

Planning and Execution of Real-World Data Evidence Analysis – AS02 (Janssen Initial Experience)

Ramesh Sundaram
Nicole Thorne

Disclaimer

Presentations are intended for informal purpose only and do not replace independent professional judgement

Statements of facts and opinions expressed are those of the presenters individually and not the opinion or position of Janssen pharmaceuticals or its affiliates.

It was prepared from the statistical analysis, programming and reporting perspective, based on our initial RWE experience

Agenda

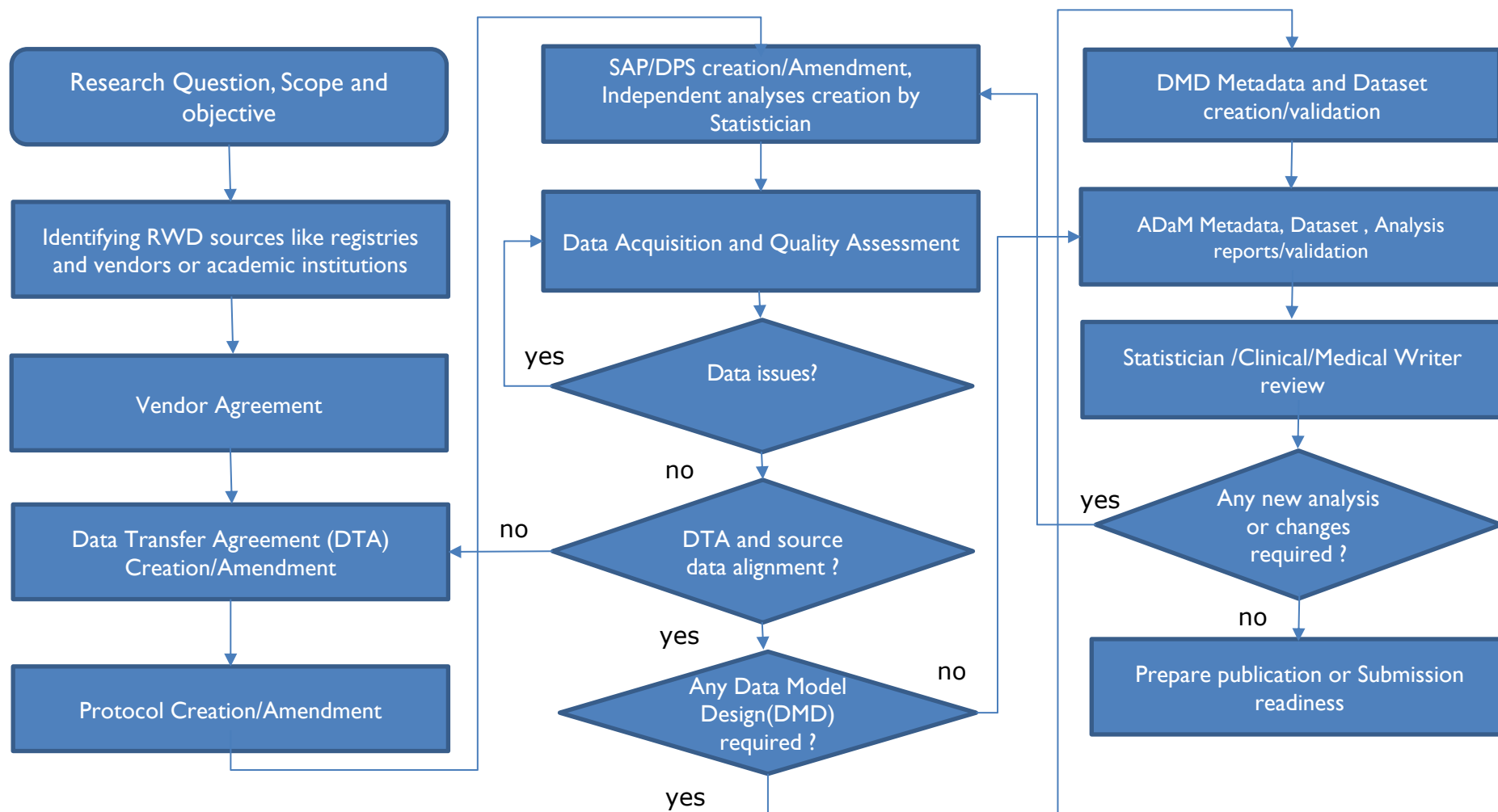
- Introduction
- Process Flow
- Key Factors Impacting Timeline
- Timeline Estimation
- Challenges/Workarounds

Introduction

- The usage of Real -World Data Evidence can optimize the cost and time of drug development process
- In recent days, the role of Real-World Data Evidence (RWDE) analyses are significantly increased in each stage of drug development and its decision-making process
- Like identifying target patient population, finding right variants, gene mutation, even comparators for an investigational product, etc..
- This presentation demonstrates a few key aspects of planning and execution of an RWE analysis and reporting based on Janssen initial experience.



Process Flow



Process Flow (cont.)

➤ Vendor Selection

- Feasibility review like research/indication background, population size, entry/selection criteria etc..

➤ Vendor Agreement

- Agreement with Real World Data (RWD) vendor(s)
- Describes research background
- Objective/Plan/Patient population/size/selection criteria
- Data source
- High level overview of data
- Budget
- Timeline for initial data transfer
- Privacy Consideration



Process Flow (cont.)

➤ Data Transfer Agreement (DTA) Creation

- Data Manager (DM) creates DTA in collaboration with Statistician and Programming lead
- Frequency of data transfer
- Source data format (.csv, .txt, .sas7bdat,.xpt)
- Method and location of data transfer
- Source database metadata (i.e.. Describes list of source files, and its attributes/code lists)
- Define Data Model Design (DMD), if any
- Audit checklist for vendor's data capture process and systems



Requirements Analysis and Design



- Protocol, Statistical Analysis Plan (SAP), Data Presentation Specification (DPS) Creation

- Clinical/Statistician/Programming lead creates these document in collaboration with Translational Research
- Collaborative agreement/DTA/Vendor meetings are key inputs
- Will be finalized and signed off, after cross functional review and confirmation

Process Flow (cont.)

➤ Draft/First source data transfer:

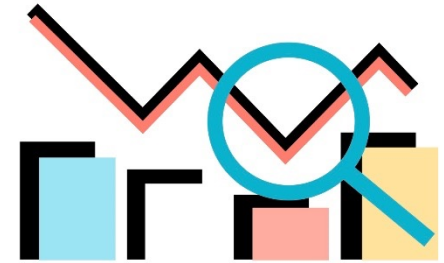
- Data vendor intimates POCs then DM process and load data into the appropriate area.
- DM converts the non-SAS source file in to .sas7bdat files
- Create & Execute Data Acquisition Quality Design Check list
- Programming/Statistician team performs source data checks



➤ DTA and Source data alignment checks

- Based on the research objectives, the cross functional team (Clinical, programming/Statistician) performs an alignment checks i.e., example if PFS is one of the objectives, then the availability of progression date, treatment/diagnosis date, the last known alive date are required.
- Simultaneously Statistician creates independent analysis datasets (Non CDISC) and exploratory reports
- These informal reports are not for submission/publication purposes but can be used for clinical protocol developments, to understand indication background, to validate/review formal outputs (from programming team)

Process Flow (cont.)

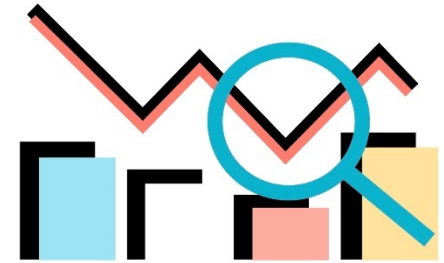


DATA ANALYSIS

➤ Analysis preparation & validation

- Protocol/SAP/DPS/DTA are inputs
- SDTMs or other Data Models (FHIR/Sentinel CDM/PCORNET CDM/OMOP CDM) are created and verified, only if required for submission
- ADaM metadata can be created using DMD (when DMDs are required)
- Otherwise, ADaM metadata can be directly created using nonstandard source raw datasets, all required dependent source variables are carefully reviewed and mapped here, i.e.. to fulfill data integrity and traceability principles.
- ADaM attributes (variable names, parameter names, derivation rules) are kept in line with dependent clinical studies standard, to ease the pooling/data integration process simple.
- Be prepared for regulatory authority Technical Meetings
- Statistician lead reviews ADaM metadata (especially key objectives related derivations)

Process Flow (cont.)



DATA ANALYSIS

➤ Analysis preparation & validation

- Formal reports are generated from standard ADaM datasets
- Programming team validates formal reports (informal datasets/reports created by Statistician can be used for validation purpose)
- Statistician lead performs a formal review of reports, In a few instances Statistician team programs and validates some reports
- Programming team incorporates Statistician review comments, if any
- After Statistician lead review and final confirmation, all reports are transferred to common location for cross functional review
- Dry run(s) can be performed on or before last data transfer, (to make sure all analyses reports as per the expectation, cross functional team will confirm this)
- Programming team incorporates cross functional review comments, if any

Key Factors Impacting Timeline

Need to consider following factors for timeline calculation

**EXPECT
DELAYS**

➤ Number of vendors

- ✓ Dealing with multiple vendors may need more time because each vendor can have their own data structures and formats, so optimizing/integrating them into required standard data model can require some extra time.
- ✓ Don't have a mechanism to validate the duplicate patients across vendors.

➤ Source file type

- ✓ Conversion and validation of non-SAS file formats (.xls,.xlsx,.csv, .txt) can take some extra time.
- ✓ SAS V5 transport format files, can save some conversion and validation time.

Key Factors Impacting Timeline (cont.)

➤ Data transfer

- ✓ Data transfers can be in batches, any delay in transfer can delay the overall timeline.
- ✓ Any structural changes in source data during batch wise transfer, i.e., addition of new variables or deletion of existing ones, can delay the overall time, because this can impact the any existing analysis programs, so reruns can take some extra time.
- ✓ Example a vendor transferred the clinical database on time but didn't transfer the Biomarker, which had the key comparison variables.

➤ Data issues/Data resolution

- ✓ Amount of data inconsistency issues is directly proportional to the volume of source data, so while working on a large sample size, keep some extra time for cleaning and resolution activity.

➤ Volume of required analyses

- ✓ Some time we may need more reports



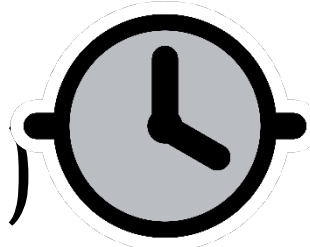
Timeline Estimation



Below table describes approximate time estimate to analyze and report RWE data for a single data vendor, where data transfer, data quality issues/resolution times are nominal and number of required analysis datasets (3-4), reports(35-40)

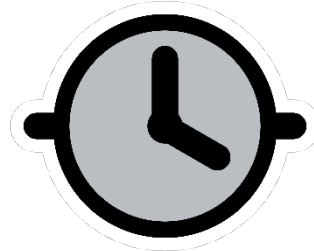
Task	Timeline	Owner(s)	Reviewer(s)	Assumptions
Vendor Agreement	Approximately 5 to 6 weeks before initial data transfer	Translational Research/Clinical	Clinical leadership at the Therapeutic Area level, Study Responsible Physician, Clinical leader.	Clinical leadership at the TA level must approve as part of the governance process for concept and funding
DTA Creation	Usually, a week before initial data transfer. (a week)	Data Manager (DM)	Clinical/Statistician/Programming leads	If no complex requirement like where to upload data etc...
Protocol/SAP/DPS	Usually, signed and finalized before the first data transfer	Clinical/Statistician/Programming lead	Clinical/Medical Writer/Translational research/ Programming leads	In parallel with above task

Timeline Estimation (cont.)



Task	Timeline	Owner	Reviewer(s)	Assumptions
Data Transfer (First)	As per DTA, usually within a week from final DTA	Vendor POC/Data Manager	Clinical/Statistician /Programming leads	Data Manager process and load the data
DTA & source data alignment, Data issues checks	Within 2 weeks from first transfer	Programming lead	Clinical/Statistician	This includes data integrity, traceability, list of source datasets, variables, parameters etc...
Statistician independent Analysis datasets & reports	Within 4-5 weeks from SAP	Statistician	Clinical/Programming leads	
Data Model Design (DMD) Metadata creation (if required)	3-4 weeks from first transfer	Programming lead	Independent programming checks/Statistician review	If SDTM required, only required domains are created (10 to 13 core domains, 8 to 10 supplement domains), 100% CDSIC compliance is not feasible 3-4 FT programming resource
Data Model Design (DMD) Dataset creation/validation (if required)	4-5 weeks from Metadata	Programming lead	Risk based Independent programming checks/Statistician review	

Timeline Estimation (cont.)



Task	Timeline	Owner	Reviewer(s)	Assumptions
Data Model Design (DMD) e-sub preparation (if required)	2 weeks	Programming lead	Independent programming Checks/Statistician review	eCTD package preparation
ADaM Metadata/Programming/Validation	within 2-3 weeks from DPS	Programming lead	Independent programming review/Statistician review	3-4 ADaM datasets, 3 FT programming resource, (ADaMs can be started once the first draft of SDTMs are ready, when SDTMs are handled by the independent team)
TLGs Programming/Validation	2-3 weeks from ADaM	Programming lead	Independent programming review/Statistician/Clinical/Medical Writer review	35-40 reports (6 to 7 unique layouts) 3 FT programming resource (TLGs can be started once the first draft of ADaMs are ready)
ADaM e-sub preparation (if required)	2 weeks from TLGs	Programming lead	Independent programming review/Statistician review	eCTD package preparation (p21 checks, Define.xml, ADRG, Supplemental data definition)

Timeline Estimation (cont.)

- In summary, without source standard data (DMD), it can be around 100 days and with source standard data (DMD) 136 days
- Multiple reruns are required, when source data are transferred in batches
 - SDTM rerun/revalidation requires 5-6 days, when there is no structural changes in source data.
 - ADaM/TLGs rerun/revalidation requires 3-4 days, when there is no structural changes in source data



Challenges / Workarounds

Unlike Clinical trials, data from real-world data sources are not robust as there are no established standards for data collection, transformation, cleaning process, etc... so following challenges are very common.

Involvement of Stats and Programming at the initial planning stage is very vital to overcome these Challenges.

Challenges		Workarounds
Non normalized database -Redundant information -Inconsistent information -Not following BDS structure for vertical data - etc.. Nonstandard database (Non CDISC)		Ask for data in a standard data model: -FHIR V4.0.1 -Sentinel CDM V6.0.2 -PCORNET CDM V4 -OMOP CDM 5.2 -SDTM Note: Probability of getting SDTM format is less

Challenges / Workarounds(cont.)

Challenges		Workarounds
Source file type -Non-SAS files -SAS incompatible variable names, special characters like nonprintable character		At vendor agreement level, demand SAS V5 transport compliance source files
Data integrity/traceability issues - Variables without its dependent source - example a data vendor transferred survival data with direct OS durations without respective death date, last known date, diagnosis date, treatment start date		Detailed review of Data Transfer Agreement are highly required at the initial stage. i.e., verify and make sure all dependent and required variables are properly defined at Agreement/DTA level.

Challenges / Workarounds(cont.)

Challenges	Workarounds
Imbalance comparison groups - Number of expected subjects can be very less for a few groups, so drawing clinically/statistically meaningful results can be non feasible	Frequency distribution of a few key variables are recommended before the vendor agreement approval/signoff
Pseudo-anonymous date shift - To preserve data privacy and patient identity, a few vendors are encrypting the source date variables by using pseudo-anonymous shift and these shifts can lead to future dates, which can raise data quality related questions example date of death =10JAN2030	On or before collaborative/vendor agreement approval, review data privacy terms and date handling methods, get them fixed if anything inappropriate for the analysis scope



Challenges / Workarounds(cont.)

Challenges		Workarounds
Clinical Study data and RWE data pooling		Pooling source database is not easy and appropriate approach ADaM pooling is more convenient approach. Note: Proactively plan & prepare all RWE ADaM structure, Metadata/attributes consistent with respective clinical study.
Partial or batch wise data transfer - A few data vendors can transfer data in batches - Sometime noncumulative datasets, this can lead to extra programming and validation task		Retrospective data collection process can be ongoing, so we can't avoid batch wise data transfer, but it's good to accept cumulative/consolidated database for every transfers. Note: If needed dummy data can be created to overcome data transfer delays

Challenges / Workarounds(cont.)

Challenges	Workarounds
Data issues/resolution - Inconsistent data points - Redundant data points - Partial dates - etc. 	- Data issues are very common - Design a few simple data checks/edit checks programs for all analysis related source variables. - Communicate vendor(s) on all finding/data issues on a timely manner - Have regular meeting with data vendor(s) - Plan one or two dry runs before the final data transfer., this can avoid last minute hiccups - When data issues cannot be resolved, add some data handling rules
Protocol/SAP/DPS amendments - Data handling and analysis sections are created using DTA	- Generalized analysis section - Define alternate analysis methods/approach - Example: Survival data with left truncation bias issue

Questions

