

# Some Sightings While Traveling the Road of ADaM Spec Review

PharmaSUG CDISC SDE - Chicago

2016-11-01

Nate Freimark

The Griesser Group

# Subtitle(s)

- Really horrible things folks have done in ADaM implementation
- When good people do bad things
- How to stretch the standards to the breaking point

# Disclaimer

- These are the opinions of the presenter only
- Comments generally fall into three buckets:
  - That is so wrong
  - I would not things that way and would suggest you do the same
  - Hmm – interesting – I can not say it is wrong/have to ponder that a bit

# Spec TOC

Dataset	Description	Structure	Class	Purpose	Keys	Documentation	IsRepeating	IsReferenceData
<a href="#">ADSL</a>	Subject-Level Analysis Dataset	One record per subject	SUBJECT LEVEL ANALYSIS DATASET	Analysis	USUBJID	See SAS program adsl.sas	No	No
<a href="#">ADAE</a>	Adverse Events Analysis Dataset	One record per record in SDTM AE domain	ADAM OTHER	Analysis	USUBJID, AESEQ	See SAS program adae.sas	Yes	No
<a href="#">ADEG</a>	ECG Test Results Analysis Dataset	One record per subject per parameter per analysis visit	BASIC DATA STRUCTURE	Analysis	USUBJID, PARAM, AVISIT, ATPT	See SAS program adeg.sas	Yes	No
<a href="#">ADLB</a>	Laboratory Tests Analysis Dataset	One record per subject per parameter per analysis visit	BASIC DATA STRUCTURE	Analysis	USUBJID, PARAM, AVISIT, ATPT	See SAS program adlb.sas	Yes	No
<a href="#">ADVS</a>	Vital Signs Analysis Dataset	One record per subject per parameter per analysis visit	BASIC DATA STRUCTURE	Analysis	USUBJID, PARAM, AVISIT, ATPT	See SAS program advs.sas	Yes	No

# Keys vs Structure

- Keys - --SEQ does not really add value as a key
- Can you give an example a variable be structure key instead of sort key? USUBJID, AEBODSYS, AEDECOD, AETERM ASTDT, AENDT vs USUBJID, AESEQ
- Is the structure or keys always be unique? In SDTM the word is “should”. In ADaM it is definitely not always possible although ADaM model document does say “ideally uniquely identifies and indexes each record in the dataset”

# Specs vs define

- Spec = What to do
  - Can be used to populate define
  - Some use to create metadata (empty datasets) for ADaM
- Define = what was done
  
- Source/derivation for spec may be programming code
- Source/derivation for define should be primarily English text and possibly SAS code also
- Consider two columns in spec – one with SAS code and one with plain English text

# Sample Spec Header

ADCR Variable Metadata Specifications	
Protocol:	ABC-DEF-01-01
SAP:	0
ADaM Version:	2.1
ADaM IG Version:	1.0
Dataset Description:	Clinical Response Analysis Dataset
Dataset Location:	adcr.xpt
Dataset Structure:	One Record Per Subject Per Clinical Response
Key Variables of Dataset:	STUDYID, USUBJID, PARAMCD, AVISIT
Class of Dataset:	BDS
Documentation:	adcr.sas

# Simple Spec

Dataset Name	Variable Name	Variable Label	Variable Type	Length	Codelist / Controlled Terms	Source / Derivation
ADCR	STUDYID	Study Identifier	Char	\$20		XR.STUDYID
ADCR	USUBJID	Unique Subject Identifier	Char	\$25		XR.USUBJID
ADCR	TRTP	Planned Treatment	Char	\$200		ADSL.TRT01P



# Define-supporting ADSL spec

Parameter Identifier	Variable Name	Variable Label	Variable Type	Origin	Length	Display Format	Codelist / Controlled Terms	Source / Derivation	Additional SAS logic
*ALL*	STUDYID	Study Identifier	text	Predecessor	\$14			DM.STUDYID	
*ALL*	USUBJID	Unique Subject Identifier	text	Predecessor	\$22			DM.USUBJID	
*ALL*	SUBJID	Subject Identifier for the Study	text	Predecessor	\$3			DM.SUBJID	
*ALL*	SITEID	Study Site Identifier	text	Predecessor	\$3			DM.SITEID	
*ALL*	AGE	Age	integer	Predecessor	8			DM.AGE	
*ALL*	AGEU	Age Units	text	Predecessor	\$5		<a href="#">AGEU</a>	DM.AGEU	

# What Should Go Into Simple Spec?

- Dataset Name - yes
- Variable Name - yes
- Variable Label - yes
- Variable Type - yes
  - SAS type or define type?
- Length – no? depends on data and would have to be updated
- Display Format – no
- Codelist / Controlled Terms - yes
- Source / Derivation – key piece of metadata

# Spec Issue – how to support PVLM

- Simple spec
  - Parameter level metadata tab
  - Tab for each dataset
- Define supporting spec
  - First column can define PARAMID

# General

- When date/time imputation are needed it is a good idea to keep the SDTM variable that contains the partial value
- date variables should have associated study day variables
- ADSL - Is -BLFL=Y always the value where ABLFL=Y?
- ADQS\_xxx- Is this data summarized? If yes then are all records summarized? Or is ANL01FL needed?
- ADVS - PARAM – uses VSSTRESU value – I would suggest using case that would go into summary/title
- ADPR - xxxxFL – SAS code (only) is not recommended as a source/derivations (if text is supposed to support define.xml)
- ADEG - I personally recommend creating these average values in SDTM and not ADaM – depends who the intended audience is
- SITEID/SUBJID in every non-ADSL dataset?
- Creating the DT, TM, and DTM triplicate just because it can be done?
- From the ADaMIG - For a given SDTM DTC variable, if only hours and minutes are ever collected, and seconds are imputed in \*DTM as 00, then it is not necessary to set \*TMF to “S”. However if seconds are generally collected but are missing in a given value of the DTC variable and imputed as 00, or if a collected value of seconds is changed in the creation of \*DTM, then the difference is significant and should be qualified in \*TMF.

# Visit/Time Point Variables #1

- If AVISIT is always the same as VISIT in the SDTM in a study, do we need to keep AVISIT in the adam or can we just keep visit without avisit? If ATPT is always the same as xxTPT in the SDTM in a study, do we need to keep xxTPT in the adam or can we just keep visit without ATPT?
- AVISIT/AVISITN/ATPT/ATPN should be the timing variables that are in datasets – VISIT/VISITNUM/-TPT/--TPTNUM should also be kept if the values/algorithms differ
- AVISITN/AVISIT should be consistent across ADaM datasets – not always feasible
- ADVS\_ATPTN – these appear to be VISIT values-----[response]: It is based SDTM.VSTPTNUM/VSTPT [response back] bad SDTM should not necessarily translate to bad ADaM – you are allowed to fix things
- ADEX\_TPTN – why are these values different than ADVS? -----[response]: Please advise whether the TPT/TPTNUM should be consistent in a study. [response back] I don't see why they would not be – AVISIT/AVISITN and ATPT/ATPTN should be based on a common codelist – truthfully I would suggest the same for SDTM – as long as it can practically be done
- Please advise whether the ATPT/ATPTN should be consistent in a study. I think it is a really good idea although it only has to be consistent (one to one) within a PARAM

## Visit/Time Point Variables #2

- ADLB - AVISIT – last value within a window is not an ENDPOINT value unless you are outputting an extra record – are you (I assume not since I do not see DTYPE present)? If not then ANLzzFL should be used
- ADLB - AVISIT – SCREENING/RESCREEN/UNSCHEDULED are not valid AVISIT values unless they are summarized on the table or defined in the SAP – if not they should be null (or windowed)
- ADxxx - VISIT/VISITNUM have to come from SDTM and cannot be derived
- ADxxx - Why are the time points captured in PARAM and not in AVISIT?
- ADQS - AVSIT/AVISITN all null - drop

# Traceability/Metadata

- --SEQ from source SDTM should be included for traceability at a minimum if record comes directly from SDTM
- When I review specs, the biggest issue I find is lack of traceability – I can't tell how a variable or a row was derived.
- Always including, in the same dataset, both --SEQ and the SRC\* triplets is not necessary
  - ADaMIG states "Table 3.3.9.1 defines additional variables useful in certain situations to facilitate datapoint traceability."
- ADEFF - No traceability variables present (QSSEQ or SRCDOM/SRCSEQ)? It appears there are many records taken directly from QS
- ADxxxx - ADEFF does not exist yet is referenced in derivation
- ADEG - PCHG – typo and should reference AVAL instead of EG
- ADEX - AVAL – don't you use parameter-level metadata instead of stuffing everything into one comment cell?
- ADxx.TRTP references ADSL.TRTP which does not exist
- ADxx.PARAMCD/PARAM only have one value and none are in the spec
- ADxx.AVALCAT1/AVALCAT2 should not reference XRSTRESC - it should reference AVALC
- ADTTE – how is "date of last contact" determined? Shouldn't it be added as a variable in ADSL?
- Intermediate datasets must be submitted if they are referenced in the metadata

# ADSL/BDS Population Flags

- ITTRFL - what is the purpose of this if it a copy of ITTFL? there is nothing "record level" about this - valid values are Y and null and not Y and N
- all BDS (at a minimum) have to have at least one population flag variable
- RANDFL – why is it tied to being dosed?
- COMPLFL -is this truly a population?
- DISCFL – this appears to be the inverse of COMPLFL – what is the purpose?
- CMPTRTFL and DSCTRTFL seem to be opposites what is the point of having both?
- NOMEDFL – why is this not covered by RANDFL and SAFFL – what is the purpose of a combination flag?
- ITTFL – same as RANDFL and not defined in protocol – why is it added?
- Some population flags in the efficacy analysis are too complicated. Can we generate these population flags by referring to each other? E.g. for the Microbiologic analysis, can we use adsl to generate admb, then use admb to generate population flag in the adsl? [response] circular logic is always a concern
- ADEG - SAFFRL – 15 records excluded –not clear why – and how can records be excluded from SAF and not PPROT? ADRG does not describe how records were excluded



# ADSL Treatment Variables

- ARM/ARMCD should not be included in every dataset – ADaM variables should be used (TRTSEQP)
- Check labels vs standard (TRTP for example does not match the standard)
- Some variables are populated with A/B/C/D and some TREATMENT A/TREATMENT B/ etc – they should be consistent
- ARMN- what is the purpose? ARM/ARMCD are brought over from SDTM for traceability and not used in analysis. ADSL.ARMN - what is the purpose of this especially since it is the exact same value as TRT01PN
- There is no variable ARMGR1 in ADaM – ARM is not an analysis variable and therefore there should be no “grouping” of ARM.
- TRT01A – do you check whether Actual differs from Planned or do you always create?
- Check TRT01P vs TRT02P – the TR01SDT/TR01EDT vs TR02SDT/TR02EDT are out of order
  - TRTSEQP – the order is reversed
- What’s the purpose of keeping ARM in ADaM. [response] it is required per the ADaMIG for traceability. Also, can ACTARM be kept as well? [response] it can always be kept – it is only perm based on SDTM v3.1.2
- Creating TRTxxP to match ARM and then creating TRTxxSP is not the intent of the ADaMIG

# ADSL Date Variables

- ADSL.TRTSTD~~TM~~ is referenced multiple times in the define – the variable is TRTSD~~TM~~
- APxxSDT/APxxEDT are paired variables – both or neither should be present
- RANDTIME – this should be RAND~~TM~~ (what is purpose in any case?)
- DMDT -purpose?
- DMDY – label changed from SD~~TM~~ – that is not allowed
- Treatment timing variables are not correct – TRTSTD~~T~~ s/b TRTSD~~T~~, etc.
- ADSL.TRTSD~~T~~/TRTED~~T~~ – add
- LSTVST – last based on what? SV has no “order” unless VISITNUM is the correct order
- TRTSD~~T~~/TRTED~~T~~ based on RF or RFX?

# ADSL #1

- SEX per SDTM is M/F – you cannot change any SDTM variable value in ADaM
- DEATHFL – why not use SDTM variable?
- HEIGHT should be HEIGHTBL or BLHEIGHT to indicate it is a baseline value (similar to label)
- WEIGHT – similar comment to HEIGHT
- Baseline variables should start or end with BL and not just B
- CBPYN – why is this changed from SCTESTCD and if changed it should end with FL
- AGEU is missing

## ADSL #2

- AEXDURN – I don't see why this is prefixed with A/Analysis – EXDURN is sufficient
- xxxDUR – DUR in SDTM is ISO8601 – I would recommend xxxDURC
- xxxxCATy – why are CATy needed instead of individual flags for each component?
- ADSL.DISIVFL/DISLFUFL - what is the purpose of this? isn't it negative of COMIVFL/COMLFUFL?

## ADSL #3

- ADSL – is there a reason you format dates as DATE9 instead of YYYYMMDD10?
- ADSL – do you have any subjects with missing disposition records? it should be null in that case and not “discontinued”

# BDS Treatment/Period Variables

- ADPC/ADPP - TRTP=ARM is incorrect
- ADPC/ADPP - TRTA=TRTSEQA is incorrect
- ADVS - APERIOD=88 is not allowed if it is not one of the treatment periods – this should presumably have been identified with APHASE instead of a “post period”
- Can we keep TRT01P, TRT02P, ARM, TRTP all in the adlb/adae?  
Because I notice some training classes say it is better to keep all of these variables. [response] you can but why would you want to?
- ARM – purpose? Why is it kept in all ADaM datasets?

# BDS – Incorrect Variables/Values #1

- ADEG - AVALU is not a valid ADaM variable – if unit is relevant it should be included in PARAM if not then just keep ORRESU/STRESU
- ADLB - LBTEST/LBTESTCD generally are not sufficient for PARAM/PARAMCD in labs since there are the same – TEST in CHEMISTRY and URINALYSIS
- ADHxx - AVALC is not left-justified for numeric values (nor does AVALC have to be populated for numeric parameters)
- Is it your standard or the sponsor's standard to capture LBCAT in PARAM instead of only the ones that are common in more than one LBCAT? -----[response]: I would like to change all of these PARAMCD/PARAM by adding LBCAT to PARAM/PARAMCD when LBCAT =" URINALYSIS" and not change them when PARAMCD ^=" URINALYSIS". Please advise whether we only need to update these param/paramcd common in more than one LBCAT or all of these PARAMCD/PARAM when LBCAT =" URINALYSIS" [response back]"need" is a strong word – either way is "legal" however I personally do not see the value added in capturing on all PARAMCD – I personally only add U to the ones that are in more than one LBCAT
- ADLB - A lot of the PARAM appear to be CRIT values and should not be added parameters (e.g ALKGR15) - for the ones that look at multiple PARAM it makes sense but not the ones that are within PARAM
- ADDV - AVALC – are these values actually analyzed/summarized? I do not see a protocol deviation summary – why was this ADaM created?

## BDS – Incorrect Variables/Values #2

- ADPE - Summary is by AVISIT – why are ABLFL/BASEC needed? Drop BASE
- ADTTE - ANLzzFL are not appropriate for ADTTE – use POOLyFL variables from ADSL
- ADEG - BASE – spec seems to indicate that this value is calculated in ADEG and not an added AVERAGE row (indicated by the absence of DTYPE) – this does not seem clear or compliant
- ADLB - PARAM/PARAMCD have to uniquely identify - using PARCAT1 is not allowed – GLUC cannot have two different PARCAT1 values
- ADLB - CRIT1 – usage is incorrect – CRIT1 to be a constant value within PARAM – this should be using AVALCAT1
- ADLB - LBLECFL/LBLEOTFL/LBLD10FL/LBLPTEFL/LBMINFL/LBMAXFL/POSTHIFL/POSTLTFL (and all the rest of the flags) – these should all be ANLzzFL (or AVISIT added rows) – see section 4.5.3.1 of the ADaMIG
- ADxxxx.CRIT2 - this appears to be a categorization of CRIT1 and not a "pre-specified criterion"
- ADxx.AVALCAT1/AVALCAT2 - these are not categorizations of AVAL



## BDS – BASE/BASETYPE

- ADLB - Basetype is not populated for all records within a PARAM
- ADLB/ADV5 - Having BASE and not CHG seems odd – what is the purpose?

## BDS – Related Variables

- ADLB - BNRIND has a value=\* which does not appear in ANRIND
- ADxx - AVALCAT1/AVALCAT2 – I do not see how these are categorizations of AVAL or AVALC
- ADLB - AVALCATy - most of these are not a categorization of AVAL or AVALC – these should be CRITy
- AVALC cannot be based on AVALCATy – it has to be the opposite
- ADEG - BASE should reference ABLFL and not AVISITN
- ADVS - BASE – derivation is incorrect – it has to be tied to ABLFL=Y record

# BDS Timing Variables

- ADPC - ATPTREF=PRE-DOSE does not make sense – the value time-point(TPT) is “pre-dose” to the dose (TPTREF)
- ADEG - ENDPOINT BASELINE is not a correct AVISIT value – this should be AVISIT=BASELINE with DTYPE=AVERAGE
- ADEG - Why are AVISIT=ENDPOINT xxxx in any case – it does not have that on the tables – does it?
- ADEG -What does ATPT=ENDPOINT 0.5 H POST-DOSE mean? If these are supposed to identify the average value then DTYPE should be used and not a changed ATPT value – it is not a different timepoint
- ADxx - TSTDTC – what is this variable? Keep MBDTC/MSDTC (define references MB.TESTDT which is not a valid SDTM variable)
- ADTC – Why? Why? Why?

# BDS Flag Variables

- Can we add free text to the label of ANLzzFL or any variable with y, zz, xx? e.g: added text “,Analysis Flag zz – Last for Dup. Records”.  
[response] you can for y and zz but not for xx
- ADEX - xxxxxFL – values of 1-4 are not allowed for an FL variable
- ADEG - AOCC – these are ADAE flags – BDS should use ANLzzFL variables
  - ANLzzFL is used in OCCDS however AOCC variables do not appear in the ADaMIG
- ADLB - no ANLzzFL variables are needed for analysis? (this is possible if there are no multiple nominal visit values and unscheduled values are not summarized)

## BDS - DTYPE

- ADEFF - DTYPE=combined is not correct
- ADQS\_xxx - DTYPE=SUM is incorrect – DTYPE is for records within a parameter and not totally derived parameters
- ADVS.DTYPE - this is incorrect usage of DTYPE - there is no indication that there is an added row with AVISIT="Maximum" - this should be identified with an ANLzzFL
- ADLB – DTYPE=EOT is not correct – it does not describe the derivation method (and whatever value you use should be used in ADEG also)
- ADLB – MCRIT variables are not in v1.0 (although you can use them as you have them in the spec)

# OCCDS

- ADAE - Creating records with ANYAE=N for APERIOD with no AEs is not allowed
- ADAE - Populating AETERM with None for ANYAE=N records is not allowed
- OCCDS does allow for repeats of Events records in certain situations
- ADAE - AEDUR is an SDTM variable that is character and IS8601 – it cannot be changed to numeric in ADaM – it is defined as character in the ADAE appendix – this should be ADURN
- ADAE - AEOUTR – why is this changed from SDTM.AE.AEOUT?
- ADAE - AEETM/AEEDTM – time is not collected – drop variables
- ADCM - CMDOSU/CMDOSFRQ/CMROUTE changed from SDTM values – this is not allowed
- ADAE - ASTDTM/AENDTM –if imputations are done then imputation flags are needed
- ADAE - AESEV – you cannot change values from SDTM – this includes changing case and label – is this supposed to be ASEV?
- ADAE.TRTEMFL - derivation may not be correct if time is imputed
- ADAE.ASTDT should be included even is ATSTDM is present
- ADMH – ASTDT/AENDT – do these add value? These appear to be mostly partial dates

# Questions?

- [nfreimark@thegriessergroup.com](mailto:nfreimark@thegriessergroup.com)
  - Nate Freimark
  - Vice President - Clinical Programming and Data Standards
  - The Griesser Group
  - West Conshohocken, PA 19428