



Incorporating Real-World Data in a Low Interventional Study: Challenges and Considerations

Sandra VanPelt Nguyen



The Clinical Data
Science Conference

US 2024

25-28 February

Bethesda North Marriott Hotel
& Conference Center



Breakthroughs that change patients' lives



Abstract

This presentation will share an example of a low interventional study that incorporates real-world data along with conventionally collected clinical data to demonstrate real-world evidence of vaccine efficacy. It will explain the use of tokenization to link RWD from external data sources with the clinical database while maintaining patient privacy, how the data was mapped to SDTM, and challenges and considerations for mapping and analyzing the data.

Study Overview

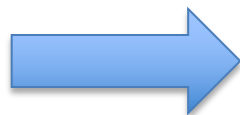
- Post approval evaluation of vaccine efficacy
- Low interventional
- Test-negative design
- Real world setting



Study Overview

conducted at investigator hospital sites

Adults ≥ 65 years of age admitted for hospitalization at a study hospital with signs, symptoms, and radiologic evidence of disease



Visit 1 Screening/Enrollment	Visit 2 Final Assessment
Day 1	Day 30
Participant Interview / Medical Record Review: demographics, medical history, substance use, risk factors, vaccination history	Hospitalization details
Clinical Assessments: vital signs, signs and symptoms, labs	Final diagnosis
Review of imaging, microbiologic test results performed as SoC	Discharge disposition
Specimen collection for viral detection	Mortality/vital status

Why Incorporate RWD?

Success of the study depends largely on adequate uptake and use of the vaccine in adults ≥ 65 years of age, accrual of cases needed to evaluate the primary endpoint, **and identification of documented vaccination history.**

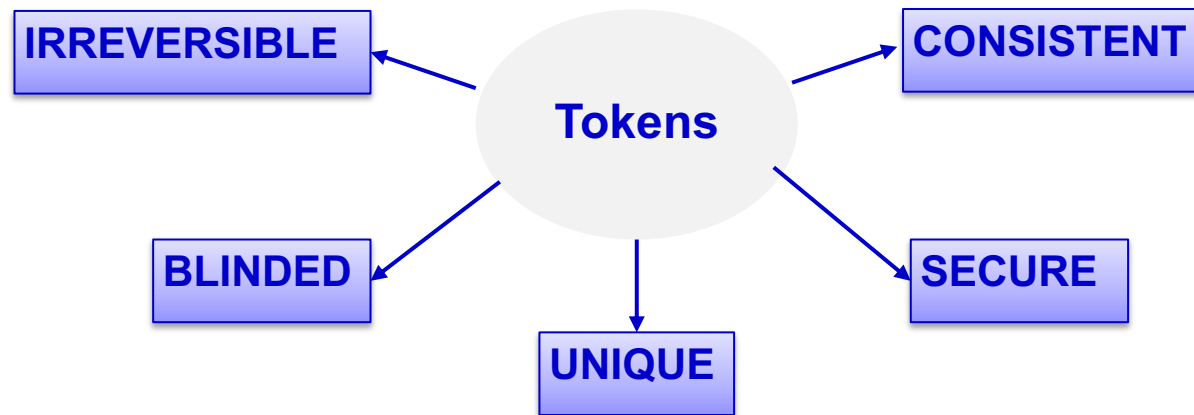
Determining vaccine efficacy (VE) requires access to complete and accurate vaccination history on all participants.

Enrolled participants without a documented vaccination history will be excluded from VE analyses.



Tokenization

- Process by which patient identifiers are de-identified through generation of a patient-specific 'token' that is encrypted
- Helps to link RWD from a patient's previous medical history from diverse sources without any breach in the patient's privacy
- All the sensitive data that can compromise a patient's identity are replaced with unique identification symbols (tokens) that retain all the essential information, but without any compromises in the confidentiality



Example Tokenization Data Flow

Tokenization

Personal Information

*First Name John
*Last Name Doe
*DOB 1/1/1970
*Gender M
Address 123 Main St
*Zip 90210

*Required Fields

Types of Token Data/
Data Tokenized
(Token 1, 2, 3, A, B, C)

Linked Data



or

Token data direct to RWE Provider

Matching & Linking



Rx Claims



Medical Claims



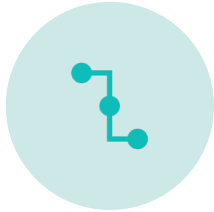
EMR

Submitting RWD

- “Currently, and absent a waiver, sponsors submitting clinical and nonclinical study data (including those derived from RWD sources) in submissions subject to section 745A(a) of the FD&C Act are required to use the formats described in the Study Data Guidance and the supported study data standards listed in the Catalog.”
- “With adequate documentation of the conformance methods used and their rationale, study data derived from RWD can be transformed into SDTM and ADaM datasets and submitted to FDA in an applicable submission.”

Reference: US Department of Health and Human Services, Food and Drug Administration. *Data Standards for Drug and Biological Product Submissions Containing Real-World Data: Guidance for Industry*. December 2023. Available at: <https://www.fda.gov/media/153341/download>

Mapping and Analysis Challenges



Incoming data structure is not well-aligned with CDISC standards



Additional operational data, including tokens, does not fit well into SDTM



Distinguishing between site-reported data and RWD



Same vaccine found more than once



Handling of “not found” RWD records



“Collisions” and “splits”

Example RWD File

*not all fields shown

Row #	HASHED_SUBJECT_ID	PRODUCT_NAME	ADMIN_DT	CODE_TYPE	CODE
1	8ir5wljBi3Xc9/qoGWARXiAl605KE3esC+jBZRfXcM=	Vaccine ABC	2023-01-01	NDC	49281012265
2	8ir5wljBi3Xc9/qoGWARXiAl605KE3esC+jBZRfXcM=	Vaccine ABC	2023-01-01	NDC	49281012265
3	8ir5wljBi3Xc9/qoGWARXiAl605KE3esC+jBZRfXcM=	Vaccine ABC	2023-01-01	CPT	90732
4	8ir5wljBi3Xc9/qoGWARXiAl605KE3esC+jBZRfXcM=	Vaccine ABC	2023-01-01	CPT	90732
5	8ir5wljBi3Xc9/qoGWARXiAl605KE3esC+jBZRfXcM=	Vaccine XYZ		N/A	N/A
6	8ir5wljBi3Xc9/qoGWARXiAl605KE3esC+jBZRfXcM=	Vaccine XYZ		N/A	N/A
7	jRyalli9NjCt1aVQZ+z/HrggBlvNke1iFm8Ge2FdZpA=	Vaccine ABC		N/A	N/A
8	jRyalli9NjCt1aVQZ+z/HrggBlvNke1iFm8Ge2FdZpA=	Vaccine ABC		N/A	N/A
9	jRyalli9NjCt1aVQZ+z/HrggBlvNke1iFm8Ge2FdZpA=	Vaccine XYZ		N/A	N/A
10	jRyalli9NjCt1aVQZ+z/HrggBlvNke1iFm8Ge2FdZpA=	Vaccine XYZ		N/A	N/A
11	H9DtCKRiQeelF/bZBbpto1oMhXm7uXW1z73O3OZu6DA=	Vaccine XYZ	2023-10-12	CPT	90732
12	H9DtCKRiQeelF/bZBbpto1oMhXm7uXW1z73O3OZu6DA=	Vaccine XYZ	2023-10-12	CPT	90732

Row #	TKN_NUM	TKN_VALUE	FOUND_FLAG	NF_REASON
1	TOKEN1	ZMFfB8B7GFqEqNksW8rsrluPCAWPBjmYxtUNbyTWGT0=	Y	
2	TOKEN2	LPhxlCrJuURIDKiJeR6JmdAde+RMA8JQ9RmRc4HEFCE=	Y	
3	TOKEN1	ZMFfB8B7GFqEqNksW8rsrluPCAWPBjmYxtUNbyTWGT0=	Y	
4	TOKEN2	LPhxlCrJuURIDKiJeR6JmdAde+RMA8JQ9RmRc4HEFCE=	Y	
5	TOKEN1	ZMFfB8B7GFqEqNksW8rsrluPCAWPBjmYxtUNbyTWGT0=	N	Token not found
6	TOKEN2	LPhxlCrJuURIDKiJeR6JmdAde+RMA8JQ9RmRc4HEFCE=	N	Token not found
7	TOKEN1	01u5ezuRo1y9/Y+wxRKE6ZVwgg3OjBsWtJ6EmZm/gRo==	N	Data excluded due to collision
8	TOKEN2	vuZKYQqOBw+KMc1kTct4vTh3x8LkNi2LgOF96ep0xzM=	N	Data excluded due to collision
9	TOKEN1	01u5ezuRo1y9/Y+wxRKE6ZVwgg3OjBsWtJ6EmZm/gRo==	N	Token was found but the specified data was not present within the requested time period
10	TOKEN2	vuZKYQqOBw+KMc1kTct4vTh3x8LkNi2LgOF96ep0xzM=	N	Token was found but the specified data was not present within the requested time period
11	TOKEN1	yZIPqWd/5oDplwxrtpNltP2OEjKJ9QpAuV1btRU1+ns=	Y	
12	TOKEN2	k6XOA2uJvXgVXamhRjyk1oX0WT7XT9kBLQDgfe6XH/g=	Y	

SDTM Example 1 (Site-collected)

*not all fields shown

Row	USUBJID	CMTRT	CMDECOD	CMPRESP	CMOCCUR	CMSTDTC	CMEVLINT	SUPPCM. CMSOINF
1	1111 1001	ABCVAX	Vaccine ABC	Y	Y	2023-01	-P5Y	SELF- REPORT
2	1111 1001	ABCVAX	Vaccine ABC	Y	Y	2024-01-07	-P5Y	PHARMACY RECORD
3	1111 1001	XYZVAX	Vaccine XYZ	Y	N		-P5Y	
4	1111 1002	ABCVAX	Vaccine ABC	Y	N		-P5Y	
5	1111 1002	XYZVAX	Vaccine XYZ	Y	N		-P5Y	
6	1111 1003	XYZVAX	Vaccine XYZ	Y	N		-P5Y	

SDTM Example 2 (RWD)

*not all fields shown

Row	STUDYID	DOMAIN	USUBJID	CMTRT	CMDECOD	CMCAT	CMRESP	CMOCCUR
1	ABC001	CM	1111 1001	ABCVAX	Vaccine ABC	VACCINATIONS	Y	Y
2	ABC001	CM	1111 1001	ABCVAX	Vaccine ABC	VACCINATIONS	Y	Y

Row	CMSTAT	CMREASND	CMSTDTC	CMEVLINT	SUPPCM. CMSOINF	SUPPCM. RWDFL	SUPPCM. TKNCDTYP
1			2023-01-02	-P5Y		Y	NDC
2			2023-01-02	-P5Y		Y	CPT

SDTM Example 3 (RWD)

*not all fields shown

Row	STUDYID	DOMAIN	USUBJID	CMTRT	CMDECOD	CMCAT	CMRESP	CMOCCUR
1	ABC001	CM	1111 1001	XYZVAX	Vaccine XYZ	VACCINATIONS	Y	
2	ABC001	CM	1111 1002	ABCVAX	Vaccine ABC	VACCINATIONS	Y	

Row	CMSTAT	CMREASND	CMSTDTC	CMEVLINT	SUPPCM. CMSOINF	SUPPCM. RWDFL	SUPPCM. TKNCDTYP
1	NOT DONE	Token not found		-P5Y		Y	
2	NOT DONE	Data excluded due to collision		-P5Y		Y	

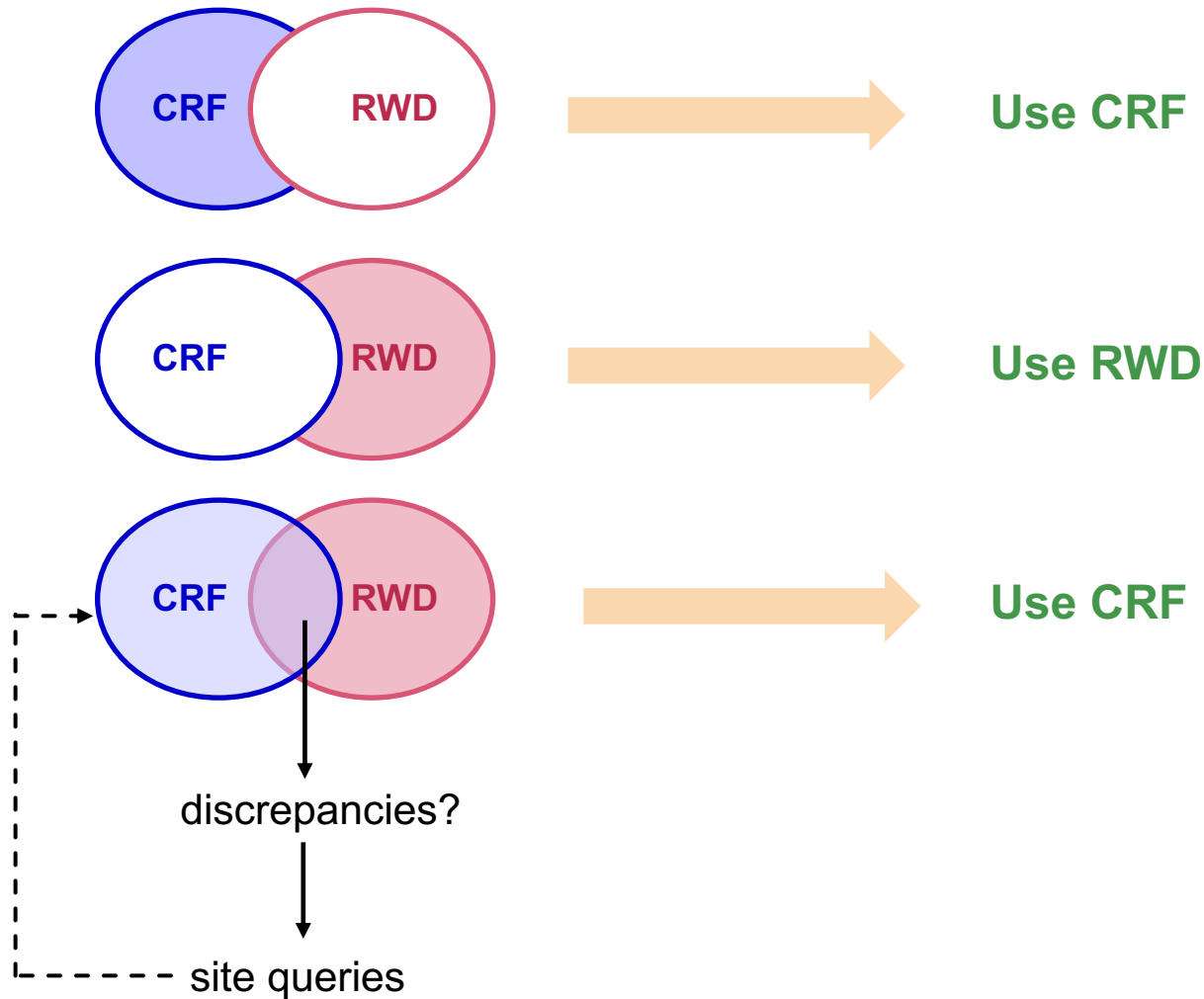
SDTM Example 4 (RWD)

*not all fields shown

Row	STUDYID	DOMAIN	USUBJID	CMTRT	CMDECOD	CMCAT	CMRESP	CMOCCUR
1	ABC001	CM	1111 1002	XYZVAX	Vaccine XYZ	VACCINATIONS	Y	N

Row	CMSTAT	CMREASND	CMSTDTC	CMEVLINT	SUPPCM. CMSOINF	SUPPCM. RWDFL	SUPPCM. TKNCDTYP
1				-P5Y		Y	

Analysis Considerations



ADaM Example - ADCM

*not all fields shown

Row #	STUDYID	USUBJID	SUBJID	SITEID	CMTRT	CMDECOD	CMRESP	CMOCCUR	CMSTAT	CMREASND
1	ABC001	1111 1001	1001	1111	ABCVAX	Vaccine ABC	Y	Y		
2	ABC001	1111 1001	1001	1111	ABCVAX	Vaccine ABC	Y	Y		
3	ABC001	1111 1001	1001	1111	ABCVAX	Vaccine ABC	Y	Y		
4	ABC001	1111 1001	1001	1111	ABCVAX	Vaccine ABC	Y	Y		
5	ABC001	1111 1001	1001	1111	XYZVAX	Vaccine XYZ	Y	N		
6	ABC001	1111 1001	1001	1111	XYZVAX	Vaccine XYZ	Y		NOT DONE	Token not found
7	ABC001	1111 1002	1002	1111	ABCVAX	Vaccine ABC	Y	N		
8	ABC001	1111 1002	1002	1111	ABCVAX	Vaccine ABC	Y		NOT DONE	Data excluded due to collision
9	ABC001	1111 1002	1002	1111	ABCVAX	Vaccine XYZ	Y	N		
10	ABC001	1111 1002	1002	1111	ABCVAX	Vaccine XYZ	Y	N		
11	ABC001	1111 1003	1003	1111	ABCVAX	Vaccine XYZ	Y	N		
12	ABC001	1111 1003	1003	1111	ABCVAX	Vaccine XYZ	Y	Y		

Row #	CMSTDTC	ASTDT	ASTDTF	RWDFL	TKNCDTYP	ANL01FL	ANL02FL	ANL03FL
1	2023-01	01JAN2023	D			Y	Y	
2	2024-01-07	07JAN2024						
3	2023-01-02	02JAN2023		Y	NDC			
4	2023-01-02	02JAN2023		Y	CPT			
5								Y
6				Y				
7						Y	Y	
8				Y				
9								Y
10				Y				
11								
12	2023-10-12	12OCT2023		Y	CPT	Y		Y

ADaM Example - ADSL

*not all fields shown

STUDYID	USUBJID	SUBJID	SITEID	ENRLFL	EVALFL	ANYVXFL	VAX1FL	VAX2FL
ABC001	1111 1001	1001	1111	Y	Y	Y	Y	N
ABC001	1111 1002	1002	1111	Y	Y	N	N	N
ABC001	1111 1003	1003	1111	Y	Y	Y		Y

Summary: Benefits and Challenges of Incorporating Tokenization/RWD

Benefits	Challenges
Helps present complete picture for certain data points of interest	Mapping to CDISC data standards
Can reduce burden on participants and sites	Collapsing and summarizing multiple data sources/matches
Eases privacy concerns	Requires clear understanding of each data field how the data needs to be used for analysis and reporting
Reduce hurdles to obtaining and integrating commercial data	Need to ensure the reliability and relevance of the data
May improve quality and completeness of the data	Needs clear and detailed documentation, assessment and planning



Thank You

For more information, contact:

sandra.vanpeltnghuyen@pfizer.com