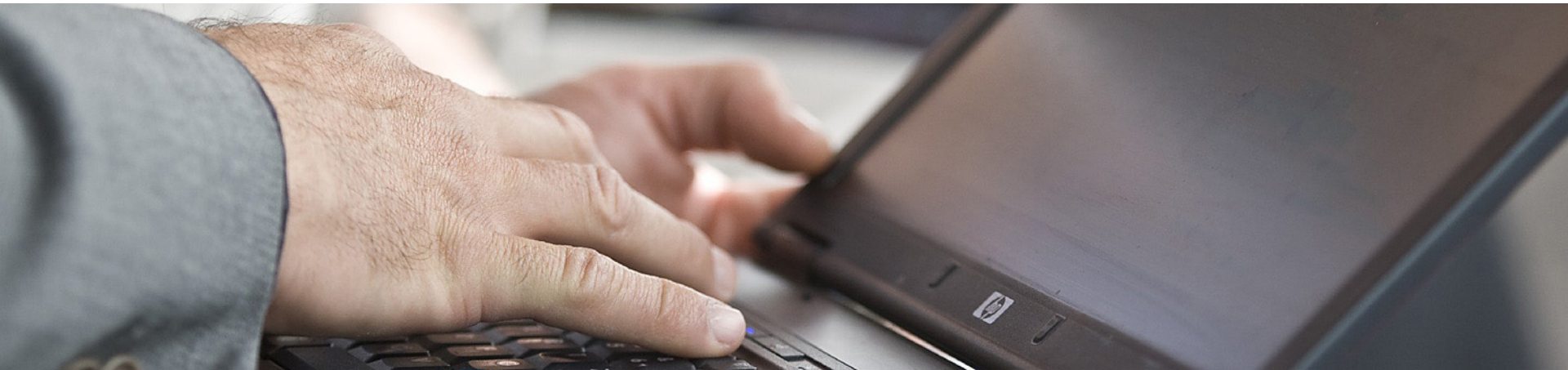

Datacut Strategies: What, Why & How

Hiren Naygandhi

Acknowledgements: Chris McKenna, Hinal Patel & Vijay Reddi



Contents

- **What is a datacut?**
- **Why do we need it?**
- **Scenarios**
- **What if there was no datacut?**
- **Who should be responsible?**
- **Conclusion**

What is a datacut?

- Data that have been **cut into subsets**, according to a **set of rules**

Sequence Number	Start Date of AE in SAS Format	End Date of AE in SAS Format	Category for Adverse Event	Reported Term for the Adverse Event
1	04MAY2011	.	GENERAL ADVERSE EVENTS	ALOPECIA
4	19OCT2011	09NOV2011	GENERAL ADVERSE EVENTS	ANXIETY
5	09NOV2011	21DEC2011	GENERAL ADVERSE EVENTS	ANXIETY
34	20JUL2011	.	GENERAL ADVERSE EVENTS	JOINT PAIN
6	09NOV2011	11JAN2012	GENERAL ADVERSE EVENTS	ARTHRALGIA

- Could be at a **date**, for a **number of patients** or significant **point in time**
- Usually
 - ✓ defined by other function e.g. Stats, Science
 - ✓ performed by Stats Programming, Data Management

Why do we need it?

- Need to ensure data is “clean”

Clinical Cut-Off

Snapshot

- Need to consider: e.g. Cutoff date: 05JAN2012, Snapshot date: 23JAN2012

Identify cut-off fields

Sequence Number	Start Date of AE in SAS Format	Start Date/Time of Adverse Event	End Date of AE in SAS Format	End Date/Time of Adverse Event	Reported Term for the Event
9	28SEP2011	2011-09-28	.		COUGH
2	04JAN2012	2012-01-04	11JAN2012	2012-01-11	ANOREXIA
3	23MAY2012	2012-05-23	20JUN2012	2012-06-20	ANOREXIA
10	01MAY2011	2011-05--	.		DEPRESSION
11	19OCT2011	2011-10-19	07DEC2011	2011-12-07	DIARRHEA

Imputation of partial dates

Rollback i.e. amend data to ongoing if after datacut

Scenarios

Scenario	Cut applied at Source or Analysis	Imputed dates	Rollback
1	Source	Retained	Yes
2	Analysis	Retained	No
3	Source	Dropped	No (flags created in analysis datasets)

Scenario 1



Scenario	Cut applied at Source or Analysis	Imputed dates	Rollback
1	Source	Retained	Yes

➤ Advantages:

- ✓ Imputations made only once
- ✓ Rollback applied, which gives a further accurate reflection of events without bias

➤ Disadvantages:

- ✓ Derivations added within source datasets
- ✓ Rollback applied, though how much of a significant impact would it have?

Scenario 2

Scenario	Cut applied at Source or Analysis	Imputed dates	Rollback
2	Analysis	Retained	No

➤ Advantages:

- ✓ Source data remains intact
- ✓ Only need to apply date imputations once

➤ Disadvantages:

- ✓ Mis-match between source and analysis data
- ✓ No Rollback applied, which could lead to bias

Scenario 3

Scenario	Cut applied at Source or Analysis	Imputed dates	Rollback
3	Source	Dropped	No (flags created in analysis datasets)

➤ Advantages:

- ✓ No rollback applied, therefore no source data changed
- ✓ Flags will identify events after datacut and allow flexibility for reporting

➤ Disadvantages:

- ✓ Date imputations would need to be performed twice

What if there was no datacut?



➤ What if we did **NOT** apply any date imputations?

➤ Use **database entry date** instead of **event date**

Record Minimum Created Datetime Stamp	Primary Adverse Event - Raw	AE Onset Date	AE Resolution Date
24FEB12:21:15:33	NEUROPATHY	09JAN2012	07FEB2012
27FEB12:22:08:28	SHORTNESS OF BREATH	08FEB2012	20FEB2012
27FEB12:22:10:42	FEVER	08FEB2012	10FEB2012
06DEC11:19:31:23	NEUROPATHY	26SEP2011	09JAN2012
06DEC11:19:33:53	INTERMITTENT CONSTIPATION	19SEP2011	.

- ✓ No need for imputations
- ✓ No need for datacut
- ✓ Accurate reflection of database

- ✓ Delay in data entry
- ✓ Inaccurate reflection of study

Who should be responsible?

- Stats Programming or Data Management?
- What if Data Management (DM) were owners for datacuts?
- **Advantages:**
 - ✓ We (Statistical Programming) don't need to worry about performing the datacut!
 - ✓ DM best placed with access and tools available to them
 - ✓ DM responsibility to provide “analysis-ready” data
- **Disadvantages:**
 - ✓ May be perceived as performing “analysis” by DM
 - ✓ Depends on DM resource/agreement
- **Need to ensure:**
 - ✓ There is a clear process
 - ✓ Stakeholders are included early

Conclusion

- Much **more to datacut** than just **subsetting**!
- Various strategies available: **No wrong or right solution**
- **Recommendation:**
 - ✓ Mixture of **multiple strategies** depending on purpose
 - ✓ Ensure this is **consistent** across all projects i.e. one process to follow
 - ✓ Ensure other **stakeholders** i.e. Statistics, Science etc. are included in **decisions**
 - ✓ As programmers we can play a **vital role** in defining process

We have reached the Cut-off!

- What **strategy** do you apply?
- Who should take **ownership** of the process?
- Any questions?



***Doing now what patients need
next***