

Data Submission Office Hours Webinar

April 2023



National Institute
of Mental Health

Purpose

- **Purpose:** Learn the general NDA Data Submission process during a Submission Cycle
- **Agenda:**
 - Who, what, when and how to submit data
 - Prerequisites to data submission
 - How to check if your Data Structures are ready for submission
 - Submission Templates
 - Mandatory Data Structures
 - Submission Exemptions

Who should be submitting data & when?

NDA expects you to submit data biannually by deadlines:

January 15 & July 15

1

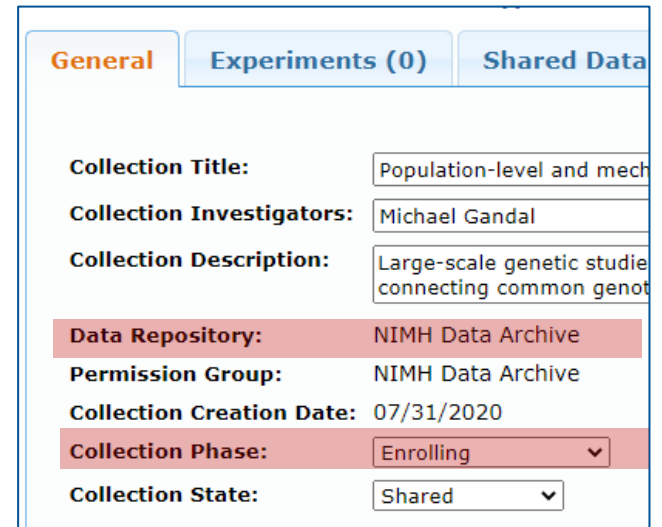
Your NDA Collection is in the NIMH Data Archive data repository

2

Your Collection phase is in Pre-Enrollment, Enrolling or Funding Completed – Data Expected

3

The Data Expected list has an initial submission date set to the deadline date or earlier



General Experiments (0) Shared Data

Collection Title: Population-level and mech

Collection Investigators: Michael Gandal

Collection Description: Large-scale genetic studie connecting common genot

Data Repository: NIMH Data Archive

Permission Group: NIMH Data Archive

Collection Creation Date: 07/31/2020

Collection Phase: Enrolling

Collection State: Shared

▼ Data Expected List: Mandatory Data Structures

These data structures are mandatory for your NDA Collection. Please update for the entire study.

Data Expected	Targeted Enrollment	Initial Submission
Research Subject and Pedigree ⓘ	46	10/01/2021

▼ Data Expected List: Data Structures per Research Aims

Who should be submitting data & when?

NDA expects you to submit data biannually by deadlines:

April 1 & October 1

1

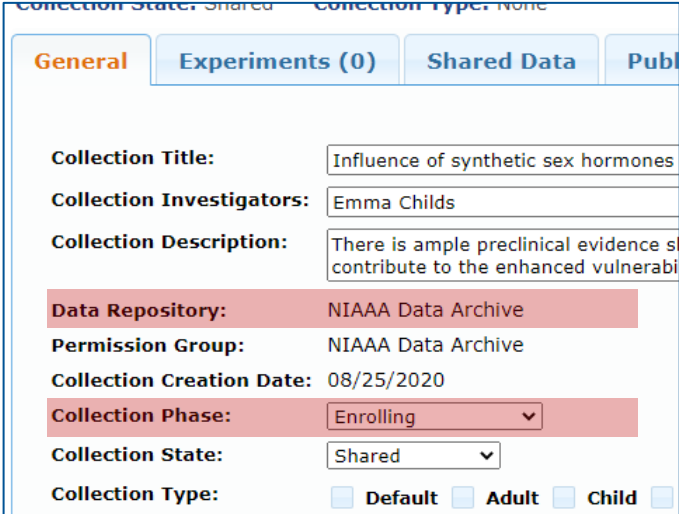
Your NDA Collection is in the NIAAA Data Archive data repository

2

Your Collection phase is in Pre-Enrollment, Enrolling or Funding Completed – Data Expected

3

The Data Expected list has an initial submission date set to the deadline date or earlier



Collection State: Shared Collection Type: None

General Experiments (0) Shared Data Publ

Collection Title: Influence of synthetic sex hormones

Collection Investigators: Emma Childs

Collection Description: There is ample preclinical evidence s contribute to the enhanced vulnerabi

Data Repository: NIAAA Data Archive

Permission Group: NIAAA Data Archive

Collection Creation Date: 08/25/2020

Collection Phase: Enrolling

Collection State: Shared

Collection Type: Default Adult Child

▼ Data Expected List: Mandatory Data Structures

These data structures are mandatory for your NDA Collection. Please update t for the entire study.

Data Expected	Targeted Enrollment	Initial Submission
Research Subject and Pedigree ①	46	10/01/2021

▼ Data Expected List: Data Structures per Research Aims

What do I need to start submitting data?

Data Submission
Agreement (DSA)

Data Expected List

Permissions to Submit

Global Unique Identifiers
(GUIDs)

Data Submission Agreement (DSA)

- Submit through the eDSA Tool
- Follow the instructions in Welcome to Data Sharing
- Access to your NDA Collection & Data Expected list

NIMH Data Archive

NDA ABCD CCF OAI NIAAQA

NDA Electronic Data Submission Agreement

Review PI Details Review NDA Collection Review Grant Details Access Types Other Repositories Institutional Official Review DSA Complete

Welcome

Welcome to the National Institute of Mental Health (NIMH) Data Archive (NDA) electronic Data Submission Agreement (eDSA) tool. NDA is an NIH-funded collaborative resource that contains human subjects research data from multiple research data repositories. The NDA Data Submission Agreement (DSA) is required to submit data to data repository within the NIMH Data Archive.

With this tool, you can populate, review, and submit the NDA Data Submission Agreement. The Agreement must be signed by the Principal Investigator of an NIH grant and an Authorized Institutional Business Official. Principal Investigators on NIH grants with NDA data sharing expectations are expected to complete an NDA DSA within 6 months of initial award.

Log in with your eRA Commons username and password to begin. Once you have completed the Data Submission Agreement you will select an Authorized Institutional Business Official (Signing Official) at your institution who will sign off on the document electronically. If you are not sure who at your institution is appropriate for you to send an NDA Data Submission Agreement for approval, the NDA Help Desk can email a list of your institution's Signing Officials.

Contact the [NDA Help Desk](#) with any questions or concerns.

Data Expected List

- List of all Data Structures you'll submit data for
- Admin permissions required to edit

General
Experiments (5)
Shared Data
Publications (25)
Data Expected (26)
Submissions
Associated Studies (0)
Permissions

Hide Data Expected tab from public view

▼ Data Expected List: Mandatory Data Structures

These data structures are mandatory for your NDA Collection. Please update the Targeted Enrollment number to accurately represent the number of subjects you expect to submit for the entire study.

Data Expected	Targeted Enrollment	Initial Submission	Subjects Submitted	Initial Share	Subjects Shared	Status	Actions
Research Subject and Pedigree ⓘ	399	01/15/2022	248	05/15/2022	0	Approved ⓘ	

▼ Data Expected List: Data Structures per Research Aims

These data structures are specific to your research aims and should list all data structures in which data will be collected and submitted for this NDA Collection. Please update the Targeted Enrollment number to accurately represent the number of subjects you expect to submit for the entire study.

+ New Data Expected ?

Data Expected	Targeted Enrollment	Initial Submission	Subjects Submitted	Initial Share	Subjects Shared	Status	Actions
ABC Community ⓘ	399	01/15/2022	235	05/15/2022	0	Approved ⓘ	
ADOS ⓘ	399	01/15/2023	0	05/15/2023	0	Approved ⓘ	
ADI-R ⓘ	399	01/15/2022	157	05/15/2022	0	Approved ⓘ	
PDD Behavior Inventory (PDDBI) ⓘ	399	01/15/2022	234	05/15/2022	0	Approved ⓘ	

Permissions to Submit

- Submission permissions required to submit data
- Your PI manages permissions via the Permissions tab

Profile **Collections (1)** Data Permissions Documents Studies (0) Data Packages Permissions Lookup

Compliance Dashboard (1) ?

This page displays a list of all NDA Collections that you have permission to access. It also indicates the compliance status of individual NDA Collections.

HIDE CLOSED COLLECTIONS

Data Sharing Status	ID	Title	Role Codes	PI List	Grant(s)	Obligated Funds	Total Subjects Submitted / Shared	Collection Phase	Permissions
Green	C3670	The Autism Biomarkers Consortium for Clinical T...		James Mcpartland	U19MH108206-05	\$23,916,153	291 / 0	Enrolling	Submission Query

Collection State: Private Collection Type: None Remote Upload Not Associated with a Collection #0

General Experiments (0) Shared Data Publications (0) Data Expected (5) Submissions Associated Studies (0) **Permissions** ?

Collection Members

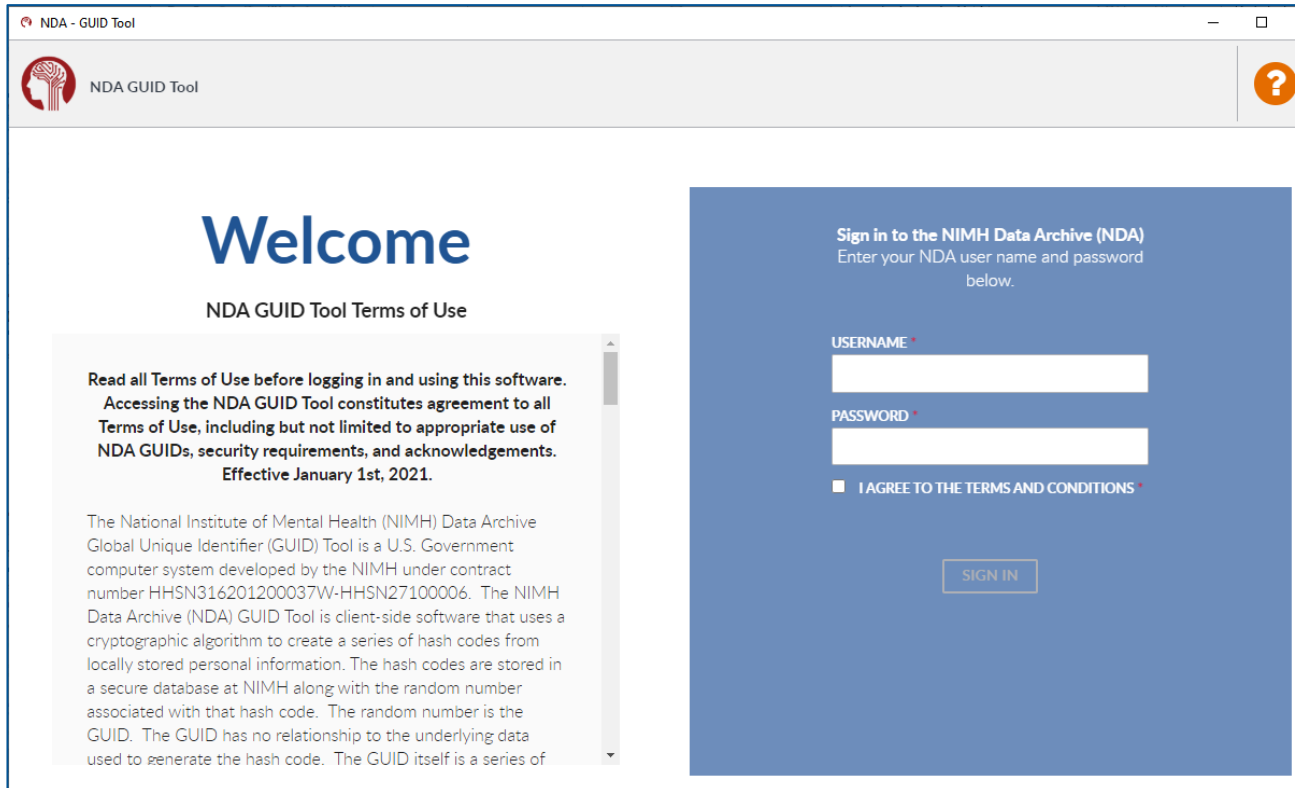
Search

Type	First Name	Last Name	Username	Email	Admin	Submission	Query	Action
User	###	@@@	test	abc@yahoo.com	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Add member by username or email

Global Unique Identifiers (GUIDs)

- Required for all data submissions
- Access to NDA GUID Tool requested via the NDA Help Desk
- Submission permissions required



NDA - GUID Tool

NDA GUID Tool

Welcome

NDA GUID Tool Terms of Use

Read all Terms of Use before logging in and using this software.
Accessing the NDA GUID Tool constitutes agreement to all Terms of Use, including but not limited to appropriate use of NDA GUIDs, security requirements, and acknowledgements.
Effective January 1st, 2021.

The National Institute of Mental Health (NIMH) Data Archive Global Unique Identifier (GUID) Tool is a U.S. Government computer system developed by the NIMH under contract number HHSN316201200037W-HHSN27100006. The NIMH Data Archive (NDA) GUID Tool is client-side software that uses a cryptographic algorithm to create a series of hash codes from locally stored personal information. The hash codes are stored in a secure database at NIMH along with the random number associated with that hash code. The random number is the GUID. The GUID has no relationship to the underlying data used to generate the hash code. The GUID itself is a series of

Sign in to the NIMH Data Archive (NDA)
Enter your NDA user name and password below.

USERNAME *

PASSWORD *

I AGREE TO THE TERMS AND CONDITIONS *

SIGN IN

Global Unique Identifier (GUID)

Alphanumeric code used as an identifier for every research participant (ex. NDARRF67XH1)

What information do I need to create a GUID?

- First, middle and last name
- Sex at birth
- Date of birth
- City/Municipality of Birth
- *Recorded as it appears on the birth certificate

The NDA Data Sharing Terms and Conditions include the expectation that you will collect the data necessary to create real GUIDs.

pseudoGUIDs

pseudoGUIDs can be requested in place of GUIDs by emailing the NDA Help Desk:

- Detailed justification for the use of pseudoGUIDs
- # of pseudoGUIDs needed
- Whether pseudoGUIDs will be needed for more than one submission cycle

What if I obtain the information needed to create a GUID, but I've already used a pseudoGUID?

Promote your pseudoGUID using the NDA GUID Tool linking the pseudoGUID to a real GUID

What do I need to start submitting data?

Data Submission Agreement (DSA)

- Submit through the eDSA Tool
- Follow the instructions in Welcome to Data Sharing
- Access to your NDA Collection & Data Expected list

Data Expected List

- List of all Data Structures you'll submit data for
- [NDA Data Dictionary](#)
- Admin permissions required to edit

Permissions to Submit

- Submission permissions required to submit data
- Your PI manages permissions via the Permissions tab

Global Unique Identifiers (GUIDs)

- Required for all data submissions
- Access to NDA GUID Tool requested via the NDA Help Desk
- Submission permissions required

Data Submission

What data should be submitted?

- Please submit all data you have collected.
- You **DO NOT** need to submit data for all the Data Structures.

Should data be submitted cumulatively?

- Clinical and phenotypic data → *cumulative* submissions
 - Your first submission includes 5 subjects at Timepoint A
 - You collect 10 new subjects at Timepoint A & original 5 subjects at Timepoint B
 - Include the 5 subjects Timepoint A & B + 10 new subjects for Timepoint A
- Imaging, EEG, Omics data → non-cumulative submissions

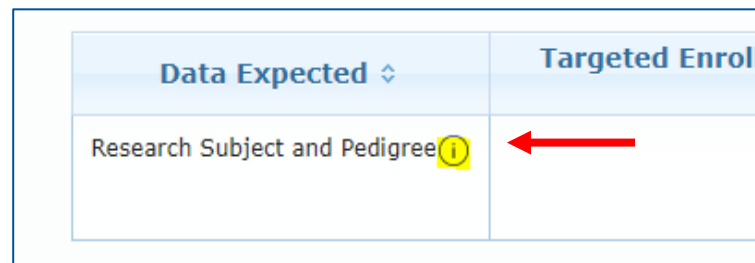
Data Submission cont.

Should I submit any data if NDA is still working on your Data Structures?

- **Yes!**
- You should submit data for Data Structures that are approved and are ready for submission!

How do I know if my Data Structure is ready?

Check the information ("i") icon next to the data structure names via your Data Expected tab!



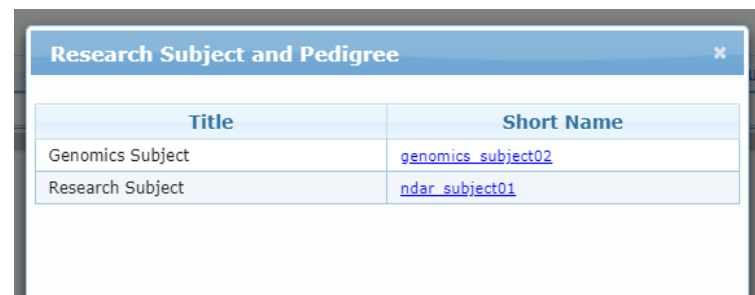
Data Expected ▾	Targeted Enroll
Research Subject and Pedigree ⓘ	

Structure not yet defined:

Data curation is still preparing your Data Structure



If the pop-up provides the title and short name of your data structure: Your data structure is ready for submission!



Research Subject and Pedigree x

Title	Short Name
Genomics Subject	genomics_subject02
Research Subject	ndar_subject01

Download Your Submission Template

Research Subject and Pedigree

Title	Short Name
Genomics	
Research	

Research Subject
404,119 Shared Subjects

DESCRIPTION
Basic information about a study subject including demographics, pedigree, and diagnosis/phenotype

DATE CREATED
04/06/2011

SHORT NAME
ndar_subject01

DOWNLOADS
[Definition](#)
[Submission Template](#)

DATA TYPE
Clinical Assessments

MOST RECENT UPDATE
05/06/2022
[View Change History](#)

VERSION
01

CATEGORIES
Resolve Identifiers

Q Refine Results

QUERY	ELEMENT NAME	DATA TYPE	SIZE	REQUIRED	CONDITION	DESCRIPTION
	subjectkey	GUID		Required		The NDAR Global Unique Identifier (GUID) for research subject

DOWNLOADS
Definition
Submission Template

Submission Template

	A	B	C	D	E	F	G	H	I	
1	ndar_subject	1								
2	subjectkey	src_subject_id	interview_date	interview_age	sex	race	ethnic_group	phenotype	phenotype_description	twins
3										
4										
5										
6										

- Do **not** change the first two rows.
- You can *hide* columns. Do **not** delete any columns.
- Each row = 1 subject at 1 timepoint
 - Example: 1 subject at 3 timepoints = 3 rows

Submission Template

The image displays two overlapping Excel spreadsheets. The top spreadsheet, titled 'ndar_subject01_template (2)', shows a data entry form with columns A through I. The bottom spreadsheet, titled 'ndar_subject01_definitions', shows a table of field definitions. A red box highlights the 'Required' column in the definitions table. A download menu is overlaid on the left, with 'Definition Submission Template' highlighted.

ElementName	DataType	Size	Required	Condition	ElementDescription
subjectkey	GUID		Required		The NDAR Global Unique Identifier (GUID) for research
src_subject_id	String	20	Required		Subject ID how it's defined in lab/project
interview_date	Date		Required		Date on which the interview/genetic test/sampling
interview_age	Integer		Required		Age in months at the time of the interview/test/sa
sex	String	20	Required		Sex of subject at birth
race	String	30	Required		Race of study subject
ethnic_group	String	255	Recommended		Ethnic group
phenotype	String	200	Required		Phenotype/diagnosis for the subject
phenotype_description	String	4000	Required		Description of the phenotype for the subject
twins_study	String	100			In this study of twins?

Required Data Elements

- Required data elements must have a value.
- Conditional data elements should have a value if applicable.
- Recommended and optional data elements can be left blank if not applicable.

ElementName	DataType	Size	Required
subjectkey	GUID		Required
src_subject_id	String	20	Required
interview_date	Date		Required
interview_age	Integer		Required
sex	String	20	Required

ElementName	Size	Required	Condition	Element
zygosity	25	Conditional	twins_study == 'Yes'	Zygos

ElementName	Size	Required
multibirthid		Recommended
tourettes_disorder		Recommended
sample		Recommended
race_other	100	Recommended
respond		Recommended
date_confirmed_phenotype		Recommended
pos_years		Recommended
position1	150	Recommended

Required Data Elements

- If you did not collect a required data element, review ValueRange & Notes for N/A or null values.

ElementName	Size	Required	ValueRange	Notes
sex	20	Required	M;F; O; NR	M = Male; F = Female; O=Other; NR = Not reported
zygosity	255	Conditional	monozygous; dizygous; trizygotic; 999	999=unknown
reg_edu		Recommended	1::34; 999	1=None; 2=Preschool; 3=Kindergarten; 4=1st grade; 5=2nd grade; 6=3rd grade; 7=4th grade; 8=5th grade; 9=6th grade; 10=7th grade; 11=8th grade; 12=9th grade; 13=10th grade; 14=11th grade; 16=High School Graduate; 17=GED; 18=Some college credit but less than 1 year; 19= Some college, no degree; 20=Associates degree (e.g., AA, AS); 21=Bachelor's degree (e.g., BA, AB, BS); 22=Masters degree (e.g., MA, MS, MEng, MEd, MSW, MBA); 23=Professional degree (e.g., MD, DDS, DVM, LLB, JD); 24=Doctorate degree (e.g., PhD, EdD); 25=One or more years of college at a 2-year program, no degree; 26=One year of college at a 4-year program, no degree; 27=Two years of college at a 4-year program, no degree; 28=Three years or more of college at a 4-year program, no degree; 29= Trade or vocational school; 30=Eighth grade or less; 999=Unknown
brain_injury		Recommended	1::3	1=Yes; 2=No; 3=NA

5 Required Data Elements in All Structures

ELEMENT NAME	DATA TYPE	SIZE	REQUIRED	DESCRIPTION

How do I submit data?

Via the [Validation and Upload Tool](#)!

- Submission permissions required
- HTML Validation and Upload Tool
 - Cumulative data (e.g. clinical and phenotypic data)
 - [Tutorial for the HTML Validation and Upload Tool](#)
- Python Client Validation and Upload Tool
 - Non-cumulative imaging files
 - Files over 2.5 GB

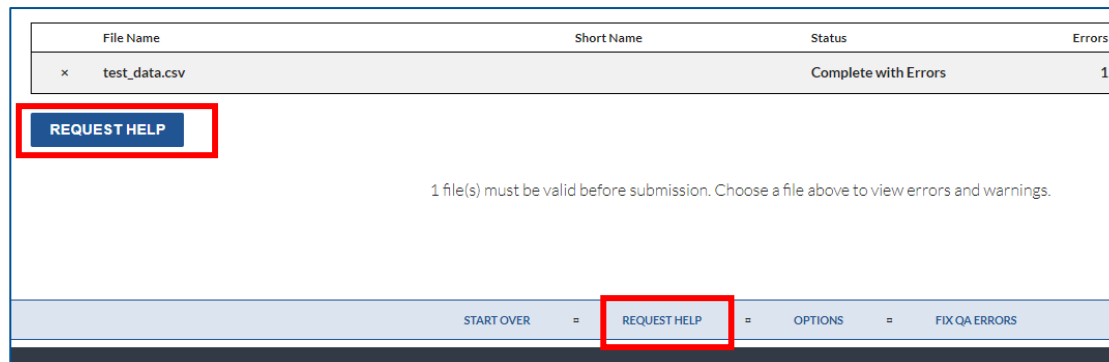
*You **must** use the Submission Templates from the NDA Data Dictionary and/or be provided by a Data Curator to submit your data.

Validation and Upload Tool

Validates data against the
NDA Data Dictionary

Ensures data is harmonized
correctly

- Use it as your personal NDA data manager
- Warnings vs Errors
 - Errors **must** be resolved
 - Users **can** submit with warnings present
- Need help? Click “Request Help”



The screenshot displays the interface of the Validation and Upload Tool. At the top, there is a table with the following columns: File Name, Short Name, Status, and Errors. The table contains one row: × test_data.csv, Complete with Errors, 1. Below the table, a blue button labeled 'REQUEST HELP' is highlighted with a red box. Below the button, a message states: '1 file(s) must be valid before submission. Choose a file above to view errors and warnings.' At the bottom of the interface, there is a navigation bar with buttons: START OVER, REQUEST HELP (highlighted with a red box), OPTIONS, and FIX QA ERRORS.

File Name	Short Name	Status	Errors
× test_data.csv		Complete with Errors	1

REQUEST HELP

1 file(s) must be valid before submission. Choose a file above to view errors and warnings.

START OVER = REQUEST HELP = OPTIONS = FIX QA ERRORS

What if none of your Structures are ready?

- If you are waiting on *NDA* to complete your data structures where you have ready-to-submit data, **don't worry!**
- There will always be at least one ready-to-submit data structure.

Mandatory Data Structures

▼ **Data Expected List: Mandatory Data Structures**

These data structures are mandatory for your NDA Collection. F

Data Expected ⚡

- Research Subject and Pedigree ⓘ
- NIMH CDE - GAD-7 ⓘ
- NIMH CDE - PHQ-9 ⓘ
- NIMH CDE - DSM-5 Crosscutting Adults ⓘ
- NIMH CDE - WHODAS 2.0 ⓘ








Research Subject & Pedigree (RS&P)

- ndar_subject01
- genomics_subject02 (omics only)

NIH Common Data Elements (CDE) (NOT-MH-20-067)

Mandatory Data Structures

- **Update** the targeted enrollment number.
- **Submit** data cumulatively.

Data Expected 	Targeted Enrollment 
Research Subject and Pedigree 	<input type="text" value="1,000"/>
NIMH CDE - GAD-7 	1,000
NIMH CDE - PHQ-9 	1,000
NIMH CDE - DSM-5 Crosscutting Adults 	1,000
NIMH CDE - WHODAS 2.0 	1,000

Don't have any data to submit / Structures not ready?

Submit a submission exemption via the Data Expected tab of your NDA Collection!

Types of Exemptions

- A. No New Data This Period: No new data to submit
- B. Deadline Extension: Data submission expected after the submission deadline
- C. No New Data - Change Phase to Data Analysis: Data submission is completed for the entire grant



Received QA Results?

If QA errors were identified from a previous submission cycle, do I still submit data?

- Please [correct the QA errors](#) and resubmit your data before submitting any new data.
- Follow the instructions in your QA results notification.

More Questions?

If you have further questions or would like a copy of the slide deck, please email the NDA Help Desk at NDAHelp@mail.nih.gov.

All information from the presentation and more can be found on our Submission Cycle FAQ: <https://nda.nih.gov/nda/submission-cycle-faq.html>.

Thank you for joining!