

Real World Evidence at Janssen

Lessons Learned using RWE for Breakthrough Designation NDA
in 2018

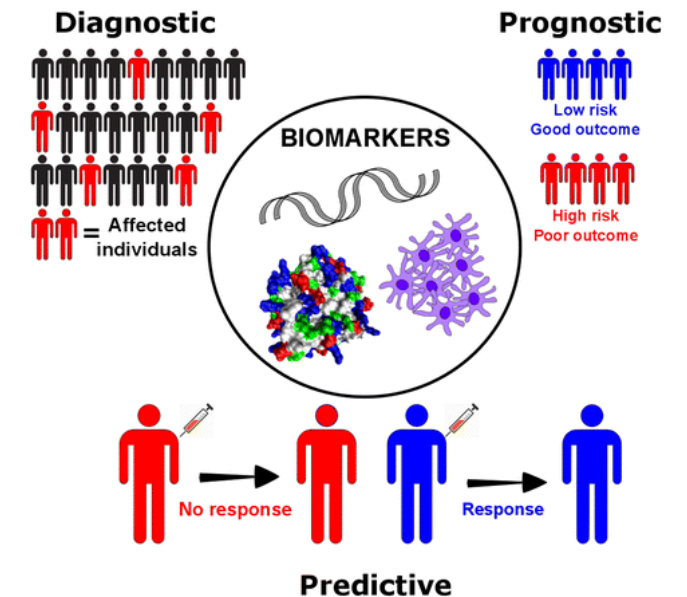
Nicole Thorne, Kathy Wu, Ramesh Sundaram

RWE at Janssen Overview

- Why Real World Evidence was used in this NDA
- Initial Presentation to FDA and request for data
- Submitting RWE data in CDISC format
- Solutions Explored
 - Mapping to SDTM
 - ADaM Creation, Structure, and Submission
- Overview of ADaM structure
- FDA questions and responses
- FDA Technical Walkthrough

Why Real World Evidence was used in this NDA

- Current standard of care therapy for the disease under study is shown to have a low response for subjects with the specific tumor biomarker alterations we are targeting with our drug
- Using a real world matched **clinical and genomic dataset** of patients with the disease under study we explored the predictive value of this specific tumor biomarker alteration to current standard of care therapy
- All real world data were derived from a vendor, containing full clinical and treatment histories and molecular data on a cohort of patients with the disease under study. Clinical data were captured from **electronic health records** by the RWE vendor.



Initial Presentation to FDA and request for data

- At the pre-NDA meeting results of this real world analyses were presented and decided to be submitted as a white paper included in the submission
- **CDISC Format** - FDA requested that the data to support this analyses be included in CDISC format
- Discussions began on whether to send ADaM, SDTM or both and an effort was started to map the RWE data to SDTM and ensure the analysis datasets were ADaM compliant



- **Data Privacy/FOIA** - *Concerns arose over disclosure of patient-level data as part of FOIA*
- Team discussed potential risk of re-identification associated with the submission of RWE to FDA and if FDA would publicly disclose RWE information in the FDA review post approval summary documents. RWE vendor highlighted the issue and the potential de-identification risk during preparation of aggregate and patient level data submission and contract negotiations with Janssen.

RWE data processing



Real World
Analytics



**Step1 – Data
Transfer Agreement
What do we need?**

**Step2 –
Agree on
data format**

**Step3 – Determine
Feasibility of the
analysis and create
an analysis plan**

**Step4 – Map
data to a
consistent
format**

**Step5 – Run the
analysis**



SDTM or ADaM?

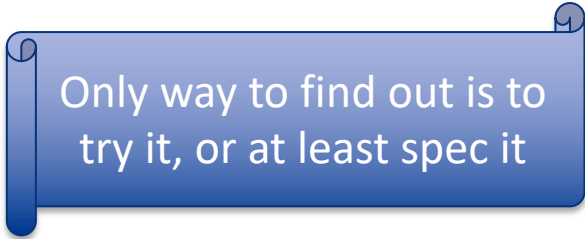
- Why not do both?



Ok, lets try!



- Can we feasibly map to SDTM?



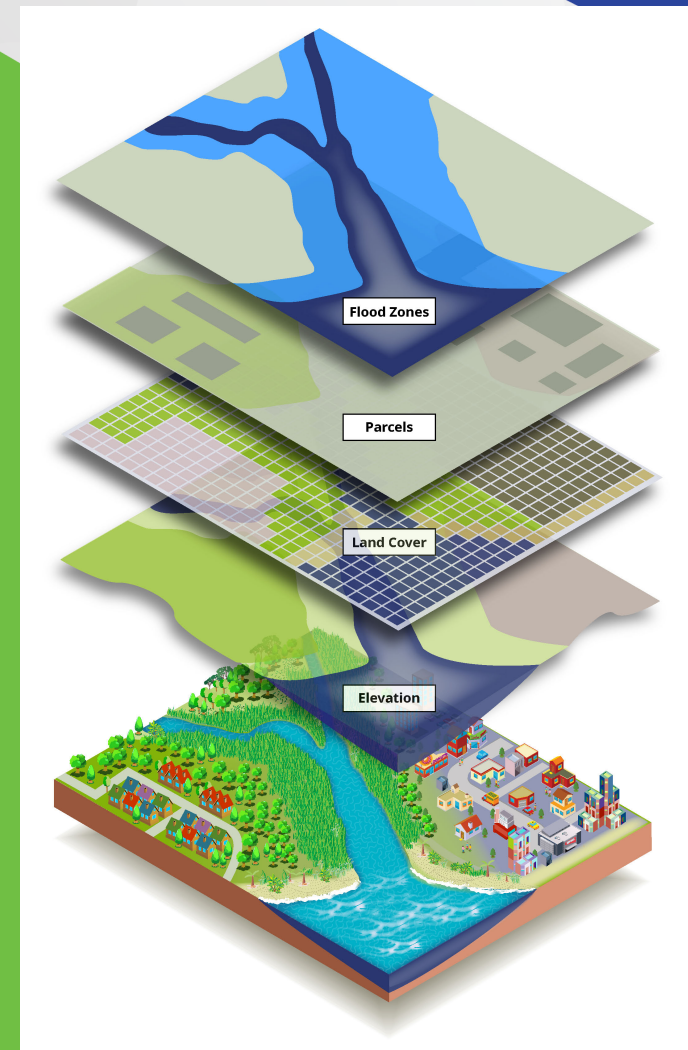
Only way to find out is to try it, or at least spec it

- What do we put in the ADaM datasets?



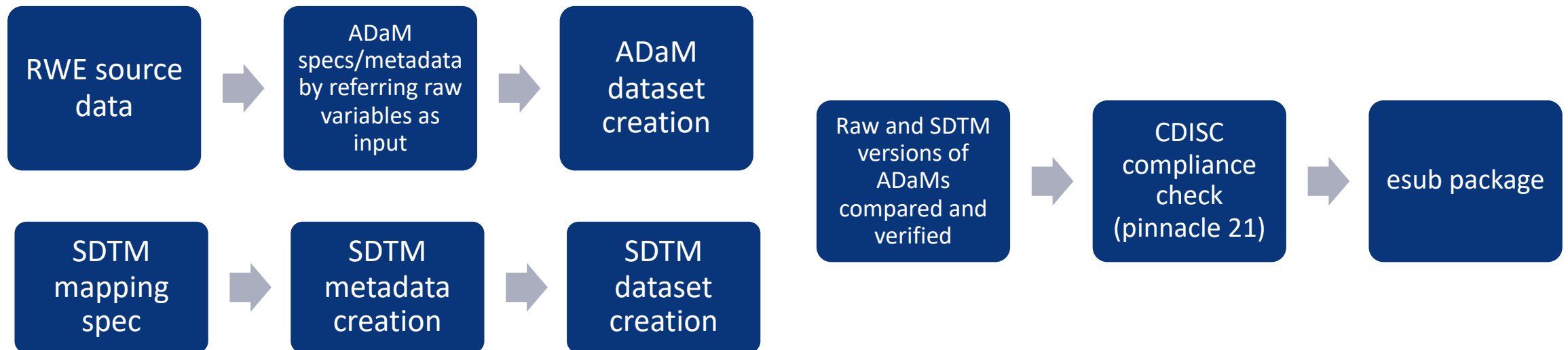
Make sure source data is included

CDISC Data Mapping



Overview of our CDISC formatting process

- To overcome delays due to data issues, we generated both SDTM conversion and ADaM creations in parallel
- ADaM are directly created from source RAW datasets and but later regenerated using mapped SDTM datasets and validated/verified against raw ADaM version.



What is in the data?

Table of content lists out files we would receive, variables in files and expected values: This normally starts the bilateral communication.

- 6 File description.....
 - 6.1 Disease History: DISHIST.....
 - 6.2 Demographics: DM_GL_900.....
 - 6.3 ECOG Performance Status and HB Level: ECOG_HB.....
 - 6.4 Molecular alteration data (FGFR gene/other molecule): FGFR
 - 6.5 PD-L1 expression by immunohistochemistry (IHC): PD_L1
 - 6.6 Systemic Therapy: SYST_TH
 - 6.7 Surgery: SURG
 - 6.8 Survival Status: SURV_STAT

Check out what really are in the source data

1.)

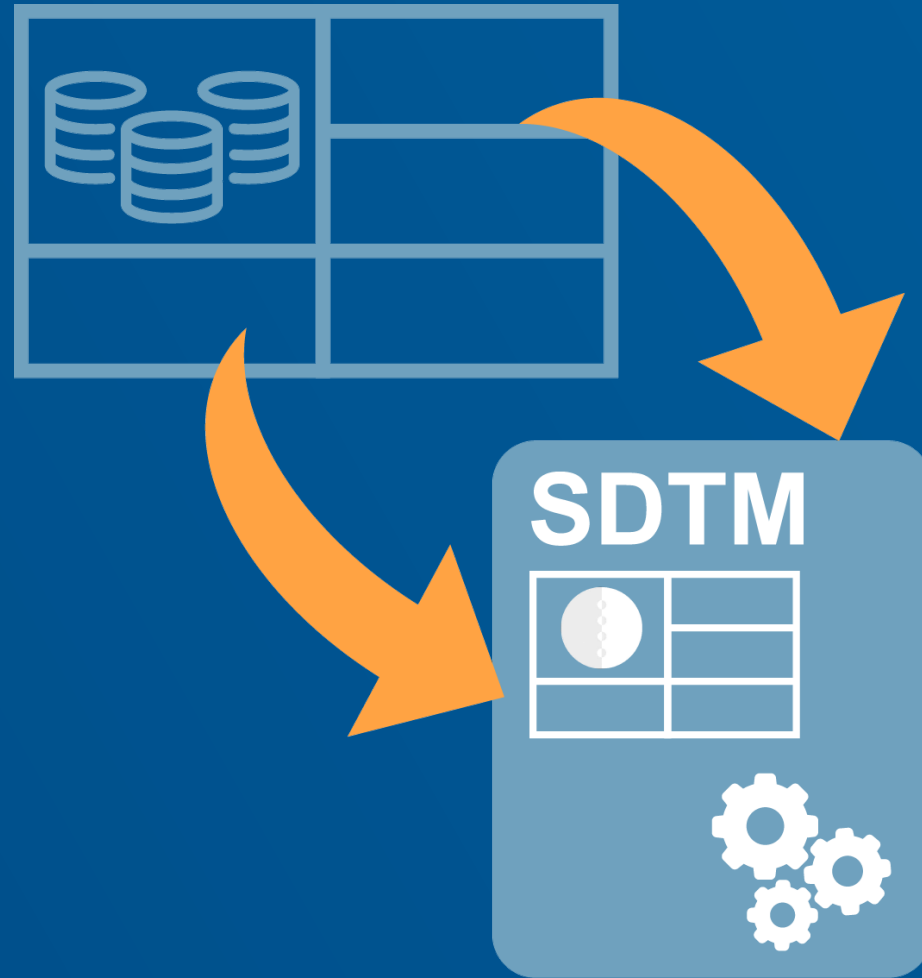
We have missing treatment start date while treatment is there. For Example below Snapshot is for Line 1 Treatment and Start date and end date, where Treatment start date is missing and we have same issue with Line 2-6 Treatment dates. How to handle missing start dates for treatment.

Can you we get complete date for all missing line start and end dates ?

	SIC	SYST_TH_LN1	SYST_TH_LN1_START	SYST_TH_LN1_END	
	LuNa8QPy	Gem_cis			C
	iQP5qy24	Other			
	U1MJ2w5p	Other			C
	gxn1oCvz	Other			
	5NldGIWH	Other			
	9UDhVjOi	Other			
	KEJqoK2o	Other			V

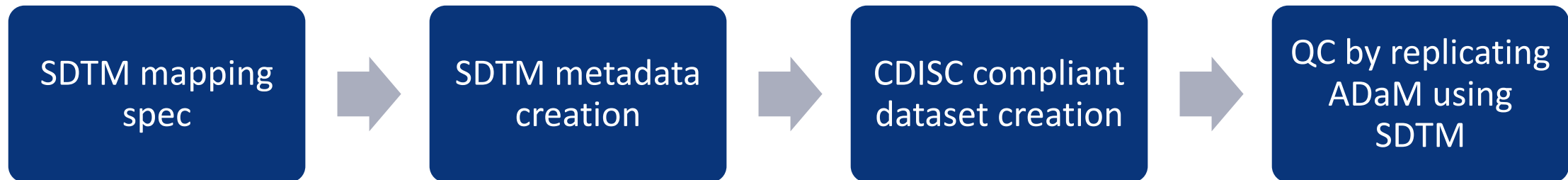
Resolved	Issue1	Issue2
No	The value in SYST_TH_OTH is "Pembrolizumab" but SYST_TH_IO_SPEC is "Nivolumab, Pembrolizumab". Is this patient took both drug or only Pembrolizumab?	Need valid date in SYST_TH_OTH_START and SYST_TH_OTH_END
No	The value in SYST_TH_OTH is "Nivolumab" but SYST_TH_IO_SPEC is "Nivolumab, Pembrolizumab". Is this patient took both drug or only Pembrolizumab?	Need valid date in SYST_TH_OTH_START and SYST_TH_OTH_END

SDTM Mapping



SDTM mapping how we tackled it

- After in-depth review of DTA and source data, Janssen SDTM domain specialist would provide one-to-one mapping specs, variable to variable, dataset to dataset;
 - First we create metadata from nothing.
 - For simple datasets with easy to understand variables, mapping may only involve labels change, variable length reset, ...
 - For complex dataset like 'pd_l1'(FGFR info), we need to reference to analysis dataset and figure out which SDTM dataset is a good fit (MI/SUPPMI).
 - Sometimes a RAW dataset can be mapped to multiple SDTM datasets like 'dishist' to MH/DM/FA.
 - It takes more time to link related domains like MH and FA via RELREC.



Concomitant Medication mapping challenge

The systemic therapy data is nonstandard and non-normalized i.e. it had redundant and inconsistent formats.

It had line wise information in horizontal structure with so many variables(i.e. line 1 name, line 1 start date, line 1 end date, line 2 name, line 2 end date, etc..) and don't have a separate placeholder variable for each line other category. The following example explains how multiple variables are mapped to CMGRIP;

The patient has "Other" for Line 2, 3, 5 and 6. In SYST_TH_OTH, though there are 9 lines of treatments. Since most subjects have up to 6 lines of therapy, we keep 6 lines. But line7 is an immune therapy drug, we decided to keep it as well.

SYST_TH_OTH

THLN2: RCT 54 Gy/6 mal Cisplatin;
THLN 3 RCT 45 Gy + Cisplatin;
THLN5: Gemcitabine/Trastuzumab;
THLN 6 Trastuzumab;
THLN7 Atezolizumab;
THLN 8 Gem/Cis;
THLN 9 Gem

	DOMAIN	USUBJID	CMSEQ	CMGRPID	CMTRT	CMSTDTC	CMENDTC
4	CM	Bridge-C7cLCETy	1	1ST LINE OF TREATMENT	Gem_cis	2010-02-23	2010-08-08
5	CM	Bridge-C7cLCETy	2	2ND LINE OF TREATMENT	RCT 54 Gy/6 mal Cisplatin	2010-11-26	2011-01-13
5	CM	Bridge-C7cLCETy	3	3RD LINE OF TREATMENT	RCT 45 Gy + Cisplatin	2013-12-16	2014-01-27
7	CM	Bridge-C7cLCETy	4	4TH LINE OF TREATMENT	Vinflunine	2014-07-09	2014-07-30
3	CM	Bridge-C7cLCETy	5	5TH LINE OF TREATMENT	Gemcitabine/Trastuzumab	2014-10-22	2016-09-28
9	CM	Bridge-C7cLCETy	6	6TH LINE OF TREATMENT	Trastuzumab	2016-11-25	2017-05-03
0	CM	Bridge-C7cLCETy	7	7TH LINE OF TREATMENT	PDL1 Inhibitor (Atezolizumab)	2017-05-08	2017-07-17

SDTM Pinnacle 21 Enterprise checking

In between, we run P21E, address WARNING, ERRORS, update SDTM dataset, till P21E runs clean.

Example in P21E:

1. In FA, FAOBJ is null of the records where FATESTCD="HISTSU", FATESTCD="Substance Use History". While FAOBJ should not be null.

So can we set FAOBJ = "SUBSTANCE USE" for these records (FATESTCD="HISTSU", FATESTCD="Substance Use History")?

FA	SD0002	NULL value in FAOBJ variable marked as Required	Error	Reject	72
----	--------	---	-------	--------	----

Expert response: Please assign value to **FAOBJ=SMOKING** when FATESTCD=HISTSU. FAOBJ cannot be indeed empty.

2. In PR, there are records PRPRES="" and PROCCUR="Y". While Occurrence (--OCCUR) may only be populated, when a given Intervention or Event has been pre-specified (--PRES = 'Y').

So can we set to PROCCUR = "" for records where PRPRES="" ?

PR	SD0041	Value for PROCCUR is populated for unsolicited Intervention or Event	Error	Warning	24
----	--------	--	-------	---------	----

Expert response: If PROCCUR is populated then PRPRES=Y, per mapping spec it if for 3 sources variables (SURG_TUR, SURG_CX, SURG_PE). So not sure what is the source data of the issue listed but I guess it is not 'pre-specified' records per the mapping spec. So, yes PROCCUR for those records should be empty.



SDTM mapping lessons learned

- SDTM mapping efforts for RWE are hard **and tedious and required extensive metadata** and specs to go with the source data;
- Issues like a lot of raw variables described in DTA document are misspelled, variables with **NO proper labels**, making it difficult to be sure of their meaning and hence difficult to map.
- Understanding **how data were collected** from patient charts to fully align with vendor's database documentation before we make decision what raw maps to which SDTM.
- Understanding medication start dates (e.g. how these were derived from patient charts), the **algorithm** for response and other fields **was not well understood** up front when the data was received.
- **Compliance checks = significant errors** because much of the information required in SDTM is just not available.
- This required **multiple phone calls** with the vendor, and explanation of the dataset to the FDA required provision of some of vendor's database documentation.

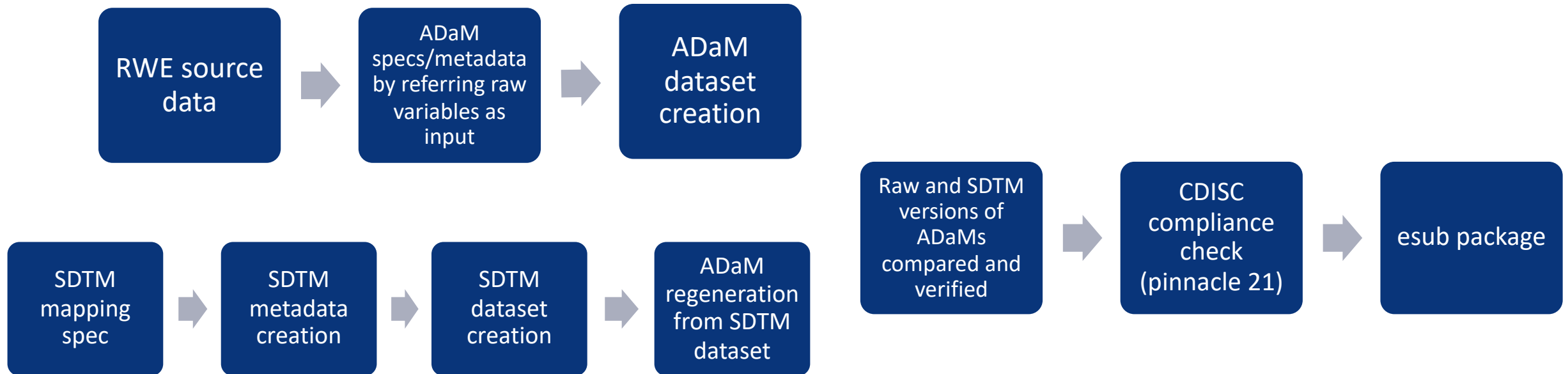


ADaM Creation, Structure, and Submission



ADaM CREATION PROCESS

- ADaM are directly created from source RAW datasets and later regenerated using mapped SDTM datasets and validated/verified against raw ADaM version.



ADaM CREATION Lessons Learned



LESSONS LEARNED

- Above parallel approach was lengthy, redundant and it took more resources and time, but this was a best feasible solution at that point.
- Had multiple meetings with our vendor to resolve data issues even though, we left with hand full of partial dates and inconsistent data points, so team decided to handle them at analysis level.
 - Example day and month of birth dates are imputed, and specimen reported dates are proposed as entry timepoint to handle the time lag issue by using a delayed entry model (left truncation model).

STRATEGY FOR ADAM DESIGN AND LIST OF ADS

Tier 1 Datasets (Source & Derived)

ADSL
Subject Level
Analysis Dataset

Contains demographics and baseline disease charac. , population variables and key date variables for all subjects in the raw demo dataset

ADBIO
Biomarkers
Analysis Dataset

Contains gene-fusion and mutation params

ADCM
Concomitant
Medications

Contains all prior meds and con meds from source med datasets w/Prior and Con med flag vars and flag vars for IO/Chemo mono and combo treatment received.

Contains best overall response parameters for IO treated subjects, because source data captured only IO associated response.

Tier 2 Datasets (Derived from Tier 1)

ADTTE
Time-to-Event
Analysis Dataset

contains OS related endpoint params

ADEFF
Efficacy Analysis
Dataset

GUIDING PRINCIPLES FOR ADaM DESIGN AND CHALLENGES FACED

- **Data integrity, traceability, and dependency** links related issues are very common in retrospective data collection models, but these are all the **fundamental principles of submission data**.
- Carefully verified these key principles and the data points, which are not meeting these key principles were escalated and resolved before analysis activity.
- **Unresolved data points are excluded from all our analysis.**
- Few examples of data integrity, traceability and dependency related issues.
 - Many survival duration data points were transferred without respective treatment start and end date (death or last known date).
 - Systemic line 1 treatment end date after line 2 start date
 - Response and response date were transferred without treatment association and few response dates were before the treatment start date

GUIDING PRINCIPLES FOR ADaM DESIGN AND CHALLENGES FACED

- Took more time and resource to understand, convert and impute the nonstandard data points to standard ADaM structure. Biostat and programming team are **worked very closely to design and implement imputation, derivation algorithms.**
- Due to the limitation of vendor's data disclosure agreement, we are directed to include only analysis required and respective dependent information in ADaMs, so cross functional team **carefully reviewed and excluded all non-required data points** from ADaMs
- Pooled analysis of **core clinical study** and RWE was well planned hence the RWE team decided to **adopt the same AD structures, attributes** (variables, labels), parameters and it's naming conventions, which was followed in the core clinical study, so the data pooling activity was easy and fast.

GUIDING PRINCIPLES FOR ADaM DESIGN AND CHALLENGES FACED

- Due to this approach we **reused wide range of reporting macros and programs** for pooled analysis. As planned, required RWE ADaMs datasets (ADSL, ADTTE, ADEFF) and parameters were directly pooled with core clinical study to compare the Overall survival and object response rate of clinical study drug with RWE drugs.



FDA Drug Approval

Regulatory Process



FDA Information Requests

1

Multiple RWE Data Sources

FDA requested that data from multiple RWE data sources be submitted, where available

2

Response Data

FDA requested response data be provided. Janssen stressed response by RECIST not typically collected outside of clinical trials. RWE vendor developed Real-World Tumor Response (rwTR).

3

Patient Narratives, Radiology Scans and Radiology Reports

Individual patient narratives requested; names of antineoplastic therapy; radiological assessment dates with reports (including baseline); and clinician assessment/plan. Radiology scans were also requested, but were not available for provision to FDA.

4

CDISC Formatting and analysis programs

FDA requested CDISC formatted data and analysis programs

5

SAP Timelines

Assess timing of statistical analysis plan in relation to initiation of data analyses to ensure analyses were pre-planned and not post-hoc

FDA Technical Walkthrough Questions on the RWE data

FDA: In RWE, X patients have been treated, why Y in the analysis?

Janssen: Reason is Patient flags to identify this population are located in the ADSL.xpt dataset. These are referenced in the ADRG.pdf document in 5.3.5.4 predictive value report which was previewed during the meeting with the FDA review team and shown below.

FDA: How reliable is the diagnosis information and line of therapy in the Electronic Health Records from the RWE vendor:

Janssen: EHR recorded diagnosis dates and all therapies received. These captured both structured records in database and unstructured records from free text clinical notes, reviewer and abstractor reviewed these data and consolidate them into the database.

FDA: Are RWE patients representative of general disease under study patients?

Janssen: The demographics seem to suggest they reflect the general disease under study patients. Note these are primarily patients from community practice and less reflective of clinical trial population.

FDA: Are there any of the disease under study patients in the RWE database excluded due to lack of determination of genomic data.

Janssen: Possibly. Since the genetic status is a key analysis variable, we did not obtain any data from patients who did not have genomic results.



Conclusion

- **SDTM mapping** efforts for RWE were **hard and tedious** and required extensive metadata and specs to go with the source data – **not worth it!**
- Only one RWE source had sufficient information to correctly map the data and the company we bought it from was **uneasy about sending all the data** to the FDA, they were more comfortable with **only sending what was used** in analyses
- Compliance checks on the **SDTM had significant errors** because much of the information required in SDTM is just not available
- **ADaM datasets were eventually the only data submitted** and all the source data used in the analyses were included in this package. E.g. the tier 1 datasets were the source data mapped to a consistent format (ADCM, ADBIO, ADSL) and the derived endpoints in tier 2 (ADEFF, ADTTE)
- Lessons Learned:
 - We **need an easier format than SDTM** to submit source RWE data, data does not conform to any traditional codelists and many required variables are not possible to create
 - We need to **ensure de-identification prior to use** in the analysis
 - We need **better information on how the data is collected** from the patient charts in order to analyze and document it properly and that information is hard to come by

