

DH07-

Enhancing Data Accuracy and Collaboration through Efficient Data Reconciliation using SAS in Clinical Trials



Piyush Makkapati & Swathi Nattam

Vita Global Sciences, Waltham, MA



Vita Global Sciences
FROM RESEARCH TO REALITY

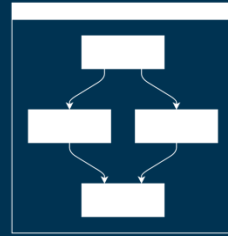
Agenda



Introduction &
Background



Stakes of
Database Locks



Process Flow
Diagram



Executing In-
House Approach



Real-World Impact
& Conclusion



Vita Global Sciences

Introduction & Background

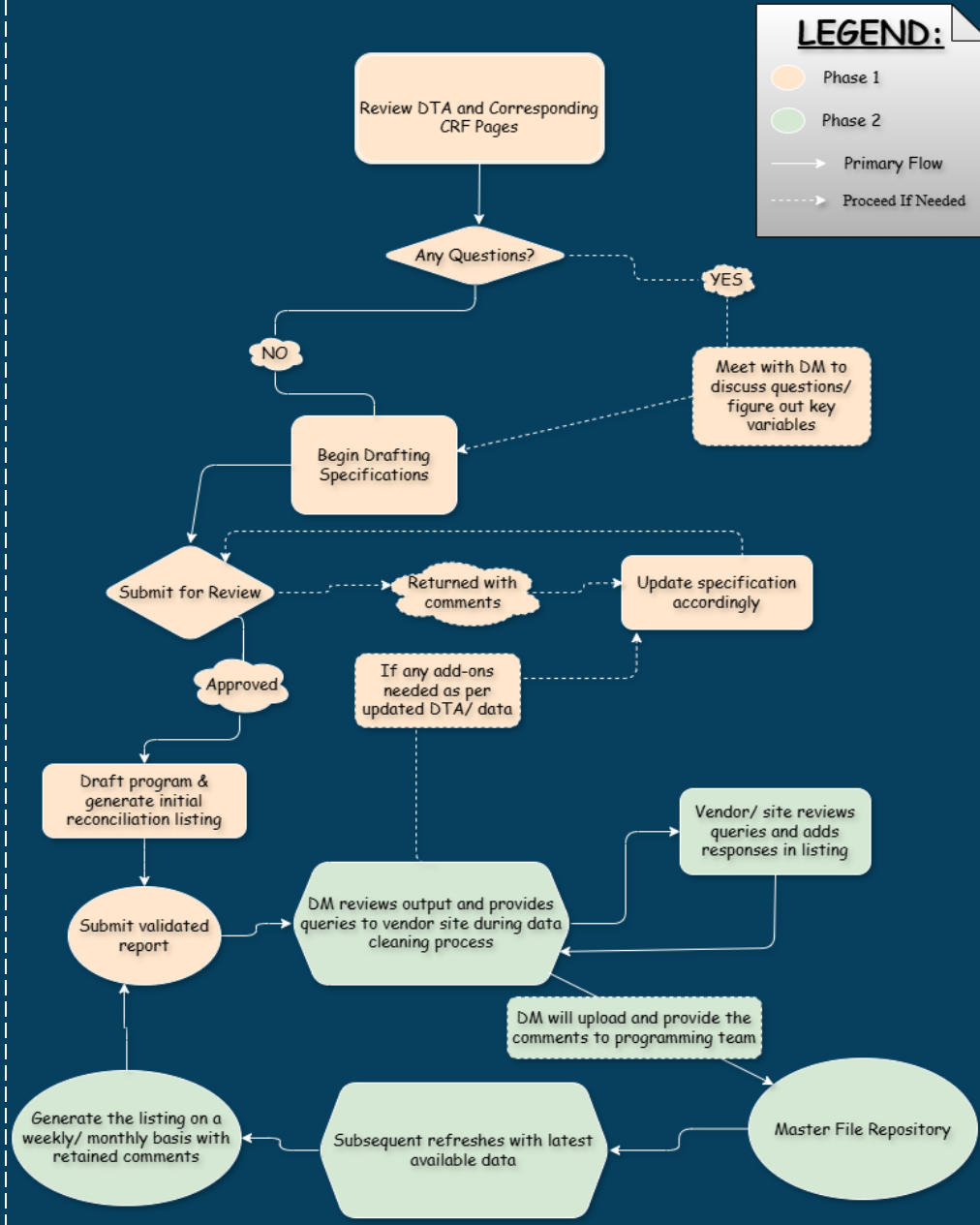
- ❖ **Definition:** In practice, data reconciliation is the process of merging and comparing key data points from EDC and external vendor datasets to identify and resolve inconsistencies in clinical trials
- ❖ **Importance:** A critical component of successfully managed clinical trials, ensuring data accuracy, consistency, and regulatory compliance
- ❖ **In-house Approach:** A SAS-based strategy that proactively detects and resolves data gaps before critical timelines

The High Stakes of Database Locks: Insights From Data Management

- ❖ A **Database Lock (DBL)** is the final stage of EDC data in clinical trials, ensuring all data entry is complete, data issues are addressed, and no further modifications made to the database
- ❖ **Problem:** Manual reconciliation for Data Management can take one or two days for basic checks
- ❖ Last-minute discrepancies can trigger a surge of data queries and urgent fixes
 - “In one study, new vendor data arrived less than two weeks away from the DBL and triggered a series of inconsistencies that delayed the planned lock by several weeks.” –Data Manager



Figure 1: SAS-Based Reconciliation Process Flow



- ❖ Flexible Two-Phase process designed to handle all types of vendor data
- ❖ Generates custom-tailored outputs for Data Management

The Role of Vendor DTAs & eCRF Pages

- ❖ The SAS-based programmatic reconciliation process starts by reviewing key documents such as Data Transfer Agreements (DTAs), Case Report Forms (CRFs), Study Protocols, and other relevant materials to understand and ensure data consistency
 - Combining these references allows teams to create specifications for matching and comparing data between the sources

VISIT	Description of timepoint	Timepoint (LBTPT)	Specimen (LBSPEC)
Day 1	Predose	0.00	Plasma
Day 1	30 min post-dose	0.50	Plasma
Day 1	1-hour post-dose	1.00	Plasma
Day 1	2-hours post-dose	2.00	Plasma
Day 1	4-hours post-dose	4.00	Plasma
Day 1	6-hours post-dose	6.00	Plasma
Day 1	8-hours post-dose	8.00	Plasma

Visit Mapping Table from DTA



Timepoint

Predose ☐

30 Minutes Postdose ☐

1 Hour Postdose ☐

2 Hours Postdose ☐

4 Hours Postdose ☐

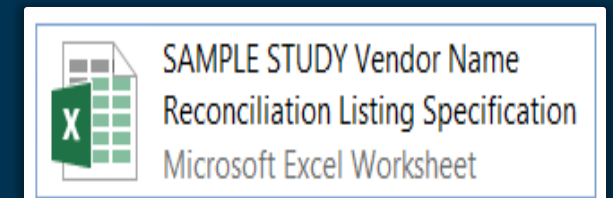
6 Hours Postdose ☐

8 Hours Postdose ☐

12 Hours Postdose ☐

16 Hours Postdose ☐

Sample Collected? Yes ☐ No ☐



Unique eCRFs
Form: Plasma PK - Serial 1
Generated On: 01Jan 2025 15:43:21



Vita Global Sciences

Specifications

- ❖ An essential step in effective automated data reconciliation is drafting well-defined specifications that draw on both the vendor DTAs and corresponding CRF pages
- ❖ Commonly compiled in an excel workbook that consists of three core sheets: Template, Programming Instructions, and Modification History
- ❖ Within the template sheet, the columns required for the final output are divided into three sections
 - The first is for Data Management entries
 - The second and third display study-specific variables

Data Management					EDC							Vendor Data File (Lab)							Data Management					
STUDYID	Issue Number	Reconciliation Date	Reviewer Name	New Issue Found?	Subject ID	Subject Sex	EDC Visit	CRF Name	Date	Time	Timepoint	Subject ID	Specimen ID	Specimen Type	EXT Visit	Vendor File	Date	Time	Timepoint	Issue Description	DM Comments or Site Query Confirmation	Vendor or Site Response (Response to the DM Query)	Status (Open or Closed or Added to Vendor Log or Irresolvable)	Screen Failure Flag

Template

Programming Instructions

Modification History

+

Figure 2: Specifications Template Tab

Specifications

❖ Programming Instructions sheet:

- Lists all EDC and vendor data files to be reconciled
- Outlines mappings, reconciliation checks, and any special conditions (e.g., screen failures)
- Determines key variables such as subject ID, visit, date, time, and time points
 - Required for accurately merging records between the two databases

❖ Once key variables are confirmed, they are used to determine discrepancy checks like below

List of Issues to be output for this Listing as below
Issue 01: Vendor Sample collection time point does not match EDC sample collection time point
Issue 02: Vendor Sample collection time does not match EDC sample collection time
Issue 03: Vendor Sample collection time point & time does not match EDC sample collection time point & time
Issue 04: Vendor Sample collection date does not match EDC sample collection date
Issue 05: Records present in edc but not in Vendor data
Issue 06: Records present in Vendor data but not in edc

Figure 3: Sample Lab Issue Descriptions



Our SAS-based approach automates data checks, detecting issues so data management and vendor teams can resolve them efficiently, ensuring study data quality and consistency.



Specifications

Modification History and Final Steps before Programming:

- ❖ Modification History sheet tracks all spec updates
- ❖ DM-approved specs are implemented by programmers to generate reconciliation listings
 - Listings highlight non-reconciled records and are sent back for resolution
- ❖ Structured process that helps maintain data quality and prevents delays



v01: Initial
Specification



v02: DM Requesting
Time Mismatch Removal



v03: Addition of New
Variables in Output



v04: New Visits Added
Based on Updated DTA



SAS-Based Reconciliation Approach: Phase I

- ❖ The programming process has two phases, each with defined steps (Figure 1)
- ❖ Phase I: Import EDC and vendor data into SAS and standardize the key variables
- ❖ Format and clean data to ensure proper merging of EDC and vendor datasets
- ❖ Merge datasets and flag any discrepancies using mismatch checks and issue descriptions
- ❖ Final output:
 - Validated Excel report of reconciliation issues
 - Contains a sheet named based on the latest EDC and vendor data used
 - Reviewed by DM and used for Phase II runs

```
/*---Merge EDC & EXT and check match/un-match records---*/  
data Edc_ext Edc_only Ext_only;  
  merge Edc(in=a) Ext(in=b);  
  by Subjid Visit Date Time Timepoint;  
  if a and b then output Edc_ext;  
  if a and ^b then output Edc_only;  
  if ^a and b then output Ext_only;  
run;
```

Figure 4: EDC and EXT Data Merging Step

SAS-Based Reconciliation Approach: Phase II

- ❖ After Data Management feedback, the reviewed listing becomes the master file
- ❖ For subsequent reconciliation runs, issues identified from using the latest EDC and vendor data are merged with the master file
 - Comments and statuses for ongoing/resolved issues are preserved
- ❖ Final output: two sheets for active (old & new) and resolved discrepancies
- ❖ The updated output is validated and resubmitted to Data Management, repeating until all discrepancies are addressed and the database is locked

Figure 5: New Listing & Master File Merging Step

```
data Curr_old Curr Old;
  merge Curfile(in=a) Master_dm(in=b);
  by Subjid Edcvisit Edcform Edcdataset Edcdate Edctime Edctimepoint
     Extvisit Extdate Exttime Exttimepoint Issuedescription;

  /*----Issues still existing (New & Old identified issues)----*/
  if a and b then output Curr_old;

  /*----Only new issues----*/
  if a and ^b then output Curr;

  /*----Only resolved issues----*/
  if b and ^a then output Old;
run;
```

Data Management					EDC							Vendor Data File (Lab)										Data Management				
STUDYID	Issue Number	Reconciliation Date	Reviewer Name	New Issue Found?	Subject ID	Subject Sex	EDC Visit	CRF Name	Date	Time	Timepoint	Subject ID	Subject Sex	Specimen ID	Specimen Type	EXT Visit	Vendor File	Date	Time	Timepoint	Issue Description	DM Comments or Site Query Confirmation	Vendor or Site Response (Response to the DM Query)	Status (Open or Closed or Added to Vendor Log or Irresolvable)	Screen Failure Flag	
Sample Study	1	3-Nov-2024		N	ADC01-001	F	Day 2	Plasma PK	3/1/2024	11:03	24H	ADC01-001	F	J027AJQ2	Plasma	Day 2	ADC_lab_01NOV2024	3/1/2024	11:08	24H	Vendor Sample collection time does not match EDC sample collection time	Queried Vendor	Sample not received yet	Open		
Sample Study	2	3-Dec-2024		N	ADC01-002	F	Day 3	Plasma PK	6/20/2024	10:26	48H	ADC01-002	F	J05HPON7	Plasma	Day 3	ADC_lab_01DEC2024	6/15/2024	10:26	48H	Vendor Sample collection date does not match EDC sample collection date for a Visit	Should resolve next round		Open		
Sample Study	3	3-Jan-2025		Y	ADC01-003	M	Day 36	Plasma PK	10/28/2024	10:05	Predose	ADC01-003					ADC_lab_01JAN2025				Records present in edc but not in Vendor data					
EDC01JAN25_EXT02JAN25_03JAN25					old_resolved_issues																					

Figure 6: Sample Lab Output

Process for more Complex Recons: SAE Reconciliation

- ❖ Phase I & II remain the foundation for EDC and vendor data reconciliation
 - SAEs require heightened scrutiny due to their critical impact on clinical trials
- ❖ CRF pages covering adverse events & fatalities are consolidated into a single EDC dataset
 - Merged with vendor safety data after aligning key variables
- ❖ Final output includes four structured sheets
 - First sheet provides a summary of SAE counts and highlights matched records that successfully pass the mismatch checks

STUDYNAME Total SAE Counts		#
Total Number of SAE's from EDC		98
Total Number of SAE's from Vendor		98
Matched SAE's		80

Total SAE Counts	All SAEs	03JAN2025_matched	EDC01JAN25_EXT02JAN25_03JAN25
------------------	----------	-------------------	-------------------------------

Figure 7: Total SAE Counts Tab

List of Issues to be output for this Listing as below	
Issue 01: Records present in edc but not in Vendor data	Issue 06: Serious Criteria DISABILITY in EDC is not matching EXT
Issue 02: Records present in Vendor data but not in edc	Issue 07: Serious Criteria CONGENITAL ANOMALY/ BIRTH DEFECT in EDC is not matching EXT
Issue 03: AEDECOD/Preferred term in EDC is not matching EXT	Issue 08: Serious Criteria OTHER IMPORTANT MEDICAL EVENT in EDC is not matching EXT
Issue 04: Serious Criteria LIFE THREATENING in EDC is not matching EXT	Issue 09: Death date in EDC is not matching EXT
Issue 05: Serious Criteria HOSPITALIZATION in EDC is not matching EXT	Issue 10: Relationship to Study Drug in EDC is not matching EXT

Figure 8: Sample SAE Issue Descriptions

Process for more Complex Recons: SAE Reconciliation

- ❖ Second sheet lists all current records from both sources.
- ❖ Third sheet shows matched/resolved events that pass mismatch checks
- ❖ Fourth sheet highlights unresolved discrepancies with reviewed issues and updated comments
- ❖ The SAE reconciliation output features a horizontal layout pairing each EDC record with its corresponding vendor record
 - Simplifies DM review and ensures accurate safety database reconciliation for reliable study outcomes

DATA	STUDYID	Recon Date	Record #	CASE_NUM	Subject ID	Reported Term for the Adverse Event	Dictionary-Derived Term	Serious Event	Is Life Threatening (AESERCR=2)	Start Date of Adverse Event	End Date of Adverse Event	Outcome	Causality	Issue Description
EDC	SAMPLE STUDY	3-Jan-25	3		ADC01-001	Left sylvian artery stroke	Cerebrovascular accident	Yes	Yes	21-Dec-24	24-Dec-24	RECOVERED	NOT RELATED	Serious Criteria LIFE THREATENING in EDC is not matching EXT, AEDECOD/Preferred term in EDC is not matching EXT
EXT	SAMPLE STUDY	3-Jan-25		2010DCO000012	ADC01-001	Left sylvian artery stroke	Middle cerebral artery stroke	Yes		21-Dec-24	24-Dec-24	RECOVERED	NOT RELATED	Serious Criteria LIFE THREATENING in EDC is not matching EXT, AEDECOD/Preferred term in EDC is not matching EXT
Total SAE Counts All SAEs 03JAN2025_matched EDC01JAN25_EXT02JAN25_03JAN25														

Figure 9: Unmatched tab

Real-World Impact & Conclusion



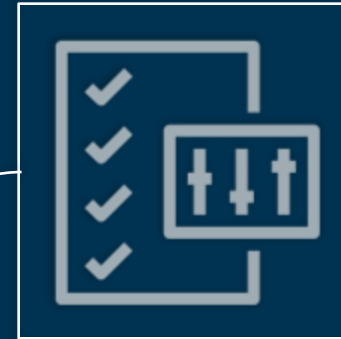
Early Detection is Key to Prevention

- ❖ Eliminates escalation near DBLs
- ❖ Reduces extra costs and resource strain



Collaboration Drives Efficiency & Accuracy

- ❖ DM teams provide clinical context
- ❖ Programmers apply these criteria into code



Automated Checks Enhance Reliability

- ❖ Identify complex mismatches often missed manually
- ❖ Saves significant time with minimal oversight



Strengthened Confidence & Credibility

- ❖ Routine reconciliations ensure data integrity
- ❖ Sponsors and regulators trust final outcomes



Vita Global Sciences

Thank You

pmakkapati@softworldinc.com snattam@softworldinc.com