

PINNACLE²¹ BY CERTARA

SUBMISSION STANDARDS FOR REAL WORLD DATA: GAPS, LIMITATIONS AND RECOMMENDATIONS

Jeffrey Abolafia, Pinnacle21
Sarah Ferko, IBM Consulting
Ingeborg Holt, IBM Consulting

DISCLAIMER

The views and opinions presented here represent those of the speakers and should not be considered to represent advice or guidance on behalf of Pinnacle 21 or IBM Consulting.

AGENDA

- Introduction
- Background
- Traceability Challenges
- Recommendation
- Conclusions



INTRODUCTION

RWD/RWE DEFINED

- ▶ Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.
- ▶ Real-World Evidence (RWE) is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD



Reference: “Framework for FDA’s Real-World Evidence Program”, December 2018.
<https://www.fda.gov/media/120060/download>

RWD DATA SOURCES

Electronic health
records (EHRs) /
medical records

Medical claims and
billing data

Product, disease, and
population-based
registries

Mobile devices/DHTs
(wearables)

Social media data

**Focus of this
presentation is
Electronic health
records (EHRs) / DHTs**

WHY RWE? RWE TRENDS

- ▶ 95% of approved NDAs and BLAs in 2021 included an RWE study
- ▶ 75% in 2019
- ▶ 58% of approvals in 2021 used RWE to support safety and/or efficacy

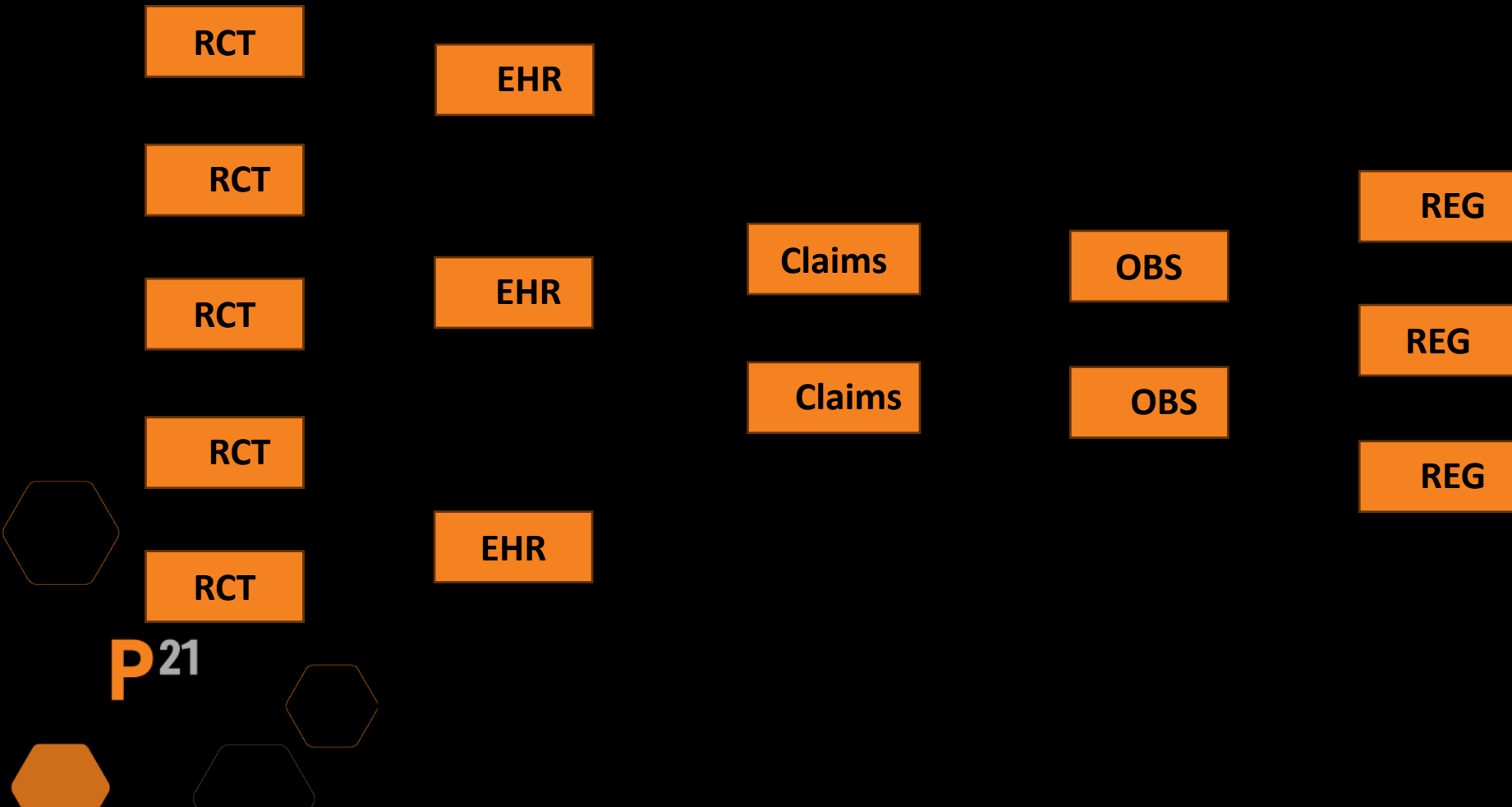
Table 3

Observed intended use of RWE in included NDAs/BLAs, January 2019 to June 2021

Included NDAs and BLAs	2019 <i>n</i> = 51 approvals	2020 <i>n</i> = 59 approvals	2021 through June 30 <i>n</i> = 26 approvals	Total <i>n</i> = 136 approvals
Incorporated RWE for any purpose	38 (75%)	53 (90%)	25 (96%)	116 (85%)
Used RWE to provide therapeutic context	25 (49%)	36 (61%)	22 (85%)	83 (61%)
Used RWE to support safety and/or effectiveness	27 (53%)	46 (78%)	15 (58%)	88 (65%)
Safety only	17 (33%)	21 (36%)	5 (19%)	43 (32%)
Effectiveness only	7 (14%)	6 (10%)	2 (8%)	15 (11%)
Safety and effectiveness	3 (6%)	19 (32%)	8 (31%)	30 (22%)

Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. Clin Pharmacol Ther. 2022 Jan;111(1):135-144. doi: 10.1002/cpt.2474. Epub 2021 Nov 22. PMID: 34726771; PMCID: PMC9299054.

CLINICAL DATA SUBMISSION TRENDS



WHAT CAN WE LEARN FROM PREVIOUS RWE SUBMISSIONS

- ▶ A lot to learn from submitted RWE studies FDA judged as inadequate for regulatory decision making
- ▶ Can be used to avoid common errors that may result in an RWE study being rejected or not considered
- ▶ Can be used to avoid delays in regulatory review of marketing applications containing RWE



BACKGROUND INFORMATION

P²¹

TRACEABILITY DEFINED

“Traceability is built by clearly establishing the path between an element and its immediate predecessor. The full path is traced by going from one element to its predecessors, then on to their predecessors, and so on, back to the SDTM datasets, and ultimately to the data collection instrument.”

TRACEABILITY: REGULATORY NEED

Trace study data from collection > Analysis

“An important component of a regulatory review is an understanding of the **provenance** of the data (e.g., traceability of the sponsor’s results back to the CRF data). Traceability permits an understanding of the relationships between the analysis results (tables, listings and figures in the study report), analysis datasets, tabulation datasets, and source data.”

TRACEABILITY: REGULATORY NEED

“Sponsor should include a data dictionary that documents the definition of every data element used and all relevant information about the element, such as its relationships to other data, origin, usage, and format.”

TRACEABILITY: REGULATORY NEED

- ▶ Many RWD databases submitted are inadequate for regulatory decision making
- ▶ Do not include required documentation and traceability
- ▶ At least the same level of traceability is required for RWD as for RCTs
- ▶ Providing same level of documentation and traceability as in RCTs can avoid
 - ▶ Delays in regulatory review of marketing applications containing RWD
 - ▶ RWE study being rejected or not considered

CURRENT SUBMISSION STANDARDS

- October 2021 FDA draft guidance

"Data Standards for Drug and Biological Product Submissions
Containing Real-World: Data Guidance for Industry"

- Guidance outlined data standards required when submitting RWD in support of a marketing application

- **Currently**

- According to the guidance, RWD must be submitted using the standards documented in the FDA Data Standards Catalog
 - For now, that means **RWD must be submitted using CDISC standards**



P²¹

CHALLENGES

RWD PRESENTS NEW TRACEABILITY CHALLENGES: RCTS VS RWD

Topic	RCT	RWD
Data Collection	Collected under a protocol	Collected in real world settings
Data Monitoring	Data monitored and cleaned	No monitoring or cleaning
Data Entry	Data collected via CRF	Data entered in EHR
Data Source	Data designed and collected by sponsor	Curated data acquired by sponsor from vendor
Availability or Source Data	Owned by sponsor	Owned by HCPs / Aggregators
Documentation	Protocol / CRFs / DMP	Aggregators
Data Uniformity	Uniform data entry across sites	Standards/formats differ by site
Terminologies/Vocabularies	Uniform across sites	May differ by site/type of RWD/region

Note: RCT (Randomized Clinical Trial)

RWD NOT COLLECTED UNDER A PROTOCOL

- ▶ RWD not collected under a protocol or to address a specific research question
- ▶ Protocol provides a blueprint for how research study is conducted
 - ▶ When conducted
 - ▶ Where conducted
 - ▶ How conducted
- ▶ Data Management Plan provides more granular details of data collection
 - ▶ Copy of aCRF
 - ▶ Description of all data elements and their attributes
 - ▶ Edit checks for each data element
 - ▶ Data monitoring and cleaning plan
- ▶ RWD collected in a real world setting and contains none of the above documentation > Not monitored or cleaned

DATA SOURCE/PROVENANCE

- ▶ RCT data comes from systems designed and collected by sponsor (or sponsor agent)
- ▶ Sponsors control entire process of collection > transformation > submission
- ▶ Most RWD not collected by sponsor
 - ▶ Collected in real world settings
 - ▶ Collected in various health settings by various clinical staff
- ▶ Most RWD acquired from vendors
 - ▶ Data curated by vendors
 - ▶ Presents challenges documenting how submitted data was created from source data
- ▶ Sponsors must be able to submit patient level data
 - ▶ Not an issue for RCTs > Sponsors control data
 - ▶ May be an issue for RWD > vendor may only have aggregate data

DATA UNIFORMITY

- ▶ RCT is collected uniformly across sites
 - ▶ Standardized CRF/data elements
 - ▶ Collected by site personnel trained in a uniform manner
 - ▶ Entered in a single/uniform EDC or other data collection system
- ▶ RWD
 - ▶ No expectation of uniformity or standardization across health care systems
 - ▶ “If you have seen 1 EHR system, you have seen 1 EHR system”

TERMINOLOGY

- Clinical research and healthcare terminology lack harmonization
- Many concepts have different names/meaning and different controlled terminology
- RCTs use same terminology and coding systems for all sites and all studies
- RCTs use MedDRA and WHO drug to code adverse events and medications
- RWD tend to use ICD and SNOMED, but that may not be the case for all healthcare EHRs
- Terminology from RWD must be recoded to meet submission required terminology
- Current submission standards (CDISC) do not contain standard data elements to represent both source and submission terminology or coding systems
- This makes source data verification a challenge



RECOMMENDATION FOR SUBMITTING RWD

VERY EARLY COMMUNICATION: FDA AND VENDORS

Challenge

- ▶ RWD not collected under a protocol or to address a specific research question
- ▶ RWD must be fit for purpose and non-bias
 - ▶ Does the RWD appropriately address the study question
 - ▶ Is the population selected in the RWD appropriate to address the study question
- ▶ Quality/provenance
- ▶ Sponsors should be able to submit patient level data for studies used to support a marketing application
- ▶ Missing data
 - ▶ How much
 - ▶ Is missingness random or biased
 - ▶ How is missing data distributed among key variables and covariates

VERY EARLY COMMUNICATION: FDA

Recommendation

- ▶ FDA guidance¹ issued in August 2023 states

“Sponsors planning to use a non-interventional study to support a marketing application should engage with FDA early in the drug development process using an appropriate regulatory pathway (e.g., requesting a Type C meeting through an existing IND for the product).”

- ▶ Submit Study plans and Statistical Analysis Plans (SAPs)
- ▶ Request FDA feedback before starting a study on
 - ▶ RWD is fit for use
 - ▶ Study design
 - ▶ Rationale for choosing data source
 - ▶ Can data source address study questions
 - ▶ Is the data reviewable
 - ▶ Data provenance / how traceability will be provided /documentation

VERY EARLY COMMUNICATION: VENDORS

Recommendation

- “Make sure you know what you are buying”
- Make sure to vet data before purchasing
- Will patient level data be available for submission
- Sample size after exclusions
- Data for all key variables/covariates available
- Information on data provenance/traceability to source
- Amount of missing data
 - for key variables including covariates
 - Is missing data random

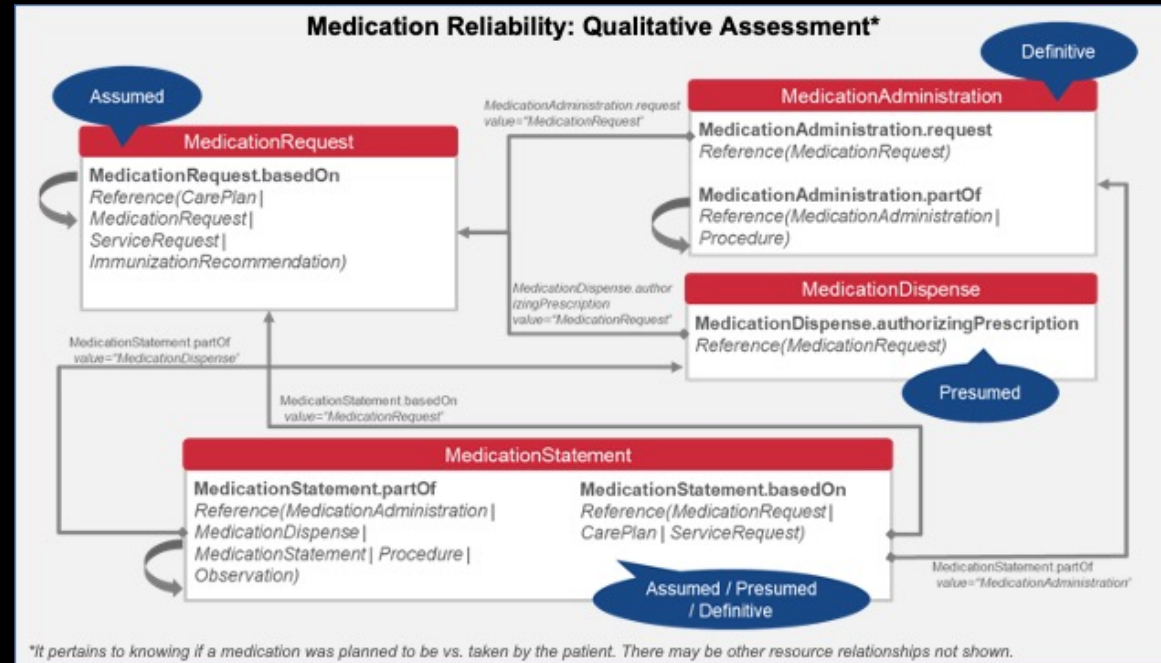
MULTIPLE SOURCES FOR CONFORMATION

Challenge

- ▶ RCTs
 - ▶ Protocol specifies how, who, and where assessment is performed
 - ▶ So, much of this info is not captured during data collection
 - ▶ Data is collected and cleaned as documented in protocol and DMP
- ▶ RWD
 - ▶ Data not collected under a protocol
 - ▶ Data not cleaned or monitored
- ▶ FDA
 - ▶ Would like more information surrounding the context of an assessment to increase confidence and understanding of the data
 - ▶ Would like multiple sources to increase confidence in RWD

MULTIPLE SOURCES FOR CONFORMATION: EXPOSURE EXAMPLE

Challenge: In RWD exposure data resides in a variety of places (record types) with varying degrees of confidence



In RWD Difficult to accurately construct exposure to a given treatment from above sources

MULTIPLE SOURCES FOR CONFORMATION: EXPOSURE EXAMPLE

Recommendation

▸ SDTM

- “As SDTM as possible”
- Submit data from as many sources as possible
- Create custom domains and supplementary variables as needed
- Document source of each record type > Include a source variable for each record

▸ ADaM

- Construct exposure dataset from all available sources
- When possible, verify from an alternative source (e.g., multiple record types, claims data)
- Provide detailed traceability to source data in define.xml and ADRG

- Derived ‘confidence’ variables can be useful to report the level of confidence in the accuracy of exposure data

TERMINOLOGY DIFFERENCES

Challenge

- Clinical research and Healthcare terminology lack harmonization
- Many concepts have different names/meaning and controlled terminology
- Coding systems differ in clinical research and health care
- EHRs use SNOMED, ICD
- Clinical research uses MedDRa and WHO Drug coding
- Source data verification is challenging

TERMINOLOGY DIFFERENCES

Recommendation

- ▶ Create new supplemental variables in SDTM to capture
 - ▶ Source coding system (i.e., SNOMED)
 - ▶ Source verbatim term
 - ▶ Source code
- ▶ Provide an explanation of the process to convert the source terminology to submission compliant terminology/coding systems
- ▶ For example > Converting SNOMED or ICD to MedDRA

TRACEABILITY FOR DHTS

- ▶ Digital Health Technologies (DHTs) are systems that use computing platforms, connectivity, software, and/or sensors for health care and related uses. They include technologies intended for use as a medical product, in a medical product, or as an adjunct to other medical products (devices, drugs, and biologics).
- ▶ DHTs often consist of sensor hardware that allows for continuous or intermittent recording of physiological and/or behavioral data (e.g., blood pressure, physical activity, glucose levels).
- ▶ DHT data includes wearables and other types of data, and is one of the many types of RWD

Source: [Framework for the Use of Digital Health Technologies in Drug and Biological Product Development](#)

TRACEABILITY FOR DHTS

- ▶ Sensor-based DHTs used in remote data acquisition can measure longitudinal data over an extended period (e.g., days, weeks, months).
 - ▶ For example, activity trackers or continuous glucose monitors can be used to capture continuous, granular measurements over time. The example below highlights the use of an activity tracker:

Evaluation of a novel orthotic device to treat knee osteoarthritis. The clinical investigation uses a general-purpose consumer activity tracker to measure step count.	
DHT	General-purpose consumer activity tracker bracelet
DHT hardware*	General-purpose consumer activity tracker bracelet with sensors
DHT software	None
General-purpose computing platform	None
Purpose of using DHT	Measure a participant's steps during the clinical investigation as part of the endpoint of interest

TRACEABILITY FOR DHTS

- ▶ Longitudinal data that is captured and summarized based on the selected DHT-derived endpoint introduces the need for reporting comprehensive traceability for each 'level' of data.
- ▶ The example below expands the step count example and highlights three 'levels' of data, resulting in a summary measure that calculated average daily step count for each subject:



Traceability for DHTs

- ▶ When data is aggregated, comprehensive traceability documentation should be provided:

Details on how data is aggregated at each 'level' is needed, including details on the algorithm(s) and software programs (code) used to summarize the data

Individual Step
Counts by
Minute



Individual Step
Counts by Day



Average Daily
Step Count

SUBMIT AN ACRF

- ▶ Sponsors can create an aCRF for the RWD to be submitted
- ▶ Facilities understanding of a given concept or question
- ▶ Shows how a data element is related to associated data elements
- ▶ Can be used to show the original and possible values of a data element
- ▶ Provides traceability for prospective and retrospective studies

Long Term Considerations/Solutions

Challenge: Current submission standards inadequate to represent RWD

- CDISC standards designed for RCTs or intervention studies
- Built on older technologies (SAS transport)
- Lack many concepts related to RWD
- Core variables not identified for RWD
- Validation rules apply to RCTs
- RWD must be repackaged and transformed to meet regulatory submission guidelines
- Data linking solutions are cumbersome and outdated

Long Term Considerations/Solutions

Recommendation: Re-examine current submission standards (CDISC)

- CDISC standards adequately represent RCTs + RWD + Observational Studies?
- Should we have a hybrid approach for collected data?
- Each data type represented in the standard for which it is optimized
 - RCTs > CDISC
 - EHR/claims > FHIR
 - Observational studies/registries > OMOP
- Should we use a more “modern” standard like FHIR for all studies?
- Trend toward more RWD submitted to support marketing applications
- Most RWD will be represented in FHIR in the future
- Will most RCT data be collected via an EHR?
- Will standards be harmonized in the future?

CONCLUSIONS

- Submitting RWD presents a number of new challenges
- There are significant gaps in the current standards for submitting RWD
- The bar/standard for assessing RWD is the same as for data from RCTs
- Need for better documentation on data provenance and source data elements
- Future is not complex transformations
- Are we asking the right questions to solve RWE challenges?
- It's time to re-evaluate the use of CDISC as the single submission standard?
- Document, Document, and Document



THANK YOU ;)

KEEP IN TOUCH!



JEFFREY ABOLAFIA

JABOLAFIA@PINNACLE21.COM

SARAH FERKO

SMFERKO@US.IBM.COM

INGEBORG HOLT

INGEBORG.HOLT@IBM.COM

P21

A blurred background image of a modern office hallway with large glass walls and a polished floor. Several people are walking through the hallway, their figures blurred to convey a sense of motion. The lighting is bright and even.

QUESTIONS?