

Summary Data Approach in NDAR

Date: May 8, 2012

Table of Contents

Data Browser	3
Data from Labs	3
Data from Papers.....	4
Appendix 1: Mechanisms and Rules used for NDAR Data Browser	5
Phenotype/Sub-Phenotype Categorization.....	5
Fragile X	6
Control Subjects	6
Non-Spectrum Control: Typical	6
Non-Spectrum Control: Neurological Disorder	7
Autism Spectrum Phenotype.....	8
Sub-Phenotype: Severely Affected	8
Sub-Phenotype: Mildly Affected.....	9
Sub-Phenotype: Affected.....	10
Precision of Definitions.....	11
Autism Sub-Types and the Minimally Verbal Categorization	12

The National Database for Autism Research (NDAR) supports data sharing across a broad array of clinical, genomic, and brain imaging autism research data. To aid those interested in applying for access, NDAR allows the general public to browse summary data now being shared. There are now three ways to browse the summary data, allowing researchers to find out how much data is available for their area of interest.

Data Browser

Data Distributions, located on the NDAR home page, provides aggregated and detailed views of shared research data that can be mined within the categories of basic data, phenotypic data, neuroimaging data and genomics data. Users can further refine results specific to a research area of interest by selecting a phenotype (e.g. autism spectrum-severely affected, typical control), sub-type (e.g. seizures, regression, verbal IQ, minimally verbal), or attributes (e.g. age, gender, genomics molecule, genomics platform, scanner manufacturer). The rules used to determine phenotype and minimally verbal are defined in the appendix below. The rules used to mine such attributes are expandable to expose and/or aggregate any combination of the 30,000+ data variables within NDAR’s data dictionary. Members of the research community are encouraged to work with NDAR staff to further define rules, exposing other phenotypes, sub-types and attributes relevant to autism research. NDAR can be contacted at ndarhelp@mail.nih.gov.

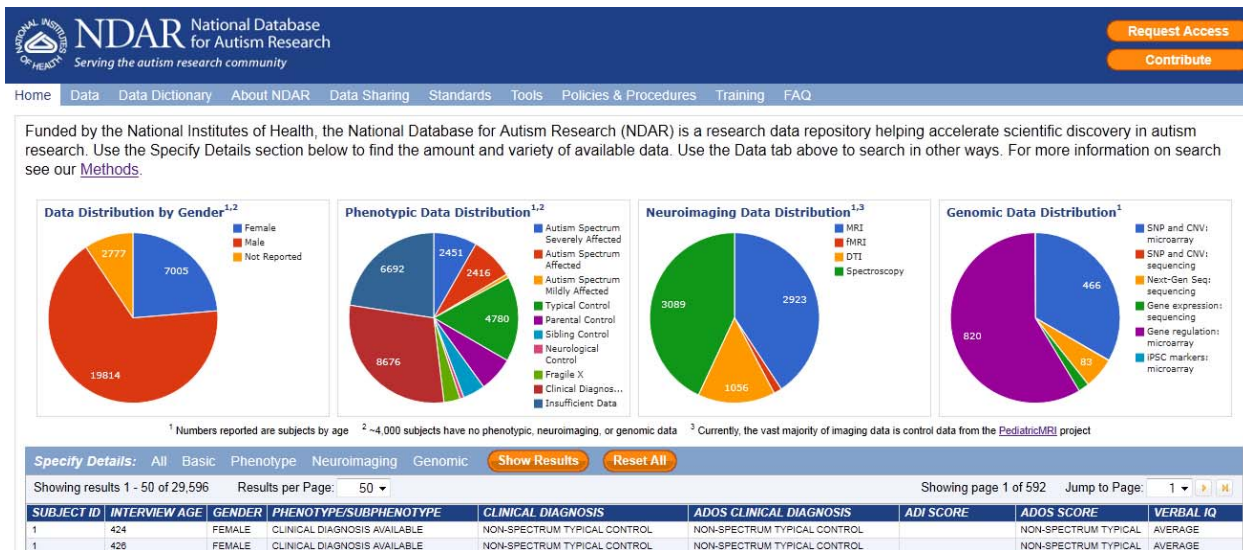


Figure 1 – Data Browser

Data from Labs

Data from Labs, located under the Data tab on the main menu, provides a snapshot of the projects who have shared data in NDAR. Listed are the investigators, publications related to the projects and the number of records and subjects shared by data structure. NDAR users with access to shared data can login to NDAR and download the specific data associated with a lab.

NDAR National Database for Autism Research
Serving the autism research community

Request Access
Contribute

Home Data About NDAR Data Sharing Standards Tools Policies & Procedures Training FAQ

Data from Labs | Data from Papers | Autism Publications | NDAR Data Dictionary

Collection Title: 1/2 Development of a Screening Interview for Research Studies of ASD **RDAP**

Investigators: Cathy Lord, Ph. D. (Owner: Lord, Catherine)

Collection Description: This collection contains phenotypic data on subjects diagnosed with ASD and subjects with non-ASD diagnoses. More: (Shared Summary Data, 2 Publications) **Collapse**

Grant Information:

Project Number	Project Title	Project Start Date	Project End Date	Organization Name
R01MH089390	1/2 Development of a Screening Interview for Research Studies of ASD	09/30/2009	08/31/2011	UNIVERSITY OF MICHIGAN AT ANN ARBOR

Huerta, Marisela; Lord, Catherine, "Pediatric clinics of North America," Diagnostic evaluation of autism spectrum disorders.
Silverman, Jill L; Yang, Mu; Lord, Catherine; Crawley, Jacqueline N, "Nature reviews. Neuroscience," Behavioural phenotyping assays for mouse models of autism.

Data Structures:

Title	Type	Records	Unique Subjects
DAS-II Differential Ability Scales 2nd Ed. School Age v4	Clinical Assessments	190	190
DAS-II Differential Ability Scales 2nd Ed. Early Years v3	Clinical Assessments	164	164
Mullen Scales of Early Learning v3	Clinical Assessments	61	61
Vineland-II - Survey Form (2005) v5	Clinical Assessments	396	396

Figure 2 – Data from Labs

Data from Papers

Data from Papers, located under the Data tab on the main menu, displays data specific to a publication. Using the NDAR Study feature within the portal, an approved researcher can define cohorts, assign research subjects to the cohorts, and specify outcome measures, methods, and results all of which can, but is not required to, be linked to a published paper. Once shared, the NDAR Study will appear in Data by Papers, allowing the data associated with an NDAR Study to be easily downloaded. For publications listed in PubMed, NDAR staff will provide a link in PubMed via the linkout feature back to the data made available in NDAR. This capability allows the autism research community to directly associate the publication with the underlying data.

NDAR National Database for Autism Research
Serving the autism research community

Request Access
Contribute

Home Data About NDAR Data Sharing Standards Tools Policies & Procedures Training FAQ

Data from Labs | **Data from Papers** | Autism Publications | NDAR Data Dictionary

Study Title: [Gastrointestinal Dysfunction in Autism: Parental Report, Clinical Evaluation, and Associated Factors](#) **RDAP**

Investigators: [Levitt, Pat](#), Gorrindo, Phillip Williams, Kent C. Lee, Evon B. Walker, Lynn S. McGrew, Susan G. Levitt, Pat

Study Abstract: The objectives of this study were to characterize gastrointestinal dysfunction (GID) in autism spectrum disorder (ASD), to examine parental reports of GID relative to evaluations by pediatric gastroenterologists, and to explore factors associated with GID in ASD. One hundred twenty-one children were recruited into three groups: co-occurring ASD and GID, ASD without GID, and GID without ASD. A pediatric gastroenterologist evaluated both GID groups. Parents in all three groups completed questionnaires about their child's behavior and GI symptoms, and a dietary journal. Functional constipation was the most common type of GID in children with ASD (85.0%). Parental report of any GID was highly concordant with a clinical diagnosis of any GID (92.1%). Presence of GID in children with ASD was not associated with distinct dietary habits or medication status. Odds of constipation were associated with younger age, increased social impairment, and lack of expressive language (adjusted odds ratio in nonverbal children: 11.98, 95% confidence interval 2.54-56.57). This study validates parental concerns for GID in children with ASD, as parents were sensitive to the existence, although not necessarily the nature, of GID. The strong association between constipation and language impairment highlights the need for vigilance by health-care providers to detect and treat GID in children with ASD. Medications and diet, commonly thought to contribute to GID in ASD, were not associated with GID status. These findings are consistent with a hypothesis that GID in ASD represents pleiotropic expression of genetic risk factors.

Results: [Results published in Autism Research](#)

Study Cohorts: **Test - ASD-GID (40 subjects)**
Age: 60 to 215 months
Gender: Both
Autism-like Developmental Disorders: Affected, Mildly Affected, Severely Affected
Control - GID-only (28 subjects)
Age: 60 to 215 months
Gender: Both
Non-ASD Control: Typical
Control - ASD-only (44 subjects)
Age: 60 to 215 months
Gender: Both
Autism-like Developmental Disorders: Affected, Mildly Affected, Severely Affected

Study Measures: **Primary Measures (2)**
Clinical Assessments: Questionnaire on Pediatric GI Symptoms: Rome III Parent - v01, Social Responsiveness Scale (SRS) - v02
Secondary Measures (5)
Clinical Assessments: Autism Diagnostic Observation Schedule - Module 1 - v02, Autism Diagnostic Observation Schedule - Module 2 - v02, Autism Diagnostic Observation Schedule - Module 3 - v02, Autism Diagnostic Observation Schedule - Module 4 - v02, Diet Diary - v02

Data Analysis: Statistical
Method: ANOVA, Chi-square test, Regression analysis
Significance
p-value: <0.05
Software: SPSS

Collapse

Figure 3 – Data from Papers

Appendix 1: Mechanisms and Rules used for NDAR Data Browser

NDAR's initial focus was to make data submission as simple as possible limiting the cost of data sharing. Through the implementation of an autism data standard and an increasingly sophisticated data validation process, NDAR ensures data is harmonized to a common data standard. This approach has kept submission costs low (see [NDAR Cost model](#)). However, receiving data at a detailed level across a variety of instruments makes it more difficult for a researcher to understand the data available for their particular areas of interest. To reduce this burden, NDAR has developed a categorization system that is applied to all shared data. The process categorizes a research subject listing the summary information known for each age that is provided in the data received. NDAR developed categories for phenotype/sub-phenotype and the minimally verbal sub-type and extracted key attributes on genomics, imaging, demographics and other areas of interest. Categorization can be modified and expanded easily allowing the community to expose relevant phenotypes and attributes appropriate for autism research.

Categorizing subjects – especially across sub-phenotypes - is subject to debate. These categories are intended to broadly define the data in NDAR helping scientists ascertain the types and quantity of data available. In a future release these categories may be used as the basis for downloading specific observations (i.e. ADOS/ADI/Images/Pedigree), if a researcher is approved for access, thus allowing a tighter integration into imaging and genomic processing pipelines.

To help understand the process NDAR used for categorization, we have provided an overview of how the phenotype/sub-phenotype and minimally verbal sub type categories were determined, as these are the most complex and thus were used as the test case for the categorization process.

Phenotype/Sub-Phenotype Categorization

For Phenotype categorization, the rules engine makes multiple passes through shared data in NDAR. Once an individual is defined with a phenotype/sub-phenotype, the process moves on to the next research subject to avoid double classification. Categorization is performed based upon the following order:

1. Fragile X
2. Controls
 - a. Non Spectrum Typical Control (e.g. typical, sibling, parent)
 - b. Non Spectrum Neurological Control
3. Autism Spectrum
 - a. Severely Affected
 - b. Mildly Affected
 - c. Affected

Subjects are evaluated by age in months. To account for observations over multiple visits, observations occurring ± 3 months are considered. So, if a subject is defined for a phenotype of autism spectrum-severely affected at age 24 months, this designation would be applied to observations made at 21 through 27 months. Unless observations are again performed at later ages, a phenotype is not provided. An option to select such subject data will likely be provided in a future version of the Data Browser.

Fragile X

Fragile X is defined according to provided genetic test results for the Fragile X mutation of the FMR1 gene.

Control Subjects

Based upon designations provided by researchers in the data, **Control Subjects** are divided into 4 sub categories:

Non-Spectrum Control:

- a) Typical
- b) Sibling
- c) Parental
- d) Neurological disorders

Typical controls are typically developing individuals. Sibling and Parental controls are family members of a research subject. The Neurological disorders sub-phenotype control group includes subjects with a learning disability, Attention Deficit Hyperactivity Disorder, developmental disability, intellectual disability/MR or other neurological disorder, excluding Fragile X and subjects with positive genetic test result for Non-Spectrum Neurological conditions. Rules for diagnosis are made based on the diagnosis data that is provided in a number of fields within NDAR.

Non-Spectrum Control: Typical

In the absence of a diagnosis, NDAR categorizes control subjects based on the following rules based on results from the ADI-R, ADOS, IQ and Vineland Survey assessments.

ADI-R: Non-Spectrum Typical

Total for Section A: Qualitative Abnormalities in Reciprocal Social Interaction	<4	<u>AND</u>
Total of Section B - Non-Verbal: Qualitative Abnormalities in Communication	<4	<u>OR</u>
Total of Section B - Verbal: Qualitative Abnormalities in Communication	<4	<u>AND</u>
Total of Section C: Restricted, Repetitive, and Stereotyped Patterns of Behavior	<2	<u>AND</u>
Total of Section D = 0		

ADOS: Non-Spectrum Typical

	<u>Module 1:</u>	<u>Module 2:</u>	<u>Module 3:</u>	<u>Module 4:</u>	
Communication Total	<1	<2	<1	<1	<u>AND</u>
Social Interaction Total	<2	<2	<2	<2	<u>AND</u>
Communication + Social Total	<3	<3	<3	<3	

IQ: Average and Higher than Average IQ>85

OR

Vineland Survey: Average and Higher than Average

Composite Domain Total Score	>85	<u>AND</u>
Communication Domain Total Score	>85	<u>AND</u>
Living skills Domain Total Score	>85	<u>AND</u>
Motor skills Domain Total Score	>85	<u>AND</u>
Socialization Domain Total Score	>85)	

Non-Spectrum Control: Neurological Disorder

ADI-R: Non-Spectrum Neurological

Total for Section A: Qualitative Abnormalities in Reciprocal Social Interaction	>4	<u>OR</u>
Total of Section B - Non-Verbal: Qualitative Abnormalities in Communication	>4	<u>OR</u>
Total of Section B - Verbal: Qualitative Abnormalities in Communication	>4	<u>OR</u>
Total of Section C: Restricted, Repetitive, and Stereotyped Patterns of Behavior	>2	<u>AND</u>
Total of Section D = 0		

ADOS: Non-Spectrum Typical or Mildly Affected

	<u>Module 1:</u>	<u>Module 2:</u>	<u>Module 3:</u>	<u>Module 4:</u>	
Communication Total	<4	<5	<3	<3	<u>AND</u>
Social Interaction Total	<4	<4	<4	<4	<u>AND</u>
Communication + Social Total	<7	<8	<7	<7	<u>AND</u>

IQ: IQ = any

OR

Vineland Survey: Less than average

Composite domain total score	<85	<u>OR</u>
Communication domain total score	<85	<u>OR</u>
Living skills domain total score	<85	<u>OR</u>
Motor skills domain total score	<85	<u>OR</u>
Socialization domain total score	<85)	

Autism Spectrum Phenotype

For all remaining subjects/ages the Autism Disorder phenotype **and sub-phenotype rules are run** according to cut offs, for each Assessment (ADI-R, ADOS, IQ and Vineland Survey). Note that a minimum of three assessments - including ADI-R and ADOS – plus one other measure (Vineland or an IQ) is needed for categorization of an autism spectrum phenotype:

Sub-Phenotype: Severely Affected

ADI-R: Severely Affected

Total for Section A: Qualitative Abnormalities in Reciprocal Social Interaction	>10	<u>AND</u>
Total of Section B - Non-Verbal: Qualitative Abnormalities in Communication	>7	<u>OR</u>
Total of Section B - Verbal: Qualitative Abnormalities in Communication	>8	<u>AND</u>
Total of Section C: Restricted, Repetitive, and Stereotyped Patterns of Behavior	>3	<u>AND</u>
Total of Section D	≥1	

ADOS: Severely Affected

	<u>Module 1:</u>	<u>Module 2:</u>	<u>Module 3:</u>	<u>Module 4:</u>	
Communication Total	>4	>5	>3	>3	<u>AND</u>
Social Interaction Total	>7	>6	>6	>6	<u>AND</u>
Communication + Social Total	>12	>12	>10	>10	

AND at least one of these tests scores:

(IQ: Any

OR

Vineland Survey: Less than average

Composite domain total score	<85	<u>AND</u>
Communication domain total score	<85	<u>AND</u>
Living skills domain total score	<85	<u>AND</u>
Motor skills domain total score	<85	<u>AND</u>
Socialization domain total score	<85)	

Sub-Phenotype: Mildly Affected

The sub-phenotype of Mildly Affected is defined as remaining subjects not falling into the Severely Affected category who also meet the following criteria:

ADI-R: Mildly Affected

Total for Section A: Qualitative Abnormalities in Reciprocal Social Interaction	<4	<u>AND</u>
Total of Section B - Non-Verbal: Qualitative Abnormalities in Communication	<4	<u>AND</u>
Total of Section B - Verbal: Qualitative Abnormalities in Communication	<4	<u>AND</u>
Total of Section C: Restricted, Repetitive, and Stereotyped Patterns of Behavior	<2	<u>AND</u>
Total of Section D	≥1	

ADOS: Mildly Affected

	<u>Module 1:</u>	<u>Module 2:</u>	<u>Module 3:</u>	<u>Module 4:</u>	
Communication Total:	1 to 4	1 to 5	1 to 3	1 to 3	<u>AND</u>
Social Interaction Total:	1 to 4	2 to 4	1 to 4	1 to 4	<u>AND</u>
Communication + Social Total:	1 to 7	2 to 8	2 to 7	2 to 7	

IQ: Average and Higher than Average IQ > 85

OR

Vineland Survey: Average and Higher than Average

Composite domain total score	> 85	<u>AND</u>
Communication domain total score	> 85	<u>AND</u>
Living skills domain total score	> 85	<u>AND</u>
Motor skills domain total score	> 85	<u>AND</u>
Socialization domain total score	> 85)	

Sub-Phenotype: Affected

The sub-phenotype of Affected is defined as subjects/ages remaining (i.e. not falling into the Severely Affected or Mildly Affected category) who also meet the following criteria:

ADI-R: Affected

Total for Section A: Qualitative Abnormalities in Reciprocal Social Interaction	>4	<u>AND</u>
Total of Section B - Non-Verbal: Qualitative Abnormalities in Communication	>4	<u>OR</u>
Total of Section B - Verbal: Qualitative Abnormalities in Communication	>4	<u>AND</u>
Total of Section C: Restricted, Repetitive, and Stereotyped Patterns of Behavior	>2	<u>AND</u>
Total of Section D	≥1	

ADOS: Affected

	<u>Module 1:</u>	<u>Module 2:</u>	<u>Module 3:</u>	<u>Module 4:</u>	
Communication Total	>4	>5	>3	>3	<u>OR</u>
Social Interaction Total	>7	>6	>6	>6	<u>OR</u>
Communication + Social Total	>12	>12	>10	>10	

IQ: Average IQ 85 to 100

OR

Vineland Survey: Average

Composite domain total score	85 to 115	<u>OR</u>
Communication domain total score	85 to 115	<u>OR</u>
Living skills domain total score	85 to 115	<u>OR</u>
Motor skills domain total score	85 to 115	<u>OR</u>
Socialization domain total score	85 to 115)	

Precision of Definitions

After applying all aforementioned rules for the phenotype parameters definition, the precision for the definition of the **Autism Spectrum Phenotype and Sub-Phenotype** is defined as follows:

I. **Ideal Phenotype**

ADI+ADOS+nviIQ+viIQ+Vineland*

II. **Very Good Phenotype**

ADI+ADOS + (nviIQ or viIQ) +Vineland*

III. **Good Phenotype**

ADI+ADOS + (nviIQ or viIQ or Vineland)*

IV. **Clinical Diagnosis**

Clinical_diagnosis is provided and categorized in the Phenotype field in Browse Data. To ascertain the specific clinical diagnosis provided, one must query the clinical diagnosis field to see specific categorizations on clinical diagnosis.

IQ measures used for categorization are:

- Differential Ability Scales – Early Years
- Differential Ability Scales – School Age
- Raven’s Coloured Progressive Matrices
- Stanford-Binet Intelligences Scales
- Leiter International Performance Scale - Revised
- Wechsler Intelligence Scale for Children
- Wechsler Abbreviated Scale of Intelligence

* nviIQ = non-verbal IQ and viIQ = verbal IQ

Autism Sub-Types and the Minimally Verbal Categorization

To address the need to further define individuals with an autism spectrum diagnosis, many autism sub-types have been defined. As with the phenotype and sub-phenotype categories, the sub-type category can be expanded and enhanced by the autism research community through collaboration with the NDAR team.

For illustration, the minimally verbal represents the most complex sub-type currently defined by NDAR. To define this sub-type, rules were created and implemented based on scores from relevant clinical measures. The minimally verbal categorization follows in a similar manner as that of phenotype where first the ideal categorization was used, which includes subjects that had a minimally verbal designation from an ADOS, an ADI and at least one OTHER validation measure (i.e. Mullen, Vineland, MacArthur-Bates) that confirm the minimally verbal designation at a particular age. Subjects – at a particular age – are noted with an “Ideal” Minimally Verbal designation. Next, we checked the presence of minimally verbal in two out of the three groups (ADI/ADOS or ADI/OTHER or ADOS/OTHER) for a minimally verbal designation. Subjects that had a minimally verbal designation were then defined with a “Good” Minimally Verbal designation. Like phenotype observations \pm 3 months were used for evaluation. The following rules were used for Minimally Verbal.

Assessment	Element	Value	Interpretation
ADI-R	Overall level of language	1; 2	2 (fewer than 5 words total or speech not used on daily basis); 1 (speech use on daily basis with min 5 words in the last month)
ADOS Module 1	Overall level of non-echoed language	2; 3; 8	8 (no words); 3 (at least one word but fewer than 5); 2 (single words only, must use min 5 during session)
ADOS Module 2	Overall level of non-echoed language	2; 3; 7	7 (all speech echoed); 3 (single words only or no spoken language); 2 (mostly single words)
Mullen	Expressive language - names objects	0	Names 0
Mullen	Expressive language - oral vocabulary	0	Names less than 5
Vineland Survey	Expressive language - answers with word	0	Never
Vineland Survey	Expressive language - names at least 3 objects	0	Never
MacArthur-Bates Words & Gestures	Productive vocabulary measure	<6	Produces less than 6 words
MacArthur-Bates Words & Sentences	Productive vocabulary measure	<6	Produces less than 6 words