Large Databases for Pediatric Research on Children with Autism Spectrum Disorder

Article *in* Journal of Developmental & Behavioral Pediatrics · December 2017
DOI: 10.1097/DBP.0000000000000534

9 authors, including:

Karen A Kuhlthau
Massachusetts General Hospital
139 PUBLICATIONS **3,378** CITATIONS
SEE PROFILE

Charles Bailey
The Children's Hospital of Philadelphia
50 PUBLICATIONS **1,322** CITATIONS
SEE PROFILE

Briella Lexene Baer
Kennedy Krieger Institute
1 PUBLICATION **1** CITATION
SEE PROFILE

Daniel Lee Coury
Nationwide Children's Hospital
118 PUBLICATIONS **1,878** CITATIONS
SEE PROFILE

Large Databases for Pediatric Research on Children with Autism Spectrum Disorder

Karen A. Kuhlthau, PhD,* L. Charles Bailey, MD, PhD,† Briella L. Baer, MHS,‡ Daniel L. Coury, MD,§ J. Kiely Law, MD, MPH,‡‖ Donna S. Murray, PhD,¶ Hanieh Razzaghi, MPH, Christopher B. Forrest, MD, PhD,† Paul H. Lipkin, MD**

ABSTRACT: ***Objective:*** **This article reviews the data available in 3 large databases for use in conducting studies of children with autism spectrum disorder (ASD).** ***Methods:*** **The article describes the data structure, data elements, and strengths and weaknesses of the 3 data sets.** ***Results:*** **Each of the 3 data sets, the Interactive Autism Network (IAN), the Autism Treatment Network (ATN), and PEDSnet have large cohorts of children with ASD. IAN has strengths in patient-reported measures, ATN in clinical characterization, and PEDSnet in health care encounters and electronic medical record data.** ***Conclusion:*** **The data sets described here have potential for further studies that could help improve the care and well-being of children with ASD and their families.**

(*J Dev Behav Pediatr* 0:1–9, 2017) **Index terms:** autism, database, epidemiology.

Autism spectrum disorder (ASD) is a neurodevelopmental disability characterized by persistent social and communication impairments, restricted interests, and repetitive patterns of behavior or activities. Reported prevalence of ASD has been increasing in recent years, currently estimated at 1 in 68 children (https://www.cdc.gov/ncbddd/autism/data.html accessed 18/30). Children and youth with ASD are commonly affected by cooccurring conditions including mental health conditions, gastrointestinal issues, sensory difficulties, sleep difficulties, epilepsy, and immune abnormalities.[1–6] The literature on these conditions among the population of children and youth with ASD is still emerging.

Applied clinical research is aimed at generating new knowledge to inform patient and clinician decision making and is challenging to do among children with ASD. Adequate sample size is a key limitation. Many studies are conducted at a single or a small number of clinical sites, which restricts generalizability because of limited diversity and representativeness. The phenotypic expression of ASD is heterogeneous, which is another reason why large samples are needed for future research. Diagnostic and behavioral assessments of children with ASD are conducted by a variety of professionals in different settings using different measures, which precludes facile aggregation of results. Lack of completion of evaluations or treatments is a common problem in clinical care, as is loss to follow-up, further threatening the validity of research done with this population.

Several trends in applied clinical research offer options for improving ASD research. First is the trend toward people-centered research, which engages patients and families in the design of research questions and methodologies, the conduct of the study, interpretation of results, feedback of results to participants, and dissemination and implementation of the evidence generated.[7] This can lead to research questions, measured outcomes, and research processes that make recruitment and long-standing participation in studies easier and more meaningful for youth and families. Second, the emergence of big data as a scientific approach in health sciences provides new opportunities to advance knowledge on the optimal approaches for diagnosing and managing childhood health conditions such as ASD.[8,9] Big data involve the integration of large amounts of clinical data obtained from electronic health records (EHRs), patient-generated data sources, and biospecimens, each aggregated across multiple institutions for thousands of children. Third, large, multi-institutional clinical registries are increasingly used as the basis for research and quality improvement and can speed the process of research and translation of research findings to clinical improvements.

This article describes 3 research networks that have produced large databases that can be used for applied clinical research for children with ASD: the Interactive

From the *Division of General Pediatrics, Department of Pediatrics, Massachusetts General Hospital and Harvard Medical School, Boston, MA; †Children's Hospital of Philadelphia and Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ‡Interactive Autism Network (IAN) at Kennedy Krieger Institute, Baltimore, MD; §The Ohio State University, Developmental and Behavioral Pediatrics, Nationwide Children's Hospital, Columbus, OH; ‖Autism Speaks, Cincinnati Children's Hospital, Division of Developmental and Behavioral Pediatrics, Boston, MA; ¶Children's Hospital of Philadelphia, Philadelphia, PA; **Interactive Autism Network (IAN) at Kennedy Krieger Institute, Johns Hopkins Medicine, Baltimore, MD.

Received June 2017; accepted October 2017.

Disclosure: The authors declare no conflict of interest.

Address for reprints: Karen A. Kuhlthau, PhD, Division of General Pediatrics, Department of Pediatrics, Massachusetts General Hospital and Harvard Medical School, 125 Nashua St, 8th Floor, Boston, MA 02114; e-mail: kkuhlthau@mgh.harvard.edu.

Autism Research Network (IAN); the Autism Treatment Network (ATN); and the PEDSnet. We describe the history of the networks, their population, ASD-related variables, the strengths of each data system, and the types of research and questions that might be addressed by each system. These networks are unique in design and can provide researchers large databases for examining key questions around clinical manifestations, history, and care of children with ASD, while providing clinicians and families new patient-centered clinical insights into ASD. We chose these databases because they are among the largest databases from the United States/North America, with clinically well-characterized populations of children with ASD. Other sources of data, for example, from the Simons Foundation and Autism BrainNet are available but are not examined.

The Interactive Autism Network

IAN was founded in 2006 by physician-parents of a child with autism, with an aim to accelerate autism research by engaging families as critical stakeholders in all steps of the research process. IAN is a family-centered research network, enrolling children and adults with ASD along with parents and siblings. Current enrollment is 55,000 participants, including 14,500 children (1.5% of the US population of children with ASD) and 7500 adults with ASD. Launched with grant funding from BLINDED, IAN is currently a partnership project of BLINDED and BLINDED. It receives additional support from the Patient-Centered Outcomes Research Institute (PCORI).

IAN is unique in its data collection through direct engagement with families, who provide valuable input about their experiences with ASD through participant-reported outcomes in ways that are not measured through medical records, researcher or clinician report, or diagnostic tools alone. As such, all IAN pediatric data are obtained directly from a parent through the IAN online research portal (www.IANresearch.org). Parents serve as a proxy for children and youth. IAN recruits individuals through community outreach by maintaining an online search presence and social media pages, attending autism walks or community events, and managing an open online community-centered website (www.IANcommunity.org), which provides a platform for families to join IAN, while also providing family-friendly articles summarizing research efforts, quarterly newsletters, webinars, and other ASD resources. IAN recruits 3000 enrollees each year through its registration portal, of which 40% are children with ASD. In addition to children and their families, IAN also enrolls both independent adults and dependent adults with ASD (by a legally authorized representative). Families are invited to update their information regularly, which provides a longitudinal resource. IAN has enrolled families from 100 countries around the world, with 95% of participants located in the United States.

All childhood probands in IAN must have parental attestation of the child's professional diagnosis of ASD, autism/autistic disorder, Asperger syndrome, childhood disintegrative disorder (CDD), pervasive developmental disorder-not otherwise specified (PDD-NOS), or pervasive developmental disorder (PDD). If the child has not received a diagnosis or has received a diagnosis of Rett Syndrome, he or she is not eligible to join IAN. All diagnostic information is based on parent-proxy or legal guardian-proxy report; however, multiple verification studies have demonstrated the validity of this type of data collection within the IAN network.[10,11] In addition, because of the English-only content, the child's parent must have the ability to understand written English to complete necessary questionnaires and research consent form. The consent form, approved by the BLINDED Institutional Review Board, is an online electronic consent on a secure platform. It is centered on the collection of protected health information and other protected data from the participant (parent or independent adult) about herself or himself or as a parent proxy for the child with ASD or siblings across a secure internet connection. The consent also grants permission for recontacting of the parent participant through email or traditional mail about other IRB-approved research projects. Through the consent for recontacting of participants, IAN allows participants repeated engagement in research, while also providing researchers access to thousands of highly motivated families and the ability to directly communicate with them about their experiences and research priorities. This ability to recontact makes IAN additionally rich as a longitudinal source of data.

Data collected during enrollment into the registry include medical history, sociodemographics, contact information, behavior and developmental history of the child, treatments and out-of-pocket costs, the child's utilization and parents' satisfaction with medical, educational, and social environments, and the family's interest or previous experience in ASD research studies. Parents complete the Social Responsiveness Scale (SRS)[12] and the Social Communication Questionnaire (SCQ)[13] on their child's behalf. The proxy-report for these measures has been verified as a valid source of ASD diagnostic information.[10,11] For example, Lee et al.[10] found that of parent-reported diagnoses (through the SCQ), 98.1% were validated by the clinician's best estimate, 99.1% by the ADI-R, 89.7% by the ADOS, and 100% by either the ADI-R or ADOS.

Among pediatric registries, a unique characteristic of IAN is that parents and siblings are also consented into the network and data are collected about each family member. Sibling data mirrors that of the child with ASD and includes developmental, medical and educational history, along with the SRS and SCQ. These data expand the utility of the registry data by providing a non-ASD comparison cohort.[14,15]

IAN research data can be accessed by IAN staff or as a deidentified data set through an application on the IAN website (www.IANcommunity.org/data_services). All researchers must provide evidence of an IRB approval

for their study protocol; a signed IAN-specific Data Access Agreement is also required.

The Autism Treatment Network

The ATN was founded in 2005 through a partnership between BLINDED and BLINDED. The ATN is funded by Autism Speaks and currently consists of 13 autism programs from academic medical centers across the United States and Canada. A core function of the network is to improve care of children with ASD by developing and disseminating a best practice model of interdisciplinary medical care. A patient registry was designed to inform the implementation of the care model and serve as a platform for research by examining the trajectory of co-occurring medical conditions and their impact on behavioral outcomes in children with ASD. The ATN serves as platform for the Health Resources and Services Administration's Autism Intervention Research Network on Physical Health (AIR-P). The ATN and registry have been described previously[16–18] and will be summarized here.

The ATN registry data are collected at each clinical site (a total of 19 sites over ATN history). Enrollment includes children and youth aged between 2 and 18 years, with a diagnosis of ASD as determined by the Autism Diagnostic Observation Schedule (ADOS), second edition (ADOS-2), the Diagnostic and Statistical Manual of Mental Disorders Fourth or Fifth Edition (DSM-IVand DSM5), and clinician consensus. Parent and clinicians contribute data at baseline and annually thereafter. Data collected in the registry include medications, laboratory results, and health and behavioral assessments. The sample includes 7029 unique individuals with medical and behavioral data, including a growing subset with longitudinal and biospecimen data.

The ATN registry data include longitudinal data on over 2000 of the 7029 individuals enrolled. The median number of visits is 2 with an average follow-up time of 2.26 years. The data set is designed to provide data for secondary analyses using only registry data. The registry data set also serves as a platform for further research by allowing researchers to select a sample of individuals with specific characteristics or by allowing researchers to add information to the already rich registry data set.

Enrollees in the ATN registry are patients at 1 of the clinical centers who plan to receive ongoing care at the center. Parents must speak English or Spanish and must speak 1 of these languages at least 75% of the time with their children. The lack of availability of all measures in the ATN battery in languages other than English and Spanish has prevented enrollment of some cultural groups. Some conditions (e.g., blindness) are exclusion criteria. Data are obtained by parent-proxy report and clinical assessment. Data are collected every year when a child returns to the clinic for ongoing care. Parents also complete assessments at home through an online parent portal or on paper to complete and mail back to the clinic.

In addition to the diagnostic assessments, registry data include sociodemographics, developmentally appropriate cognitive assessment, the Child Behavior Checklist, the Vineland Adaptive Behavior Scale, the PedsQL, the Aberrant Behavior Checklist, and a communication measure. ATN data have the advantage of including both clinician-completed assessments and those completed by parents, depending on the measure. In addition, information is collected on whether the individual has a co-occurring health condition (obesity based on height and weight, seizures, mental health conditions, and gastrointestinal conditions). Data can be accessed through the ATN website (http://asatn.org/request/data). Initial queries of the registry can be conducted through http://asatn.org/asatn-query.

PEDSnet

PEDSnet (http://pedsnet.org) was founded in 2014 with support from the Patient-Centered Outcomes Research Institute, as a collaboration among 8 large academic pediatric health systems.[19,20] The network conducts observational research and clinical trials across multiple pediatric specialties in both inpatient and outpatient settings, and has produced reusable and expandable governance, logistical, informatics, regulatory, scientific, and training resources, organized as a Pediatric Research Commons. PEDSnet is governed by parents and senior leaders of the 8 founding health systems.

PEDSnet has created a longitudinal data resource comprising 5 million children, which cuts across pediatric disease specialties. On a quarterly basis, electronic health records for all children seen at a member institution since 2009 are extracted, transformed to a common data model, and merged into the data network. Data domains include demographics, encounter data for primary care, specialty care, emergency department and inpatient visits, procedures ordered, medications prescribed, laboratory results, diagnoses assigned, anthropometrics, and vital signs. The data are used to evaluate the feasibility of clinical studies, conduct large-scale observational research, and recruit patients into surveys, clinical research studies, and clinical trials.

Within this large database, PEDSnet has identified numerous patients aged 2 to 24 years with ASD . Inclusion diagnoses are shown in Table 1; a child was included if a qualifying diagnosis was assigned at 2 separate visits or once at a developmental/behavioral pediatrics visit. This approach yielded a prevalence of 15 per 1000 in the PEDSnet population, with a male:female ratio of 3.8, both of which align well with current epidemiologic estimates for childhood ASD.

The PEDSnet core data resource provides detailed information about utilization of pediatric health services. Because it spans specialties and care settings, it offers a broad view that can illuminate the range and impact of comorbidities on the health of children with ASD. It also includes a number of primary observations, such as test results and vital signs that can be incorporated directly

Table 1. Inclusion Diagnoses for PEDSnet ASD Cohort

Diagnosis Name	SNOMED-CT Diagnostic Code
Active infantile autism	191689008
Asperger disorder	23560001
Autistic disorder	408856003
Childhood disintegrative disorder	71961003
Pervasive developmental disorder	35919005
Residual disintegrative psychoses	191693002
Residual infantile autism	191690004
Rett disorder	68618008

into phenotypes rather than relying solely on assigned diagnoses. The data standardization and quality assessment process used to construct the resource facilitates the conduct of analyses across multiple sites and times. Finally, PEDSnet's core data can be linked to records at member institutions, allowing retrieval of additional information from medical records or contact with families for participation in clinical studies. The network contains similar information for large groups of healthy children or children with other chronic diagnoses, providing the opportunity to assemble control cohorts for comparison to children with ASD.

For children meeting the inclusion criteria described above, electronic health record data standardized in the PEDSnet Common Data Model (https://pedsnet.org/data/common-data-model/) are available through the PEDSnet core data resource (https://pedsnet.org/research/access-pedsnet/). Demographic data include date of birth, sex, race/ethnicity, home ZIP code, and broad class of health insurance coverage. Encounter records describe dates, visit types (e.g., ambulatory, inpatient, or ED) and clinic/provider specialty. Both visit diagnoses and problem list entries are available, standardized to SNOMED-CT. Information about hospital-administered and prescribed medications are included, and for a subset of patients, outpatient dispensing data are also available. Similarly, procedure codes for inpatient and outpatient care, including provider level of service codes, are collected. Finally, anthropometrics, blood pressure, and results of approximately 400 common laboratory tests are included in the data set.

To limit risk to privacy, data are stored as a limited data set, and free text such as clinical notes are not included. However, data such as notes, survey instruments, and clinical assessments can be reached when needed by linking core data records back to institutional EHRs. The entry point for requesting data is https://pedsnet.org/research/access-pedsnet/.

Contrasts Among the 3 Large Databases

The 3 databases share several commonalities: a commitment to parent and patient involvement; large sample size; longitudinal records; and a commitment to supporting applied clinical research. Each has diverse racial, ethnic, and geographic representation with broad coverage across the United States. Although each of the data sets has different process for sharing, they each offer researchers some access to the data for the conduct of new studies and the replication of existing findings. They offer a great opportunity to learn more about the population with over 40 articles published with IAN data and 30 from ATN registry data as of the end of 2016. (Publication lists can be found at https://www.iancommunity.org/cs/ian_research_publications/scientific_journal_publications and http://airpnetwork.org/what-we-do/publications). PEDSnet has not had sufficient time to generate ASD-related publications.

The data sets differ in the source of data: IAN obtains direct parent-report data from outside a clinical setting; ATN is derived from clinical samples with children well characterized by clinical- and parent-reported measures; and PEDSnet comprises longitudinal electronic health records from children's hospitals. Each sample has potential biases. For IAN, families need to know about IAN and be motivated to join the network by submitting core data elements and be willing to participate in additional studies. IAN thus only includes families who volunteer to participate and does not include those who do not have computer literacy or have barriers to inputting data through an internet platform, including socioeconomic or educational barriers. In addition, as noted earlier, as an English-only research site, IAN data do not include non-English speakers. Some researchers also have concerns about bias resulting from parent report, which is based primarily on home and community experience. However, in contrast to other observers, a parent offers a longer duration of daily and lifelong experience with the child and provides insights into a child's functioning at home, in the community, and therapy settings. Families in the ATN data need to receive care in 1 of the affiliated specialty care clinics, and thus, it excludes those without access to those specialty care centers. PEDSnet is similarly constrained to a sample of children who use care affiliated with a children's hospital, but it also includes families who only receive primary care/those who do not seek specialty autism care. PEDSnet data lack key developmental data and are limited to items that get recorded in medical records.

Each has key strengths that suggest opportunities for different types of research. IAN has exceptionally strong patient and family engagement, and parents provide the data included in the IAN database. IAN is thus a strong platform for examining issues where the parent is the appropriate reporter such as its study of the previously unreported problems of wandering and elopement[21] and its study of bullying.[22] It also offers sibling data for use as controls or comparison, as in these previously noted studies or its twin studies[23]; ATN has strong clinical data, including clinically validated diagnostic data, partnered with clinical assessments and questionnaires that are

Table 2. Available Data Elements Across 3 Unique Data Networks Containing Clinical Data From Pediatric Patients Diagnosed with Autism

Available Data Across All Research Networks	PEDSnet	Interactive Autism Network	Autism Treatment Network
Demographics			
Age	✓	✓	✓
Gestational age	✓	✓	✓
Race	✓	✓	✓
Ethnicity	✓	✓	✓
Sex	✓	✓	✓
Death information (date and cause of death)	✓	✓	
Location (3-digit zip)	✓	✓	✓
Total number living in household		✓	
Parent or household education		✓	✓
Parent or household income		✓	✓
Parent or household employment status		✓	
Respondent sex		✓	✓
Parent or household military status		✓	
Parent marital status		✓	
Family autism spectrum disorder status (multiplex vs simplex)		✓	
Use of providers by type			
Primary care providers	✓		✓
Specialty care providers	✓		
Other provider (e.g., education services, chirpractic, etc)	✓		✓
Health care utilization			
Inpatient encounters	✓		✓
Emergency department encounters	✓		
Outpatient encounters	✓		
Dates of visits	✓		✓
Diagnoses			
ASD diagnosis	✓	✓	✓
Other diagnoses	✓	✓	✓
Dates of diagnosis	✓		✓
Parental concerns		✓	✓
Prescribed medications			
All drug exposures	✓		✓
Date prescribed	✓		
Route of administration	✓		
Dose	✓		
Primary indication			✓
Dispensed medications			
All drug exposures	✓		
Date dispensed	✓		
Route of administration	✓		
Procedures			
Procedure code	✓		
Date of procedures	✓		
Laboratory Values			

(*Table continues*)

Table 2. Continued

Available Data Across All Research Networks	PEDSnet	Interactive Autism Network	Autism Treatment Network
Common laboratory tests	✓		
Organisms related to laboratory cultures taken from patient	✓		✓
Date of laboratory tests	✓		✓
Health indicators			
Height/weight/BMI	✓		✓
Blood pressure (systolic/diastolic)	✓		
Tobacco use	✓		
Birth weight		✓	✓
Head circumference	✓		✓
Tanner stage			✓
Neurologic evaluation			✓
Dysmorphology examination			✓
Payer			
Plan type (Health maintenance organization, preferred provider organization, and point of service)	✓		✓
Plan class (private, Medicaid, Medicare, etc)	✓	✓	✓
Standardized instruments commonly used in ASD research			
Social Responsiveness Scale (SRS)		✓	✓
Social Communication Questionnaire (SCQ)—lifetime		✓	✓
DSM-IV/DSM-5 checklist			✓
Clinical global impression			✓
Autism impact measure			✓
Caregiver strain			✓
Patient activation measure and developmental disabilities (PAM 13 DD)			✓
Gastrointestinal (GI) Symptoms Inventory			✓
Aberrant Behavior Checklist			✓
Autism Diagnostic Interview–Revised (ADIR)			limited
Autism Diagnostic Observation Schedule (ADOS)			✓
Cognitive assessments			✓
Communication assessments			✓
Child Behavior Checklist			✓
Pediatric Quality of Life Inventory (PedsQL)			✓
Short Sensory Profile			✓
Vineland			✓
Patient-Reported Outcomes Measurement Information System (PROMIS) Global Health		✓	
Developmental and autism-specific data			
Age walked independently		✓	✓
Age at first words		✓	✓
Age at 2 to 3 words of meaningful speech		✓	✓
National Database for Autism Research Global Unique Identifier Elements			
First name		✓	
Middle name (y/n) (if Y, as on birth certificate)		✓	
Last name (as on birth certificate)		✓	
Sex (as on birth certificate)		✓	✓
City or municipality at birth		✓	

ASD, autism, spectrum disorder, BMI, body mass index.

Table 3. Database Overview and Participant Demographic Information Across 3 Unique Data Networks Containing Clinical Data From Pediatric Patients Diagnosed with Autism

	PEDSnet		Interactive Autism Network		Autism Treatment Network[b]	
Database Overview	**Count**	**% of Total**	**Count**	**% of Total**	**Count**	**% of Total**
Total no. of unique participants, by year						
pre-2009	12,315	14.5	10,150	55.2	519	7.4
2009	9135	10.8	1455	7.9	1228	17.5
2010	9593	11.3	1463	8.0	1294	18.4
2011	8617	10.2	1166	6.3	1040	14.8
2012	8636	10.2	438	2.4	1344	19.1
2013	8912	10.5	288	1.6	1060	15.1
2014	9264	10.9	1745	9.5	426	6.1
2015	9950	11.7	860	4.7	43	0.6
2016	8429	9.9	832	4.5	75	1.1
Total	84,851	100	18,397	100.0	7029	100.0
Median no. of observations per participant	21	NA	NA	NA	2	NA
Median duration of patient follow-up, yr	5.2	NA	NA	NA	2.26	NA
Age, yr						
Current age						
0–<2	76	0.1	1	0.0	0	0.0
2–<5	6473	7.6	307	1.7	27	0.38
5–<10	23,101	27.2	2190	11.5	2259	32.1
10–<15	25,843	30.5	6554	35.6	3102	44.1
15–<20	19,227	22.7	6120	33.3	1113	15.8
20–<25	7985	9.4	2675	14.5	297	4.2
25+	2146	2.5	633	3.4	8	0.11
Median age	12.2	NA	15.1	NA	11.5	NA
Not reported	0	0.0	0	0.0	223	3.2
Age at diagnosis						
0–<2	2306[a]	2.8	1651	9.0	266	3.8
2–<5	28,584[b]	33.7	6761	36.8	3940	56.1
5–<10	28,374[b]	33.4	2272	12.4	1941	27.6
10–<15	17,644[b]	20.8	538	2.9	610	8.7
15–<20	6960[b]	8.2	32	0.2	104	1.5
20–<25	688[b]	0.8	0	0.0	0	0.0
25+	295[b]	0.3				
Median age	6.75[b]	NA	3.1	NA	4.19	NA
Not reported	NA	NA	7143	38.8	168	2.4
Sex						
Male	67,194	79.2	14,973	81.4	5835	83.0
Female	17,657	20.8	3424	18.6	1150	16.4
Other	0	0.0	0	0.0	0	0.0
Not reported	0	0.0	0	0.0	44	0.63
Race						
White	57,620	68.0	15,175	82.5	5245	74.6
Black or African-American	9541	11.2	796	4.3	496	7.1
Asian	2996	3.5	299	1.6	345	4.9

(*Table continues*)

Table 3. Continued

Database Overview	PEDSnet		Interactive Autism Network		Autism Treatment Network[b]	
	Count	% of Total	Count	% of Total	Count	% of Total
Multiple race	2148	2.5	943	5.1	434	6.2
Other	7511	8.9	898	4.9	55	0.8
Unknown	4243	5.0	36	0.2	0	0.0
Not reported	792	0.9	250	1.4	454	6.5
Ethnicity						
Not Hispanic or Latino	69,042	81.3	16,345	88.9	5943	84.6
Hispanic or Latino	7906	9.4	1903	10.3	705	10.0
Other	1302	1.5	0	0.0	0	0.0
Not reported	6601	7.8	149	0.8	381	5.4

[a]For PEDSnet, this is the age at the earliest diagnosis recorded in the data; the actual date of initial diagnosis may be earlier, if it occurred outside the network. [b]

based on parent report. These data are well suited to studies that examine the clinical characteristics of children along or in combination with an examination of treatment- or parent-reported behavior. Example studies include a study of sleep difficulties and medication use,[24] studies of the medical and behavioral characteristics of children and how they are associated with quality of life,[25] and assessment and treatment of anxiety in youth with ASD.[26] PEDSnet has strong health service utilization. It is best used to examine the utilization and quality of care. These differences offer opportunities to address clinically relevant questions from different perspectives, which hold the potential for expanding the literature, ultimately contributing to improved care for individuals with ASD. Each of these data sets has the potential of being expanded by adding in additional data or additional sources of data. They can also be used as a platform for identifying patients for clinical trials.

Table 2 shows the variables available in each data set. Pediatric-specific variables collected by IAN include those that can be obtained from parents. ATN includes family-reported measures (e.g., reports of behaviors) and measures collected in a clinical setting (e.g., IQ). PEDSnet's strength is in the data collected from medical records, including health status and health care utilization. The IAN registry of consented participants makes recontact for research relatively direct. Individuals in the ATN registry can be recontacted through the clinical sites. Recruitment of participants in the PEDSnet data set can be done by identifying participants eligible for a study in the limited data set, reidentifying those participants in the local institution, and then recruiting participants locally in collaboration with care centers.

Information on co-occurring conditions is collected from families in IAN, from families and/or clinician report in ATN, and from clinical records in PEDSnet. Families may know of conditions that have not been recorded in medical records, for example, because they were diagnosed outside of a health care center-based system. At the same time, parental reports can be limited by memory or recall bias. Clinician reports outside of the medical record are also subject to clinician's memory, recall bias, or ability to cull the information from the medical record. Similarly, clinical workflows, terminologies available in EHRs, and nonclinical drivers such as regulatory mandates or reimbursement may drive differential collection of data in medical records. Although co-occurring condition data exist in all 3 of these databases, each has strengths and weaknesses and the use of 1 versus another may be driven by the research question or specific conditions of interest.

Table 3 shows the characteristics of samples from each of the databases. There is a large difference in total ASD subject size: about 7000 children in ATN, 18,000 in IAN, and 85,000 in PEDSnet. About 80% of each population is male, but the percentage of children who are nonwhite differs. The amount of follow-up data is about 5 years for PEDSnet and an average of 2 years for ATN. For IAN, follow-up occurs based on specific studies and thus is not reported.

CONCLUSION

As these networks grow in number of subjects and in terms of data quality, it will be important to assure that the data are used efficiently and well. Given the partial geographic overlap of the samples and the likelihood that the individuals in the samples also overlap, conversations about complementary or joint analyses are under way. Such efforts, of course, need to be mindful of the consent obtained and the human subject issues raised by potentially merging data sources. Improvements in access to clinical and family-reported data, the quality and speed with which they can be analyzed, and the quality of research questions and methodologies could allow these data sources to further improve the knowledge about and treatment of children with ASD.

ACKNOWLEDGMENTS

IAN Acknowledgments: The IAN is a partnership project of the Kennedy Krieger Institute and the Simons Foundation. IAN is also

partially supported by the Patient-Centered Outcomes Research Institute (PCORI) as a PPRN member of the PCORnet national research network, under contract PPRN-1501-26462. IAN was initially established through the support of Autism Speaks. The IAN thanks Alison Marvin, PhD, for her help with data analytics. In addition, IAN is grateful to the participating IAN families who have provided their valuable input and experiences. ATN Acknowledgments: This network activity was supported by Autism Speaks and cooperative agreement UA3 MC11054 through the U.S. Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Research Program to the Massachusetts General Hospital. This work was conducted through the Autism Speaks Autism Treatment Network. PEDSnet Acknowledgments: PEDSnet is a collaboration of St. Louis Children's Hospital, Seattle Children's Hospital, Nemours Children's Health System, Nationwide Children's Hospital, Children's Hospital of Philadelphia, Children's Hospital Colorado, Children's Hospital Boston, and Cincinnati Children's Hospital and Medical Center, and the children and families who receive care at these institutions. PEDSnet is supported by the Patient-Centered Outcomes Research Institute (PCORI) as a CDRN member of the PCORnet national research network, under contract CDRN-1306-01556.

REFERENCES

1. Viscidi EW, Triche EW, Pescosolido MF, et al. Clinical characteristics of children with autism spectrum disorder and co-occurring epilepsy. *PLoS One.* 2013;8:e67797.
2. Gottfried C, Bambini-Junior V, Francis F, et al. The impact of neuroimmune alterations in autism spectrum disorder. *Front Psychiatry.* 2015;6:121.
3. Levy SE, Giarelli E, Lee LC, et al. Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States. *J Dev Behav Pediatr.* 2010;4:267–275.
4. Coury DL, Ashwood P, Fasano A, et al. Gastrointestinal conditions in children with autism spectrum disorder: developing a research agenda. *Pediatrics.* 2012;130:S160–S168.
5. Reynolds AM, Malow BA. Sleep and autism spectrum disorders. *Pediatr Clin North Am.* 2011;58:685–698.
6. Simonoff E, Pickles A, Charman T, et al. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry.* 2008;47:921–929.
7. Frank L, Basch E, Selby JV; For the Patient-Centered Outcomes Research Institute. The PCORI perspective on patient-centered outcomes research. *JAMA.* 2014;312:1513–1514.
8. Currie J. "Big data" versus "big brother": on the appropriate use of large-scale data collections in pediatrics. *Pediatrics.* 2013;131: S127–S132.
9. Berger ML, Lipset C, Gutteridge A, et al. Optimizing the leveraging of real-world data to improve the development and use of medicines. *Value Health.* 2015;18:127–130.
10. Lee H, Marvin AR, Watson T, et al. Accuracy of phenotyping of autistic children based on internet implemented parent report. *Neuropsychiatr Genet.* 2010;153B:1119–1126.
11. Daniels AM, Rosenberg RE, Anderson C, et al. Verification of parent-report of child autism spectrum disorder diagnosis to a web-based autism registry. *J Autism Developmental Disord.* 2012;42: 257–265.
12. Constantino JN, Gruber CP. Social Responsiveness Scale (SRS). *Western Psychological Services Publishing Web Site 2005*. Available at https://www.wpspublish.com/store/p/2993//srs-social-responsiveness-scale-by-john-n-constantino-md. Accessed August 30, 2017.
13. Rutter M, Bailey A, Lord C. *The Social Communication Questionnaire: Manual [Western Psychological Services Publishing Web Site] 2003*. Available at https://www.wpspublish.com/store/p/2954/social-communication-questionnaire-scq. Accessed August 30, 2017.
14. Frazier TW, Youngstrom EA, Embacher R, et al. Demographic and clinical correlates of autism symptom domains and autism spectrum diagnosis. *Autism.* 2014;18:571–582.
15. Constantino JN, Todorov A, Hilton C, et al. Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. *Mol Psychiatry.* 2013;18:137–138.
16. Lajonchere C, Jones N, Coury DL, et al. Leadership in health care, research, and quality improvement for children and adolescents with autism spectrum disorders: autism Treatment Network and Autism Intervention Research Network on Physical Health. *Pediatrics.* 2012;130:S62–S68.
17. Perrin JM, Coury DL, Klatka K, et al. The autism intervention research network on physical health and the autism speaks autism treatment network. *Pediatrics.* 2016;137:S67–S71.
18. Murray DS, Fedele A, Shui A, et al. The autism speaks autism treatment network registry data: opportunities for investigators. *Pediatrics.* 2016;137:S72–S78.
19. Forrest CB, Margolis PA, Bailey LC, et al. PEDSnet: a national pediatric learning health system. *J Am Med Inform Assoc.* 2014;21: 602–606.
20. Forrest CB, Margolis P, Seid M, et al. PEDSnet: how a prototype pediatric learning health system is being expanded into a national network. *Health Aff (Millwood).* 2014;33:1171–1177.
21. Anderson C, Law JK, Daniels A, et al. Occurrence and family impact of elopement in children with autism spectrum disorder. *Pediatrics.* 2012;130:870–877.
22. Anderson C. *IAN Research Report: Bullying and Children With ASD. [Interactive Autism Network Web Site]*; 2012. Available at: https://iancommunity.org/cs/ian_research reports/ian research report bullying. Accessed August 15, 2017.
23. Frazier TW, Thompson L, Youngstrom EA, et al. A twin study of heritable and shared environmental contributions to autism. *J Autism Dev Disord.* 2014;44:2013–2025.
24. Malow BA, Katz T, Reynolds AM, et al. Sleep difficulties and medications in children with autism spectrum disorders: a registry study. *Pediatrics.* 2016;137:S98–S104.
25. Kuhlthau KA, McDonnell E, Coury D, et al. Associations of quality of life with health-related characteristics among children with ASD. *Autism.* 2017 [epub ahead of print].
26. Vasa RA, Mazurek MO, Mahajan R, et al. Assessment and treatment of anxiety in youth with Autism spectrum disorders. *Pediatrics.* 2016;137:S115–S123.

