trends of additional prevalence in the ASD cohort (Figure 6). GI and immune-related disorders were found to both have a drop in additional prevalence during the first year after the index date (5.5% to 3.4% and 4.8% to 4.5%, respectively), followed by an overall increase until three-and-a-half years (immune, 8.2%) or four years (GI, 8.1%) after the index date. Seizure and sleep disorders also clustered together, starting at 1.4% and 2.0% additional prevalence, respectively, and reaching a maximum (5.1% and 3.2%, respectively) around four years after the index date. Auditory disorders, DDs, and psychiatric disorders followed distinct patterns of additional prevalence and did not cluster with other categories.

Discussion

All COC categories were more prevalent in the ASD cohort than in the POP cohort, agreeing with previous studies that have observed an increased prevalence of COCs in ASD (Muskens et al.,

2017). However, our estimated proportion of children with ASD affected by at least one COC (99.1%) is higher than other studies' estimates; for example, Soke et al. (2018), Wiggins et al. (2015), Lundström et al. (2015), and Levy et al. (2010) found approximately 95%, 81%, 96%, and 83% of children with ASD to be affected, respectively. Our observed COC prevalence in the POP cohort is also higher than what would be expected given previously-reported differences between children with and without ASD. This discrepancy may arise from differences in defining what constitutes a COC, specifically our inclusion of conditions likely to affect children regardless of ASD status, such as ear infections and allergies. The estimate of children affected by at least one COC may thus be inflated, and a better indicator of relative COC prevalence is given by the number of children affected by four or more COC categories (58.2% ASD versus 13.0% POP). Additionally, many children with ASD without COCs may never be coded for ASD in their medical charts, as ASD is not uncommonly diagnosed in the educational setting where behavioral and educational therapies can be provided with a doctor's prescription. Thus, unless the child engaged the medical system multiple times (at least twice as part of the inclusion criteria), they may not have been included in this cohort. Furthermore, many subspecialists may not code for ASD, but rather code for the medical disorder being treated, especially if the child's ASD symptoms are mild. This may bias the inclusion of more severe children with ASD in this cohort.

The seven COC categories were defined based on ICD-9-CM codes used in previous ASD clustering studies (Doshi-Velez et al., 2014; Lingren et al., 2016) and in other studies of COC prevalence (L. J. Meltzer, Johnson, Crosette, Ramos, & Mindell, 2010; Zerbo et al., 2015). Our clustering analysis is also similar to that performed by other studies clustering individuals with ASD using COC diagnosis codes. Doshi-Velez et al. (2014) used hierarchical clustering to identify four ASD clusters characterized by (1) seizures, (2) auditory and GI disorders, (3) psychiatric disorders, and (4) low COC prevalence. The analysis by Lingren et al. (2016) yielded four clusters describing high rates of seizures, high rates of psychiatric disorders, high rates of DDs, and lower overall COC prevalence. The mid-prevalence cluster in our study, which was closest to the ASD cohort average, was also characterized by a high prevalence of DDs and the low-prevalence cluster was a low overall prevalence group, which is generally consistent with the patterns found in these studies. Our high-prevalence cluster, however, featured high prevalence of most COC categories and was not specific to any particular category. Identifying a greater number of clusters would yield groups more closely associated with specific categories, although at the expense of seeing fewer clinically meaningful differences in temporal trajectories between clusters. It is worth noting that these studies featured validation sets while ours does not, limiting our ability to say whether these results would generalize to other study samples.

It is not surprising that there were ascertainable differences in the clusters' distributions for time of first ASD diagnosis. What is surprising is that the mid-prevalence cluster tended to be diagnosed earlier than the high-prevalence cluster, as it was expected that a greater number of COCs would make ASD more easily recognized. Clearly, considering only the number of COCs does not explain the whole story and the specific COCs diagnosed likely influence ASD diagnosis timing. In their study, Soke et al. (2018) found that certain COCs affected the average age of ASD diagnoses in children, with developmental regression/disabilities lowering the diagnosis age and psychiatric disorders delaying the diagnosis. These findings could explain the timing of diagnoses for the high- and mid-prevalence clusters, as the high-prevalence cluster had higher rates of psychiatric disorders and lower rates of DDs compared to the mid-prevalence cluster, which points

towards the high-prevalence cluster having the later diagnoses. It is also worth noting that DD diagnoses may have been used as placeholders for ASD and do not necessarily represent a true underlying COC. Although it is difficult to deduce this with claims data, these results should be interpreted with this point in mind.

Our finding that immune disorders followed a temporal pattern of prevalence similar to GI conditions is not unexpected, as the current understanding of gut activity (and the gut microbiome) is that it can modulate immune system activity (Vuong & Hsiao, 2017). There is evidence for perturbed gut microbiome homeostasis in ASD (H. K. Hughes, Rose, & Ashwood, 2018) and it is possible this dysbiosis is associated with altered immune function, which is also believed to influence ASD pathophysiology (A. Meltzer & Water, 2017). The association between sleep disorders and epilepsy is also well-recognized and has been undergoing active investigation to improve the lives of children with epilepsy. Both of these conditions may arise from abnormal neuronal connectivity (Jülich & Sahin, 2014). Individuals with epilepsy commonly have poorer sleep quality and are frequently diagnosed with sleep disorders; at the same time, the threshold for seizures may be lowered in individuals with epilepsy who are sleep-deprived (Accardo & Malow, 2015). Antiepileptic medications may also affect sleep quality (Jülich & Sahin, 2014). Thus, some of the associations found may not be specific to ASD, but may represent the interplay between other common disorders. Some studies have also reported an association between GI and sleep disorders (Aldinger et al., 2015; McCue, Flick, Twyman, & Xian, 2017) that our study did not detect. Since this analysis only examined population-level associations between COCs, we cannot form any specific conclusions about individual-level biological interactions between COCs, although previous studies have indicated the possibility of these interactions.

This study benefits from its large, heterogeneous sample of privately-insured individuals across the United States, whereas previous studies using medical claims or records to cluster ASD based on COCs have generally been restricted to data from a few sites (Doshi-Velez et al., 2014; Lingren et al., 2016). Instead of relying on diagnoses made only when individuals visited a particular site, our study provides a complete snapshot of each child's medical diagnosis history during the first five years of insurance coverage. Furthermore, our study focuses on young children to provide greater insight into how COC presentation and/or diagnosis may vary through early childhood in different subsets of children, which could not be easily obtained using a wider age range. To our knowledge, this study provides the most complete illustration to date of COC diagnosis and clustering patterns in young children with ASD using medical claims data. Although these clustering results require validation in other data sets, they represent a significant addition to the list of potential ASD subtypes that may eventually aid with more targeted COC screening.

Strengths aside, our study is inevitably limited by certain aspects of the data source. For one, ICD-9-CM codes may be misleading in that their presence/absence in a claim may not necessarily indicate that an individual was affected or unaffected by a condition. With ASD diagnoses, for example, it may be difficult to evaluate the validity of diagnoses made without gold-standard behavioral assessments. Doctors' notes, although unavailable in this study, would be useful for validating diagnoses of ASD as well as COCs. Since only an individual's year, and not exact date, of birth is available in the data set, our study analyzes timing of diagnoses from the beginning of health plan enrollment rather than from birth. This introduces uncertainty into the clinical and biological interpretations when comparing diagnosis patterns between individuals; however, our

requirement that the beginning of enrollment be within one year of the birth year is intended to reduce this uncertainty. Furthermore, despite having cohort definitions that were identical to a previous study using the OLDW (Vargason et al., 2019), our sample numbers changed slightly due to the absence of an additional exclusion criterion and updates to OLDW contents. Diagnosis patterns in our study population of privately insured individuals may also not be representative of those in the United States population at large. Since all children analyzed in this study (both ASD and POP) were enrolled with the same insurance provider, they likely had similar access to healthcare providers as far as healthcare coverage is concerned. However, other discrepancies in socioeconomic status, such as differences in access to specialists due to geographic location, may have influenced observed diagnosis patterns and cannot be addressed by our study.

Other limitations arise from our study design. An important limitation is that some conditions we consider (such as certain psychiatric disorders) do not typically manifest during early childhood, so we cannot analyze these conditions as comprehensively as other studies. On top of this, our use of COC categories instead of individual COCs limits our ability to draw conclusions about specific conditions of particular interest. We also did not include an independent sample of individuals to validate our findings, as has been done by other studies (Doshi-Velez et al., 2014; Lingren et al., 2016). Furthermore, our estimated distributions for ASD diagnosis timing are likely right-skewed. Individuals born later in the study period have less time to receive an ASD diagnosis during the defined period, giving them a smaller chance of being included in the ASD cohort. Those with delayed diagnoses would only be included if born earlier in the study period, meaning these individuals are likely under-represented in the study sample compared to the true population distribution. Children born at the end of the sample identification period (close to 2010), in particular, are less likely to be diagnosed with ASD within the window of our study since they would only be approximately two years old at the end of the ASD identification period. That being said, the size of the ASD cohort is only about 1.2% of the size of the POP cohort, so a child with ASD misclassified as belonging to the POP cohort (due to not being diagnosed within the study period) would have minimal influence in the conclusions drawn about the POP cohort, although it would yield a smaller sample size for the ASD cohort.

The choices of statistical methodologies may also have influenced the findings of this study. Since *k*-means clustering relies on the calculation of cluster means to partition samples, it would be inappropriate to use solely binary variables for computing these means. Rather than employ a clustering algorithm employing a distance metric suitable for binary variables, we chose to use principal component analysis on our data to produce a set of continuous variables from the binary data. If principal component analysis had not been used, the results of *k*-means clustering would have been different from those presented here. Additionally, the results of *k*-means clustering are typically dependent on the choices of initial cluster centers, and since the cluster centers are initialized randomly for each run, different runs of *k*-means do not always return the same solution. We aimed to address this by performing 100 trials of *k*-means with random cluster centers (for each tested number of clusters) and then accepted the solution yielding the minimum within-cluster sum of squares. Even with multiple trials, the optimal returned solutions were not always identical and occasionally showed a small number of samples shifting between clusters; however, the overall patterns/conclusions determined by each solution did not change significantly. Additionally, although we chose to use *k*-means clustering to identify subtypes of ASD, other

clustering approaches such as latent class analysis would also have been appropriate for this purpose.

Finally, since ASD is a heterogeneous disorder and the ASD diagnoses of many children in our retrospective study were likely not determined until late in their five-year enrollment periods (or beyond this period), there is little information offered in terms of understanding these children's individual risks of being diagnosed with COCs during early childhood prior to their ASD diagnoses. However, the goal of this study was not necessarily to develop predictive models for diagnoses given a known ASD diagnosis, but rather to help add to the base of knowledge regarding how COC screening should change depending on whether a child has been diagnosed with ASD or not. The population-level trajectories investigated and characterized by our study are important for developing this understanding and may eventually aid with correlating COC diagnoses with early developmental/behavioral milestones. Knowing which COCs are more likely to co-occur with other COCs will also assist with estimating individuals' risks of new COC diagnoses given prior diagnoses of other COCs. Earlier and targeted detection of such conditions will allow for earlier treatment and/or management of their symptoms and contribute to improved quality of life for affected individuals and their families. Since the general population without an ASD diagnosis is also very heterogeneous, future work could investigate subtypes of COC diagnoses in this population to further understand their clinical trajectories.

Conflict of Interest

The authors have no conflicts of interest to report.

References

- Accardo, J. A., & Malow, B. A. (2015). Sleep, epilepsy, and autism. *Epilepsy & Behavior*, 47, 202–206. <https://doi.org/10.1016/j.yebeh.2014.09.081>
- Aldinger, K. A., Lane, C. J., Veenstra-VanderWeele, J., & Levitt, P. (2015). Patterns of risk for multiple co-occurring medical conditions replicate across distinct cohorts of children with autism spectrum disorder. *Autism Research*, 8(6), 771–781. <https://doi.org/10.1002/aur.1492>
- Alexeeff, S. E., Yau, V., Qian, Y., Davignon, M., Lynch, F., Crawford, P., ... Croen, L. A. (2017). Medical conditions in the first years of life associated with future diagnosis of ASD in children. *Journal of Autism and Developmental Disorders*, 47(7), 2067–2079. <https://doi.org/10.1007/s10803-017-3130-4>

- 492 American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental*
493 *Disorders: DSM-IV-TR* (Fourth edition, text revision). Washington, D.C.: American
494 Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890420249.dsm-iv-tr>
- 495 American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental*
496 *Disorders: DSM-V*. Washington, D.C.: American Psychiatric Publishing.
- 497 Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., ... Dowling, N.
498 F. (2018). Prevalence of autism spectrum disorder among children aged 8 years - Autism
499 and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014.
500 *Morbidity and Mortality Weekly Report Surveillance Summaries*, 67(6), 1–23.
501 <https://doi.org/10.15585/mmwr.ss6706a1>
- 502 Buescher, A. V. S., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of autism spectrum
503 disorders in the United Kingdom and the United States. *JAMA Pediatrics*, 168(8), 721–
504 728. <https://doi.org/10.1001/jamapediatrics.2014.210>
- 505 Buie, T., Campbell, D. B., Fuchs, G. J., Furuta, G. T., Levy, J., VandeWater, J., ... Winter, H.
506 (2010). Evaluation, diagnosis, and treatment of gastrointestinal disorders in individuals
507 with ASDs: A consensus report. *Pediatrics*, 125(Supplement 1), S1–S18.
508 <https://doi.org/10.1542/peds.2009-1878C>
- 509 Burke, J. P., Jain, A., Yang, W., Kelly, J. P., Kaiser, M., Becker, L., ... Newschaffer, C. J.
510 (2013). Does a claims diagnosis of autism mean a true case? *Autism*, 18(3), 321–330.
511 <https://doi.org/10.1177/1362361312467709>
- 512 Cheng, N., Rho, J. M., & Masino, S. A. (2017). Metabolic dysfunction underlying autism
513 spectrum disorder and potential treatment approaches. *Frontiers in Molecular*
514 *Neuroscience*, 10. <https://doi.org/10.3389/fnmol.2017.00034>

- 515 Coleman, K. J., Lutsky, M. A., Yau, V., Qian, Y., Pomichowski, M. E., Crawford, P. M., ...
516 Croen, L. A. (2015). Validation of autism spectrum disorder diagnoses in large healthcare
517 systems with electronic medical records. *Journal of Autism and Developmental*
518 *Disorders*, 45(7), 1989–1996. <https://doi.org/10.1007/s10803-015-2358-0>
- 519 Diaz-Beltran, L., Esteban, F. J., Varma, M., Ortuzk, A., David, M., & Wall, D. P. (2017). Cross-
520 disorder comparative analysis of comorbid conditions reveals novel autism candidate
521 genes. *BMC Genomics*, 18(1), 315. <https://doi.org/10.1186/s12864-017-3667-9>
- 522 Díaz-Román, A., Zhang, J., Delorme, R., Beggiano, A., & Cortese, S. (2018). Sleep in youth with
523 autism spectrum disorders: systematic review and meta-analysis of subjective and
524 objective studies. *Evidence-Based Mental Health*, ebmental-2018-300037.
525 <https://doi.org/10.1136/ebmental-2018-300037>
- 526 Doshi-Velez, F., Ge, Y., & Kohane, I. (2014). Comorbidity clusters in autism spectrum
527 disorders: An electronic health record time-series analysis. *Pediatrics*, 133(1), e54–e63.
528 <https://doi.org/10.1542/peds.2013-0819>
- 529 Emerson, N. D., Morrell, H. E. R., & Neece, C. (2016). Predictors of age of diagnosis for
530 children with autism spectrum disorder: The role of a consistent source of medical care,
531 race, and condition severity. *Journal of Autism and Developmental Disorders*, 46(1),
532 127–138. <https://doi.org/10.1007/s10803-015-2555-x>
- 533 Findon, J., Cadman, T., Stewart, C. S., Woodhouse, E., Eklund, H., Hayward, H., ... McEwen, F.
534 S. (2016). Screening for co-occurring conditions in adults with autism spectrum disorder
535 using the strengths and difficulties questionnaire: A pilot study. *Autism Research*, 9(12),
536 1353–1363. <https://doi.org/10.1002/aur.1625>

Frye, R. E. (2015). Metabolic and mitochondrial disorders associated with epilepsy in children with autism spectrum disorder. *Epilepsy & Behavior*, 47, 147–157.

<https://doi.org/10.1016/j.yebeh.2014.08.134>

Fulceri, F., Morelli, M., Santocchi, E., Cena, H., Del Bianco, T., Narzisi, A., ... Muratori, F. (2016). Gastrointestinal symptoms and behavioral problems in preschoolers with autism spectrum disorder. *Digestive and Liver Disease*, 48(3), 248–254.

<https://doi.org/10.1016/j.dld.2015.11.026>

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.

<https://doi.org/10.2307/2346830>

Herlihy, L. E., Brooks, B., Dumont-Mathieu, T., Barton, M. L., Fein, D., Chen, C.-M., & Robins, D. L. (2014). Standardized screening facilitates timely diagnosis of autism spectrum disorders in a diverse sample of low-risk toddlers. *Journal of Developmental and Behavioral Pediatrics : JDBP*, 35(2), 85–92.

<https://doi.org/10.1097/DBP.0000000000000014>

Hirata, I., Mohri, I., Kato-Nishimura, K., Tachibana, M., Kuwada, A., Kagitani-Shimono, K., ... Taniike, M. (2016). Sleep problems are more frequent and associated with problematic behaviors in preschoolers with autism spectrum disorder. *Research in Developmental Disabilities*, 49–50, 86–99. <https://doi.org/10.1016/j.ridd.2015.11.002>

Hirota, T., Veenstra-VanderWeele, J., Hollander, E., & Kishi, T. (2014). Antiepileptic medications in autism spectrum disorder: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders*, 44(4), 948–957. <https://doi.org/10.1007/s10803-013-1952-2>

- 560 Hirvikoski, T., Mittendorfer-Rutz, E., Boman, M., Larsson, H., Lichtenstein, P., & Bölte, S.
561 (2016). Premature mortality in autism spectrum disorder. *The British Journal of*
562 *Psychiatry*, 208(3), 232–238. <https://doi.org/10.1192/bjp.bp.114.160192>
- 563 Holingue, C., Newill, C., Lee, L.-C., Pasricha, P. J., & Daniele Fallin, M. (2018).
564 Gastrointestinal symptoms in autism spectrum disorder: A review of the literature on
565 ascertainment and prevalence. *Autism Research*, 11(1), 24–36.
566 <https://doi.org/10.1002/aur.1854>
- 567 Horlin, C., Falkmer, M., Parsons, R., Albrecht, M. A., & Falkmer, T. (2014). The cost of autism
568 spectrum disorders. *PLOS ONE*, 9(9), e106552.
569 <https://doi.org/10.1371/journal.pone.0106552>
- 570 Hughes, H. K., Rose, D., & Ashwood, P. (2018). The gut microbiota and dysbiosis in autism
571 spectrum disorders. *Current Neurology and Neuroscience Reports*, 18(11), 81.
572 <https://doi.org/10.1007/s11910-018-0887-6>
- 573 Hughes, H., Mills Ko, E., Rose, D., & Ashwood, P. (2018). Immune dysfunction and
574 autoimmunity as pathological mechanisms in autism spectrum disorders. *Frontiers in*
575 *Cellular Neuroscience*, 12. <https://doi.org/10.3389/fncel.2018.00405>
- 576 Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietiläinen, O. P. H., ...
577 Clair, D. M. S. (2011). Copy number variations of chromosome 16p13.1 region
578 associated with schizophrenia. *Molecular Psychiatry*, 16(1), 17–25.
579 <https://doi.org/10.1038/mp.2009.101>
- 580 Jülich, K., & Sahin, M. (2014). Mechanism-based treatment in tuberous sclerosis complex.
581 *Pediatric Neurology*, 50(4), 290–296. <https://doi.org/10.1016/j.pediatrneurol.2013.12.002>

- 582 Kohane, I. S., McMurry, A., Weber, G., MacFadden, D., Rappaport, L., Kunkel, L., ...
583 Churchill, S. (2012). The co-morbidity burden of children and young adults with autism
584 spectrum disorders. *PLOS ONE*, 7(4), e33224.
585 <https://doi.org/10.1371/journal.pone.0033224>
- 586 Lavelle, T. A., Weinstein, M. C., Newhouse, J. P., Munir, K., Kuhlthau, K. A., & Prosser, L. A.
587 (2014). Economic burden of childhood autism spectrum disorders. *Pediatrics*, 133(3),
588 e520–e529. <https://doi.org/10.1542/peds.2013-0763>
- 589 Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., ... Rice, C. E.
590 (2010). Autism spectrum disorder and co-occurring developmental, psychiatric, and
591 medical conditions among children in multiple populations of the United States. *Journal*
592 *of Developmental & Behavioral Pediatrics*, 31(4), 267.
593 <https://doi.org/10.1097/DBP.0b013e3181d5d03b>
- 594 Lingren, T., Chen, P., Bochenek, J., Doshi-Velez, F., Manning-Courtney, P., Bickel, J., ...
595 Savova, G. (2016). Electronic health record based algorithm to identify patients with
596 autism spectrum disorder. *PLoS ONE*, 11(7), e0159621.
597 <https://doi.org/10.1371/journal.pone.0159621>
- 598 Lundström, S., Reichenberg, A., Melke, J., Råstam, M., Kerekes, N., Lichtenstein, P., ...
599 Anckarsäter, H. (2015). Autism spectrum disorders and coexisting disorders in a
600 nationwide Swedish twin study. *Journal of Child Psychology and Psychiatry*, 56(6), 702–
601 710. <https://doi.org/10.1111/jcpp.12329>
- 602 Mannion, A., & Leader, G. (2016). An investigation of comorbid psychological disorders, sleep
603 problems, gastrointestinal symptoms and epilepsy in children and adolescents with

autism spectrum disorder: A two year follow-up. *Research in Autism Spectrum Disorders*, 22, 20–33. <https://doi.org/10.1016/j.rasd.2015.11.002>

Mannion, A., Leader, G., & Healy, O. (2013). An investigation of comorbid psychological disorders, sleep problems, gastrointestinal symptoms and epilepsy in children and adolescents with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 7(1), 35–42. <https://doi.org/10.1016/j.rasd.2012.05.002>

Margolis, K. G., Buie, T. M., Turner, J. B., Silberman, A. E., Feldman, J. F., Murray, K. F., ... Winter, H. S. (2019). Development of a brief parent-report screen for common gastrointestinal disorders in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 49(1), 349–362. <https://doi.org/10.1007/s10803-018-3767-7>

McCue, L. M., Flick, L. H., Twyman, K. A., & Xian, H. (2017). Gastrointestinal dysfunctions as a risk factor for sleep disorders in children with idiopathic autism spectrum disorder: a retrospective cohort study. *Autism*, 21(8), 1010–1020. <https://doi.org/10.1177/1362361316667061>

Meltzer, A., & Water, J. V. de. (2017). The role of the immune system in autism spectrum disorder. *Neuropsychopharmacology*, 42(1), 284–298. <https://doi.org/10.1038/npp.2016.158>

Meltzer, L. J., Johnson, C., Crosette, J., Ramos, M., & Mindell, J. A. (2010). Prevalence of diagnosed sleep disorders in pediatric primary care practices. *Pediatrics*, 125(6), e1410–e1418. <https://doi.org/10.1542/peds.2009-2725>

Murdoch, J. D., & State, M. W. (2013). Recent developments in the genetics of autism spectrum disorders. *Current Opinion in Genetics & Development*, 23(3), 310–315. <https://doi.org/10.1016/j.gde.2013.02.003>

- 627 Muskens, J. B., Velders, F. P., & Staal, W. G. (2017). Medical comorbidities in children and
628 adolescents with autism spectrum disorders and attention deficit hyperactivity disorders:
629 A systematic review. *European Child & Adolescent Psychiatry*, 26(9), 1093–1103.
630 <https://doi.org/10.1007/s00787-017-1020-0>
- 631 OptumLabs. (2018). *OptumLabs and OptumLabs Data Warehouse (OLDW) Descriptions and*
632 *Citation*. Cambridge, MA: n.p.
- 633 Pickett, J., Xiu, E., Tuchman, R., Dawson, G., & Lajonchere, C. (2011). Mortality in individuals
634 with autism, with and without epilepsy. *Journal of Child Neurology*, 26(8), 932–939.
635 <https://doi.org/10.1177/0883073811402203>
- 636 Rose, D. R., Yang, H., Serena, G., Sturgeon, C., Ma, B., Careaga, M., ... Ashwood, P. (2018).
637 Differential immune responses and microbiota profiles in children with autism spectrum
638 disorders and co-morbid gastrointestinal symptoms. *Brain, Behavior, and Immunity*.
639 <https://doi.org/10.1016/j.bbi.2018.03.025>
- 640 Rossignol, D. A., & Frye, R. E. (2011). Melatonin in autism spectrum disorders: A systematic
641 review and meta-analysis. *Developmental Medicine & Child Neurology*, 53(9), 783–792.
642 <https://doi.org/10.1111/j.1469-8749.2011.03980.x>
- 643 Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of
644 cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
645 [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- 646 Rubenstein, E., Schieve, L., Wiggins, L., Rice, C., Van Naarden Braun, K., Christensen, D., ...
647 Lee, L.-C. (2018). Trends in documented co-occurring conditions in children with autism
648 spectrum disorder, 2002–2010. *Research in Developmental Disabilities*, 83, 168–178.
649 <https://doi.org/10.1016/j.ridd.2018.08.015>

- Rydzewska, E., Hughes-McCormack, L. A., Gillberg, C., Henderson, A., MacIntyre, C., Rintoul, J., & Cooper, S.-A. (2018). Prevalence of sensory impairments, physical and intellectual disabilities, and mental health in children and young people with self/proxy-reported autism: Observational study of a whole country population. *Autism*, 1362361318791279. <https://doi.org/10.1177/1362361318791279>
- Sacco, R., Lenti, C., Saccani, M., Curatolo, P., Manzi, B., Bravaccio, C., & Persico, A. M. (2012). Cluster analysis of autistic patients based on principal pathogenetic components. *Autism Research*, 5(2), 137–147. <https://doi.org/10.1002/aur.1226>
- Smoller, J. W., Craddock, N., Kendler, K., Lee, P. H., Neale, B. M., Nurnberger, J. I., ... Sullivan, P. F. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet*, 381(9875), 1371–1379. [https://doi.org/10.1016/S0140-6736\(12\)62129-1](https://doi.org/10.1016/S0140-6736(12)62129-1)
- Soke, G. N., Maenner, M. J., Christensen, D., Kurzius-Spencer, M., & Schieve, L. A. (2018). Prevalence of co-occurring medical and behavioral conditions/symptoms among 4- and 8-year-old children with autism spectrum disorder in selected areas of the United States in 2010. *Journal of Autism and Developmental Disorders*, 48(8), 2663–2676. <https://doi.org/10.1007/s10803-018-3521-1>
- Souders, M. C., Zavodny, S., Eriksen, W., Sinko, R., Connell, J., Kerns, C., ... Pinto-Martin, J. (2017). Sleep in children with autism spectrum disorder. *Current Psychiatry Reports*, 19(6), 34. <https://doi.org/10.1007/s11920-017-0782-x>
- Spain, D., Sin, J., Chalder, T., Murphy, D., & Happé, F. (2015). Cognitive behaviour therapy for adults with autism spectrum disorders and psychiatric co-morbidity: A review. *Research in Autism Spectrum Disorders*, 9, 151–162. <https://doi.org/10.1016/j.rasd.2014.10.019>

- 673 Tilford, J. M., Payakachat, N., Kuhlthau, K. A., Pyne, J. M., Kovacs, E., Bellando, J., ... Frye, R.
674 E. (2015). Treatment for sleep problems in children with autism and caregiver spillover
675 effects. *Journal of Autism and Developmental Disorders*, 45(11), 3613–3623.
676 <https://doi.org/10.1007/s10803-015-2507-5>
- 677 Tye, C., Runicles, A. K., Whitehouse, A. J. O., & Alvares, G. A. (2019). Characterizing the
678 interplay between autism spectrum disorder and comorbid medical conditions: An
679 integrative review. *Frontiers in Psychiatry*, 9. <https://doi.org/10.3389/fpsyt.2018.00751>
- 680 Vargason, T., McGuinness, D. L., & Hahn, J. (2019). Gastrointestinal symptoms and oral
681 antibiotic use in children with autism spectrum disorder: Retrospective analysis of a
682 privately insured U.S. population. *Journal of Autism and Developmental Disorders*,
683 49(2), 647–659. <https://doi.org/10.1007/s10803-018-3743-2>
- 684 Vuong, H. E., & Hsiao, E. Y. (2017). Emerging roles for the gut microbiome in autism spectrum
685 disorder. *Biological Psychiatry*, 81(5), 411–423.
686 <https://doi.org/10.1016/j.biopsych.2016.08.024>
- 687 Wiggins, L. D., Levy, S. E., Daniels, J., Schieve, L., Croen, L. A., DiGuseppi, C., ... Schendel,
688 D. (2015). Autism spectrum disorder symptoms among children enrolled in the Study to
689 Explore Early Development (SEED). *Journal of Autism and Developmental Disorders*,
690 45(10), 3183–3194. <https://doi.org/10.1007/s10803-015-2476-8>
- 691 Woolfenden, S., Sarkozy, V., Ridley, G., Coory, M., & Williams, K. (2012). A systematic
692 review of two outcomes in autism spectrum disorder - epilepsy and mortality.
693 *Developmental Medicine and Child Neurology*, 54(4), 306–312.
694 <https://doi.org/10.1111/j.1469-8749.2012.04223.x>

Co-occurring condition clusters in ASD

- 695 Zerbo, O., Leong, A., Barcellos, L., Bernal, P., Fireman, B., & Croen, L. A. (2015). Immune
696 mediated conditions in autism spectrum disorders. *Brain, Behavior, and Immunity*, 46,
697 232–236. <https://doi.org/10.1016/j.bbi.2015.02.001>
- 698 Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., ...
699 Wetherby, A. (2015). Early identification of autism spectrum disorder: Recommendations
700 for practice and research. *Pediatrics*, 136(Supplement 1), S10–S40.
701 <https://doi.org/10.1542/peds.2014-3667C>
702

Table 1. Categories of co-occurring conditions used in this study, specific conditions comprising these categories, and their *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) codes.

Category of Condition	Specific Condition	ICD-9-CM Codes
Auditory disorders	Disorders of external ear	380.xx
	Otitis media	381.xx, 382.xx
	Hearing loss	389.xx
Developmental delays	Speech/language disorder	315.3x
	Coordination disorder	315.4
	Mixed development disorder	315.5
	Other or unspecified delays	315.8, 315.9
Gastrointestinal disorders/symptoms	Non-infective enteritis and colitis	555.xx, 556.xx, 558.xx
	Constipation	564.0x
	Irritable bowel syndrome	564.1x
	Diarrhea	787.91
	Abdominal pain	789.0x
Immune disorders	Food allergy	V15.01 – V15.05, 692.5, 693.1, 995.6x, 995.7
	Rhinitis	477.xx
	Asthma	493.xx
	Dermatitis	691.8x, 692.xx
	Other allergies	V14.xx, V15.06 – V15.09, 287.0, 372.05, 372.14, 708.0, 995.2, 995.20, 995.29, 995.3
Psychiatric disorders	Episodic mood disorders	296.xx, 311.xx
	Anxiety (and related)	300.xx
	Oppositional defiant disorder	313.81
	Hyperkinetic syndrome (incl. attention deficit disorders)	314.xx
Seizure disorders	Epilepsy	345.xx
	Other convulsions	780.39
Sleep disorders	Insomnia	307.41, 307.42, 327.0x, 780.52
	Hypersomnia	307.43, 307.44, 327.1x, 780.54
	Parasomnia	307.46, 307.48, 327.4x, 780.56
	Sleep-disordered breathing	327.2x, 780.51, 780.53, 780.57, 786.03
	Infant apnea	770.81, 770.82
	Circadian rhythm disorder	780.55, 307.45, 327.3x
	Sleep-related movement disorder	327.5x, 333.94, 780.58
	Not otherwise specified	307.40, 307.47, 307.49, 327.8x, 780.50, 780.59

Table 2. Characteristics of the study cohorts describing children with autism spectrum disorder (ASD) and children from the general population with no ASD diagnosis (POP cohort) (Vargason et al., 2019), and of the three ASD clusters. When counts are given, the percentage of the respective cohort/cluster is provided in parentheses. Mean numbers of co-occurring condition (COC) categories diagnosed during the five-year enrollment period are reported with the standard deviation in parentheses.

Characteristic	POP Cohort	ASD Cohort	ASD Cluster 1 (High-Prevalence)	ASD Cluster 2 (Mid-Prevalence)	ASD Cluster 3 (Low-Prevalence)
Overall	279,693	3,278	776	870	1,632
Gender					
Male	141,800 (50.7%)	2,673 (81.5%)	657 (84.7%)	699 (80.3%)	1,317 (80.7%)
Female	137,893 (49.3%)	606 (18.5%)	119 (15.3%)	171 (19.7%)	315 (19.3%)
Race/ethnicity					
White	211,368 (75.6%)	2,489 (75.9%)	570 (73.5%)	653 (75.1%)	1,266 (77.6%)
Asian	17,467 (6.2%)	236 (7.2%)	61 (7.9%)	82 (9.4%)	93 (5.7%)
Black	21,739 (7.8%)	230 (7.0%)	65 (8.4%)	49 (5.6%)	116 (7.1%)
Hispanic	29,119 (10.4%)	323 (9.9%)	80 (10.3%)	86 (9.9%)	157 (9.6%)
Census Division					
New England	8,694 (3.1%)	127 (3.9%)	27 (3.5%)	39 (4.5%)	61 (3.7%)
Mid Atlantic	24,221 (8.7%)	430 (13.1%)	130 (16.8%)	101 (11.6%)	199 (12.2%)
East North Central	45,259 (16.2%)	445 (13.6%)	86 (11.1%)	113 (13.0%)	246 (15.1%)
West North Central	37,948 (13.6%)	455 (13.9%)	79 (10.2%)	128 (14.7%)	248 (15.2%)
South Atlantic	70,194 (25.1%)	908 (27.7%)	253 (32.6%)	213 (24.5%)	442 (27.1%)
East South Central	8,704 (3.1%)	88 (2.7%)	22 (2.8%)	26 (3.0%)	40 (2.5%)
West South Central	41,970 (15.0%)	362 (11.0%)	101 (13.0%)	97 (11.1%)	164 (10.0%)
Mountain	23,750 (8.5%)	204 (6.2%)	31 (4.0%)	69 (7.9%)	104 (6.4%)
Pacific	18,839 (6.7%)	259 (7.9%)	47 (6.1%)	84 (9.7%)	128 (7.8%)
Other	114 (< 0.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Mean number of COC categories diagnosed	2.30 (1.15)	3.80 (1.39)	4.69 (1.13)	4.22 (1.17)	3.14 (1.27)

Figure Captions

Figure 1. Summary of dates relevant to the study's design and cohort definitions. The total study period spanned from January 1, 2000 to September 30, 2015, and included children must have been born between the years 2000–2010. The first date of a child's coverage enrollment (index date) must have been within one year of the child's birth year, and at least five years of continuous coverage (medical, pharmacy, and behavioral) must have followed the index date and been reached before the end of the study period. Autism spectrum disorder (ASD) diagnoses were only considered if they were made at least two years after the birth year and prior to the end of the year 2012.

Figure 2. Prevalence of each of the seven co-occurring condition categories across the five-year enrollment period in (a) the autism spectrum disorder (ASD) cohort and general population with no ASD diagnosis (POP) cohort, and (b) the three ASD clusters. Dots signify the prevalence is at least as great as the indicated value.

Figure 3. Proportions of the autism spectrum disorder (ASD) cohort, general population with no ASD diagnosis (POP) cohort, and three ASD clusters with a diagnosis in each co-occurring condition category during each six-month interval of the five-year enrollment period: (a) auditory disorders; (b) developmental delays; (c) gastrointestinal symptoms; (d) immune-related conditions; (e) psychiatric disorders; (f) seizures; (g) sleep disorders.

Figure 4. Frequency distributions for the number of co-occurring condition categories diagnosed during the five-year enrollment period in an individual in the autism spectrum disorder (ASD) cohort, general population with no ASD diagnosis (POP) cohort, and three ASD clusters.

Figure 5. Cumulative distribution functions for the time of first autism spectrum disorder (ASD) diagnosis in the ASD cohort and its three clusters.

Figure 6. Differences in prevalence of the seven co-occurring condition categories between the autism spectrum disorder (ASD) cohort and general population with no ASD diagnosis (POP) cohort over time, with categories following similar patterns grouped together: (a) auditory disorders; (b) developmental delays; (c) gastrointestinal conditions clustered with immune-related disorders; (d) psychiatric disorders; (e) seizures clustered with sleep disorders.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6