Clinical Development

JNJ-629330

12-101

Anonymisation Data Derivation Specification Document

Document Type	Reference document
Document Version	Final
Date	Aug 17, 2015

Table of contents

Status and Version	Release Date	Summary of Key Changes

1. Datasets

1.1. Specifications Introduction

This specification for each dataset will be in two parts

- Dataset description
- Variables within dataset

Part I: Dataset description

Dataset	Name of dataset
Creating Program	The program that created the dataset
Description	Short description
Unique Identifier	Unique key
Sorted by	Sort key
Notes	Any useful notes

Part II: Variables within dataset

Variable	SAS variable name
Type	Character or Numeric
Label	SAS variable label
Codes	Codelist name
Comments	Variable source derivation explanation if variable derived.

1.2. Guidelines for Preparing Data

The data will be provided according to the De-identified/ Anonymisation data guidelines standards with the following exceptions:

- Subject initials will not be provided
- Subject and center/site numbers will be assigned in a random manner so they are not matching the subject and center/site numbers that were used in the actual trial
- Date of birth will not be provided, only age in years and grouped to protect PII as per HIPAA rules (ages above 89 will be assigned to 90+).
- Remove the free text verbatim terms.
- Remove “Other” free text terms.
- Drug Record Number will not be provided.
- Drug Sequence Number will not be provided.
- Accession Number will not be provided.
- Vial, Bottle, lot, kit number will not be provided.
- Central Lab Specimen Label Number will not be provided.
- Lab Identifier information will not be provided.

- Vendor Panel Comments will not be provided.
- Vendor Test Specific Comments will not be provided.
- Lab Name information will not be provided.
- All original dates relating to individuals subject will be removed. Instead a Relative study day would be provided.
- Complete missing value variables will be removed.
- Partial date's relative day cannot be calculated.
- INVDATA Dataset will not be submitted, since it contains sensitive information regarding Investigator.
- COMMENTS dataset will be submitted with zero observation.

1.3. Data Files

The 12-101 Clinical Study Report (CSR) data should be used for converting to de-identification. Use the 12-101 CSR data from the following folders.

1.4. Data Domains

1.4.1. Demographics– DEMOG

Dataset	DEMOG
Creating program	demog.sas
Description	Demographics
Unique identifier	DINVNO,DPATNO
Sorted by	DINVNO,DPATNO
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines PTINIT,RACESP,DOBMM,DOBDD,DOBYY,DOBDT,VISMM,VISDD,VISYY,VISDT

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity
VISNO	Char	Visit Number		Collected at CRF.
CONSYN	Char	Written Consent (y/n)		Collected at CRF.
INCEXCYN	Char	Met Incl/Excl criteria (y/n)		Collected at CRF.
SEX	Char	Gender		Collected at CRF.

Variable	Type	Label	Codes	Comments
RACE	Char	Race		Group element to protect PII.
TRTGRP	Char	Study Drug Treatment Assignment		Collected at CRF.
AGE	Num	Age in Years		Collected at CRF.

1.4.2. Adverse Events – ADVERSE

Dataset	ADVERSE
Creating program	adverse.sas
Description	Adverse Events
Unique identifier	DINVNO,DPATNO,LLT_NAME, AESTRTDY, AESTOPDY,SEQ
Sorted by	DINVNO,DPATNO,LLT_NAME, AESTRTDY, AESTOPDY,SEQ
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines: PTINIT,SYMPREF,AESTRTMM,AESTRTDD,AESTRTYY,AESTRTDT,AESTOPMM,AESTOPDD,AESTOPYY,AESTOPDT,SIGNMM,SIGNDD,SIGNYY,SIGNDT,SYMPTOM

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity

Variable	Type	Label	Codes	Comments
AENONE	Char	AE None Check Box		Collected at CRF.
SERIOUS	Char	Serious AE?		Collected at CRF.
AECONT	Char	AE Continuing		Collected at CRF.
AEOUTC	Char	AE Outcome		Collected at CRF.
AESEV	Char	AE Severity		Collected at CRF.
AEFREQ	Char	AE Frequency		Collected at CRF.
SDREL	Char	Relationship to Study Drug		Collected at CRF.
ACTNONE	Char	Action: None		Collected at CRF.
ACTDISC	Char	Action: Discontinued Study		Collected at CRF.
ACTDRUG	Char	Action: Drug Therapy Initiated		Collected at CRF.
ACTOTRT	Char	Action: Other Therapy Initiated		Collected at CRF.
ACTEVAL	Char	Action: Additional Evaluation		Collected at CRF.
SIGNYN	Char	AE Page Signed (y/n)		Collected at CRF.
SEQ	num	Sequence number		Collected at CRF.
LLT_CODE	num	MedDRA Code		Collected at CRF.
LLT_NAME	char			Collected at CRF.
PT_CODE	num			Collected at CRF.
PT_NAME	char			Collected at CRF.
SOC_CODE	num			Collected at CRF.
SOC_NAME	char			Collected at CRF.

Variable	Type	Label	Codes	Comments
SOC_ABBR	char			Collected at CRF.
AESTRTDY	num	Relative AE Start Day		If AESTRTDT and VISDT not missing then perform below logic to calculate AESTRTDY, If AESTRTDT less than VISDT then (AESTRTDT - VISDT).Else if AESTRTDT is greater than equal to VISDT then (AESTRTDT- VISDT) +1.
AESTOPDY	num	Relative AE Stop Day		If AESTOPDT and VISDT not missing then perform below logic to calculate AESTOPDY, If AESTOPDT less than VISDT then (AESTOPDT - VISDT).Else if AESTOPDT is greater than equal to VISDT then (AESTOPDT- VISDT) +1.
SIGNDY	num	Relative Signature Day		If SIGNDT and VISDT not missing then perform below logic to calculate SIGNDY, If SIGNDT less than VISDT then (SIGNDT - VISDT).Else if SIGNDT is greater than equal to VISDT then (SIGNDT- VISDT) +1.

1.4.3. Comments – COMMENTS

Dataset	COMMENTS
Creating program	comments.sas
Description	Comments
Unique identifier	Not Applicable
Sorted by	Not Applicable
Notes	Comments data is sensitive data, contains free text information. Empty dataset will be submitted.

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity
VISNO	char	Visit Number		Collected at CRF.
DATASET	char	Dataset for Linking		Collected at CRF.
FIELD	char	Dataset field for Linking		Collected at CRF.
SEQ	num	Record Tieback Sequence #		Collected at CRF.
COMMSEQ	num	Comment Sequence #		Collected at CRF.

1.4.4. Concomitant Medications – CONMED

Dataset	CONMED
Creating program	conmed.sas
Description	Concomitant Medications
Unique identifier	DINVNO,DPATNO,MEDPREF, VISNO
Sorted by	DINVNO,DPATNO,MEDPREF ,VISNO
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines: PTINIT,CONMED,CMIND

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
PATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity
VISNO	char	Visit Number		Collected at CRF.
SEQ	num	Sequence #		Collected at CRF.
CMXINFO	char	Additional Conmed Info		Collected at CRF.
MEDPREF	char	Preferred Term		Collected at CRF.
MEDCLSTX	char			Collected at CRF.

1.4.5. Medical History – MEDHIST

Dataset	MEDHIST
Creating program	medhist.sas
Description	Medical History
Unique identifier	DINVNO,DPATNO
Sorted by	DINVNO,DPATNO
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines or due to missing values: PTINIT, DIAGMM,DIAGYY,DIABETES,LIVER,REPRO,OTHERS1,OTHERS2,OTHERS3,OTHERS4,OTHERS5,OTHERS6,OTHERS7,OTHERS8,OTHERS9,OTHERS10,OTHERS11, OTHERS12,OTHERS13,OTHERS14,OTHERS15,VISMM,VISDD,VISYY,VISDT

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity
VISNO	char	Visit Number		Collected at CRF.
HXADHDYN	Char	Family History of ADHD?		Collected at CRF.
PGCOMPYN	Char	Complications w/ Pregnancy/Birth		Collected at CRF.
PREMEDYN	Char	Prior Treatment for ADHD?		Collected at CRF.
CINONE	Char	ConIll: None		Collected at CRF.

Variable	Type	Label	Codes	Comments
ALLERGY	Char	ConIll: Allergy/Ashma		Collected at CRF.
CV	Char	ConIll: Cardiovascular		Collected at CRF.
DERM	Char	ConIll: Dermatological		Collected at CRF.
EENT	Char	ConIll: Eyes/Ears/Nose/Throat		Collected at CRF.
GI	Char	ConIll: Gastrointestinal		Collected at CRF.
MUSC	Char	ConIll: Musculoskeletal		Collected at CRF.
RENAL	Char	ConIll: Renal/Urological		Collected at CRF.
OTHER	Char	ConIll: Other		Collected at CRF.
TRTGRP	Char	Study Drug Treatment Assignment		Collected at CRF.
VISDY	num	Relative Visit Day		If VISDT and VISDT not missing then perform below logic to calculate VISDY, If VISDT less than VISDT then (VISDT - VISDT).Else if VISDT is greater than equal to VISDT then (VISDT- VISDT) +1.

1.4.6. Prior ADHD Medications – PREMED

Dataset	PREMED
Creating program	premed.sas
Description	Prior ADHD Medications
Unique identifier	DINVNO,DPATNO,MEDPREF, PMLDOSDY,SEQ
Sorted by	DINVNO,DPATNO,MEDPREF ,PMLDOSDY,SEQ
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines : PTINIT,PREMED,PMLDOSMM,PMLDOSDD,PMLDOSYY,PMLDOSDT

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity
VISNO	char	Visit Number		Collected at CRF.
SEQ	num	Sequence Number		Collected at CRF.
PMDOSE	char	PreMed Dose		Collected at CRF.
PMXINFO	char	PreMed Extra Info		Collected at CRF.
MEDPREF	char	Preferred Term		Collected at CRF.

Variable	Type	Label	Codes	Comments
MEDCLSTX	char			Collected at CRF.
PMLDOSDY	num	Relative PreMed Last Dose Day		If PMLDOSDT and VISDT not missing then perform below logic to calculate PMLDOSDY, If PMLDOSDT less than VISDT then (PMLDOSDT - VISDT).Else if PMLDOSDT is greater than equal to VISDT then (PMLDOSDT- VISDT) +1.

1.4.7. ADHD Rating Scale – RATINGS

Dataset	RATINGS
Creating program	ratings.sas
Description	ADHD Rating Scale
Unique identifier	DINVNO,DPATNO,COMPSCOR, VISNO
Sorted by	DINVNO,DPATNO,COMPSCOR ,VISNO
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines: PTINIT,SIGNMM,SIGNDD,SIGNYY

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity

Variable	Type	Label	Codes	Comments
ADHDTYPE	Char	Predominant Type of ADHD		Collected at CRF.
TSCORE	num	V1: Total Score (From Form)		Collected at CRF.
SIGNYN	Char	Rating Page Signed (y/n)		Collected at CRF.
Q1	num	Attention to detail		Collected at CRF.
Q1C	Char	Attention to detail(Char)		Collected at CRF.
Q2	num	Fidgets or squirms		Collected at CRF.
Q2C	Char	Fidgets or squirms(Char)		Collected at CRF.
Q3	num	Difficulty sustaining attention		Collected at CRF.
Q3C	Char	Difficulty sustaining attention(Char)		Collected at CRF.
Q4	num	Leaves seat		Collected at CRF.
Q4C	Char	Leaves seat(Char)		Collected at CRF.
Q5	num	Does not seem to listen		Collected at CRF.
Q5C	Char	Does not seem to listen(Char)		Collected at CRF.
Q6	num	Runs or climbs excessively		Collected at CRF.
Q6C	Char	Runs or climbs excessively(Char)		Collected at CRF.
Q7	num	Does not follow through		Collected at CRF.
Q7C	Char	Does not follow through(Char)		Collected at CRF.
Q8	num	Difficulty playing quietly		Collected at CRF.
Q8C	Char	Difficulty playing quietly(Char)		Collected at CRF.
Q9	num	Difficulty organizing tasks		Collected at CRF.

Variable	Type	Label	Codes	Comments
Q9C	Char	Difficulty organizing tasks(Char)		Collected at CRF.
Q10	num	On the go		Collected at CRF.
Q10C	Char	On the go(Char)		Collected at CRF.
Q11	num	Avoids sustained mental tasks		Collected at CRF.
Q11C	Char	Avoids sustained mental tasks(Char)		Collected at CRF.
Q12	num	Talks excessively		Collected at CRF.
Q12C	Char	Talks excessively(Char)		Collected at CRF.
Q13	num	Loses things		Collected at CRF.
Q13C	Char	Loses things(Char)		Collected at CRF.
Q14	num	Blurts out answers		Collected at CRF.
Q14C	Char	Blurts out answers(Char)		Collected at CRF.
Q15	num	Is easily distracted		Collected at CRF.
Q15C	Char	Is easily distracted(Char)		Collected at CRF.
Q16	num	Has difficulty awaiting turn		Collected at CRF.
Q16C	Char	Has difficulty awaiting turn(Char)		Collected at CRF.
Q17	num	Is forgetful		Collected at CRF.
Q17C	Char	Is forgetful(Char)		Collected at CRF.
Q18	num	Interrupts or intrudes		Collected at CRF.
Q18C	Char	Interrupts or intrudes(Char)		Collected at CRF.

Variable	Type	Label	Codes	Comments
COMPSCOR	num	Computed Total Score		Collected at CRF.
VISNO	Char			Collected at CRF.
SIGNDY	num	Relative Signature Day		If SIGNDT and VISDT not missing then perform below logic to calculate SIGNDY, If SIGNDT less than VISDT then (SIGNDT - VISDT).Else if SIGNDT is greater than equal to VISDT then (SIGNDT- VISDT) +1.

1.4.8. Parent Satisfaction Survey – SATISF

Dataset	SATISF
Creating program	satisf.sas
Description	Parent Satisfaction Survey
Unique identifier	DINVNO,DPATNO,DAY
Sorted by	DINVNO,DPATNO,DAY
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines : PTINIT,SATMM,SATDD,SATYY,SATDT

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity
VISNO	char	Visit Number		Collected at CRF.
DAY	char	Study Day		Collected at CRF.
SQ1	num	Behavior Improved		Collected at CRF.
SQ1C	char	Behavior Improved		Collected at CRF.
SQ2	num	Has been paying more attention		Collected at CRF.
SQ2C	char	Has been paying more attention(Char)		Collected at CRF.

Variable	Type	Label	Codes	Comments
SQ3	num	Played better with friends		Collected at CRF.
SQ3C	char	Played better with friends(Char)		Collected at CRF.
SQ4	num	Fidgets or cannot sit still		Collected at CRF.
SQ4C	char	Fidgets or cannot sit still(Char)		Collected at CRF.
SQ5	num	Concentrates better		Collected at CRF.
SQ5C	char	Concentrates better(Char)		Collected at CRF.
SQ6	num	Completed tasks on time		Collected at CRF.
SQ6C	char	Completed tasks on time(Char)		Collected at CRF.
SQ7	num	Waits turn/does not interrupt		Collected at CRF.
SQ7C	char	Waits turn/does not interrupt(Char)		Collected at CRF.
SQ8	num	Less Forgetful		Collected at CRF.
SQ8C	char	Less Forgetful(Char)		Collected at CRF.
SQ9	num	Interacts appropriately		Collected at CRF.
SQ9C	char	Interacts appropriately(Char)		Collected at CRF.
SQ10	num	How well is today going		Collected at CRF.
SQ10C	char	How well is today going(Char)		Collected at CRF.
SQ11	num	Now compared to previous meds		Collected at CRF.

Variable	Type	Label	Codes	Comments
SQ11C	char	Now compared to previous meds(Char)		Collected at CRF.
SQ12	num	Overall satisfaction with med		Collected at CRF.
SQ12C	char	Overall satisfaction with med(Char)		Collected at CRF.
CONTSM	Char	Continue child on study med?		Collected at CRF.
WKDAY	Char	Day of Week		Collected at CRF.
SATDY	num	Relative Satisfaction Day		If SATDT and VISDT not missing then perform below logic to calculate SATDY, If SATDT less than VISDT then (SATDT - VISDT).Else if SATDT is greater than equal to VISDT then (SATDT- VISDT) +1.

1.4.9. Visits Data – VISITS

Dataset	VISITS
Creating program	visits.sas
Description	Visits Data
Unique identifier	DINVNO,DPATNO,VISNO
Sorted by	DINVNO,DPATNO,VISNO
Notes	Below listed variables will be dropped from dataset to protect PII as per HIPAA and EMA guidelines or due to missing values: PTINIT,VISMM,VISDD,VISYY,VISDT,DCRSN1,DCRSN2,DCRSN3,DCRSN4,DCRSN5,PHNVISMM,PHNVISDD,PHNVISYY,PHNVISDT,NXTVISMM,NXTVISDD,NXTVISYY,NXTVISDT,LDOSEMM,LDOSEDD,LDOSEYY,LDOSEDT,SIGNMM,SIGNDD,SIGNYY,SIGNDT

Variable	Type	Label	Codes	Comments
DINVNO	num	Investigator ID assigned for De-identity		Randomly assigned Investigator ID for De-identity
DPATNO	num	Patient ID assigned for De-identity		Randomly assigned Patient ID for De-identity
VISNO	char	Visit Number		Collected at CRF.
TRTGRP	char	Study Drug Treatment Assignment		Collected at CRF.
CMYN	char	Concomitant Meds (y/n)		Collected at CRF.
AEYN	char	Adverse Event (y/n)		Collected at CRF.

Variable	Type	Label	Codes	Comments
STDRUG	char	Study Drug Prescribed		Collected at CRF.
STDOSE	char	Study Drug Dose Prescribed		Collected at CRF.
COMPLYN	Char	Compliant with dosing regimen?		Collected at CRF.
TITRYN	Char	Dose titrated to next level?		Collected at CRF.
CONTYN	Char	Continue child in study?		Collected at CRF.
DCCODE	num	Discontinuation Reason Code		Collected at CRF.
DIARYN	Char	Was study diary collected?		Collected at CRF.
SIGNYN	Char	Visit Page Signed (y/n)		Collected at CRF.
HEIGHT	num	Height		Collected at CRF.
WEIGHT	char	Weight		Group element to protect PII.
SYSBP	num	Systolic Blood Pressure		Collected at CRF.
DIABP	num	Diastolic Blood Pressure		Collected at CRF.
HR	num	Heart Rate		Collected at CRF.
CGI	num	CGI: Severity of Illness		Collected at CRF.
CGIC	char	CGI: Severity of Illness(Char)		Collected at CRF.
VISDY	num	Relative Visit Day		If VISDT and VISDT not missing then perform below logic to calculate VISDY, If VISDT less than VISDT then (VISDT - VISDT).Else if VISDT is greater than equal to VISDT then (VISDT- VISDT) +1.

Variable	Type	Label	Codes	Comments
PHNVISDY	num	Relative Phone Visit Day		If PHNVISDT and VISDT not missing then perform below logic to calculate PHNVISDY, If PHNVISDT less than VISDT then (PHNVISDT - VISDT).Else if PHNVISDT is greater than equal to VISDT then (PHNVISDT- VISDT) +1.
NXTVISDY	num	Relative Next Visit Day		If NXTVISDT and VISDT not missing then perform below logic to calculate NXTVISDY, If NXTVISDT less than VISDT then (NXTVISDT - VISDT).Else if NXTVISDT is greater than equal to VISDT then (NXTVISDT- VISDT) +1.
LDOSEDY	num	Relative Last Dose Day		If LDOSEDT and VISDT not missing then perform below logic to calculate LDOSEDY, If LDOSEDT less than VISDT then (LDOSEDT - VISDT).Else if LDOSEDT is greater than equal to VISDT then (LDOSEDT- VISDT) +1.
SIGNDY	num	Relative Signature Day		If SIGNDT and VISDT not missing then perform below logic to calculate SIGNDY, If SIGNDT less than VISDT then (SIGNDT - VISDT).Else if SIGNDT is greater than equal to VISDT then (SIGNDT- VISDT) +1.