NeuroNEXT Network

Standard Operating Procedure (SOP)

Generation and Validation of Analysis Data Sets

Version 2.0

SOP NN BIO 904

Originators: NeuroNEXT CCC and DCC Personnel

Reviewed and Approved by:

Christopher S. Coffey, PhD (DCC Principal Investigator)

Merit E. Cudkowicz, MD MSc (CCC Principal Investigator)

Marianne Kearney Chase, BA (CCC Director of Clinical Operations)

Dixie J. Ecklund, RN MSN MBA (DCC Associate Director)

Katherine B. Gloer, PhD (DCC Quality Management Lead)

Janice Cordell, RN MPH (NINDS, NeuroNEXT Program Official)

September 21, 2016
Issue Date

October 21, 2016
Effective Date (*30 calendar days after the Issue Date*)

NN BIO 904

NEURONEXT NETWORK STANDARD OPERATING PROCEDURE FOR GENERATION AND VALIDATION OF ANALYSIS DATA SETS

SOP: NN BIO 904 Version No: 2.0 Effective Date: 21Oct2016	GENERATION AND VALIDATION OF ANALYSIS DATA SETS	Supercedes Document: Version 1.0 Effective Date: 13May2012

1. POLICY

This SOP describes NeuroNEXT Data Coordinating Center (DCC) procedures for creating and validating analysis data sets for:

- critical study analyses (analyses of the primary and secondary endpoints as described in the Statistical Analysis Plan [SAP]; interim analyses; and analyses for the final study report and primary manuscript; and
- analyses that are requested by study investigators or other authorized researchers and approved by the Sponsor and/or CTSDMC Leadership (e.g. for abstracts, presentations, or secondary analyses after the conclusion of the study).

Before data sets are created, all DCC databases are carefully developed, and the SOPs related to database development and validation (e.g., NN CS 702, NN CS 703, NN DM 1004),data collection and management (e.g., NN DM 1001, NN DM 1005), and routine monitoring visits (NN SS 403) are followed. The procedures described in these SOPs help to ensure that all data stored in DCC electronic databases are complete, and that all targeted data have been validated against source documents at the Clinical Study Sites (CSS).

Before the data sets are created, a written Statistical Analysis Plan (SAP) is developed that includes plans for final analyses as well as an *a priori* defined plan for interim analyses (see SOP NN BIO 902). Study endpoints that will be analyzed in the final and interim analyses are defined in the SAP. Analysis data sets are created so that the endpoints may be determined and analyzed according to the parameters defined in the SAP.

The SAS® statistical package is the primary tool used for analyses conducted at the DCC, and all analysis data sets are initially created in the SAS® format. If specific analysis methods require the use of another statistical format, the SAS® files are exported to that format.

2. SCOPE

This SOP has been developed to be in alignment with federal regulations and Good Clinical Practices (GCP) as set forth in the 1996 ICH E6 Consolidated Guidance. The policies and procedures described in this SOP apply to the NeuroNEXT Clinical Coordinating Center (CCC) and DCC within the context of their oversight and advisory roles for the NeuroNEXT Network, and to all NeuroNEXT investigators, staff, subcontractors, External Biostatisticians (if applicable), or other entities associated with the NeuroNEXT Network who manage, oversee, and conduct research regulated by FDA and/or applicable review committees.

3. ROLES AND RESPONSIBILITIES

DCC Biostatisticians are responsible for adhering to the procedures outlined in this SOP, and for ensuring that the procedures are followed by other DCC staff members who may participate in biostatistics projects.

The Study Biostatistician is responsible for generating specifications for analysis data sets (in consultation with the DCC Lead Biostatistician, the DCC PI, and the PPI as necessary), for creating and validating the SAS® data set, and for converting the SAS® data set to another format (if required for the analysis).

4. APPLICABLE REGULATIONSAND GUIDELINES

ICH E3	Structure and Content of Clinical Study Reports
ICH E6, 4.5	Compliance with Protocol
ICH E6, 4.9	Records and Reports
ICH E6, 5.1	Quality Assurance and Quality Control
ICH E6, 5.5	Trial Management, Data Handling and Record Keeping
ICH E8	General Considerations for Clinical Trials (December 1997)
ICH E9	Statistical Principles for Clinical Trials (September 1998)

5. REFERENCES TO OTHER APPLICABLE SOPS

NN CS 702	Application Development and Validation
NN CS 703	IT Environments
NN BIO 905	Validating Statistical Programs and Deliverables
NN DM 1001	Clinical Data Management
NN DM 1004	Specifications Development, Testing Plans, and Validation Documentation
NN DM 1005	Data Collection and Data Handling

6. ATTACHMENTS AND REFERENCES

NN BIO 904 - A	Document History

7. TERMS AND ABBREVIATIONS

The following terms and abbreviations are used in this document:

CCC	Clinical Coordinating Center at Massachusetts General Hospital
DCC	Data Coordinating Center at The University of Iowa
DM	Data Management
DM Lead	Leader of the Data Management team at the DCC
External Biostatistician	A Biostatistician who is not a member of the DCC Biostatistics Team
IT	Information Technology
IT Lead	Leader of the Information Technology team at the DCC
Lead Biostatistician	Primary Biostatistician for a study
PPI	Protocol Principal Investigator
QA	Quality Assurance
QA Officer	Quality Assurance officer for a study
SAP	Statistical Analysis Plan
Study Biostatistician	A member of the DCC Biostatistics team. At least one Study Biostatistician will be assigned to each study. The Lead

Biostatistician may also act as a Study Biostatistician for a specific study.

8. SPECIFIC PROCEDURES

A. Generating Specifications for Analysis Data Sets

#	*Who*	*Task*	*Attachment/ Reference*	*Related SOP*
1.	DCC PI, DCC Lead Biostatistician, and DCC Study Biostatistician	For critical study analyses, thoroughly review the SAP and other relevant documents, and discuss the analysis with the PPI.		
2.	DCC Study Biostatistician	For critical study analyses designated in the SAP, obtain an in-depth understanding of study design, endpoint definitions, and the relevant statistical methods.		
3.	DCC Study Biostatistician	Consult with the PPI to resolve questions or to clarify ambiguities before generating analysis data sets.		
4.	DCC Study Biostatistician	Understand the structure of study database so that any necessary data manipulations may be performed.		
5.	DCC Lead Biostatistician and DCC Study Biostatistician	Develop specifications for all derived variables that will be used in the analyses.		
6.	DCC Study Biostatistician	If an analysis will be based on locked data, coordinate with all NeuroNEXT DCC and CCC staff to ensure that the electronic study database is complete and ready to be locked before a data lock is performed.		

B. Creating Analysis Data Sets

#	*Who*	*Task*	*Attachment/ Reference*	*Related SOP*
1.	DCC IT Lead or designee	Perform data lock, if applicable, using agreed-upon specifications.		
2.	DCC Study Biostatistician	If a data lock is not conducted, create a frozen snapshot of the source data to be used in generating the analysis data set, and store the frozen source data in the appropriate location.		
3.	DCC Study Biostatistician	Perform necessary data manipulations, such as data table merges.		

#	Who	Task	Attachment/ Reference	Related SOP
4.	DCC Study Biostatistician	Create derived variables based on specifications.		
5.	DCC Study Biostatistician	Validate the derived variables: • Develop test code to create derived variables. • For critical study analyses, assign a second Biostatistician to independently code and create derived variables. • Compare values for derived variables and identify all values where there is disagreement. • Resolve cases where there is disagreement.		NN BIO 905

C. Validating Analysis Data Sets

#	Who	Task	Attachment/ Reference	Related SOP
1.	DCC Study Biostatistician	Validate the analysis data set: • Ensure that the appropriate observations are included in each table • Ensure that the appropriate variables are included in each table. • Ensure that treatment assignments are accurate. • Examine each variable in a data table to ensure that there are no unexpected missing values • Examine each variable to identify possible outliers.		
2.	DCC Study Biostatistician	Work with DCC IT and DCC Coordinators to determine if unusual values are real or should be corrected.		

D. Converting SAS® Data Sets to other Formats

#	Who	Task	Attachment/ Reference	Related SOP
1.	DCC Study Biostatistician	If required for the analysis, convert SAS® data sets to the desired format.		

#	*Who*	*Task*	*Attachment/ Reference*	*Related SOP*
2.	DCC Study Biostatistician	Perform QC checks to verify the integrity of the converted file and to ensure that values have been imported accurately and completely: • Spot-check the converted data set against the original data set. • Tabulate simple descriptive statistics • Compare the results in the two environments • Maintain documentation of the comparison.		

Attachment NN BIO 904 - A. Document History

NeuroNEXT Network Standard Operating Procedure (SOP) Generation and Validation of Analysis Data Sets SOP NN BIO 904				
Version	**Description of Modification**	**Reason or Justification for Modification**	**Issue Date**	**Effective Date**
1.0	New	N/A	13Apr2012	13May2012
2.0	Expanded Policy section and certain specific procedures (Section 8.B) to distinguish between critical study analyses and additional analyses that are requested by investigators. Added description of data freezes used to generate certain analysis data sets, and outlined QC checks that are performed after converting SAS data sets to other formats.	Updates for Version 2.0.	21Sep2016	21Oct2016