



PARTNERS IN RESEARCH



Handling Interim and Immature Data In a Clinical Trials Setting

Paul Stutzman
Axio Research

22-Oct-2015

Introduction

Outline

- Motivation
- Definition
- Preparation
- Potential pitfalls
- Reassess
- Tools and processes
- Concluding remarks

Motivation

Given:

- ❖ Variety of statistical projects
- ❖ Early involvement
- Maintain quality/accuracy
- Minimize cost/effort
- Timely deliverables

Definition

Context

- Preparatory programming - standards based
Raw → SDTM → ADaM
- Medical monitoring
- Interim analyses
- Data Monitoring Committee (DMC) / Data Safety Monitoring Board (DSMB)

Definition

What is interim and incomplete data?

- Interim:
 - ❖ Before database lock
 - ❖ Often iterative – multiple snapshots over time
 - ❖ Unclean
- Incomplete:
 - ❖ Missing records (subjects, visits, parameters...)
 - ❖ Missing values

Definition

Dirty data

Pick any 2:



Good: Accurate and Complete

Fast: Timely

Cheap: Cost & Effort

Axio

PARTNERS IN RESEARCH

Preparation

What can we know ahead of time?

- Protocol
- Annotated CRFs
- SAP
- DMC/DSMB Charter

- Look at the actual data!

Preparation

What can we know ahead of time?

- Visit structure
 - ❖ The protocol is your friend.
 - ❖ However...

“Patients randomized to both study arms will have 21-day chemotherapy cycles until disease progression, unacceptable toxicity, withdrawal of consent, or protocol specified parameters to stop treatment.”

Preparation

What can we know ahead of time?

- All permissible values

 - ❖ Annotated CRF

 - ❖ Data dictionary

- Standards constraints

Raw → Analysis datasets → TLFs

Raw → SDTM → ADaM → TLFs (metadata is critical)

Preparation

Use defensive programming/coding techniques

- Your program should not fail
- Handle all values (anticipated or not)
- Unanticipated values should be flagged
- Code for all variables – even if insufficient data

- Use the SAS log!

Potential pitfalls

Initial considerations

All missing

- In sufficient data, no records meet criteria yet
- For example: No deaths
 - ❖ In DS?
 - ❖ In AE?
 - ❖ Separate deaths dataset?
 - ❖ Will they need to be reconciled?
- Should be noted in the SAS log

Potential pitfalls

Initial considerations

Not all values are present. Unsure of:

- What actual values will be (categorical)
- The relationship between variables
 - ❖ xxCAT, xxSCAT
 - ❖ xxTEST, xxTESTCD
 - ❖ Numeric coded categorical variables
- Visits

When in doubt, note in the SAS log

Potential pitfalls

Initial considerations

Invalid/incorrect data

- Out of range
- Mismatched related variables' values
- Invalid categorical values

Should be noted in the SAS log

Reassess as you go

Hey, I got a new data transfer!

- New or missing datasets
- How did the number of records in each dataset change
- New or missing variables
- Variable attributes have changed
- New values in categorical variables
- New ranges (or outliers) in continuous variables
- Changes in related variables (xxTEST vs. xxTESTCD)

Reassess as you go

Hey, I got a new data transfer!

- Scrutinize lab values and units
- Visits
- Sort orders no longer provide uniqueness
- What does any of this mean for my metadata?

Tools and processes

Defensive programming/coding defined

Typical definition (www.drdobbs.com)

- “Defensive programming is a practice where you anticipate failures in your code, then add supporting code to detect, isolate, and in some cases, recover from the anticipated failure.”

In this context, however...

- Writing programs that handle all input data (both anticipated and unanticipated) gracefully, and notify the user/programmer when data or processing issues arise.

Tools and processes

Defensive programming/coding: Notify!

- Look for all possible known values
- Make unanticipated values obvious
- Use the SAS log to highlight unexpected values

- Use the SAS log to identify variables that can't be calculated yet

Tools and processes

Defensive programming/coding techniques

- For unanticipated values: flag and notify
 - ❖ Final ELSE on IF/ELSE IF constructs
 - ❖ OTHERWISE on SELECT statements
 - ❖ ELSE on SQL CASE statements
 - ❖ PROC FORMAT OTHER

- Make unanticipated values to something obvious

- Use the SAS log to highlight unexpected values

Tools and processes

Defensive programming/coding SAS log messages

- Standardize and differentiate from SAS' messages

```
!!! Warning:
```

```
### Note:
```

- Make the code invisible

```
put "!!" "! War" "ning: message text";  
put "##" "# Note: message text";
```

Tools and processes

Misc: %FindDups macro

```
%finddups(InDS=ds1 ds2 ds3, SortOrd=USUBJID);
```

```
data combined;  
  merge ds1 ds2 ds3;  
  by USUBJID;
```

```
### Note: No duplicates found in ds1 with SortOrd=USUBJID  
!!! Warning: Duplicates in ds2: USUBJID=111-123-0001  
!!! Warning: Duplicates in ds2: USUBJID=111-345-0001  
### Note: No duplicates found in ds3 with SortOrd=USUBJID
```

```
%finddups(InDS=ADLB, SortOrd=USUBJID PARAMCD AVISITN ADT);
```

Tools and processes

Process an iterative data transfer

- Archive old data – but keep available
- Load new data
- Compare

Tools and processes

Process an iterative data transfer

- Comparison criteria
 - ❖ New or missing datasets
 - ❖ How did the number of records in each dataset change
 - ❖ New or missing variables
 - ❖ Variables with different attributes
 - ❖ New values in categorical variables
 - ❖ New ranges (or outliers) in continuous variables
 - ❖ Changes in related variables (xxTEST vs. xxTESTCD)
 - ❖ Visits!!!

Tools and processes

Data transfer comparison macro

- Compare current data to previous transfer
 - ❖ Datasets: presence, record counts
 - ❖ Variables: presence, attributes
 - ❖ Values (optional): categorical, continuous
- Output Spreadsheet (ODS EXCELXP Tagset)

```
%mLibComp(Folder=SDTM) ;
```

```
%mLibComp(Folder=SDTM, Values=_all_) ;
```

```
%mLibComp(Folder=SDTM, Values=DM EX, MaxCats=25) ;
```

Tools and processes

Datasets comparison spreadsheet

		Record Counts			
				Change	
Dataset	Status	Old	New	Count	Percent
AE		14095	14298	203	1.40%
AM		6913	6962	49	0.70%
CM		26015	24532	-1483	(5.7%)
DM		797	795	-2	(0.3%)
DS		10312	5542	-4770	(46.3%)
DV		4242	4717	475	11.20%
EX		11087	11090	3	0.00%
...					
PD		7650	7742	92	1.20%
PR	Not in Old		2093		
QS		371242	380121	8879	2.40%

Tools and processes

Variables and attributes spreadsheet

Dataset	Variable	Old Transfer			New Transfer		
		Label	Type	Length	Label	Type	Length
AE	AEACN	Action Taken with Study Treatment	Char	200	Action Taken with Study Treatment	Char	40
	AEACNOTH	Other Action Taken	Char	200	Other Action Taken	Char	200
	AEBDSYCD	Body System or Organ Class Code	Num	8	Body System or Organ Class Code	Num	8
	AEBODSYS	Body System or Organ Class	Char	80	Body System or Organ Class	Char	80
	AECAT	Category for Adverse Event	Char	40	Category for Adverse Event	Char	40
	AECONTRT	Concomitant or Additional Trtmnt Given	Char	1	Concomitant or Additional Trtmnt Given	Char	2
	AEDECOD	Dictionary-Derived Term	Char	200	Dictionary-Derived Term	Char	200
	AEENDTC	End Date/Time of Adverse Event	Char	64	End Date/Time of Adverse Event	Char	64
	AEENDY	Study Day of End of Adverse Event	Num	8	Study Day of End of Adverse Event	Num	8
	AEENRF	End Relative to Reference Period	Char	20			
	AEENRTPT	End Relative to Reference Time Point	Char	40	End Relative to Reference Time Point	Char	40

Tools and processes

Values comparison (categorical) spreadsheet

Variable	Values	Counts			Percentages		
		Old	New	Difference	Old	New	Difference
DOMAIN	QS	371242	380121	8879	100.00%	100.00%	0.00%
EPOCH	FOLLOW-UP	37731	45953	8222	10.00%	12.00%	1.90%
	SCREENING	26117	28354	2237	7.00%	7.50%	0.40%
	TREATMENT	307394	305645	-1749	83.00%	80.00%	(2.4%)
	Missing		169			0.00%	
QSCAT	BPI SHORT FORM	105732	104697	-1035	28.00%	28.00%	(0.9%)
	FACT-P	265510	268639	3129	72.00%	71.00%	(0.8%)
	KPS Scale		6785			1.80%	
QSDTC	More than 20 discrete values						
QSDY	More than 20 discrete values						

Tools and processes

Values comparison (continuous) spreadsheet

Variable	N		Minimum		Maximum		Mean		STD	
	Old	New	Old	New	Old	New	Old	New	Old	New
QSDY	369651	378165	-742	-742	716	778	105.0908	108.6304	107.7306	112.5284
QSSEQ	371242	380121	1	1	1479	1507	359.2554	370.0487	254.9425	264.1602
QSSTRESN	353946	364065	0	0	155	155	7.86212	9.207232	19.8587	22.10026
VISITNUM	371122	379889	-1	-1	600120	600140	132786.2	138146.6	201805.8	206520.2

Tools and processes

Examine the relationships

PROC FREQ is your friend

- VISITNUM → VISIT (across all domains)
- xxTESTCD → xxTEST (what about units?)
- PARAMN → PARAMCD → PARAM

Conclusions

- Understand potential impact
- Prepare
- Reassess
- Tools and processes

Axio

PARTNERS IN RESEARCH