

Question 7: How Do We Expand and Enhance Research Infrastructure Systems to Meet the Needs of the Autism Community?

Introduction

Aspirational Goal: Develop, enhance, and support research infrastructure and statistical data gathering systems that advance the speed, efficacy, and dissemination of autism research and services.

Appropriate research infrastructure is critically important to the success of the *IACC Strategic Plan*. This includes repositories for biological materials and data, systems for gathering data on autism prevalence, and enhancing the research workforce. Progress toward the Aspirational Goal has been rapid in recent years. New databases are being built to leverage recent genetics findings, and efforts to share biospecimens among multiple research efforts are intensifying. Funders such as the NIH are putting increased focus on data sharing by integrating the sharing requirements into funding announcements. This has increased availability of resources has advanced the efficacy and speed of ASD research. Meanwhile, many funders are making efforts to increase the breadth and diversity of the research workforce, ensuring that the field is equipped to address the most pertinent issues of the entire autism community. Additionally, prevalence data monitoring systems, with new efforts focused on streamlining methods and expanding the age groups that are monitored.

Biorepository Infrastructure

Biological materials repositories collect, process, store, and distribute biospecimens to support scientific investigation. In the autism research community, biorepositories have been developed to support collection and dissemination of brain tissue, fibroblasts, and other tissues. Greater participation in brain and tissue banking is needed from members of the autism community in order to obtain enough samples to meet research requests. Outreach campaigns to encourage families to donate brain and other tissue need to be expanded and enhanced, especially among underrepresented groups.

Brain Banking

The [NIH NeuroBioBank](#) was formed in 2013 to address the increasing demand for postmortem human brain tissue for research purposes¹. Although this resource provides tissues for wide-ranging neurological and neurodevelopmental disorders, there is high demand for tissue from donors diagnosed with autism spectrum disorders. The NIH NeuroBioBank supports six independent brain and tissue repositories. The collection has been highly sampled over the years and continues to grow through outreach activities and collaborations with other organizations.

A more autism-focused effort was undertaken in 2015 by the [Autism BrainNet](#), managed and supported by the Simons Foundation Autism Research Initiative (SFARI). Autism BrainNet is focused exclusively on creating a collection of ASD and control brain tissue for research. The program supports four nodes throughout the United States (New York, Massachusetts, Texas, and California) and one in the United Kingdom that share standardized protocols for tissue harvesting, storage, and tissue dissemination. Autism BrainNet has a robust public awareness campaign to encourage donation, led by the Autism

Science Foundation. The NIH NeuroBioBank and Autism BrainNet work closely together to ensure that tissue acquisition, processing, and distribution from both resources are conducted with the highest standards possible.

Tissue Banking

The [NIMH Repository and Genomics Resource \(NRGR\)](#) provides a centralized national biorepository that plays a key role in facilitating ASD research. The repository contains over 22,000 biospecimens in its autism collection, and accompanying genotypic and phenotypic data are available to qualified researchers worldwide. Biomaterials are stored at Sampled (formerly Infinity BiologiX), supported through a cooperative agreement from the National Institute of Mental Health (NIMH). Other partners include Rutgers University, the Information Sciences Institute at the University of Southern California, and the Batelle Institute for Mathematical Medicine at the Nationwide Children's Hospital. Clinical projects funded by NIMH that propose to collect biospecimens are required to submit the samples to NRGR. The Repository collects and stores several types of biomaterials, including DNA and immortalized cell lines. Recently, the cell services have expanded to include the NIMH Stem Cell Resource, which provides banking and validation of reprogrammed cells (e.g., iPSCs) and source cells (e.g., fibroblasts) derived from postnatal-to-adult human patients and controls.

Data Infrastructure

Data infrastructure refers to data collection, storage, sharing, and consumption to support autism research, services, and policy development. Autism is a highly heterogeneous condition requiring large sample sizes to make significant findings. Thus far, tens of thousands of research subjects have consented to make their genomics, imaging, and clinical research data available to scientists in the hope that those data will help lead to important research discoveries. These datasets have become very large and have grown exponentially with the rapid advances in technology, new methods of data acquisition, and the integration of patient-directed reporting applications. Other research communities have established related data repositories and funded data sharing initiatives making those datasets broadly available for use by the autism research community. Given the size of these data and the complexity of the software, algorithms, and analytic methods used, it is essential that all the data and associated metadata be shared when a result is published or a significant finding is announced. Ensuring that all data is shared will increase the rigor and reproducibility of findings, a core responsibility of publicly funded research.

Data Banks

New findings, technologies, and research methods have emerged that can drive autism research forward, capitalizing on advances in participant engagement through electronic portals and the collection of large data sets. Together, these participant-powered and clinical data networks can be further leveraged for rapid research on large numbers of participants throughout the country, offering the potential for a broad and rich view of the health and well-being of those with ASD and their families.

The NIMH Data Archive (NDA) houses, harmonizes, and shares all human-subjects data collected as part of NIMH-funded projects with the goal of accelerating progress in the research of mental health. The NDA uses a de-identified research subject identifier (the Global Unique Identifier, or GUID) and a precise method for associating research data with publications/results². NDA also supports common data

definitions, a standardized set of data collection measures ensuring that results across studies can be accurately combined or compared. The NDA infrastructure was initially established to support the NIH Autism Centers of Excellence, but it has grown into an informatics platform that facilitates data sharing across all mental health and other research communities, making data available from each of these repositories combined into a single resource with a single process for gaining access to all shared data. Today, research data from over 2,000 research projects are being shared through the NDA. Investment is still needed to extend this infrastructure to support big data analytics better and to integrate with other biobanks and genomics data repositories more fully.

Another mechanism for data sharing is the [Autism Sequencing Consortium \(ASC\)](#), an international group of scientists who share autism samples and genetic data³. Currently, ASC has whole exome sequencing (WES) data for over 35,000 samples⁴, many of which are derived from DNA samples in the NIMH Repository. Summary data is available for all samples, as is raw and called data for samples with appropriate consents. Permission to re-contact research participants from completed studies exists for many of the samples within the ASC, managed by the contributing site.

In 2016, the Simons Foundation launched [SPARK \(Simons Foundation Powering Autism Research for Knowledge\)](#) to recruit, engage and retain a cohort of 100,000 individuals with ASD, as well as 175,000 of their family members, to participate in autism research. To participate in SPARK, families enroll online, provide saliva samples for genetic analysis, and agree to be re-contacted for future research opportunities. SPARK participants are being sequenced and genotyped to identify new genes associated with autism risk. Clinical, behavioral, and genetic data on the SPARK cohort are available to all qualified investigators, and SPARK participants can be invited to participate in other ASD research studies. SPARK partners with more than 30 U.S. medical schools and autism research centers to help recruit autistic individuals and their family members.

Data Sharing

When all research projects share their data, the quality of the accumulated data increases. For example, when a new research participant is enrolled in a research study, that person may also have registered previously with one or more data or biorepositories. If the data are linked and widely accessible to researchers (with appropriate privacy protections in place), the potential richness of the information available on that participant is thereby enhanced. Sharing of data will also reduce the burden of participation in research studies on the autistic community; if an individual's data can be used for more than one study in a coordinated manner, then the participant can be assured that the yield of their time and energy commitment to research participation are maximized.

In 2022, The White House Office of Science and Technology Policy (OSTP) issued guidance on Desirable Characteristics of Data Repositories for Federally Funded Research⁵. This guidance includes several recommendations, ensuring that the data is freely and easily accessible, uses unique identifiers, provides clear instructions for how it is to be used, and is stored securely in a common format. The guidance also includes special considerations for the sharing of human data, including additional security safeguards and standards for participant consent. Most repositories already comply with these guidelines, but it will be important to continue harmonizing repository standards in the future in order to increase access, maintain privacy, and ensure interoperability. Responsibly sharing high quality data will increase the return on the collective research investment, protect intellectual contributions, and

2021-2022 IACC Strategic Plan for Autism Research, Services, and Policy Draft Chapter
July 2022

help accelerate research discovery. Collectively, open data sharing offers the best opportunity to reach the sample sizes that are likely needed to improve understanding of autism and related disorders.

Care should be taken to ensure that all stakeholders across the research enterprise understand the importance of data sharing and that those sharing the most used and highest quality datasets be credited for their contributions. To facilitate data sharing in research involving human participants, an identifier or code is used to identify and link each individual to his or her specimens and perhaps also to associated medical information; use of a de-identified code (i.e., a code that does not reveal the identity of the individual) maintains privacy of the individual's information. The GUID was developed to provide an easy method of identifying the same research participant across various data repositories and biobanks while maintaining the privacy of their personal information. The advantage of the GUID is that it enables linkage of data and specimens for a given individual over multiple studies, which can enrich the data set and prevent unnecessarily repeating the collection of the same types of samples from a given individual for multiple studies. While most data repositories have standardized identification of research participants using the GUID, adoption of this method has been less consistent across biobank repositories. Compounding this problem is the fact that most of the biobanks hold samples that are consented for restricted use (e.g., a study of autism and schizophrenia would require separate access) and are shared in separate repositories with different access restrictions and policies. The result is that it is often easier to request a tissue or sample from a biobank, re-sequence or re-analyze it, and then share the data with a new and different identifier, causing unnecessary (and often undetectable) duplication. For genomics, tools have been developed to eliminate this duplication, and attempts have been made to provide similar safeguards for imaging data. Though these additional tools exist, it is strongly encouraged that all data and biobank repositories maintain the use of the GUID and that those publishing genomics- or biobank-related studies provide a publicly available manifest of subject GUIDs and links to phenotypic data locations when publishing, even if the data are only available as restricted use datasets. This action will provide standardization allowing data from the same individual to be linked across repositories, eliminate data duplication, and help minimize redundant sample and tissue requests, thereby conserving precious resources.

Several national surveys and administrative efforts collect information about people with ASD. Many of these surveys are Federally funded through agencies such as CDC [National Health Interview Survey (NHIS)], the Health Resources and Services Administration (HRSA) [National Survey of Children's Health (NSCH)], and the Department of Education [National Longitudinal Transition Study 2012 (NLTS 2012)]. Although each responsible agency may focus on its own research priorities when collecting and analyzing the data, synchronization of the national data sources will maximize their utility. Concordance of questions and sampling across surveys and administrative data could add greatly to the comparability of research undertaken across these national platforms. Additionally, infrastructure for linking these surveys to other sources of data is essential. For example, the CDC links the NHIS to administrative records from the Department of Housing and Urban Development (HUD), which allows for the addition of detailed housing information for those NHIS participants who use HUD services. The CDC recently launched and updated an [interactive autism data visualization website](#) which presents the most up-to-date state-based autism prevalence information from four major data sources: CDC ADDM Network (discussed below), Department of Education administrative data, CMS administrative claims data, and HRSA's National Survey of Children's Health (discussed below). Additionally, Federal Statistical Research

2021-2022 IACC Strategic Plan for Autism Research, Services, and Policy Draft Chapter
July 2022

Data Centers make national data from the Census bureau, CDC, and the Agency for Healthcare Research and Quality (AHRQ) available to researchers in one place. More projects like these, and additional means of capitalizing on the data that has already been collected and funded, are a key priority in order to generate an expansion of the information available on autism to a nationally representative sample.

Human Infrastructure

Human infrastructure refers to the development of human resources necessary to support autism research and services. Human infrastructure for research includes developing a professional workforce to conduct research and provide services, as well as encouraging individuals with autism and their family members to participate in autism research. In addition, systems must be developed to share research findings with community stakeholders and translate research findings into policy and practice. Human infrastructure for services is discussed in more detail in Chapter 5.

Research Training and Workforce Development Efforts

There are several efforts underway to enhance research training and workforce development. Several federal and private funders support research training opportunities including, but not limited to, training and career development grants and travel awards for early career investigators to attend research conferences. In many cases, these awards emphasize building relationships with experienced mentors and on encouraging multidisciplinary avenues of exploration. In recent years, the NIH has also offered funding for mid-career investigators from other research fields to transition into autism services research. This program was created in response to the recommendations from the IACC to expand the research workforce that studies autism services.

Workforce development is an area of immense need as the number of identified individuals with autism continues to grow. While progress has been made in the area of early detection and intervention, and in the support of children on the spectrum, much less effort has been expended on research and services relevant to adults, as tens of thousands of children with autism transition to adulthood. Further, there is a dearth of trained medical professionals that are knowledgeable in providing care to the autism community, particularly the adult community. Increased attention should be devoted to the development of early career researchers, particularly those from diverse backgrounds. It is also important to increase recruitment and training of autistic researchers, as they are intimately aware of important issues and thus well suited to address the needs of the autistic community. Given that Federal funding for workforce development and training is limited, it is important to ensure the workforce development and training efforts are evidence-based and that their delivery is standardized across communities. The development of best-practices guidelines will enhance the implementation of such training programs.

Participation in Research Studies

The participation of autistic individuals in research studies is crucial in order to build knowledge about the autistic experience across the lifespan, to build the evidence base for interventions, and to identify the most effective and efficient services and supports. Individuals on the autism spectrum and their families participate in research studies at relatively low rates, often due to unawareness of opportunities, not perceiving research as relevant to their needs, apprehension about the methods involved or the goal of the research, and/or concerns about how data will be used, stigma or bias.

Coordinated efforts are needed to educate stakeholders from diverse backgrounds on the personal and community benefits of participating in research, the level of risk associated with the study (if any), and the privacy protections in place to maintain anonymity. Simultaneously, researchers should be open to the designing their studies around the stated needs and wants of autistic individuals and their family members and to ensuring that studies are inclusive of diverse populations. This can be best achieved using Community-Based Participatory Research (CBPR) methods, engaging stakeholders in the earliest stages of research study design and ensuring that diverse communities are included in outreach efforts. Particular efforts should be to include individuals across the full spectrum of autism, especially those with intellectual disabilities or communication challenges who often exclusion from research based on study criteria. Research should also be conducted to understand the barriers that discourage participation.

Dissemination of Research Findings

Increasing and improving mechanisms for dissemination of research findings after publication should be a priority for the autism community. It is vital that findings and data become more accessible to researchers, practitioners, families, and the general public. Training to improve science communication skills should be more readily available to researchers who wish to share their work with lay audiences. Particularly important is benefit and/or risk communication in the interpretation of research findings, as the information disseminated to the public is sometimes contradictory, oversimplified, overstated, or sensationalized. This misinformation can have a negative impact on research participation. Whenever possible, it is important the research information communicated to the public is written in plain or lay-friendly language. When appropriate, information should also be translated into multiple languages, in order to increase the accessibility of information to non-English speaking populations. Mechanisms that allow for the summation of the evidence base into actionable recommendations such as systematic reviews and meta-analysis are encouraged. Much of this work will be more feasible as the data sharing infrastructure further develops and expands.

NDA provides an infrastructure to make data broadly accessible to the research community through a universal platform and federation with other data sources. To make NDA the most useful resource possible, autism researchers must improve both the consistency and quality of data shared, especially those data supporting published results, allowing the infrastructure to be better utilized and supporting the dissemination of scientific advances. NDA and similar data sharing efforts can help maximize the return on Federal and private investment in autism research made over the last decade by providing the research community with richer datasets and opportunities for research that would not have been possible without the coordination of these data.

International Collaboration

Most ASD research is currently published by researchers in the US, Canada, Europe, Australia, and China, with only a small proportion representing international collaborations and or studying autism in low- and middle-income countries (LMICs)⁶. Thus, many of the benefits of autism research may not be reaching populations or be applicable in LMICs. In addition, the research community may be missing opportunities to include and learn from diverse cultures and settings, to diversify their study samples, and to study populations with different kinds of challenges. Therefore, researchers and funding organizations should seek opportunities to collaborate with researchers from other regions of the world,

particularly those with lower research capacity. International research collaborations not only present opportunities to disseminate and implement evidence-based science and services in diverse settings around the world, but also allow the ASD research community to learn about how diverse populations, including those from low-resource settings, have addressed issues such as limited research infrastructure and large service gaps. For these reasons, it is imperative that international research efforts and collaborations continue to be promoted and supported.

Prevalence Monitoring

Population-based statistical data gathering, or surveillance, for autism spectrum disorder is essential for monitoring prevalence over time, assessing patterns by demographic factors and level of support necessary, characterizing co-occurring conditions, estimating resource needs, and stimulating research into potential environmental factors. For the data provided to be used effectively, surveillance efforts should be as complete and valid as possible. Population-based studies of the prevalence and characteristics of autism spectrum disorder among children in the United States has been ongoing since the year 2000, and more recent studies have also estimated the prevalence of autism among adults. However, sustained investment is necessary to continue and expand these efforts.

There are several different methodologies currently used for estimating the prevalence and characteristics of autism spectrum disorder among children, including: 1) use of administrative records; 2) parent or caregiver surveys; 3) expert review of records from multiple sources; and 4) screening and examination of children. Each of these methodologies has strengths and limitations. Administrative records are readily available and cost-effective to use, but are collected for other purposes and do not always contain adequate and pertinent information. Health surveys are nationally representative, generate data relatively quickly, include extensive questions on service needs and utilization, include a comprehensive age range of children, and are cost-effective in terms of the marginal cost of adding ASD-related questions; however, the validity of parent/caregiver-reported ASD has not been established, and low response rates have raised concerns about bias. Expert review of records from multiple sources, including healthcare and education records, can ascertain records-based data on a number of factors such as demographics, educational placement, intellectual and adaptive function, and behavioral phenotype. However, this methodology is dependent on data in children's records, focuses on a few specific ages, and is resource- and time-intensive and so currently cannot be done at a national level. Finally, screening and examination of children using a standardized and validated ASD diagnostic tool is a rigorous methodology that attempts to give all children in the selected population an opportunity for ascertainment. However, this methodology is also resource- and time-intensive and cannot currently be done on a national level. In addition, low response rates in previous studies suggest a potential for bias.

ASD prevalence monitoring systems should be complementary, offering unique strengths and contributions that will further the understanding of the population of individuals with ASD. Where appropriate, data collection should be designed to allow comparisons across systems. Further linkage of surveillance data with other state and Federal datasets should be encouraged to leverage the surveillance efforts and expand the scope and utility of the information collected.

While many research studies are focused on understanding and meeting the needs of children with ASD, much less research has been done on adults. Using state-based data on children with ASD, the CDC recently estimated the prevalence of adults with ASD at 2.21%⁷. There is an urgent need to expand ASD

2021-2022 IACC Strategic Plan for Autism Research, Services, and Policy Draft Chapter
July 2022

surveillance efforts to adults to fully understand prevalence, adolescent/young adult transition needs, employment and housing, co-occurring conditions, premature mortality, and other lifespan issues. In particular, investigating ASD prevalence in adults will help researchers understand how the interaction of ASD and co-occurring conditions impacts the ability to adults with ASD to live and work.

Current Surveillance Programs

Autism and Developmental Disabilities Monitoring Network

The [Autism and Developmental Disabilities Monitoring \(ADDM\) Network](#) is a population-based surveillance program for ASD and other developmental disabilities. CDC has been conducting surveillance for ASD among 8-year-old children through the ADDM Network every 2 years since 2000 at sites throughout the United States; the most recent prevalence estimate for 2018 was 23.0 per 1,000 8-year-old children^{8,9}. In 2010, the ADDM Network was expanded to include surveillance for ASD among 4-year-old children. The program was further expanded in 2018 to conduct follow-up studies on 16 year-olds initially included in earlier ADDM Network surveillance. There are currently 11 sites monitoring 8- and 4-year-old children; 5 of those sites are also performing follow-up studies on 16-year-olds. Data have been linked to various sources such as environmental pollutant monitoring, juvenile justice records, and others. Recently, the ADDM Network has revised its methodology to estimate prevalence by monitoring records of autism diagnoses from clinicians, special-education classifications of autism, and hospital billing codes for autism services¹⁰. By making use of these existing tools rather than conducting assessments of clinician and educational records, ADDM Network researchers have been able to reduce the amount of time needed to make prevalence estimates.

National Survey of Children's Health

The [National Survey of Children's Health \(NSCH\)](#) is currently administered by the Maternal and Child Bureau of HRSA. This nationally representative telephone survey of children's health and development based on parent/ caregiver report includes questions on whether the child currently had an ASD as well as whether a healthcare provider ever informed the parent or caregiver that the child had an ASD. Data are also collected on a variety of topics including the child's health, health as an infant, recent healthcare service, experiences with healthcare providers, health insurance coverage, sociodemographic factors, and the child's learning, home, and family environment. In the most recently published dataset (collected in 2020), ASD prevalence was 2.7% for children aged 3-17 years. Beginning in 2016, this survey was moved to a mail-invitation, online survey based on a US Census Bureau sampling platform. This survey has been combined with the previously fielded National Survey of Children with Special Healthcare Needs. State-level estimates are available for many variables.

National Health Interview Survey

CDC conducts the [National Health Interview Survey \(NHIS\)](#), a nationally representative survey of parents/caregivers that provides data on the health of children in the United States, including information on whether a healthcare provider ever informed the parent or caregiver that the child had an ASD. The US Census Bureau is the data collection agent and the data are collected through personal household interviews. The content and structure of the NHIS were updated in 2019 to better meet the needs of data users. Aims of the questionnaire redesign were to improve the measurement of covered health topics, reduce respondent burden, harmonize overlapping content with other federal health surveys, establish a long-term structure of ongoing and periodic topics, and incorporate advances in

survey methodology and measurement. Data are collected on children aged 0-17 years every year on a variety of topics including the child's health status, healthcare access and utilization, and a mental health screener (the Strengths and Difficulties Questionnaire), as well as family factors, including sociodemographic factors. ASD prevalence was 2.76% for children aged 3-17 years in 2016¹¹. As with NSCH, linkages to data from other Federal agencies should be encouraged to expand the scope and usefulness of the data.

Summary

Continuing to build the infrastructure necessary for autism research is an important priority. In particular, researchers continue efforts to standardize data collection and share with others in order to build higher-powered studies across multiple areas of research. Research institutions must continue to support biobanks and databanks, and to work towards integrating common collection and processing methods. Efforts to increase the participation of individuals with autism and their families in research and contributions to biorepositories are important, as information and samples gathered have the potential to make significant contributions to our understanding of ASD. Inclusion of people on the autism spectrum and their families in the research process, as well as recruiting and training autistic researchers, will help ensure that studies maintain a focus on issues that matter most to those who are impacted by ASD. Continued optimization of prevalence data monitoring efforts, including expansion to gather data on adults, will better inform research and service priorities. Finally, funding agencies should continue to devote resources to ensuring dissemination of research findings and best practices, gaining better understanding of ASD prevalence across the lifespan, and training the next generation autism researchers, clinicians, and care providers.

Objectives

OBJECTIVE 1: Promote growth, linkage, coordination, and security of biorepository and data repository infrastructure systems, equitable access to these systems, and inclusion of diverse samples.

Examples:

- Promote biological sample donation to ensure that demand for research studies is met.
- Make efforts to standardize data collection, and responsibly share all the data supporting any findings when those findings are announced.
- Develop and expand programs and outreach campaigns to encourage families from diverse backgrounds to participate in autism research, join registries, and donate biological samples.
- Ensure equitable access to federal data sets so that all communities can benefit from these resources.

OBJECTIVE 2: Expand and enhance the research workforce, with attention to diversity and inclusion, and accelerate the pipeline from research to practice.

Examples:

2021-2022 IACC Strategic Plan for Autism Research, Services, and Policy Draft Chapter
July 2022

- Expand and enhance programs that provide funds to train current and future researchers on innovative research techniques.
- Support programs to train autistic researchers and researchers from diverse communities to conduct research related to autism.
- Develop programs to translate and disseminate autism research findings into actionable recommendations and real-world practice.

OBJECTIVE 3: Strengthen statistical data gathering systems to advance understanding of the autistic population, while allowing comparisons and linkages across systems as much as possible.

Examples:

- Continue to expand prevalence data monitoring activities to gain a better understanding of needs and concerns over the lifespan.
- Expand data monitoring efforts to collect more descriptive data regarding co-occurring conditions, including intellectual disability, seizure disorders, anxiety, and depression to increase understanding of the prevalence of these conditions in the autistic population.
- Support inclusion of autism and disability research through large surveys, including those conducted by federal agencies such as CDC, HRSA, and the Department of Education.
- Promote efforts to ensure that diverse samples are captured in survey data.

2021-2022 IACC Strategic Plan for Autism Research, Services, and Policy Draft Chapter
July 2022

References

¹ Nichols L, Freund M, Ng C, Kau A, Parisi M, Taylor A, Armstrong D, Avenilla F, Joseph J, Meinecke D, Wagner A, Roger Little A. The National Institutes of Health Neurobiobank: a federated national network of human brain and tissue repositories. *Biol Psychiatry*. 2014 Jun 15;75(12):e21-2. doi: 10.1016/j.biopsych.2013.07.039. Epub 2013 Sep 25. PMID: 24074636.

² Johnson SB, Whitney G, McAuliffe M, Wang H, McCreedy E, Rozenblit L, Evans CC. Using global unique identifiers to link autism collections. *J Am Med Inform Assoc*. 2010 Nov-Dec;17(6):689-95. doi: 10.1136/jamia.2009.002063. PMID: 20962132; PMCID: PMC3000750.

³ Buxbaum JD, Daly MJ, Devlin B, Lehner T, Roeder K, State MW; Autism Sequencing Consortium. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron*. 2012 Dec 20;76(6):1052-6. doi: 10.1016/j.neuron.2012.12.008. PMID: 23259942; PMCID: PMC3863639.

⁴ Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, Peng M, Collins R, Grove J, Klei L, Stevens C, Reichert J, Mulhern MS, Artomov M, Gerges S, Sheppard B, Xu X, Bhaduri A, Norman U, Brand H, Schwartz G, Nguyen R, Guerrero EE, Dias C; Autism Sequencing Consortium; iPSYCH-Broad Consortium, Betancur C, Cook EH, Gallagher L, Gill M, Sutcliffe JS, Thurm A, Zwick ME, Børglum AD, State MW, Cicek AE, Talkowski ME, Cutler DJ, Devlin B, Sanders SJ, Roeder K, Daly MJ, Buxbaum JD. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. 2020 Feb 6;180(3):568-584.e23. doi: 10.1016/j.cell.2019.12.036. Epub 2020 Jan 23. PMID: 31981491; PMCID: PMC7250485.

⁵ The National Science and Technology Council, Desirable Characteristics of Data Repositories for Federally Funded Research, 2022, DOI: <https://doi.org/10.5479/10088/113528>

⁶ Office of Autism Research Coordination (OARC), National Institute of Mental Health and Thomson Reuters, Inc. on behalf of the Interagency Autism Coordinating Committee (IACC). IACC/OARC Autism Spectrum Disorder Research Publications Analysis Report: The Global Landscape of Autism Research. July 2012. Retrieved from the Department of Health and Human Services Interagency Autism Coordinating Committee website: <http://iacc.hhs.gov/publications-analysis/july2012/index.shtml>

⁷ Dietz PM, Rose CE, McArthur D, Maenner M. National and State Estimates of Adults with Autism Spectrum Disorder. *J Autism Dev Disord*. 2020 Dec;50(12):4258-4266. doi: 10.1007/s10803-020-04494-4. PMID: 32390121; PMCID: PMC9128411.

⁸ Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, Esler A, Furnier SM, Hallas L, Hall-Lande J, Hudson A, Hughes MM, Patrick M, Pierce K, Poynter JN, Salinas A, Shenouda J, Vehorn A, Warren Z, Constantino JN, DiRienzo M, Fitzgerald RT, Grzybowski A, Spivey MH, Pettygrove S, Zahorodny W, Ali A, Andrews JG, Baroud T, Gutierrez J, Hewitt A, Lee LC, Lopez M, Mancilla KC, McArthur D, Schwenk YD, Washington A, Williams S, Cogswell ME. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. *MMWR Surveill Summ*. 2021 Dec 3;70(11):1-16. doi: 10.15585/mmwr.ss7011a1. PMID: 34855725; PMCID: PMC8639024.

⁹ Shaw KA, Maenner MJ, Baio J; EdS1, Washington A, Christensen DL, Wiggins LD, Pettygrove S, Andrews JG, White T, Rosenberg CR, Constantino JN, Fitzgerald RT, Zahorodny W, Shenouda J, Daniels JL, Salinas A, Durkin MS, Dietz PM. Early Identification of Autism Spectrum Disorder Among Children Aged 4 Years - Early Autism and Developmental Disabilities Monitoring Network, Six Sites, United States, 2016. *MMWR Surveill Summ*. 2020 Mar 27;69(3):1-11. doi: 10.15585/mmwr.ss6903a1. PMID: 32214075; PMCID: PMC7119643.

¹⁰ Maenner MJ, Graves SJ, Peacock G, Honein MA, Boyle CA, Dietz PM. Comparison of 2 Case Definitions for Ascertaining the Prevalence of Autism Spectrum Disorder Among 8-Year-Old Children. *Am J Epidemiol*. 2021 Oct 1;190(10):2198-2207. doi: 10.1093/aje/kwab106. PMID: 33847734; PMCID: PMC9022042.

¹¹ Zablotsky B, Black LI, Blumberg SJ. Estimated prevalence of children with diagnosed developmental disabilities in the United States, 2014–2016. NCHS Data Brief, no 291. Hyattsville, MD: National Center for Health Statistics. 2017.

DRAFT