



ADaM Categorization: Groups, Categories, and Criteria. Which Way Should I Go?

Jack Shostak



Duke Clinical Research Institute



Agenda

- Review categorization needs
- Review the various ADaM categorization variables and methods
- Look at a few examples
- Examine method pros and cons
- Provide author recommendations



Disclaimer

The opinions expressed in this presentation are solely the fault of the author and his imagination.

Statements presented here as factual should be found in the CDISC ADaM Implementation Guide.



What does it mean to categorize?

Simple definition of categorize from Merriam-Webster:

to put (someone or something) into a group of similar people or things



Why categorize in ADaM?

- For categorical data analysis
- For model covariates
- For subpopulation determination
- For record selection for an analysis
- For simple presentation ordering purposes



Scope of talk

The focus of the talk is primarily on categorization of ADaM ADSL and BDS values

- Will ignore BDS SHIFTy variables used for shift tables.



Scope of talk

The focus of the talk is primarily on categorization of ADaM ADSL and BDS values

- Will ignore OCCDS
 - Will ignore Standardized MedDRA Query Variables SMQ*. This is a special case of OCCDS AE categorization.
 - Will ignore the OCCDS special ACATy variable
“Category used in analysis. May be derived from --CAT and/or --SCAT. Examples include records of special interest like prohibited medications, concomitant medications taken during an infusion reaction, growth factors, antimicrobial medications ...”



ADaM categorization variables to explore

- PARCATy parameter categorization
- *GRy grouping variables
- *CATy analysis variable categorization variables
- (M)CRITy criteria record selection variables
- Custom user defined BDS variables



PARCATy parameter categorization

PARAM to PARCATy is a many-to-one mapping; any given PARAM may be associated with at most one level of PARCATy.

This is fine.....

PARAM	PARCAT1
Secondary One	Secondary Endpoints
Secondary Two	

This is not.....

PARAM	PARCAT1
Secondary One	Subtype 1
	Subtype 2



*GRy and *GRyN variables

From ADaM Implementation Guide section 3.1.1
General Variable Conventions:

Rule #9 states

Variables whose names end in GRy, Gy, or CATy are grouping variables, where y refers to the grouping scheme or algorithm.

Within this document, **CATy** is the suffix used for **categorization of ADaM-specified analysis variables** (e.g., CHGCATy categorizes CHG).



*GRy and *GRyN variables

From ADaM Implementation Guide section 3.1.1
General Variable Conventions:

Rule #10 states

It is recommended that producer-defined grouping or categorization variables begin with the name of the variable being grouped and end in GRy (e.g., variable ABCGRy is a character description of a grouping or categorization of the values from the ABC variable for analysis purposes). **If any grouping of values from an SDTM variable is done**, the name of the derived ADaM character grouping variable should begin with the SDTM variable name and end in GRy.



*GRy and *GRyN variables

ADaM Implementation Guide defined ADaM *GRy variables:

- SITEGRy
- RACEGRy
- AGEGRy
- DTHCGRy (based on ADaM DTHCAUS variable)



*GRy and *GRyN example

Using *GRy and *GRyN to group AGE

USUBJID	AGE	AGEGR1	AGEGR1N
101	20	18 – 65	1
102	65	>= 65	2
103	42	18 – 65	1
104	18	18 – 65	1



*GRy and *GRyN variables

- *GRy variables are often used to group SDTM content, but they can be used for non-AVAL based ADaM variables as well.
- *GRy variables are inherently self-descriptive by nature.



*CATy variables

These *CATy variables include BDS:

- AVALCATy
- BASECATy
- CHGCATy
- PCHGCATy

These categorize AVAL, AVALC, BASE, CHG, and PCHG ADaM variables respectively, and are generally used to categorize the AVAL/BASE/CHG/PCHG continuous analysis values



*CATy variables

Extrapolated definition from the ADaM Implementation Guide for *CATy variables:

- A categorization of the variable (e.g., AVAL/AVALC) **within a parameter.**
- **Intended to be a many to one mapping, not a one to many as in subcategorization of an AVAL value.**



AVALCATy example

Categorizing AVALC:

AVALC	AVALCAT1
None	None or Mild
Mild	

AVALC	AVALCAT1
Moderate	Moderate or Severe
Severe	

USUBJID	PARAM	AVALC	AVALCAT1
101	Pain Severity	None	None or Mild
102	Pain Severity	Severe	Moderate or Severe
103	Pain Severity	Moderate	Moderate or Severe
104	Pain Severity	Mild	None or Mild



(M)CRITy and associated flag variables

The (M)CRITy variable set contains:

- A text string identifying a pre-specified criterion within a parameter (CRITy or MCRITy) and...
- For CRITy, its associated boolean flag CRITyFL or...
- For MCRITy, its associated multichotomous result in MCRITyML

The original intent behind (M)CRITy was to select subgroups of subjects that met a given criteria



(M)CRITy flag variables

CRITyFL and MCRITyML are defined in Implementation Guide table 3.3.4.2.

Character flag variable indicating whether the criterion defined in (M)CRITy **was met by the data on the record.**



(M)CRITy variables row dependence

Also from section 4.7 in the Implementation Guide:

- “The definition of **CRITy** can use any variable(s) located on the row, and the definition must stay constant across all rows within the same value of PARAM. **A complex criterion which draws from multiple rows (different parameters or multiple rows for a single parameter) will require a new PARAM be created.**”
 - “CRITy for one parameter can be different than CRITy for a different parameter in the same dataset.”
- “MCRITy is populated with a text description identifying the criterion being evaluated. The definition of **MCRITy** can use any variable(s) located on the row and the definition must stay constant across all rows within the same value of PARAM. **A complex criterion which draws from multiple rows will require a new PARAM be created.**”



CRITy example

Applying CRITy to systolic blood pressure

USUBJID	PARAM	AVAL	CRIT1	CRIT1FL
101	Systolic Blood Pressure (mm Hg)	163	SBP > 160	Y
102	Systolic Blood Pressure (mm Hg)	133	SBP > 160	N
103	Systolic Blood Pressure (mm Hg)	120	SBP > 160	N
104	Systolic Blood Pressure (mm Hg)	165	SBP > 160	Y
105	Systolic Blood Pressure (mm Hg)	140	SBP > 160	N



MCRITy example

Applying MCRITy to systolic blood pressure

USUBJID	PARAM	AVAL	MCRIT1	MCRIT1ML
101	Systolic Blood Pressure (mm Hg)	163	SBP Classification	SBP \geq 160
102	Systolic Blood Pressure (mm Hg)	133	SBP Classification	120 \geq SBP \geq 139
103	Systolic Blood Pressure (mm Hg)	120	SBP Classification	120 \geq SBP \geq 139
104	Systolic Blood Pressure (mm Hg)	165	SBP Classification	SBP \geq 160
105	Systolic Blood Pressure (mm Hg)	140	SBP Classification	140 \geq SBP \geq 159



(M)CRITy variable summary

- (M)CRITy is nice in that it codifies the criteria into the dataset as a data element. It essentially places the definition of the flag variable CRITyFL/MCRITyML into the dataset itself.
- You cannot create CRITyFL/MCRITyML results based on information across multiple BDS rows. In that case, you likely need to create a new PARAM.



Case Study: Clinical Response

- Nootropic drug study and the BDS AVAL contains the cognitive score response value.
- Goal is to create a BDS clinical response variable containing “Not effective”, “Effective”, or “Very effective” *which is dependent on the subject’s AGE.*

AGE 18-50

AVAL	RESULT
<15	Not Effective
15-30	Effective
>30	Very Effective

AGE > 50

AVAL	RESULT
<10	Not Effective
10-20	Effective
>20	Very Effective



Case Study: Clinical Response

Raw BDS data of the cognition scores

USUBJID	AVISIT	PARAM	AVAL	AGE
101	Month 1	Cognition	15	20
101	Month 2	Cognition	25	20
101	Month 3	Cognition	29	20
102	Month 1	Cognition	15	65
102	Month 2	Cognition	25	65
102	Month 3	Cognition	26	65



Case Study: Clinical Response

Can I use AVALCATy ?

- Per the IG, “A categorization of AVAL or AVALC within a parameter. ”
- Since there is a dependency on AGE, AVALCATy may not be the best approach. The IG text doesn't preclude AVALCATy having a dependency on something other than AVAL, but it is implied by the text and the variable name itself.

USUBJID	AVISIT	PARAM	AVAL	AGE
101	Month 1	Cognition	15	20
101	Month 2	Cognition	25	20
101	Month 3	Cognition	29	20
102	Month 1	Cognition	15	65
102	Month 2	Cognition	25	65
102	Month 3	Cognition	26	65



Case Study: Clinical Response

Can I use (M)CRITy?

- Yes because all needed data is on the row.
- Would need to use MCRITy due to multi-level response.
- Would also need an MCRITy for each age group

So.....

USUBJID	AVISIT	PARAM	AVAL	AGE
101	Month 1	Cognition	15	20
101	Month 2	Cognition	25	20
101	Month 3	Cognition	29	20
102	Month 1	Cognition	15	65
102	Month 2	Cognition	25	65
102	Month 3	Cognition	26	65



Case Study: Clinical Response

Using MCRITy (noting that this structure might make table production difficult)

USUBJID	AVISIT	PARAM	AVAL	AGE	MCRIT1	MCRIT1ML	MCRIT2	MCRIT2ML
101	Month 1	Cognition	15	20	Clinical Response (Age 18-50)	Effective	Clinical Response (Age over 50)	
101	Month 2	Cognition	25	20	Clinical Response (Age 18-50)	Effective	Clinical Response (Age over 50)	
101	Month 3	Cognition	29	20	Clinical Response (Age 18-50)	Effective	Clinical Response (Age over 50)	
102	Month 1	Cognition	15	65	Clinical Response (Age 18-50)		Clinical Response (Age over 50)	Effective
102	Month 2	Cognition	25	65	Clinical Response (Age 18-50)		Clinical Response (Age over 50)	Very Effective
102	Month 3	Cognition	26	65	Clinical Response (Age 18-50)		Clinical Response (Age over 50)	Very Effective



Case Study: Clinical Response

Can I use PARAM?

- Absolutely, as you can always create a new PARAM.

USUBJID	AVISIT	PARAM	AVAL	AGE
101	Month 1	Cognition	15	20
101	Month 2	Cognition	25	20
101	Month 3	Cognition	29	20
102	Month 1	Cognition	15	65
102	Month 2	Cognition	25	65
102	Month 3	Cognition	26	65



Case Study: Clinical Response

Creating a new PARAM

USUBJID	AVISIT	PARAM	AVAL	AVALC	AGE
101	Month 1	Cognition	15		20
101	Month 1	Clinical Response		Effective	20
101	Month 2	Cognition	25		20
101	Month 2	Clinical Response		Effective	20
101	Month 3	Cognition	29		20
101	Month 3	Clinical Response		Effective	20



Case Study: Clinical Response

Creating a new PARAM actually works pretty well to produce a table like this:

Parameter	Treatment A (n=xxx)	Treatment B (n=xxx)	p-value
Cognition			xxxx.x
N	xxx	xxx	
Mean	xxx.x	xxx.x	
Std	xxx.xx	xxx.xx	
Min-Max	xxx-xxx	xxx-xxx	
Clinical Response			xxxx.x
Not Effective	xxx (xxx.x%)	xxx (xxx.x%)	
Effective	xxx (xxx.x%)	xxx (xxx.x%)	
Very Effective	xxx (xxx.x%)	xxx (xxx.x%)	



Case Study: Clinical Response

Hey, if I can do this

USUBJID	AVISIT	PARAM	AVAL	AVALC	AGE
101	Month 1	Cognition	15		20
101	Month 1	Clinical Response		Effective	20
101	Month 2	Cognition	25		20
101	Month 2	Clinical Response		Effective	20
101	Month 3	Cognition	29		20
101	Month 3	Clinical Response		Effective	20

Why can't I just collapse and make AVALC then like this?

USUBJID	AVISIT	PARAM	AVAL	AVALC	AGE
101	Month 1	Cognition	15	Effective	20
101	Month 2	Cognition	25	Effective	20
101	Month 3	Cognition	29	Effective	20



Case Study: Clinical Response

Because AVAL to AVALC isn't 1-1 within the PARAM

USUBJID	AVISIT	PARAM	AVAL	AVALC	AGE
101	Month 1	Cognition	15	Effective	20
101	Month 2	Cognition	25	Effective	20
101	Month 3	Cognition	29	Effective	20
102	Month 1	Cognition	15	Effective	65
102	Month 2	Cognition	25	Very Effective	65
102	Month 3	Cognition	26	Very Effective	65



Case Study: Clinical Response

Could I use ANLzzFL here?

- No, primarily because ANLzzFL is intended to be an additional record selection flag and not an analysis result.

USUBJID	AVISIT	PARAM	AVAL	AGE
101	Month 1	Cognition	15	20
101	Month 2	Cognition	25	20
101	Month 3	Cognition	29	20
102	Month 1	Cognition	15	65
102	Month 2	Cognition	25	65
102	Month 3	Cognition	26	65



Case Study: Clinical Response

Could I create a custom BDS variable such as CRESP here to indicate clinical response?

- Per ADaM IG section 4.2 it says “Rule 1: A parameter-invariant function of AVAL and BASE on the same row that does not involve a transform of BASE should be added as a new column.”
- So, probably not because of the dependency on AGE.

USUBJID	AVISIT	PARAM	AVAL	AGE
101	Month 1	Cognition	15	20
101	Month 2	Cognition	25	20
101	Month 3	Cognition	29	20
102	Month 1	Cognition	15	65
102	Month 2	Cognition	25	65
102	Month 3	Cognition	26	65



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create an ADSL patient level flag that identifies subjects with Systolic BP ≥ 160 and Diastolic BP ≥ 100 at baseline.

How can we do this with categorical variables in ADaM?



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP \geq 160 and Diastolic BP \geq 100 at baseline.

Can I just create a new flag variable in ADSL like this?

USUBJID	HBP2FL
101	Y
102	N
103	Y

Sure, but where is the traceability?

It is within the algorithm metadata for HBP2FL.

Is there another way?



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP ≥ 160 and Diastolic BP ≥ 100 at baseline.

Can I add two supportive binary ADSL flags to help?

USUBJID	HBP2FL	SYSBPFL	DIABPFL
101	Y	Y	Y
102	N	Y	N
103	Y	Y	Y

Now we have three flags in ADSL. We have the one desired flag plus the two composite flags.

For further transparency, you could also keep baseline systolic and diastolic BP values.



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP ≥ 160 and Diastolic BP ≥ 100 at baseline.

For further traceability, it might be better to show the classification derivation in a BDS dataset...

USUBJID	AVISIT	PARAM	AVAL
101	Baseline	Systolic Blood Pressure (mm Hg)	165
101	Baseline	Diastolic Blood Pressure (mm Hg)	100

So, how can I categorize those two records?

Use AVALCATy?

Use CRITy variables?

Create new BDS flag variables?



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP \geq 160 and Diastolic BP \geq 100 at baseline.

Using AVALCATy:

USUBJID	AVISIT	PARAM	AVAL	AVALCAT1
101	Baseline	Systolic Blood Pressure (mm Hg)	165	Systolic BP \geq 160
101	Baseline	Diastolic Blood Pressure (mm Hg)	100	Diastolic BP \geq 100



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP \geq 160 and Diastolic BP \geq 100 at baseline.

Using CRITy:

USUBJID	AVISIT	PARAM	AVAL	CRIT1	CRIT1FL
101	Baseline	Systolic Blood Pressure (mm Hg)	165	Systolic BP \geq 160	Y
101	Baseline	Diastolic Blood Pressure (mm Hg)	100	Diastolic BP \geq 100	Y



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP \geq 160 and Diastolic BP \geq 100 at baseline.

Can you create new BDS flag variables?

USUBJID	AVISIT	PARAM	AVAL	SYSFL	DIAFL
101	Baseline	Systolic Blood Pressure (mm Hg)	165	Y	
101	Baseline	Diastolic Blood Pressure (mm Hg)	100		Y

This would get past the Pinnacle validator, but it is a stretch as these new flags are PARAM dependent.



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP \geq 160 and Diastolic BP \geq 100 at baseline.

Assuming we used CRITy:

USUBJID	AVISIT	PARAM	AVAL	CRIT1	CRIT1FL
101	Baseline	Systolic Blood Pressure (mm Hg)	165	Systolic BP \geq 160	Y
101	Baseline	Diastolic Blood Pressure (mm Hg)	100	Diastolic BP \geq 100	Y

We now need that information combined, which is readily done with a new PARAM.



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP ≥ 160 and Diastolic BP ≥ 100 at baseline.

CRITy with a new PARAM:

USUBJID	AVISIT	PARAM	AVAL	AVALC	CRIT1	CRIT1FL
101	Baseline	Systolic Blood Pressure (mm Hg)	165		Systolic BP ≥ 160	Y
101	Baseline	Diastolic Blood Pressure (mm Hg)	100		Diastolic BP ≥ 100	Y
101	Baseline	Systolic Blood Pressure ≥ 160 and Diastolic Blood Pressure ≥ 100		Y		

This shows the categorical CRITy variables being used to populate a new PARAM.



Case Study: High Blood Pressure (Stage 2)

In this case, we want to create a patient level categorization that identifies subjects with Systolic BP ≥ 160 and Diastolic BP ≥ 100 at baseline.

Now, how this new BDS PARAM.....

USUBJID	AVISIT	PARAM	AVAL	AVALC	CRIT1	CRIT1FL
101	Baseline	Systolic Blood Pressure ≥ 160 and Diastolic Blood Pressure ≥ 100		Y		

Gets back into the ADSL equivalent like this:

USUBJID	HBP2FL
101	Y
102	N
103	Y

Is another conversation entirely



Summary thoughts for ADaM categorical variables



Things to do with ADaM categorical variables

- Keep ADaM as simple as you can
 - You want ADaM to be end user friendly
 - Allow for traceability, but remember usability
 - There are often multiple legal ways to do the same categorization
- Try to use CATy variables to categorize ADaM analysis value variables and GRy variables to group other variable content.
- If CATy or (M)CRITy doesn't work for you, then consider creating a new PARAM instead.
- For complex categorizations, consider using (M)CRITy with a new PARAM to combine the composite information.



Things to do with ADaM categorical variables

- Consider a new BDS variable for additional categorizations
 - Traceability can be limited to the derivation metadata.
 - You have to follow the rules for adding new BDS variables.
- A new PARAM is often a very clean solution and easy to “see” in a BDS dataset.



Things not to do with ADaM categorical variables

- Don't create new variables for categorization when predefined ADaM categorization variables such as SITEGRy or SAFFL exist.
- Don't use AVALC as a categorization of AVAL. That must be a 1-1 relationship.
- Don't cram analysis value concepts into ANLzzFL as that is meant as a special record selection flag. Some people do this to avoid Pinnacle 21 errors.



Things not to do with ADaM categorical variables

- Don't use *AVALCAT* to subcategorize *AVAL* in a one to many way. *AVALCAT* is meant to categorize many to one. If you need one to many, then:
 - If data on one row, you can use *(M)CRITy* for this
 - If data on one row and it is a parameter invariant function of *AVAL/BASE*, you can create a new custom BDS variable
 - Otherwise, create a new *PARAM*
- Don't create *(M)CRITy* variables in a way that they are defined based on multiple rows. *(M)CRITy* must be defined on the content found on the data row per the ADaM Implementation Guide.



ADaM Categorization: Groups, Categories, and Criteria. Which Way Should I Go?

- Often times the most simple solution is the best one.
- There may be more than one ADaM legal solution.
- Examine the reporting needs to pick the best ADaM variable solution. An analysis dataset structure that is similar to output structure is often the best.
- Study the ADaM implementation guide for detailed variable rules.



Questions?

jack.shostak@duke.edu