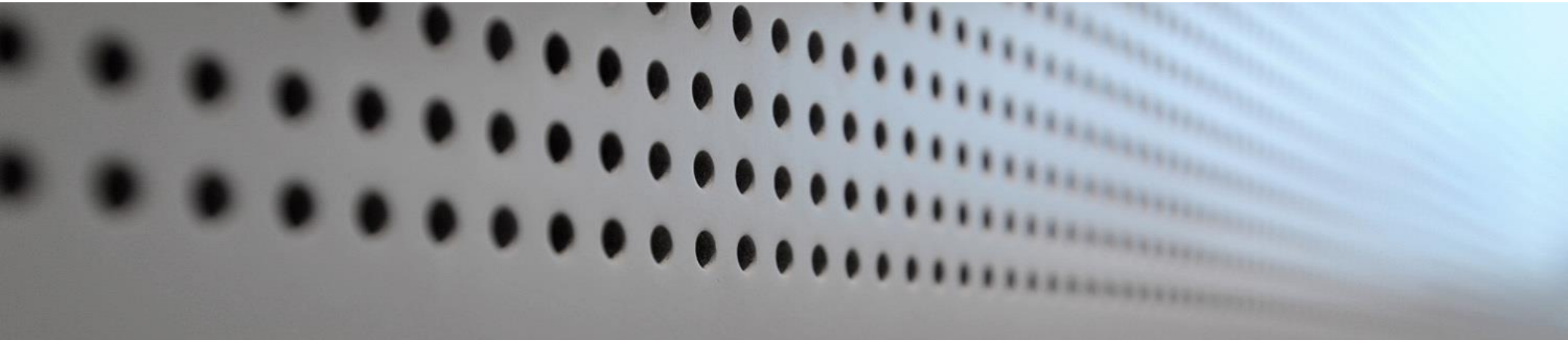
Standardizing Data Cuts

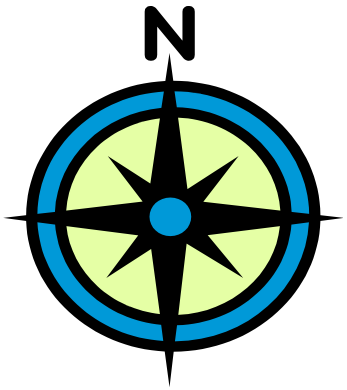
Mercidita T. Navarro

December 2, 2015



Outline

- Background
- Data Cut Standards
- Technical Implementation
- Summary



Background

Motivation

What is a Data Cut?

- A data cut is a predefined subset of the clinical study database
- Typically, all available data at a pre-specified point in time, or a subset of data collected up to that point in time

Why Use a Data Cut?

- Interim analyses are prepared to support the objectives of a study and enable decisions for a project
- A data cut provides a static set of data on which to perform an interim analysis
  - allows for traceability and reproducibility of results while the study is ongoing and data continue to be collected
- With a predefined data cutoff, relevant data can be “cleaned”

Why Standardize?

- Consistency leads to efficiency in resources
- Opportunity for workflow automation

Types of Data Cut

1. **Snapshot** – All data **available** on the date of the snapshot
   Example: Snapshot taken after enrolling 25% of target
2. **Fixed-cutoff data cut** – All data **collected** on or before a study milestone
   Examples:
   - Date of Week 12 visit of the last patient enrolled
   - Date of 100th event (e.g., death or disease progression)
3. **Variable-cutoff data cut** – All data **collected** as of a **subject-specific** milestone
   Examples: All data collected
   - On or before the Week 12 visit
   - On or before completion of 12 weeks of dosing
   - Before crossing over to the next treatment in a crossover-design study

DSMB = Data & Safety Monitoring Board

Choosing the Type of Data Cut

- The type of data cut to use will depend on the purpose of the interim analysis
- Possible scenarios:

Purpose	Example	Applicable Data Cut
Safety monitoring report	DSMB	Snapshot
Annual safety update	DSUR/PBRER	Fixed-cutoff = DLP
Efficacy: Time-to-event	Overall survival endpoint	Fixed-cutoff = date when required no. events have occurred
Efficacy/safety over a pre-specified exposure interval	Efficacy after 12 weeks of treatment	Variable-cutoff = date of milestone visit

- Determination of clinical cutoff date is study-specific

DSUR=Development Safety Update Report; PBRER=Periodic Benefit-Risk Evaluation Report; DLP=Data Lock Point

Data Cut Standards

Rules for Fixed- and Variable-Cutoff Data Cuts

- **Which subjects to include?**
  - All subjects enrolled on or before the clinical cutoff date
- **Which records to include?**
  - Records with assessment/milestone/start date on or before the clinical cutoff date
- **Which data elements to truncate?**
  - **Dosing** data and **time in study** will be truncated at the clinical cutoff date as part of the analysis
  - Allows for accurate documentation of
    a) duration of exposure to study drug
    b) time in study
    as of the clinical cutoff date

Addressing Potential Discrepancies

- Missing/Partial Dates: Include records in the data cut unless it is clear that the relevant date is after the clinical cutoff date, for example:

Assessment/Start Date	Clinical Cutoff Date	Include?
CM start date = 2014	30 NOV 2013	No
Lab test date = JAN 2015	01 DEC 2014	No
AE start date = missing	15 JUN 2014	Yes

- For AE and CM records: All data elements will be kept as tabulated from raw data in the snapshot, i.e., **no "rollback"** (e.g., AE end date, outcome, most extreme intensity)
  - Provides as much information as possible for data interpretation at the subject level
- Any inconsistencies in the following key data displays will be explained:
  - Summary/listing of AEs leading to study drug discontinuation/death
  - Time-to-event analyses

Adverse Event and Outcome

- Potential inconsistencies between summary tables/listings

	Topic	Source	Included in the Data Cut
1	Subjects with **AE** resulting in **death**	AE domain	All outcomes for AEs with start date on or before clinical cutoff
2	Subjects who **discontinued** from the study due to **death**	DS domain	All DS records with study discontinuation date on or before clinical cutoff

- The number of subjects may differ between (1) and (2) after applying the data cut rules
- Clearly indicate in a table/listing footnote what are included in the analysis
- Note discrepancies when interpreting results
- Similarly, review results for inconsistencies associated with AE outcome = study drug discontinuation

Adverse Event and Outcome – Example

- Clinical cutoff date = March 31, 2009
- Snapshot date = April 28, 2009
- As of the snapshot date, 4 patients experienced serious adverse events

Patient	AE	Onset Date	End Date	Reason Classified as Serious
10003	Diarrhea	**Sept. 14, 2008**	Sept. 28, 2008	Hospitalization
11105	Thrombocytopenia	**Dec. 15, 2008**	Dec. 29, 2008	Life threatening
10001	Pneumonia	**Jan. 1, 2009**	Feb. 28, 2009	Resulted in death
12345	Coronary artery disease	**March 1, 2009**	April 15, 2009	Resulted in death

- All **4 SAEs** will be included in the data cut
- All outcomes including **2 deaths** will be reflected in the data cut
- Two (2) deaths due to AE are included in the snapshot

Patient	Date of Death	Reason
10001	Feb. 28, 2009	Pneumonia
~~12345~~	~~**April 15, 2009**~~	Coronary artery disease

- Only **1 death record** (Patient 10001) will be included in the data cut

Discrepancy between the SAE listing and death listing must be noted in the report

Additional Considerations

- No records with assessment/milestone/start date after the clinical cutoff date will be included in any analysis
- For time-to-event analyses, death status and date last known to be alive will be updated in the VAD to be consistent with the clinical cutoff date, for example:

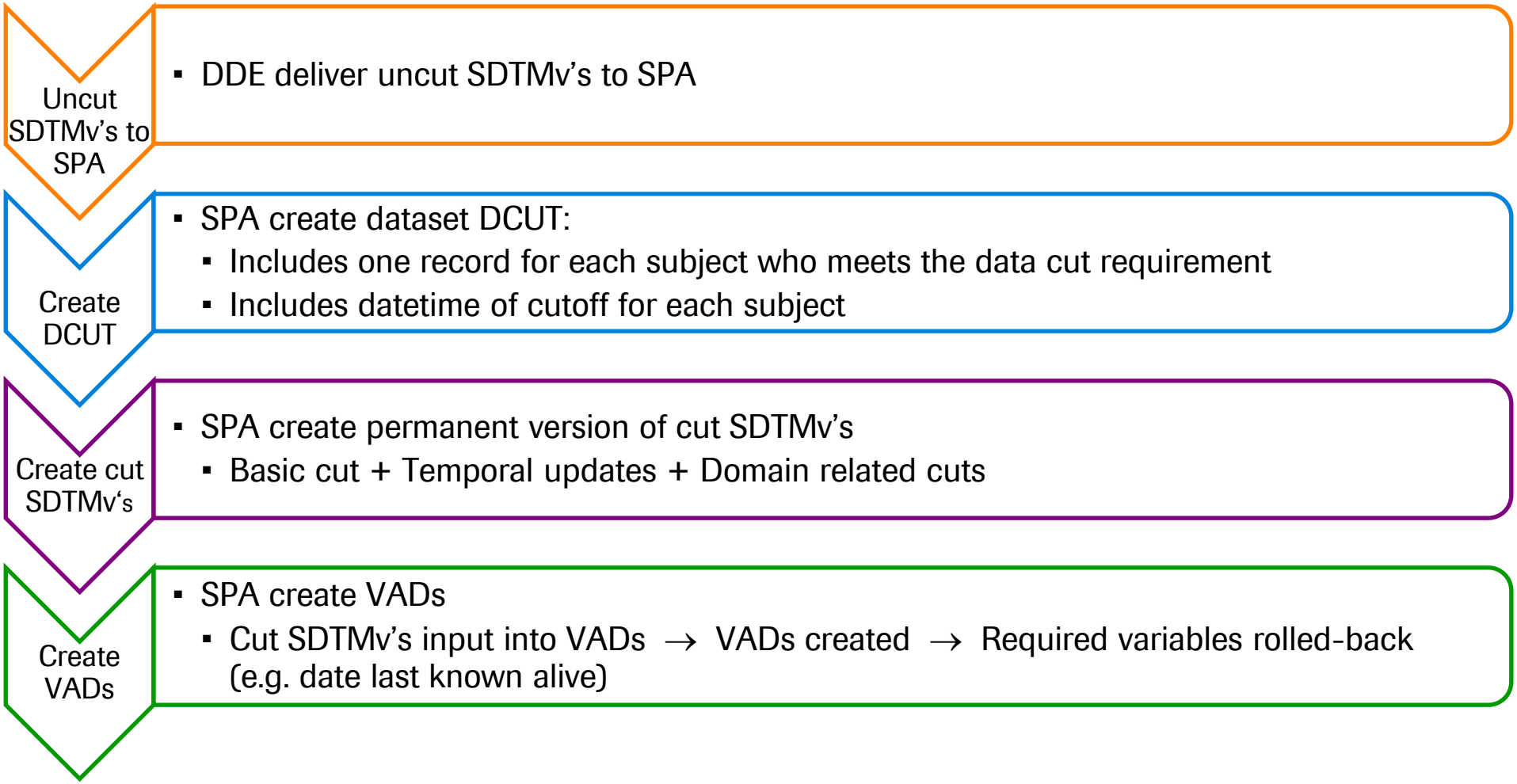
Analysis Variable	Derived Value	Source	Clinical Cutoff	Analysis Value
Patient died? (date)	"Y" (13 JUN 2014)	Fatal AE end date	30 MAY 2014	null[a]
Date last known alive	30 JUN 2014	CM end date	30 MAY 2014	30 MAY 2014[b]

[a]Delete record from Death VAD
[b]Create an analysis flag to indicate that value is truncated at the clinical cutoff date

- Additional steps may be required when analyzing other variables derived from end dates (e.g., duration of AE)
  - Note that data elements will be kept in the data cut as tabulated from raw data (i.e., no "rollback")

Technical Implementation

Process Flow



DDE=Database Development & Extraction; SPA=Statistical Programming & Analysis; VAD=Value-Added Dataset (analysis file)

Key Steps

- DCUT dataset is used to create cut SDTMv's which are the basis of VADs
- SDTMv's with no date component will be used as transferred from DDE (e.g., TS and other non-patient level data)
- Several levels of cutting data:
  1. **Basic cut** – Delete records with assessment/start date after the clinical cutoff date
  2. **Temporal update** – Update temporal derivations in DM and other affected domains (e.g., DM.ACTARM based on cut EX records)
  3. **Relational update** – Ensure that child-domain records link to a parent-domain record (e.g., RELREC, FA)
     - Some RELREC records may become meaningless after the basic cut—decision to delete these records is study-specific
  4. **VAD-specific update** – Roll back key derived variables (e.g., date last known alive)
- Important to validate all data cut steps at "high-risk" level

Summary

Main Takeaways

- A data cut is a predefined subset of the clinical study database that is used for an interim analysis
- Standard definitions and rules for creating a data cut allow for a consistent process flow
- Discrepancies will arise as a consequence of the data cutting steps
  - Need to carefully address inconsistencies with high impact
- Technical implementation is multi-faceted
  -- it will take several steps to achieve a robust data cut

Acknowledgments

Data Cut Working Group	Global Information Standards Governance Committee Data Analysis Core Teams
Tracy Burgess	Claire Hughes
Beki Finch	Ian Whatmough
Brandon Arnieri	Robin Koeger
Peter Gerber	Tim Barnett
Chris McKenna	Ravinder Bahia
Hiren Naygandhi	Mark Luff
Melanie Bennett	Coen Bernaards
Tarik Khaznadar	Cornelia Irl
Simon Eng	Barbara Tong
	Betty Nelson
	Elina Asikanius
	Irene Alobwede
	Chris Price

… and special thanks to subject matter experts and GIS-GC extended team members who provided input during the review process, and the early adopters whose feedback helped shape the final implementation workflow

Doing now what patients need next

Backup

From Concept to Reality

3–4 mos.

Data Cut Working Group

- Cross-functional team with experience in various therapeutic areas proposed standard definitions and rules for data cuts
- Membership from Biostatistics, Statistical Programming, Clinical Science, Safety Science, Standards Group

2–3 mos.

Data Standards Governance Committee – Definitions

- Data Analysis Definitions Core Team reviewed proposal from Data Cut Working Group
- Key stakeholders consulted as needed
- Rules and definitions approved and added to Global Data Standards Repository

4–6 mos.

Data Standards Governance Committee – Technical Specifications

- Technical details published following focused discussion within the Technical Specifications Core Team

~12 mos.

Technical Implementation/Systems Development

- Pending development of standard data cut macros, teams were advised to apply the data cut standards following workflow recommended by Governance Committee
- Early adopters shared their experience including those from recent filings
- Dedicated SPA team developed a comprehensive, technical implementation guidance