Using Machine Learning to Identify Patterns of Lifetime Health Problems in Decedents with Autism Spectrum Disorder

Lauren Bishop-Fitzpatrick, PhD[1,2], **Arezoo Movaghar, MS**[1], **Jan S. Greenberg, PhD**[1,2], **David Page, PhD**[3], **Leann S. DaWalt, PhD**[1], **Murray H. Brilliant, PhD**[4], and **Marsha R. Mailick, PhD**[1,2]

[1]Waisman Center, University of Wisconsin-Madison, 1500 Highland Ave., Madison WI 53705

[2]School of Social Work, University of Wisconsin-Madison, 1350 University Ave., Madison, WI 53706

[3]Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, 600 Highland Ave., Madison, WI 53792

[4]Marshfield Clinic Research Institute, 1000 N. Oak Ave., Marshfield, WI 54449

Abstract

Very little is known about the health problems experienced by individuals with autism spectrum disorder (ASD) throughout their life course. We retrospectively analyzed diagnostic codes associated with de-identified electronic health records using a machine learning algorithm to characterize diagnostic patterns in decedents with ASD and matched decedent community controls. Participants were 91 decedents with ASD and 6,186 sex and birth year matched decedent community controls who had died since 1979, the majority of whom were middle aged or older adults at the time of their death. We analyzed all ICD-9 codes, V-codes, and E-codes available in the electronic health record and Elixhauser comorbidity categories associated with those codes. Diagnostic patterns distinguished decedents with ASD from decedent community controls with 75% sensitivity and 94% specificity solely based on their lifetime ICD-9 codes, V-codes, and E-codes. Decedents with ASD had higher rates of most conditions, including cardiovascular disease, motor problems, ear problems, urinary problems, digestive problems, side effects from long-term medication use, and non-specific lab tests and encounters. In contrast, decedents with ASD had lower rates of cancer. Findings suggest distinctive lifetime diagnostic patterns among decedents with ASD and highlight the need for more research on health outcomes across the lifespan as the population of individuals with ASD ages. As a large wave of individuals with ASD diagnosed in the 1990s enters adulthood and middle age, knowledge about lifetime health problems will become increasingly important for care and prevention efforts.

Keywords

Corresponding author: Lauren Bishop-Fitzpatrick, PhD, Waisman Center, University of Wisconsin-Madison, 1500 Highland Ave., Madison, WI 53705, bishopfitzpa@wisc.edu.

Introduction

Autism spectrum disorder (ASD) is a chronic, lifelong neurodevelopmental disorder that by current estimates affects 1 in 68 children (Christensen et al., 2016) and will cost the United States $461 billion annually by 2025 (Leigh & Du, 2015). Throughout life, individuals with ASD experience marked and debilitating challenges with social functioning and exhibit restricted and repetitive patterns of behaviors (American Psychiatric Association, 2013). Long-term outcomes in ASD are poor in that most affected individuals do not live independently, become self-supporting, or develop reciprocal friendships or romantic relationships in adulthood (Magiati, Tay, & Howlin, 2014; Seltzer, Shattuck, Abbeduto, & Greenberg, 2004). High rates of psychiatric comorbidities (Caamaño et al., 2013; Hofvander et al., 2009) and co-occurring medical conditions (Croen et al., 2015) may further hinder outcomes and reduce quality of life. While a previous study suggests that individuals with ASD have a heightened risk of mortality compared to the general population (Hirvikoski et al., 2016), disparities in their health problems throughout life have not been investigated (Bishop-Fitzpatrick & Kind, 2017). A better understanding of their lifetime health problems is important for developing health prevention efforts as the population of adults with ASD diagnosed at the beginning of the diagnostic boom in the 1990s enters midlife over the coming decades (King & Bearman, 2009).

Research on children, adolescents, and young adults with ASD suggests the presence of a number of health risk factors that may lead to excess morbidity. Health risk factors identified in young people with ASD include poor eating habits (Ho, Eaves, & Peabody, 1997), obesity (Croen et al., 2015; Ho et al., 1997), and limited physical activity (Ho et al., 1997), and many individuals with ASD take multiple psychotropic medications (Esbensen, Greenberg, Seltzer, & Aman, 2009), all of which may increase morbidity and inhibit healthy aging. Indeed, a small number of preliminary studies indicate that adults with ASD have high rates of health problems, including epilepsy (Woolfenden, Sarkozy, Ridley, Coory, & Williams, 2012), early parkinsonism (Starkstein, Gellar, Parlier, Payne, & Piven, 2015), hypertension, and diabetes (Croen et al., 2015). Some individuals with ASD may also have co-occurring rare genetic disorders such as 22q11.2 deletion syndrome (Vorstman et al., 2006), methylenetetrahydrofolate reductase gene polymorphisms (Pu, Shen, & Wu, 2013), or Timothy syndrome (Splawski et al., 2004) that are associated with a range of health problems including motor anomalies and cardiovascular problems. However, no studies have examined health problems throughout the life course of individuals with ASD using representative, population-level data.

To date, one representative study conducted in Sweden (Hirvikoski et al., 2016) has investigated mortality in individuals with ASD compared to controls. This study analyzed all-cause mortality for individuals with ASD registered in the population-based Swedish National Patient Registry from 1987–2009. Findings indicate that, compared those without ASD, individuals with ASD die at younger ages (nearly 20 years younger) and experience increased mortality from accidents (i.e., choking, suffocation, drowning), epilepsy, congenital malformations, malignancy, respiratory conditions, circulatory conditions, and suicide.

The present study sought to identify health problems that distinguish decedents with ASD from matched decedent community controls using a data-driven approach based on information recorded in their electronic health records (EHRs). Follow-up analyses examined group differences in established comorbidity categories.

Methods

Data and Participants

We retrospectively analyzed diagnostic codes associated with EHRs from the Marshfield Clinic, a multi-specialty group practice with more than 700 physicians providing integrated, comprehensive care to more than one million people across more than 50 locations in northern, central, and western Wisconsin. The Marshfield Clinic covers approximately 97% of residents, and captures 99% of deaths, 95% of hospital discharges, and 90% of outpatient visits in the region (Greenlee, 2003). For research purposes, a sub-set of the area covered by the Marshfield Clinic, termed the Marshfield Clinic Epidemiologic Study Area (MESA) consisting of 100,000 patients, has been identified as representative of the population in northern, central, and western Wisconsin. Previous research has validated EHR data from MESA (Greenlee, 2003). Incidence and prevalence rates of clinically detected diseases in Marshfield Clinic EHR data compare well to previously reported data in the medical literature (Greenlee, 2003). The Marshfield Clinic converted their health records to EHRs in 1979.

Participants were decedents with ASD and decedent community controls. Decedents with ASD (1) had a diagnosis, as specified by ICD-9 codes, of autism (299.0), Asperger's disorder (299.8) or pervasive developmental disorder not otherwise specified (299.9); and (2) had died since 1979. In order to rule out potential participants who received an ASD diagnosis in error, to be included in this study, decedents with ASD had to have had at least two diagnoses of ASD on different days recorded in their EHR. There were 131 decedents with ASD who met these criteria and were therefore eligible for inclusion in this study. Decedents with ASD were 61.8% male. They were born between 1903 and 2000 (M=1951.8, SD = 22.4) and were between the ages of 4 and 89 (M = 56.1, SD = 21.8) at the time of their death.

The comparison group was drawn from a 10% random sample (N=16,981) of the total number of decedents who had been patients at the Marshfield Clinic (N_{total}=169,810) who had (1) died since 1979; and (2) did not have a code for ASD, Down syndrome, or intellectual disability recorded in their EHR. Decedent community controls were 52.1% male (N = 8,837). They were born between 1891 and 2015 (M = 1931.5, SD = 13.1) and were between the ages of 0 and 89 (M = 75.2, SD = 15.4) at the time of their death. The distribution of age of death in the sample of decedents with ASD and the 10% random sample of decedent community controls is displayed in Figure 1, and indicates that individuals with ASD died at a younger age (nearly 20 years younger) than their counterparts without ASD.

Because the distribution of sex ($\chi^2 = 51.6, p < 0.001$) and birth year ($t = 16.3, p < 0.001$) differed significantly between our full samples of decedents with ASD and decedent

community controls, we frequency matched decedents with ASD to a 1:68 ratio of decedent community controls on birth year (within five years) and sex. This ratio represents the current estimated prevalence of ASD in the United States (Christensen et al., 2016). Our matching procedure also takes into account both known sex differences in certain diseases (e.g., heart disease; Lerner & Kannel, 1986) and potential birth cohort effects in medical and diagnostic practices (Kuh & Shlomo, 2004). Due to differences in the sex distribution of decedents with ASD (i.e., a larger proportion of the decedents with ASD were male), we were unable to match all decedents with ASD to 68 decedent community controls of the same sex and birth year (within five years). These matching procedures resulted in a final analytic sample of 91 decedents with ASD and 6,186 decedent community controls. The analytic sample of 91 decedents with ASD were 51.2% male (N=53). They were born between 1903 and 1980 (M = 1940.3, SD = 14.0) and were between the ages of 23 and 89 (M = 67.3, SD = 13.5) at the time of their death. All but two decedents with ASD were aged 40 or older at the time of their death. The resulting sample of 6,186 decedent community controls were 58.2% male (N=3,602). They were born between 1898 and 1984 (M = 1936.4, SD = 15.3) and were between the ages of 0 and 89 (M = 68.7, SD = 15.1) at the time of their death.

Variables

Age at death was calculated by subtracting year of birth from year of death. The Marshfield Clinic caps age of death at 89 in order to de-identify records.

Sex was recorded by clinicians.

Length of EHR was calculated by subtracting date of first encounter with the Marshfield clinic from date of death.

ICD-9 codes, V-codes, and E-codes included all codes recorded in the EHR during health services encounters. As part of their normal data management procedures, the Marshfield Clinic has converted all codes from other versions of the International Classification of Diseases (e.g., ICD-8 or ICD-10) to ICD-9 codes. For the current analysis, we did not select codes based on either relevance to ASD or functionally group codes that were similar to each other. However, diagnoses related to developmental disabilities and mental health conditions (i.e., Chapter 5: Mental Disorders) were excluded from our analyses because of their co-occurrence with ASD. All other codes that were recorded in the EHR were used in analyses.

Comorbidities were defined using the Elixhauser method (Elixhauser, Steiner, Harris, & Coffey, 1998), a comprehensive set of 30 comorbidity measures that are extracted from ICD-9 codes. Elixhauser comorbidities have been shown to be predictive of mortality in critically ill patients (Ladha et al., 2015).

We additionally identified decedents with ASD with co-occurring intellectual disability and Down syndrome using the EHR given the high co-occurrence of these conditions with ASD and excluded decedents with these diagnoses from the decedent community control group.

Analyses

We used a machine learning algorithm (random forest) to classify participants into two groups based on their ICD-9 codes, V-codes, and E-codes. Random forest, a robust and reliable classification method with low generalization error and high predictive performance, fits multiple decision trees and chooses a class that best aggregates the results of these trees (Breiman, 2001). In contrast to linear methods (e.g., logistic regression), the use of decision trees can discover important multivariate interactions, but at the increased risk of overfitting the training data. The random forest algorithm combats overfitting by learning multiple decision trees; diversity among the trees is obtained by using different bootstrap samples of the original data and considering only random subsets of variables (features) at various nodes in the tree. The algorithm splits each subtree using the best predictor among a subset of randomly chosen predictors for each node. To accurately estimate performance of the model on future, unseen data, we use 10-fold cross-validation, a widely-used form of hold-out testing and evaluation. We trained and tested the model with RandomForest in Weka version 3-6-11. In-fold feature selection with cross validation, which reduces the computational cost and improves the performance and speed of prediction models, was used to ensure the selection of a subset of relevant attributes and reduce the possibility of overfitting the model to our data (Yu & Liu, 2004). Information gain scores, which take into account information entropy for each class (ASD versus community control) and feature (e.g., ICD-9 code), were used to measure the amount of information in each feature with respect to target class.

Follow-up analyses extracted comorbidities from all ICD-9 codes using R package icd (Wasey, 2016) and R version 3.2.0. T or chi-square statistics compared age of death, sex, length of EHR, and number of codes between decedents with ASD and our sample of matched decedent community controls. Separate binary logistic regression models, controlling for age and sex, estimated the effect size (odds ratio) of group differences in Elixhauser comorbidities.

Results

Our random forest algorithm identified 10,142 ICD-9 codes, V-codes, and E-codes with information gain scores greater than zero. Model fit did not substantively differ after reducing the number of codes to the top 50 ICD-9 codes, V-codes, and E-codes. The final algorithm was thus fitted on the ICD-9 codes, V-codes, and E-codes with the 50 highest information gain scores (Table 1). The area under the ROC curve was 0.88 (Figure 2), suggesting good model fit. Sensitivity (0.75), specificity (0.94), and accuracy (0.93) were also high (Figure 3), indicating that our model correctly predicted whether a case was a decedent with ASD or a decedent community control 93% of the time.

Table 1 presents the top 50 ICD-9 codes, V-codes, and E-codes that distinguish decedents with ASD from decedent community controls, ordered by information gain. Descriptively, decedents with ASD had higher rates of nearly all ICD-9 codes, V-codes, and E-codes with the 50 highest information gain scores. These groupings with two or more codes include: **epilepsy** (other convulsions; epilepsy, unspecified, without mention of intractable epilepsy; generalized convulsive epilepsy); **long-term medication use and side effects** (long-term

(current) use of other medications; orofacial dyskinesia); **developmental problems** (lack of expected normal physiological development in childhood; chromosomal anomolies); **ear problems** (impacted cerumen; infective otitis externa); **non-specific lab tests and encounters** (routine general medical examination at a health care facility; observation and evaluation for other suspected conditions; other malaise and fatigue; screening examination for pulmonary tuberculosis); **urinary problems** (urinary incontinence, unspecified; other symptoms involving urinary system; other disorders of urethra and urinary tract); **digestive problems** (functional digestive disorders not elsewhere classified; megacolon, other than Hirschsprung's); **motor problems** (abnormality of gait; other cerebral degenerations; lack of coordination; other and unspecified extrapyramidal diseases and abnormal movement disorders); and **choking** (pneumonitis due to solids and liquids; foreign body in respiratory tree, unspecified; foreign body in larynx).

We then conducted follow-up analyses using Elixhauser comorbidity categories to assess group differences in conditions that are predictive of mortality in the general population (Elixhauser et al., 1998). Analysis of comorbidities using Elixhauser categories (Table 2) indicates that decedents with ASD are distinguished from matched decedent community controls on 12 of 28 (42.9%) comorbidities after controlling for age and sex. Decedents with ASD had an *increased* risk of 122% for coagulopathy, 90% for congestive heart failure, 163% for deficiency anemia, 179% for fluid and electrolyte disorders, 376% for hypothyroidism, 218% for paralysis, 133% for valvular disease, and 112% for weight loss compared to decedent community controls. Decedents with ASD had a *decreased* risk of 253% for alcohol abuse, 68% for hypertension, and 439% for metastatic cancer compared to decedent community controls. We also found an expected very large (661%) increased risk of having a diagnosis of other neurological disorders given the high prevalence of epilepsy in ASD (Woolfenden et al., 2012).

Discussion

A single, previous population-based study suggests that individuals with ASD are at risk of premature mortality compared to the general population (Hirvikoski et al., 2016). Our findings based on the full sample from the Marshfield Clinic that individuals with ASD died at a much younger age (approximately 20 years) than decedent community controls is consistent with this previous study. However, we know very little about lifetime health problems in ASD, which may put individuals with ASD at risk for high health care utilization and costs and early death. This study is the first to characterize patterns of health problems using a machine learning approach based on a population-based sample of decedents with ASD, most of whom were middle aged or older at the time of their death.

Analyses related to our central research question revealed that decedents with ASD can be distinguished from a matched sample of decedent community controls with 75% sensitivity, 94% specificity, and 93% accuracy solely based on their lifetime ICD-9 codes, V-codes, and E-codes. Diagnostic patterns that distinguished decedents with ASD from matched controls included higher rates of epilepsy, long-term medication use and side effects, developmental problems, ear problems, non-specific lab tests and encounters, urinary problems, digestive problems, motor problems, and choking. When we grouped ICD-9 codes by comorbidity

categories, we found that decedents with ASD were distinguished from matched decedent community controls by increased risk of coagulopathy, congestive heart failure, deficiency anemia, fluid and electrolyte disorders, hypothyroidism, paralysis, valvular disease, weight loss, and other neurological disorders compared to decedent community controls. Decedents with ASD had a decreased risk of alcohol abuse, hypertension, and metastatic cancer compared to decedent community controls. Of note, the majority of our analytic sample of decedents with ASD (97.8%) and decedent community controls (95.6%) were aged 40 or older at the time of their death, which is a major departure from previous studies that examine physical health problems in adults with ASD that include mostly young adults (e.g., Cashin et al., 2016; Croen et al., 2015). Unlike these previous studies, our study sheds light on the health problems that differentiate decedents with ASD from decedent community controls in midlife and old age. In order to fully study health in a population that is hypothesized to have more health problems at younger ages than the general population, it is necessary to examine health in midlife and beyond when health problems are likely to emerge.

Although we were unable to explore causal factors in the current study, it may be that a combination of underlying biological vulnerability, coupled with lifestyle factors and difficulties interacting with the health care system, lead to differential diagnostic patterns in decedents with ASD compared to decedent community controls. Our findings confirm well-established reports of heightened epilepsy (Woolfenden et al., 2012) in individuals with ASD. In addition, the pattern of heightened cardiovascular problems identified by our analysis of comorbidities is consistent with a potential increased biological vulnerability related to broad cardiac parasympathetic hypofunction in ASD found in previous literataure (Ming, Patel, Kang, Chokroverty, & Julu, 2016). Previous studies have suggested heightened cardiovascular risk factors in ASD (Cashin, Buckley, Trollor, & Lennox, 2016), but this is the first study, to our knowledge, that identifies heightened rates of cardiovascular disease, including higher rates of coagulopathy, congestive heart failure, and valvular disease, in individuals with ASD compared to controls.

It should be noted that findings also revealed lower risk of metastatic cancer, hypertension, and alcohol abuse. Decreased risk of some health problems has been documented in other populations with developmental disabilities and mental health problems. For instance, individuals with Down syndrome have a lower risk of hypertension and some cancers (Esbensen, 2010), while individuals with schizophrenia have a lower risk of cancer (Jablensky & Lawrence, 2001). It is possible that our findings signal lower risk for these physical health problems in ASD. It is also possible that individuals with ASD are screened less frequently for these comorbidities, thus limiting opportunities for timely and potentially life-saving diagnoses. Potential protective effects and their mechanisms, as well as the possibility of disparities in screening and diagnosis, should be probed in future research.

This study has three key limitations. First, as is the case with all EHR data, EHR data from the Marshfield Clinic may contain inaccurate or incomplete data (Hersh et al., 2013; Weiskopf & Weng, 2013) and may more accurately capture serious diseases (i.e., diabetes, colon cancer, epilepsy) than other disorders that are more difficult to diagnose or that are retained in sequestered health records (i.e., psychiatric conditions; Weiskopf & Weng, 2013).

Second, the population analyzed in this study should not be seen as representative of the current population of children and adolescents with ASD, both with respect to diagnostic criteria and access to services and supports. Lack of lifetime access to services and supports may bias the emergence of health problems in our sample. Knowledge about the current generation of middle aged and older adults with ASD is nevertheless valuable given the growing population of individuals with ASD who have reached midlife and beyond (Howlin & Magiati, 2017). Third, data are only indicative of the ICD-9 codes, V-codes, and E-codes assigned to decedents and may not represent the actual extent of health problems experienced by individuals with ASD.

Despite these limitations, this study is the first study to characterize health problems in a representative sample of decedents with ASD, and to use a machine learning algorithm to differentiate decedents with ASD from decedent community controls with a high level of accuracy based on ICD-9 codes, V-codes, and E-codes identified in EHRs. It is also the first study to our knowledge to examine health problems in a largely middle aged and older sample of individuals with ASD. This analysis found distinctive lifetime profiles of health problems among decedents with ASD compared to decedent community controls. While preliminary, these findings suggest the need for more research focused on understanding health problems, and their antecedents, in individuals with ASD across the life course.

Acknowledgments

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5. Washington, DC: American Psychiatric Association; 2013.

Bishop-Fitzpatrick L, Kind AJ. A scoping review of health disparities in autism spectrum disorder. Journal of Autism & Developmental Disorders. 2017; 47(11):3380–3391. [PubMed: 28756549]

Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.

Caamaño M, Boada L, Merchán-Naranjo J, Moreno C, Llorente C, Moreno D, … Parellada M. Psychopathology in children and adolescents with ASD without mental retardation. Journal of Autism and Developmental Disorders. 2013; 43(10):2442–2449. [PubMed: 23468070]

Cashin A, Buckley T, Trollor JN, Lennox N. A scoping review of what is known of the physical health of adults with autism spectrum disorder. Journal of Intellectual Disabilities. 2016

Christensen DL, Baio J, Braun KVN, Bidler D, Charles J, Constantino JN, … Yeargin-Allsopp M. Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. MMWR Surveillance Summaries 2016. 2016; 65(SS-3):1–23.

Croen LA, Zerbo O, Qian Y, Massolo ML, Rich S, Sidney S, Kripke C. The health status of adults on the autism spectrum. Autism. 2015; 19(7):814–823. [PubMed: 25911091]

Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Medical Care. 1998; 36(1):8–27. [PubMed: 9431328]

Esbensen AJ. Health conditions associated with aging and end of life of adults with Down syndrome. International review of research in mental retardation. 2010; 39:107–126. [PubMed: 21197120]

Esbensen AJ, Greenberg JS, Seltzer MM, Aman MG. A longitudinal investigation of psychotropic and non-psychotropic medication use among adolescents and adults with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2009; 39(9):1339–1349. [PubMed: 19434487]

Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, … Cimino JJ. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical Care. 2013; 51(8 0 3):S30. [PubMed: 23774517]

Hirvikoski T, Mittendorfer-Rutz E, Boman M, Larsson H, Lichtenstein P, Bölte S. Premature mortality in autism spectrum disorder. The British Journal of Psychiatry. 2016; 208(3):232–238. [PubMed: 26541693]

Ho HH, Eaves LC, Peabody D. Nutrient intake and obesity in children with autism. Focus on Autism and Other Developmental Disabilities. 1997; 12(3):187–192.

Hofvander B, Delorme R, Chaste P, Nydén A, Wentz E, Ståhlberg O, … Gillberg C. Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. BMC psychiatry. 2009; 9(35):1–9. [PubMed: 19133132]

Howlin P, Magiati I. Autism spectrum disorder: Outcomes in adulthood. Current Opinion in Psychiatry. 2017; 30(2):69–76. [PubMed: 28067726]

Jablensky A, Lawrence D. Schizophrenia and cancer: Is there a need to invoke a protective gene? Archives of General Psychiatry. 2001; 58(6):579–580. [PubMed: 11386987]

King M, Bearman P. Diagnostic change and the increased prevalence of autism. International Journal of Epidemiology. 2009; 38(5):1224–1234. [PubMed: 19737791]

Kuh D, Shlomo YB. A life course approach to chronic disease epidemiology. New York: Oxford; 2004.

Ladha KS, Zhao K, Quraishi SA, Kurth T, Eikermann M, Kaafarani HM, … Lee J. The Deyo-Charlson and Elixhauser-van Walraven Comorbidity Indices as predictors of mortality in critically ill patients. BMJ open. 2015; 5(9):e008990.

Leigh JP, Du J. Brief Report: Forecasting the Economic Burden of Autism in 2015 and 2025 in the United States. Journal of Autism and Developmental Disorders. 2015; 45(12):4135–4139. [PubMed: 26183723]

Lerner DJ, Kannel WB. Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. American heart journal. 1986; 111(2):383–390. [PubMed: 3946178]

Magiati I, Tay XW, Howlin P. Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. Clinical Psychology Review. 2014; 34(1):73–86. [PubMed: 24424351]

Ming X, Patel R, Kang V, Chokroverty S, Julu PO. Respiratory and autonomic dysfunction in children with autism spectrum disorders. Brain and Development. 2016; 38(2):225–232. [PubMed: 26235973]

Pu D, Shen Y, Wu J. Association between MTHFR gene polymorphisms and the eisk of autism spectrum disorders: A meta-analysis. Autism Research. 2013; 6(5):384–392. [PubMed: 23653228]

Seltzer MM, Shattuck PT, Abbeduto L, Greenberg JS. Trajectory of development in adolescents and adults with autism. Mental Retardation and Developmental Disabilities Research Reviews. 2004; 10(4):234–247. [PubMed: 15666341]

Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, Bloise R, … Condouris K. Ca v 1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell. 2004; 119(1):19–31. [PubMed: 15454078]

Starkstein S, Gellar S, Parlier M, Payne L, Piven J. High rates of parkinsonism in adults with autism. Journal of neurodevelopmental disorders. 2015; 7(29):1–15. [PubMed: 25972975]

Vorstman JA, Morcus ME, Duijff SN, Klaassen PW, Beemer FA, Swaab H, … van Engeland H. The 22q11. 2 deletion in children: High rate of autistic disorders and early onset of psychotic symptoms. Journal of the American Academy of Child & Adolescent Psychiatry. 2006; 45(9): 1104–1113. [PubMed: 16926618]

Wasey JO. icd: Tools for Working with ICD-9 and ICD-10 Codes, and Finding Comorbidities. R package version 2.0.1. 2016. http://cran.r-project.org/package=icd

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the American Medical Informatics Association. 2013; 20(1):144–151. [PubMed: 22733976]

Woolfenden S, Sarkozy V, Ridley G, Coory M, Williams K. A systematic review of two outcomes in autism spectrum disorder–epilepsy and mortality. Developmental Medicine & Child Neurology. 2012; 54(4):306–312. [PubMed: 22348343]

Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy. Journal of machine learning research. 2004; 5(Oct):1205–1224.

Lay Summary

This study looked at patterns of lifetime health problems to find differences between people with autism who had died and community controls who had died. People with autism had higher rates of most health problems, including cardiovascular, urinary, respiratory, digestive, and motor problems, in their electronic health records. They also had lower rates of cancer. More research is needed to understand these potential health risks as a large number of individuals with autism enter adulthood and middle age.
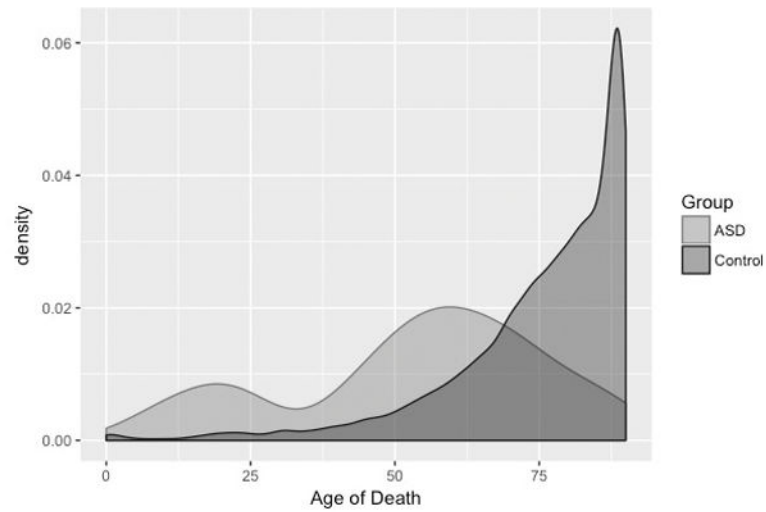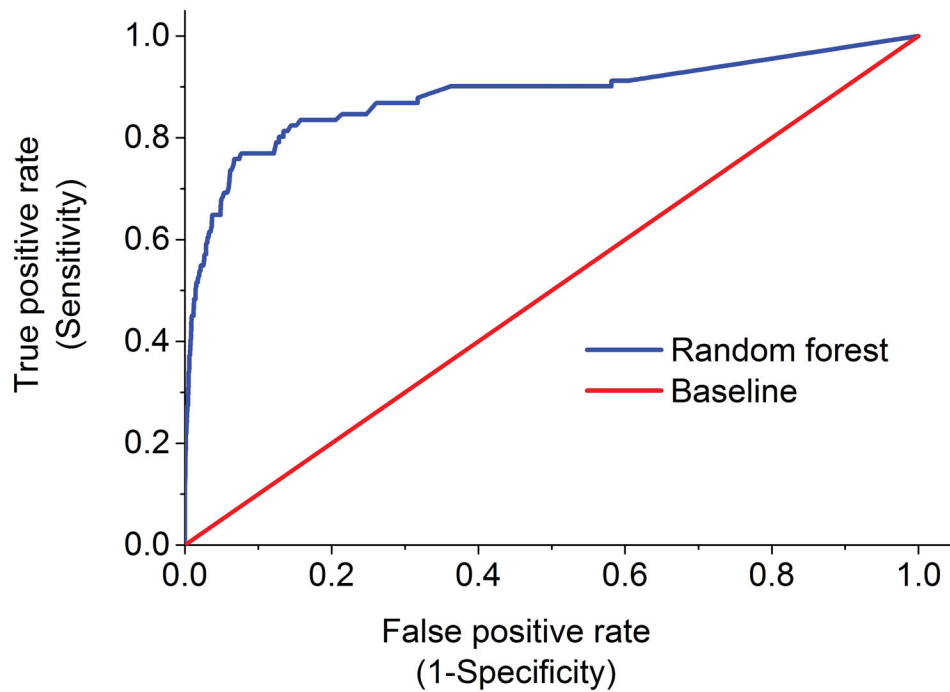
Figure 1. Age of death in decedents with autism spectrum disorder (ASD; N=131) and decedent community controls (N=16,981)

This figure displays the distribution (kernel density estimate) of age of death for the full sample of decedents with ASD and decedent community controls.

Figure 2. Random forest classifier performance based on lifetime ICD-9 codes, V-codes, and E-codes

This receiver operating characteristic (ROC) curve provides a comprehensive visualization of the performance of our predictive model. The area under the ROC curve (AUC) illustrates how well our random forest algorithm can distinguish between decedents with ASD and matched decedent community controls. The ROC curve displays the false-positive rate, or 1 – specificity versus sensitivity. The current classifier has an AUC of 0.88 which is significantly higher than the baseline AUC of 0.5.
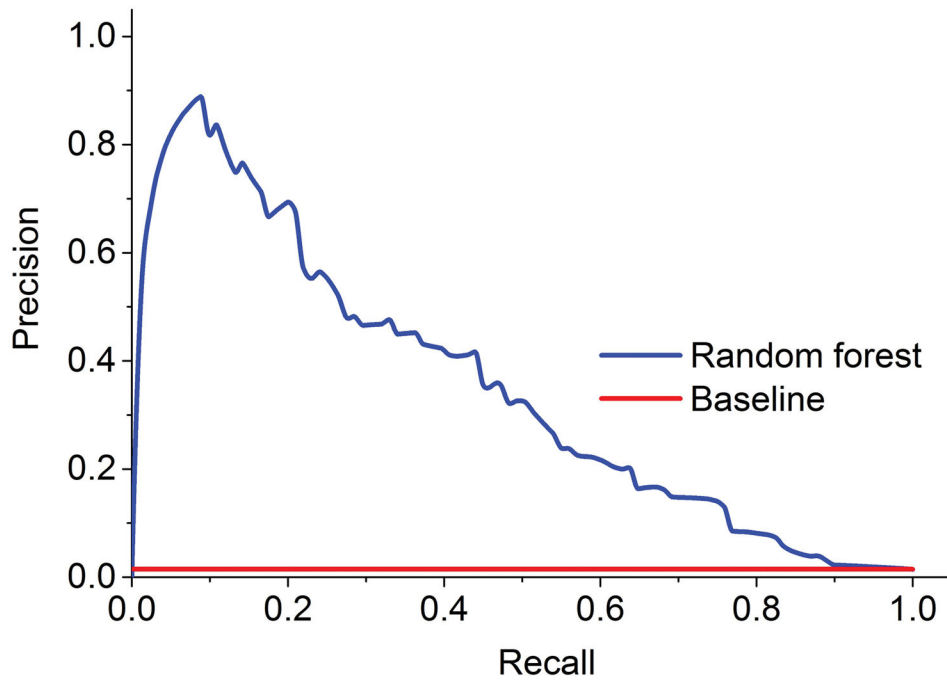
Figure 3. Random forest classifier precision-recall curve
This precision-recall curve displays the model-wide relationship between precision and recall. Precision, also called positive predictive value, represents the ratio of correctly predicted positive cases to all cases that have been predicted as positive by the classifier. Recall (sensitivity) represents the ratio of correctly predicted target cases in relation to all cases in the target class.

Table 1

ICD-9 Codes, V-Codes, and E-Codes with Fifty Highest Information Gain Scores

Code	IG	Control	Autism	Description
345.9	0.01200	138 (2.2)	29 (31.9)	Epilepsy, unspecified, without mention of intractable epilepsy
780.39	0.01158	264 (4.3)	35 (38.5)	Other convulsions
507	0.01020	189 (3.1)	29 (31.9)	Pneumonitis due to solids and liquids
380.4	0.00878	843 (13.6)	48 (52.7)	Impacted cerumen
V58.69	0.00836	2242 (36.2)	73 (80.2)	Long-term (current) use of other medications
333.82	0.00826	3 (0.0)	10 (11.0)	Orofacial dyskinesia
783.4	0.00733	25 (0.4)	13 (14.3)	Lack of expected normal physiological development in childhood
V74.1	0.00717	244 (3.9)	26 (28.6)	Screening examination for pulmonary tuberculosis
788.3	0.00602	427 (6.9)	30 (33.0)	Urinary incontinence, unspecified
788.9	0.00572	326 (5.3)	26 (28.6)	Other symptoms involving urinary system
999.9	0.00549	83 (1.3)	15 (16.5)	Other and unspecified complications of medical care, not elsewhere classified
486	0.00509	1493 (24.1)	52 (57.1)	Pneumonia, organism unspecified
345.1	0.00506	63 (1.0)	13 (14.3)	Generalized convulsive epilepsy
758	0.00502	5 (0.1)	7 (7.7)	Chromosomal anomalies
331	0.00496	290 (4.7)	23 (25.3)	Other cerebral degenerations
343.9	0.00481	12 (0.2)	8 (8.8)	Infantile cerebral palsy, unspecified
782.1	0.00478	688 (11.1)	34 (37.4)	Rash and other nonspecific skin eruption
345.91	0.00470	13 (0.2)	8 (8.8)	Epilepsy, unspecified, with intractable epilepsy
385.89	0.00465	74 (1.2)	13 (14.3)	Other disorders of middle ear and mastoid
V71.89	0.00449	764 (12.4)	35 (38.5)	Observation and evaluation for other specified suspected conditions
564	0.00439	908 (14.7)	38 (41.8)	Functional digestive disorders not elsewhere classified
599	0.00411	1558 (25.2)	50 (54.9)	Other disorders of urethra and urinary tract
934.9	0.00409	43 (0.7)	10 (11.0)	Foreign body in respiratory tree, unspecified
V81.2	0.00408	325 (5.3)	22 (24.2)	Screening for other and unspecified cardiovascular conditions
372.3	0.00404	360 (5.8)	23 (25.3)	Other and unspecified conjunctivitis
521	0.00401	466 (7.5)	26 (28.6)	Diseases of hard tissues of teeth
525.9	0.00385	123 (2.0)	14 (15.4)	Unspecified disorder of the teeth and supporting structures
780.6	0.00382	1343 (21.7)	45 (49.5)	Fever and other physiologic disturbances of temperature regulation
V67.59	0.00382	642 (10.4)	30 (33.0)	Other follow-up examination
369	0.00380	24 (0.4)	8 (8.8)	Blindness and low vision
345.11	0.00376	15 (0.2)	7 (7.7)	Generalized convulsive epilepsy, with intractable epilepsy
V70.3	0.00370	357 (5.8)	22 (24.2)	Other general medical examination for administrative purposes
V04.81	0.00368	1532 (24.8)	48 (52.7)	Need for prophylactic vaccination and inoculation against influenza
244.9	0.00363	708 (11.4)	31 (34.1)	Unspecified acquired hypothyroidism
959.6	0.00362	158 (2.6)	15 (16.5)	Hip and thigh injury
380.1	0.00355	305 (4.9)	20 (22.0)	Infective otitis externa
596.54	0.00331	83 (1.3)	11 (12.1)	Neurogenic bladder NOS

Code	IG	Control	Autism	Description
372	0.00330	104 (1.7)	12 (13.2)	Disorders of conjunctiva
780.79	0.00324	1859 (30.1)	52 (57.1)	Other malaise and fatigue
781.2	0.00322	638 (10.3)	28 (30.8)	Abnormality of gait
787.03	0.00311	380 (6.1)	21 (23.1)	Vomiting alone
V70.0	0.00301	1793 (29)	50 (54.9)	Routine general medical examination at a health care facility
781	0.00300	356 (5.8)	20 (22.0)	Symptoms involving nervous and musculoskeletal systems
781.3	0.00296	361 (5.8)	20 (22.0)	Lack of coordination
733	0.00291	516 (8.3)	24 (26.4)	Other disorders of bone and cartilage
276	0.00286	152 (2.5)	13 (14.3)	Disorders of fluid electrolyte and acid-base balance
564.7	0.00273	10 (0.2)	5 (5.5)	Megacolon, other than Hirschsprung's
933.1	0.00269	68 (1.1)	9 (9.9)	Foreign body in larynx
333.9	0.00266	21 (0.3)	6 (6.6)	Other and unspecified extrapyramidal diseases and abnormal movement disorders
742.1	0.00265	4 (0.1)	4 (4.4)	Microcephalus
345.9	0.01200	138 (2.2)	29 (31.9)	Epilepsy, unspecified, without mention of intractable epilepsy

Note. Frequencies represent the number of cases and controls with each diagnosis. IG = information gain; ASD = autism spectrum disorder; N = number

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Age of Death and Comorbidities in Decedents with Autism Spectrum Disorder and Matched Decedent Community Controls

	ASD (n = 91)	Control (n = 6186)	t or χ^2	p-value
Intellectual disability, *N (%)*	69 (75.8)	0 (0.0)	4741.86	<0.001
Down syndrome, *N (%)*	7 (7.7)	0 (0.0)	476.30	<0.001
Male, *N (%)*	53 (58.2)	3602 (58.2)	0.00	0.99
Length of EHR, *mean (SD)*	16.8 (8.1)	12.4 (9.7)	−4.29	<0.001
Adjusted lifetime codes, *mean (SD)*[A]	82.9 (59.2)	71.1 (58.7)	−4.16	<0.001
Year of birth, *mean (SD)*	1940.3 (14.0)	1936.4 (15.3)	−2.41	0.02
Age of death, *mean (SD)*	67.3 (13.5)	68.7 (15.1)	0.86	0.39
Elixhauser category, *N (%)*			OR[B]	p-value
AIDS/HIV	2 (2.2)	218 (3.5)	0.61	0.49
Alcohol abuse	3 (3.3)	640 (10.3)	0.28	0.03
Blood loss anemia	4 (4.4)	176 (2.8)	1.64	0.34
Cardiac arrhythmias	40 (42.9)	2742 (44.3)	1.04	0.87
Chronic pulmonary disease	37 (40.7)	2304 (37.3)	1.18	0.45
Coagulopathy	24 (26.4)	854 (13.8)	2.22	<0.001
Congestive heart failure	34 (37.4)	1603 (25.9)	1.90	0.004
Deficiency anemia	50 (54.9)	2020 (32.7)	2.63	<0.001
Diabetes, complicated	7 (7.7)	741 (12.0)	0.63	0.24
Diabetes, uncomplicated	8 (8.8)	880 (14.2)	0.60	0.17
Drug abuse	1 (1.1)	205 (3.3)	0.32	0.26
Fluid and electrolyte disorders	56 (61.5)	2306 (37.3)	2.79	<0.001
Hypertension	35 (38.5)	3209 (51.9)	0.60	0.02
Hypothyroidism	32 (35.2)	717 (11.6)	4.76	<0.001
Liver disease	5 (5.5)	410 (5.5)	0.80	0.64
Lymphoma	1 (1.1)	243 (3.9)	0.27	0.19
Metastatic cancer	4 (4.4)	1202 (19.4)	0.19	<0.001
Obesity	19 (20.9)	1576 (25.5)	0.77	0.33
Other neurological disorders	56 (61.5)	1079 (17.4)	7.61	<0.001
Paralysis	16 (17.6)	393 (6.4)	3.18	<0.001
Peptic ulcer disease	9 (9.9)	492 (8.0)	1.34	0.41
Peripheral vascular disorders	12 (13.2)	1312 (21.2)	0.60	0.10
Pulmonary circulation disorders	3 (3.3)	483 (7.8)	0.42	0.15
Rheumatoid arthritis	4 (3.3)	487 (7.9)	0.55	0.25
Renal failure	15 (16.5)	1148 (18.6)	0.91	0.75
Solid tumor without metastasis	14 (15.4)	1115 (18.0)	0.88	0.65
Valvular disease	26 (28.6)	995 (16.1)	2.33	<0.001

	ASD (n = 91)	Control (n = 6186)	t or χ^2	p-value
Weight loss	17 (18.7)	602 (9.7)	2.12	0.006

Note. ASD = autism spectrum disorder; OR = odds ratio; EHR = electronic health record.

[A]Least square means estimate adjusted for length of EHR.

[B]Odds ratio adjusts for age and sex.