www.ispor.org

ISPOR

Improving healthcare decisions

Session 1: From Data Quality to Qualities

ISPOR Real-World Evidence Summit 2023

May 7, 2023

Moderator: Rachele Hendricks-Sturrup, DHSc, MSc, MA, Duke-Margolis Center for Health Policy

Speakers:

- **Sebastian Schneeweiss, MD, ScD,** Harvard Medical School; Brigham and Women's Hospital
- **Andre Araujo, PhD,** GSK
- **Jaclyn L. F. Bosco, PhD, MPH, FISPE,** IQVIA

Session 1: From Data Quality to Qualities

ISPOR Real-World Evidence Summit 2023

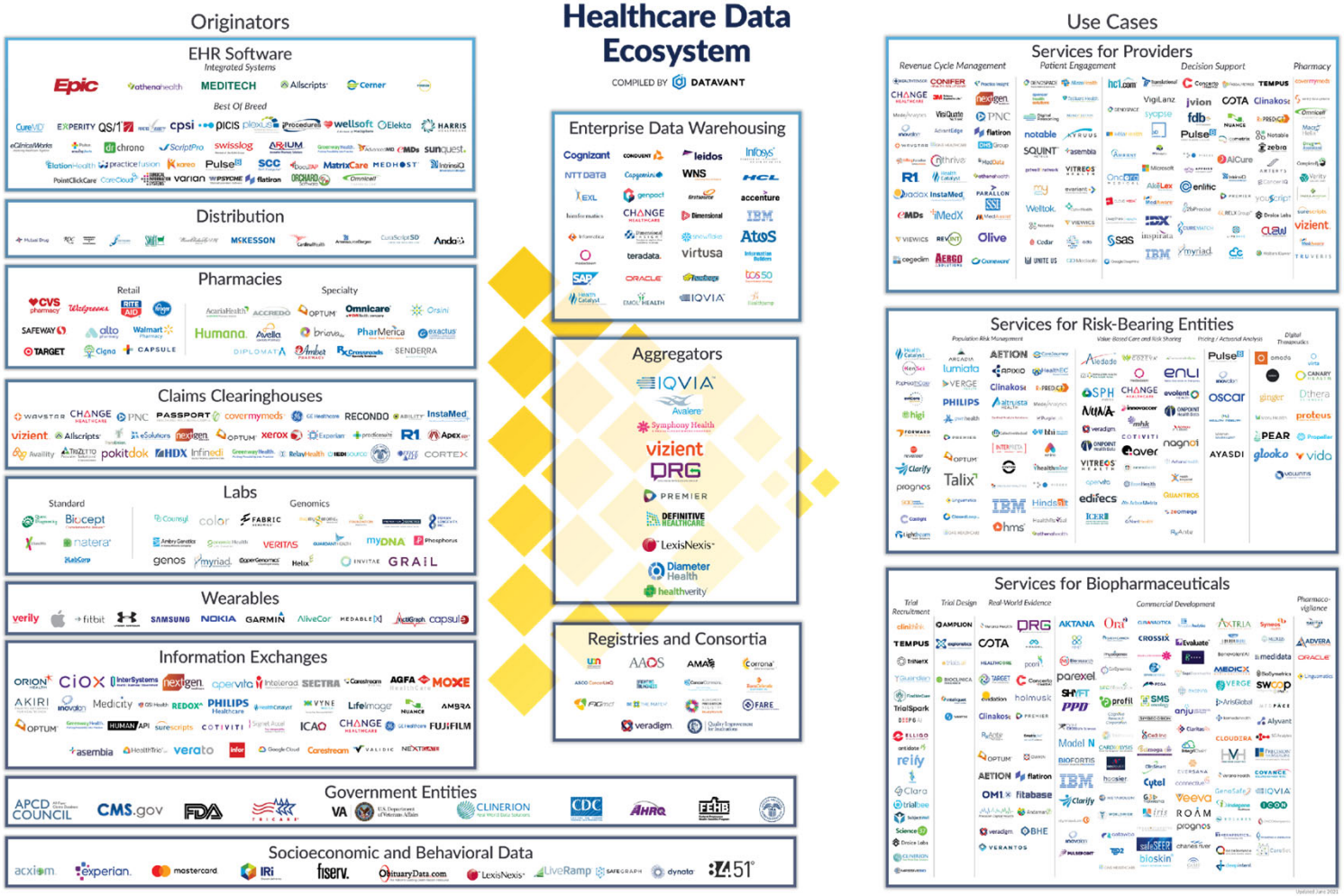
Rachele Hendricks-Sturrup, DHSc, MSc, MA

Research Director, Real-World Evidence

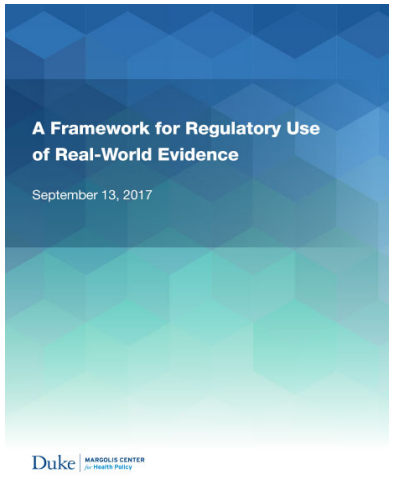
Duke-Margolis Center for Health Policy

Duke | MARGOLIS CENTER *for* Health Policy

Vast and Growing Health Data That Could Inform Care



Compiled by Datavant: https://medium.com/datavant/the-fragmentation-of-health-data-8fa708109e13



Generating RWE Fit for Regulatory Purpose

Regulatory Context

What specific decision is FDA considering?

- New indication
- Labeling revision
- Safety revision
- Benefit-risk profile

Clinical Context

Can the clinical question be reliably addressed with RWE?

- Prevalence of the disease
- Clinical equipoise
- Expected treatment effect size
- Relevant prior evidence

Data Considerations

Is the real-world dataset fit for regulatory purpose?

1. **Is the data relevant?**
   - Representative of the population of interest
   - Contains key variables and covariates
2. **Is the data of adequate quality?**
   - Minimal missing data
   - Data reliability and validity is satisfactory for study purpose
   - Known provenance and transparency of data processing

Methods Considerations

Are the methodological approaches of sufficient rigor?

1. **Are the methods credible?**
   - Appropriate analytic approach
2. **Can the approach produce actionable evidence?**
   - Interplay of body of clinical evidence and tolerance for uncertainty

Fit-for-purpose RWE

Matching data sources and methods to answer specific clinical and regulatory questions determines applicability of RWE for different regulatory uses

Guidance for Industry

DRAFT GUIDANCE

FDA Draft RWE Guidance – Sep-Dec 2021

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products

Data Standards for Drug and Biological Product Submissions Containing Real-World Data

Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products

https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

Duke | MARGOLIS CENTER *for* Health Policy

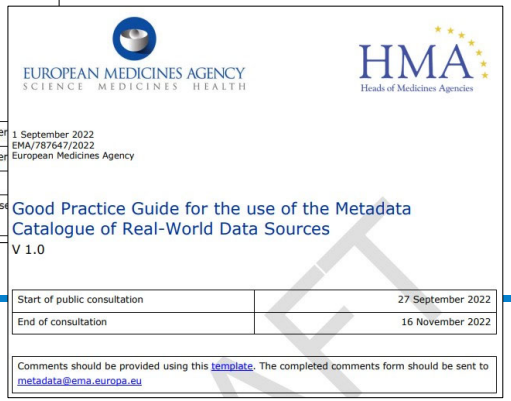
Current landscape of relevant EU guidance on RWE

Oct. 2021

Dec. 2021

May 2022

June 2022



Consultations launched in Sept. 2022

For revision in 2023

2022-2025

NICE recently published its RWE framework, which describes best practices for the planning, conduct, and reporting of RWE studies

Overarching Principals for NICE RWE Framework[1]

NICE National Institute for Health and Care Excellence

1. Generating evidence in a transparent way "with integrity from study planning to study conduct and reporting"
2. Ensuring that the data is of good provenance, that it is trustworthy and fit for purpose
3. The use of appropriate analytical methods that minimize the risk of bias and characterize uncertainty

- Enhancing use of real-world data to resolve gaps in knowledge and drive forward access to innovative medicines for patients was noted as a strategic focus in the NICE Strategy 2021 to 2026[2]
- As a result, NICE published the RWE framework in June 2022 to help deliver on this ambition by[2]:
  - Identifying when real-world data can be used to reduce uncertainties and improve guidance
  - Clearly describing best-practices for planning, conduct, and reporting RWE studies to improve the quality and transparency of evidence
- The framework provides transparency by advising clear specification of research questions, early planning of studies, and clear descriptions of data sources and data curation data sources[1]
- It is intended to be a "living document" that will broaden overtime according to need[2]

Abbreviations: NICE - National Institute for Health and Care Excellence; RWE - Real-world evidence

References: 1) Bruce, Franchesca. England. HTA Body NICE Makes Big RWE Push. Pink Sheet Pharma Intelligence. Available at https://pink.pharmaintelligence.informa.com/PS146414/England-HTA-Body-NICE-Makes-Big-RWE-Push. Accessed 3 November 2022. 2) National Institute for Health and Care Excellence. NICE real-world evidence framework. Corporate document [ECD9]. 2022.

Preliminary Takeaways from Emerging Landscape of *RWD* Regulatory Frameworks

- Generally, support for and broad understanding of potential regulatory use cases for regulatory RWE (safety and effectiveness) and definitions/considerations surrounding fit-for-purpose datasets.
- However, there appears to be early potential divergence in terminologies and concepts:
  - Defining fitness-for-use datasets:
    - FDA: Reliability and Relevance
      - Reliability: data accuracy, completeness, provenance, and traceability
      - Relevance: availability of key data elements (exposure, outcomes, covariates) and sufficient numbers of representative patients for the study
    - EMA: Reliability, Relevance, Extensiveness, Coherence, and Timeliness
      - Reliability: precision, accuracy, and plausibility
      - Relevance: covers how closely the data reflects the aspects of reality that we intend to measure
    - NICE: Quality and Relevance
      - Reliability: completeness and accuracy
      - Relevance: data content, coverage, and characteristics
  - Challenging to apply and operationalize these concepts as part of fit-for-purpose assessment frameworks (e.g., validation approaches, quality checks, and documentation needs)
- Interactions with HTA for fit-for-purpose RWD guiding clinical practice and payment decisions are becoming clearer – similar issues for supporting evidence on comparative effectiveness and "label deepening" (clinical guidelines, increasing impact in care delivery)

ICMRA Statement on RWE Collaboration – Jul 2022

The June 2022 ICMRA workshop on RWE identified four areas of opportunities for regulatory collaboration which could help address common challenges and further enable the integration of RWE into regulatory decision-making.

- **Harmonisation of RWD and RWE terminologies**:
  - Generate common operational definitions of RWD and RWE, with clear scope and level of granularity (e.g., pertaining to RCTs and observational studies);
  - Leverage existing ICH activities, such as M14 on "*General principles on planning and designing pharmacoepidemiological studies that utilize real-world data for safety assessment of a medicine*".
- **Convergence on RWD and RWE guidance and best practice, including**:
  - Common principles for RWD quality;
  - Metadata to enable characterisation and discoverability of RWD;
  - Suitable scenarios where RWE may contribute to regulatory decision-making, building on existing use-cases;
  - Templates for study protocols/reports that can be used in multiple regulatory jurisdictions.

Side presented by Dr. John Concato, US FDA, 2022

Panelist Introductions

- Rachele Hendricks-Sturrup, DHSc, MSc, MA, **Duke-Margolis Center for Health Policy** - Moderator
- Jaclyn L. F. Bosco, PhD, MPH, FISPE, **IQVIA** - Panelist
- Andre Araujo, PhD, **GSK** - Panelist
- Sebastian Schneeweiss, MD, ScD, **Harvard Medical School; Brigham and Women's Hospital** - Panelist

Thank You!

Contact Us

POC: Rachele Hendricks-Sturrup, DHSc, MSc, MA

Rachele.hendricks.sturrup@duke.edu

1201 Pennsylvania Avenue, NW, Suite 500
Washington, DC 20004

DC office: 202-621-2800
Durham office: 919-419-2504

Follow Us

DukeMargolis

@DukeMargolis

@DukeMargolis

Duke Margolis

Real-world data quality for causal inference

Sebastian Schneeweiss, MD, ScD
Professor of Medicine and Epidemiology, Harvard Medical School, Boston

Chief, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine
Brigham and Women's Hospital

May 2023

Funding

- In parts by FDA (HHSF223201710186C and HHSF…46C)
- In parts by the FDA Sentinel Innovation Center (75F40119F19002)
- In parts by the NHLBI (R01-HL141505), NIAMS (R01-AR080194)
- In parts by the Burroughs Wellcome Fund
- Additional funding came from PCORI

Disclosures

- PI, Sentinel Innovation Center (FDA)
- Co-Chair, Partners Center for Integrated Healthcare Data Research
- PI of grants and contracts from NIH, AHRQ, PCORI, FDA, IMI, Arnold Foundation
- PI of research grants awarded to BWH by Bayer, Vertex, Boehringer Ingelheim
- Consulting fees from Aetion, Inc. (incl. equity)

Causal inference from RWD

EVERYBODY wants the most accurate evidence possible

– some need to make more compromises than others

Bias as an obstacle to causal inference

1) Selection bias

2) Information bias

3) Confounding

- Outcome
  misclassification/
  measurement error
  - Random
  - Differential
- Exposure
  missclass/m.e.
- Confounder
  missclass/m.e.

Characterization of MC/me:

Binary data
- Sensitivity
- Specificity
- PPV

Continuous data
- % missing
- Mean squared deviation

Time-to-event data
- Accuracy of onset

Porta M. A Dictionary of Epidemiology 5th ed. 2008

Study features:	Examples of ways to improve measurement characteristics	Typical proxies for data quality in secondary data	Actual measurement characteristics*
1) Study pop^{n},	Require two diagnosis codes to increase specificity of underlying condition		
2) Exposure measurement	Use dispensing information instead of prescribing data to increase completeness		
3) Outcome measurement	Use serious events, e.g. that require hospitalizations to increase specificity of outcome measurement		
4) Confounder measurement	Screen a wide range of potential confounders and their proxies to limit unobserved confounding		

* These metrics are relevant for quantifying potential bias and assessing the likelihood of a causal drug-outcome relationship vs. spurious findings

Schneeweiss S, Patorno E. Endocr Rev. 2023

Study features:	Examples of ways to improve measurement characteristics	Typical proxies for data quality in secondary data	Actual measurement characteristics*
1) Study pop^{n} identification	Require two diagnosis codes to increase specificity of underlying condition	Data relevance Data accrual Data provenance	
2) Exposure measurement	Use dispensing information instead of prescribing data to increase completeness		
3) Outcome measurement	Use serious events, e.g. that require hospitalizations to increase specificity of outcome measurement		
4) Confounder measurement	Screen a wide range of potential confounders and their proxies to limit unobserved confounding		

* These metrics are relevant for quantifying potential bias and assessing the likelihood of a causal drug-outcome relationship vs. spurious findings

Schneeweiss S, Patorno E. Endocr Rev. 2023

Study features:	Examples of ways to improve measurement characteristics	Typical proxies for data quality in secondary data	Actual measurement characteristics*
1) Study pop^{n} identification	Require two diagnosis codes to increase specificity of underlying condition	Prior experience with a data source, publications Availability of validation studies Detailed documentation of data generation mechanism Detailed description of data curating process Detailed description of mapping to medical constructs (if any) Documentation of coding shift over time	
2) Exposure measurement	Use dispensing information instead of prescribing data to increase completeness		
3) Outcome measurement	Use serious events, e.g. that require hospitalizations to increase specificity of outcome measurement		
4) Confounder measurement	Screen a wide range of potential confounders and their proxies to limit unobserved confounding		

* These metrics are relevant for quantifying potential bias and assessing the likelihood of a causal drug-outcome relationship vs. spurious findings

Schneeweiss S, Patorno E. Endocr Rev. 2023

Study features:	Examples of ways to improve measurement characteristics	Typical proxies for data quality in secondary data	Actual measurement characteristics*
1) Study pop^{n},	Require two diagnosis codes to increase specificity of underlying condition	Prior experience with a data source, publications Availability of validation studies Detailed documentation of data generation mechanism Detailed description of data curating process Detailed description of mapping to medical constructs (if any) Documentation of coding shift over time	**Binary data** e.g. diagnostic codes: • Sensitivity • Specificity • PPV **Continuous data** e.g. lab test values: • % missing • Mean squared deviation **Time-to-event data** • Accuracy of onset
2) Exposure measurement	Use dispensing information instead of prescribing data to increase completeness		
3) Outcome measurement	Use serious events, e.g. that require hospitalizations to increase specificity of outcome measurement		
4) Confounder measurement	Screen a wide range of potential confounders and their proxies to limit unobserved confounding		

* These metrics are relevant for quantifying potential bias and assessing the likelihood of a causal drug-outcome relationship vs. spurious findings

Schneeweiss S, Patorno E. Endocr Rev. 2023

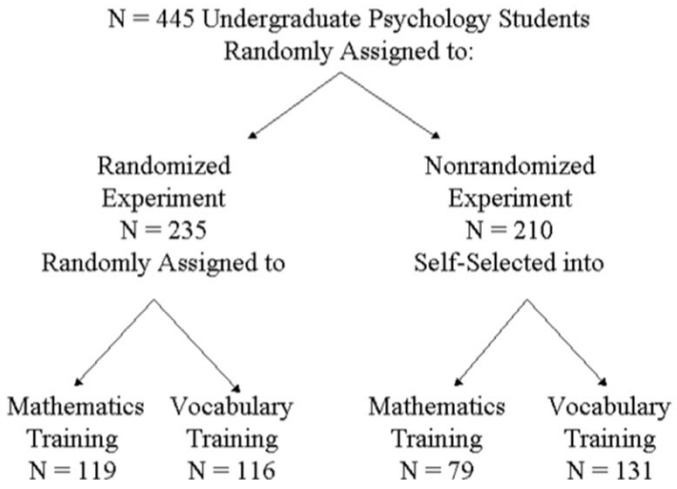
How good is good enough?

Study features:	Examples of ways to improve measurement characteristics	Typical proxies for data quality in secondary data	Actual measurement characteristics*
1) Study pop^{n},	Require two diagnosis codes to increase specificity of underlying condition	Prior experience with a data source, publications Availability of validation studies	**Binary data** e.g. diagnostic codes: • Sensitivity • Specificity • PPV **Continuous data** e.g. lab test values: • % missing • Mean squared deviation **Time-to-event data** • Accuracy of onset
2) Exposure measurement	Use dispensing information instead of prescribing data to increase completeness	Detailed documentation of data generation mechanism	
3) Outcome measurement	Use serious events, e.g. that require hospitalizations to increase specificity of outcome measurement	Detailed description of data curating process Detailed description of mapping to medical constructs (if any)	
4) Confounder measurement	Screen a wide range of potential confounders and their proxies to limit unobserved confounding	Documentation of coding shift over time	

* These metrics are relevant for quantifying potential bias and assessing the likelihood of a causal drug-outcome relationship vs. spurious findings

Schneeweiss S, Patorno E. Endocr Rev. 2023

Food for thought



Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments

William R. SHADISH, M. H. CLARK, and Peter M. STEINER

Table 1. Percent bias reduction in quasi-experimental results by various adjustments

	Mean difference (standard error)	Absolute bias (Δ)	Percent bias reduction
Vocabulary Outcome			
Covariate-adjusted randomized experiment	$8.25_{(.37)}$		
Unadjusted quasi-experiment	$9.00_{(.51)}$	.75	
PS stratification	$8.15_{(.60)}$	.11	86%
Plus covariates with strata	$8.32_{(.49)}$	.07	91%
PS linear ANCOVA	$8.07_{(.49)}$	.18	76%
Plus covariates	$8.07_{(.47)}$	.18	76%
PS nonlinear ANCOVA	$8.03_{(.50)}$	.21	72%
Plus covariates	$8.03_{(.48)}$	.22	70%
PS weighting	$8.22_{(.66)}$	.03	96%
Plus covariates	$8.19_{(.51)}$	.07	91%
PS stratification with predictors of convenience	$8.77_{(.48)}$	.52	30%
Plus covariates	$8.68_{(.47)}$	.43	43%
ANCOVA using observed covariates	$8.21_{(.43)}$	.05	94%

-> If we have excellent measurements, we get the same findings as RCTs

Shadish et al. JASA 2008;103:1334-43

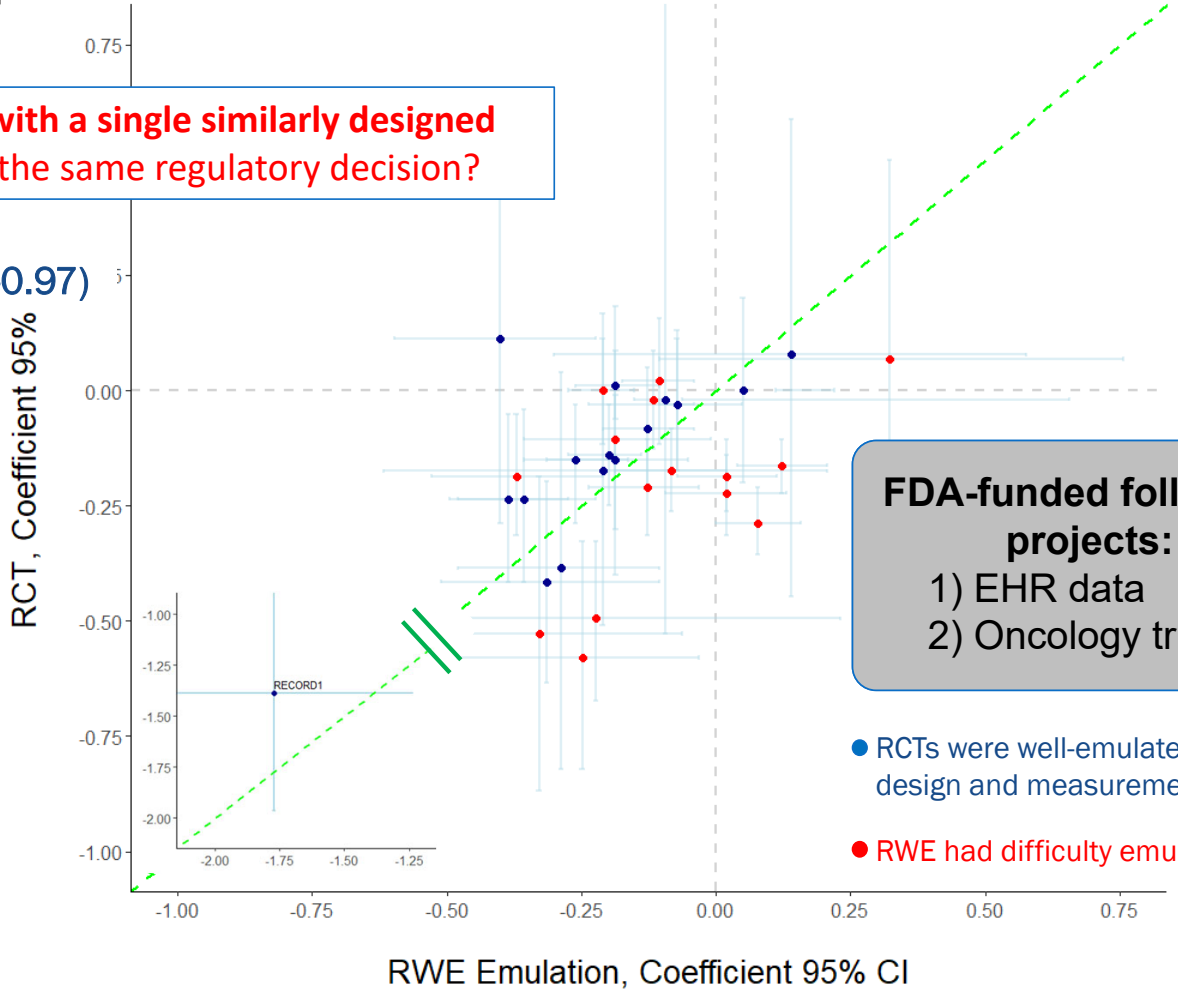
JAMA | Original Investigation

Emulation of Randomized Clinical Trials With Nonrandomized Database A
Results of 32 Clinical Trials

Question: Had we replaced an RCT with a single similarly designed RWE study, would we have come to the same regulatory decision?

All trials r = 0.80 (0.63-0.90)

Well-emulated trials r = 0.93 (0.79-0.97)



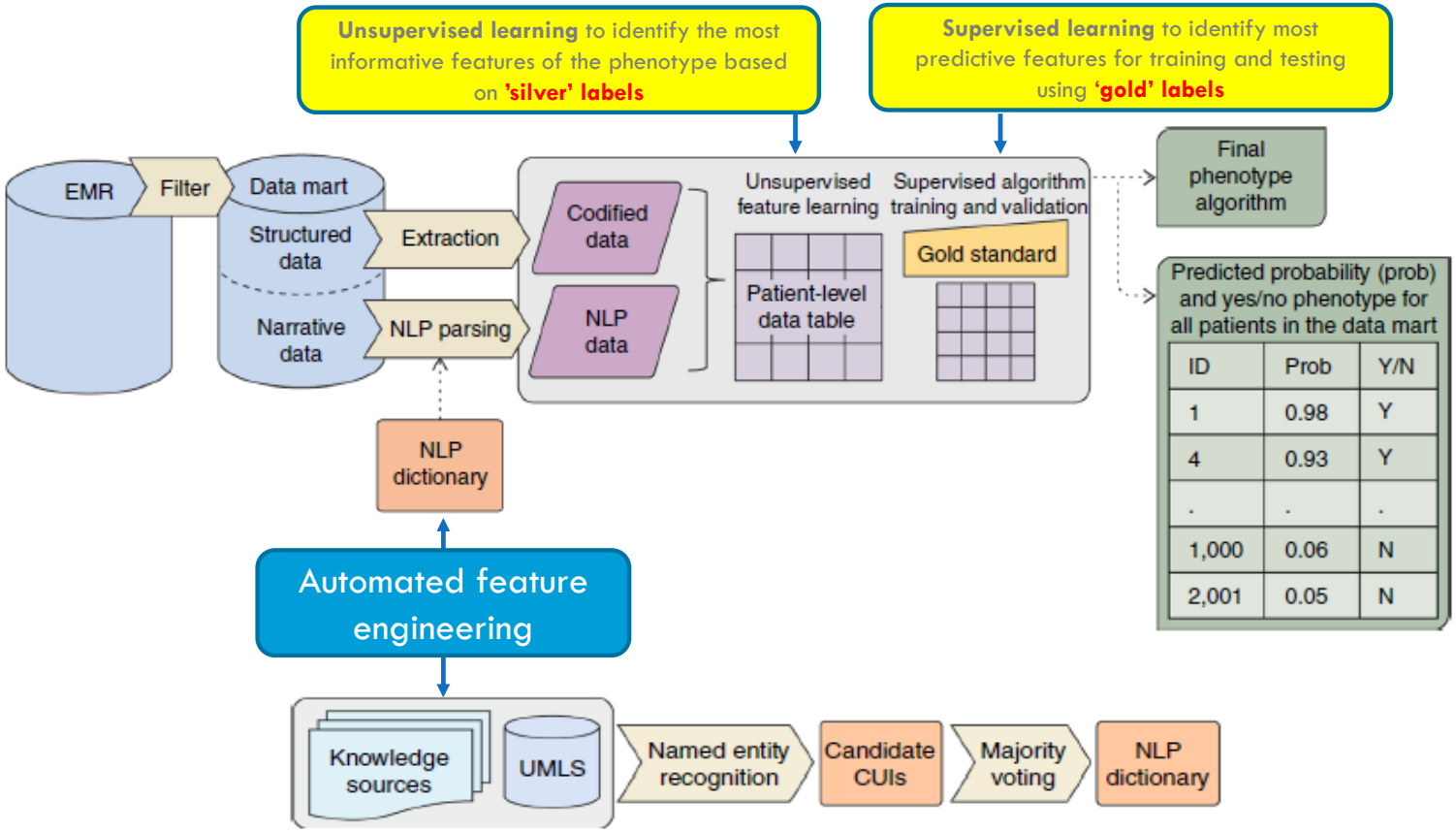
RWE studies and RCTs come to the same conclusions if they emulate an RCT design well and data are fit-for-purpose

FDA-funded follow-on projects:
1) EHR data
2) Oncology trials

- RCTs were well-emulated by RWE in design and measurement (n=16)
- RWE had difficulty emulating (n=16)

Wang SV, Schneeweiss S, et al. JAMA 2023

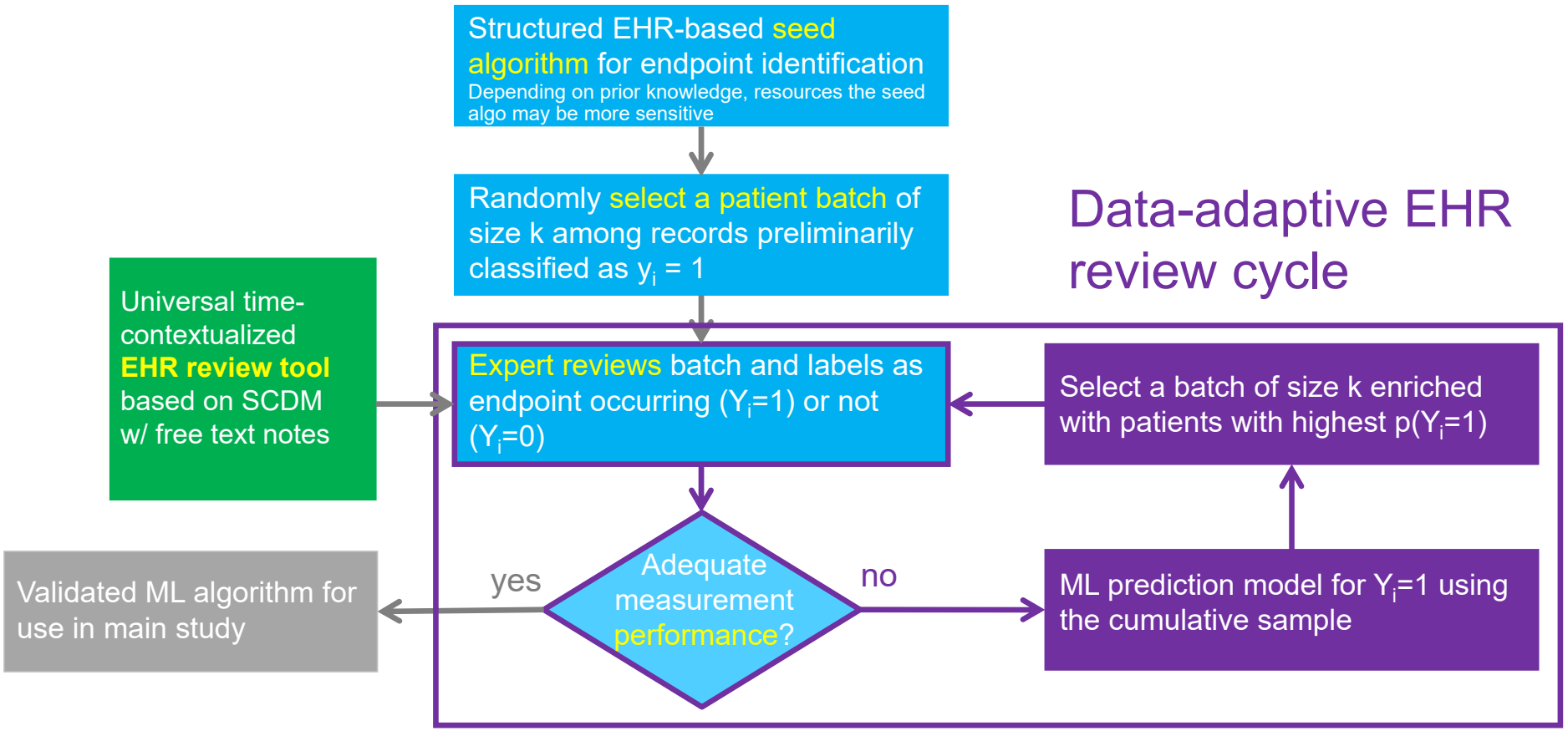
FDA Sentinel Innovation Ctr: A pipeline for ML-augmented variable definitions and labeling based on linked EHR+claims



Zhang et al. *Nat protocols*. 2019

FDA Sentinel Innovation Ctr: **Expedited labeling**

"Gold" standard labels are time consuming and costly to derive in medicine



Take home

- How accurate and complete are key measurements?
- Is that good enough?

Real-World Data Audit Readiness Considerations

TransCelerate BioPharma

Session "From Data Quality to Qualities"
ISPOR RWE Summit 2023
Boston, MA

May 2023

Speaker

Andre B. Araujo, PhD

TransCelerate
Workstream Lead, Real-World Data, Audit Readiness Initiative

GlaxoSmithKline (GSK)
Head of Real-World Analytics, Value Evidence & Outcomes

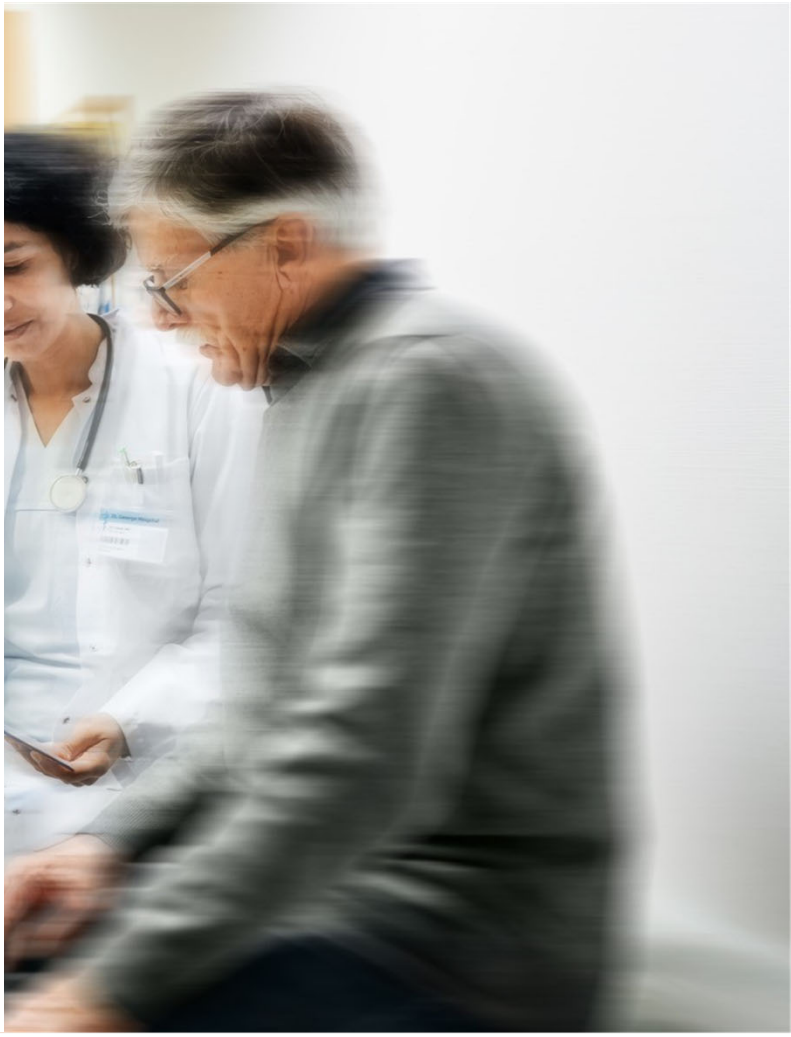
Disclosures

Full-time employee of GSK and minor stockholder; Workstream lead for TransCelerate; No confidential or proprietary data are included in this presentation

TransCelerate BIOPHARMA INC.

TransCelerate is a not-for-profit entity created to foster collaboration.

Our mission is to improve the health of people around the world by accelerating and simplifying the research and development of innovative new therapies.

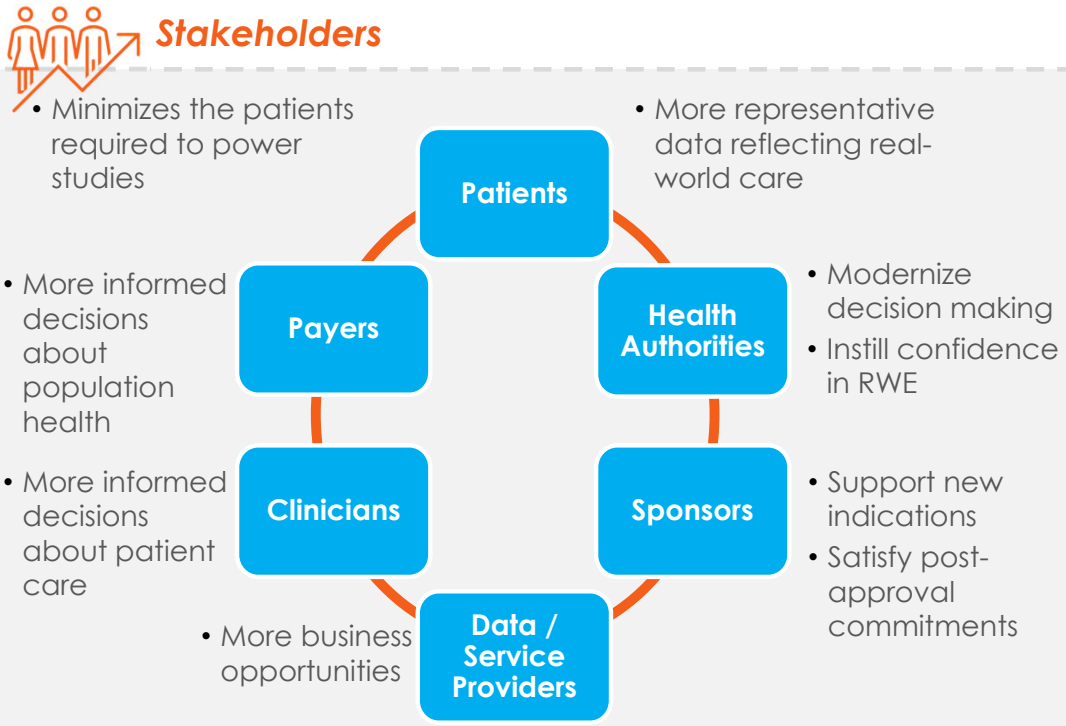


TransCelerate BIOPHARMA INC.

Today's Real-World Data Landscape

Multiple stakeholder groups can benefit from greater use of RWD, but several challenges limit uptake

Stakeholders



Challenges

- **Regulator caution** in evaluating RWD for decision-making
- **Lack of clarity** on how to engage with regulators on specific RWD/RWE use cases
- Lack of clarity on what constitutes RWD **relevance and reliability**
- Quality management approaches for clinical trial data are not **fit-for-purpose** when applied to RWD
- Sponsor companies (generally) **do not own or control RWD**

TransCelerate's RWD Audit Readiness Initiative

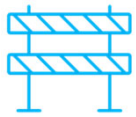
Focus

Operationalize the thought leadership stemming from **Duke Margolis/FDA** and many others on the use of RWD in regulatory decision-making.

The team will leverage Health Authority and Data/Service Provider interactions to **develop documentation that supports quality management (QA, QC, and audit) for RWD sources**, resulting in an **"Audit Readiness Tool"** targeting data relevance and reliability.

Desired Outcomes

Build Trust



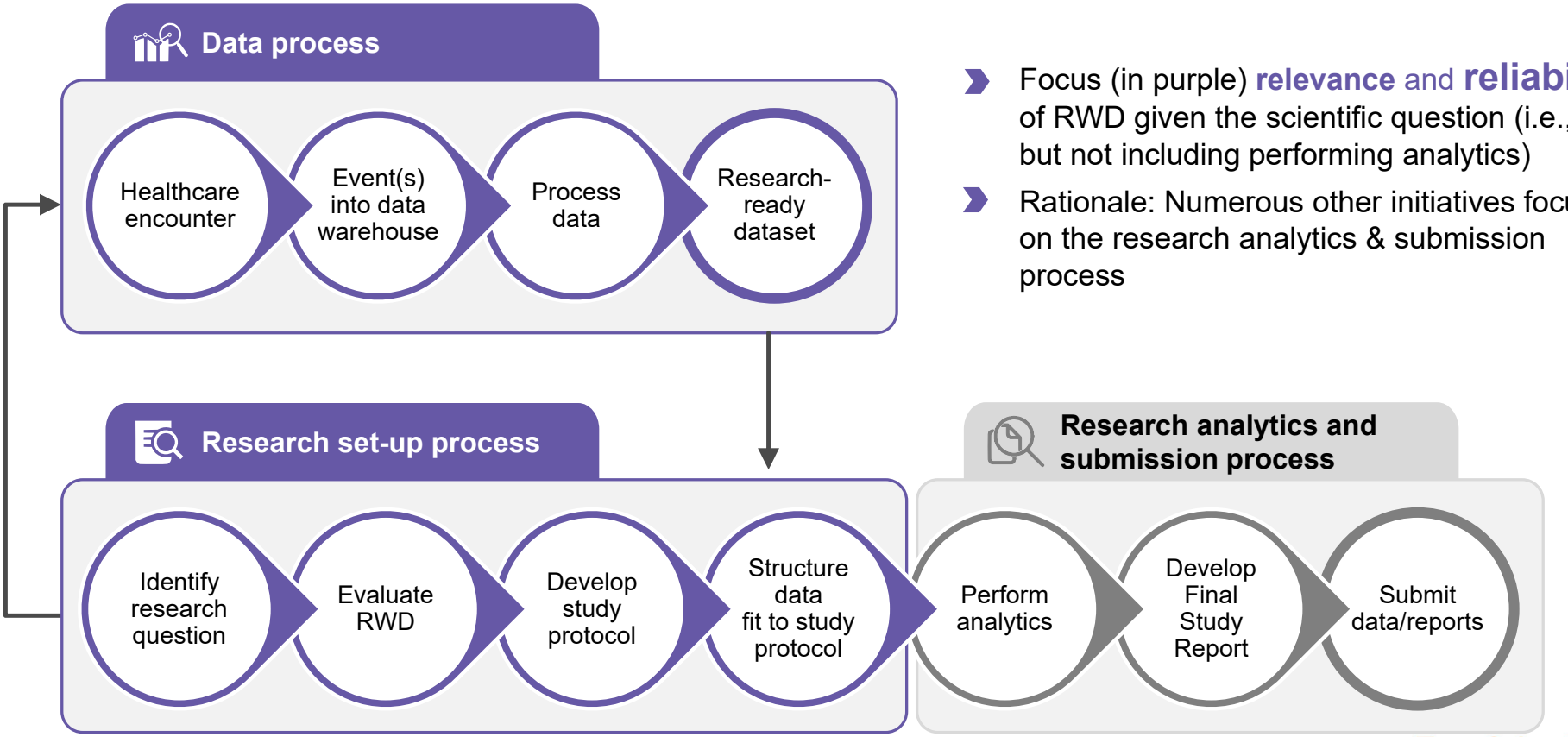
Reduce Barriers

Demonstrate Fit-for-Purpose Use

*The **Audit Readiness Considerations** will help operationalize best practices in order to aid quality management oversight of RWD, including inspection readiness, in a manner suitable for regulatory decision making.*

TransCelerate BIOPHARMA INC.

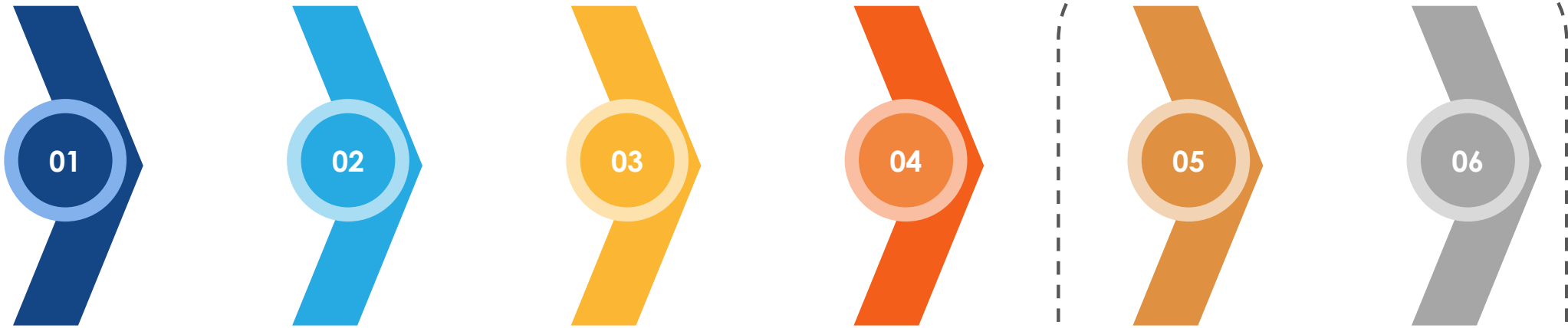
Scope of RWD Audit Readiness Initiative



- Focus (in purple) **relevance** and **reliability** of RWD given the scientific question (i.e., up to but not including performing analytics)
- Rationale: Numerous other initiatives focused on the research analytics & submission process

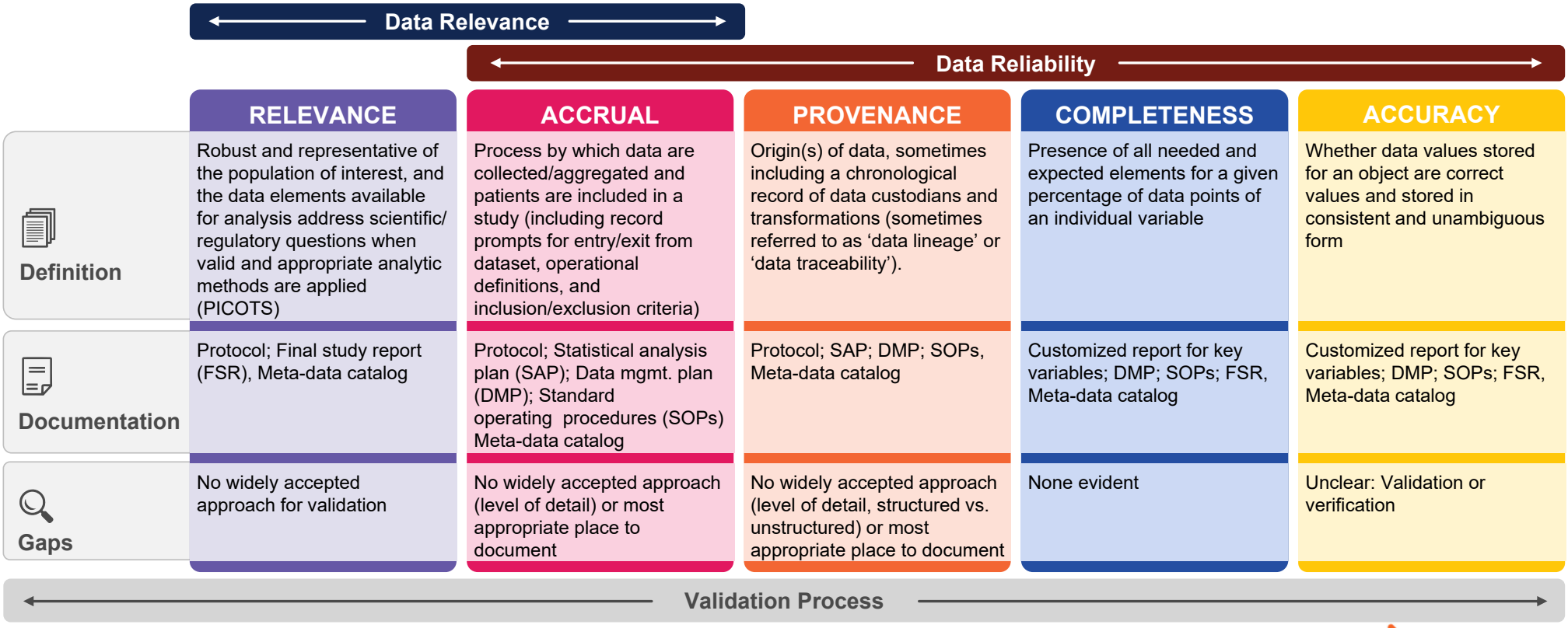
TransCelerate BIOPHARMA INC.

RWD Audit Readiness High-Level Process



TransCelerate BIOPHARMA INC.

RWD Audit Readiness Initiative: Landscape Assessment Insights Framework



* Insights gathered from targeted literature review, including the following sources: Daniel et al. Characterizing RWD Quality and Relevancy for Regulatory Purposes. Oct 2018; Franklin et al. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. Clin Pharmacol Ther 2019;105:867; Kahn et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. Egems 2016;4:1244; Mahendraratnam et al. Determining Real-World Data's Fitness for Use and the Role of Reliability. Sep 2019; US FDA. Framework for FDA's Real-World Evidence Program. Dec 2018; Data Quality Framework for EU medicine regulation. October 2022; EMA Technical workshop on real-world metadata for regulatory purposes. September 2021

RWD Audit Readiness Considerations Public Review

Industry Outreach & Feedback Update

OUTREACH

Stakeholder Outreach

- Data/Service Providers (50+ Organizations)
- RWE Alliance
- EHR Trade Association
- Reagan-Udall
- Duke-Margolis
- FDA
- EMA
- DKMA/BfArM
- ICH
- PhRMA
- Basel Biometric Society
- ISPOR
- GetReal Institute
- Sigma
- ISPE
- OHDSI
- Friends of Cancer Research & ASCO
- SCDM
- QA Leadership Team
- ACRO
- TransCelerate Member Companies
- LinkedIn

INTEREST

Meetings

- RWE Alliance
- Duke Margolis
- PhRMA

LinkedIn Sponsored Post

- Impressions: 188,808
- Clicks: 1,185
- Website visits: 1,182

Website Statistics Q4 2023

- Website visits: 1,419
- RWD AR Considerations Downloads: 148
- Top Countries To Solution Page: U.S. (60.52%), India (14.8%), U.K. (3.17%), Japan (2.26), Germany (1.9%)

RESULTS

26 Stakeholders Provided Feedback

Count	Type of Organization
11	BioPharma Company
7	Data / Service Provider
2	Academic Research Organization
2	Clinical Research Organization
2	Industry Group or Consortium
2	Investigator or Site Staff
26	**Grand Total**

TransCelerate BIOPHARMA INC.

Select Themes from Public Review Feedback

Audit Readiness Considerations [Draft]

Theme	Potential Action
Overlap of 'considerations' between pillars	Reduce overlap where possible, but not eliminate to encourage 'modular use'
Link to broader and relevant scientific concepts, e.g., PICOTS	Will update, important to situate considerations in larger scientific context
Provide additional examples of considerations	Will provide examples where possible, without complete coverage as this is expected to be a 'living document' that will mature over time
Clarify where and how specific considerations should be documented	Will provide clarity where possible, but intent is allow for flexibility and not to create a 'standard' as to where/how information is documented
The role of 'validation' is unclear in terms of considerations	Will provide additional considerations related to validation of key variables
Purpose of tool not explicitly clear	Additional text to be added (e.g., study planning, QMS)

TransCelerate BIOPHARMA INC.

THANK YOU

TransCelerate BIOPHARMA INC.

From Data Quality to Qualities

ISPOR RWE Summit 2023

Jaclyn Bosco, PhD, MPH, FISPE
VP, Global Head of Epidemiology
7 May 2023
ISPOR RWE Summit, Boston, MA

Jaclyn Bosco, PhD, MPH, FISPE

Vice President

Global Head, Epidemiology

jaclyn.bosco@iqvia.com

Disclosures

- I am a full-time employee of IQVIA and perform no research or consultancy outside of that employment
- I am the co-lead discussions around data access and quality for the RWE Alliance
- I accept no personal consulting fees
- I have participated in the design and conduct of studies of the safety, effectiveness, and use of drug, devices, and biologics. None of my research activities are described here
- No confidential or proprietary data are included in these slides

Recent article about the RWE Alliance

INTRODUCING THE REAL-WORLD EVIDENCE ALLIANCE:

A COALITION DEDICATED TO HARNESSING REAL-WORLD EVIDENCE TO IMPROVE THE LIVES OF PATIENTS

Thomas Brown (Syapse), Marni Hall (IQVIA), Tara Isherwood (Syneos Health), Michelle Leavy (OM1), Irene Nunes (Flatiron Health), Lowell Schiller (Aetion), Lauren Silvis (Tempus), Aracelis Torres (Verana Health)

Biopharmaceutical Report of the American Statistical Association, Volume 29, No. 2, Summer 2022, pages 3-4.
Published online at: BioPharm_summer2022_FINAL.pdf (higherlogicdownload.s3.amazonaws.com)

FDA expectations have become clearer

Highlights of "non-binding – not for implementation" recent draft guidance documents 2021

- **Sept 2021** — Assessing EHR & Medical Claims Data to Support Regulatory Decision-Making for Drug & Biological Products
- **Oct 2021** — Data Standards for Drug & Biological Product Submissions Containing RWD
- **Nov 2021** — RWD: Assessing Registries to Support Regulatory Decision-Making for Drug & Biological Products
- **Dec 2021** — Considerations for the Use of RWD and RWE to Support Regulatory Decision-Making for Drugs & Biological Products

- ✓ Know your data – explain the source, any manipulations, matching, curation, transformations, etc. Some validation may be needed.
- ✓ Explain why the data is sufficiently fit for study purpose.
- ✓ Use common recommended coding frameworks, as feasible. Explain differences.
- ✓ Prepare an analytic plan that addresses impact of misclassification and potential sources of bias. Stick to the plan or be prepared to defend changes.
- ✓ Share plans with the FDA prior to execution to enhance likelihood of acceptance

Specific Expectations

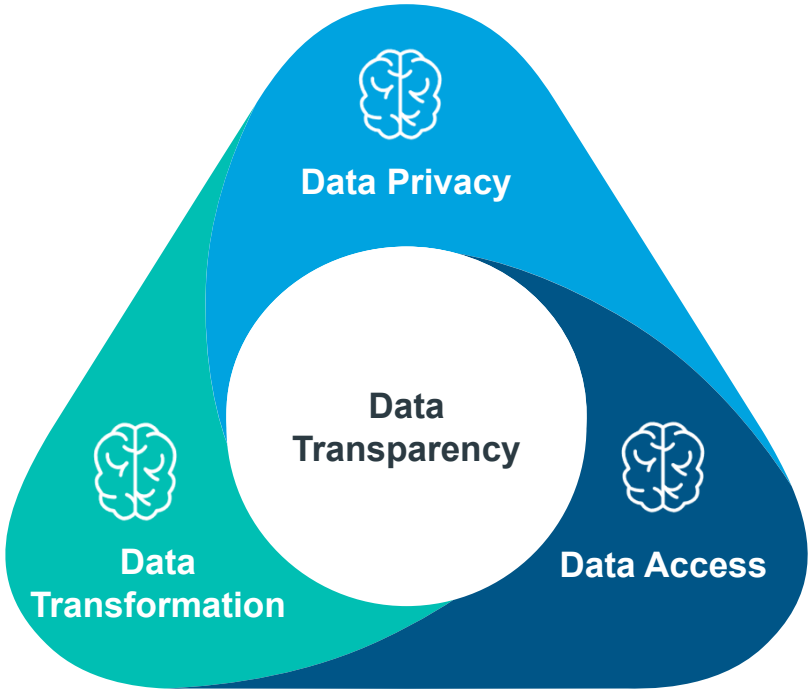
Sponsors need to…

- Ensure RWD associated programming codes and algorithms are **“documented, well-annotated and complete”**
- Implement policies and procedures that enable FDA and persons interested in using the registry’s data to **assess the quality of the data**
- Demonstrate whether and how data from different sources “can be obtained and integrated with **acceptable quality**”
- “**Document all analyses** performed on the data during the study design phase, including **feasibility evaluation** and **exploratory analyses**”
- “Describe in the **study protocol all the data sources accessed** when designing the study, as well as **results from feasibility evaluations** or **exploratory analyses of those data sources**”

Turning expectations into practice raises challenges

1. Access to patient-level source data
2. Conversion of RWD to a supported standard
3. Data quality assessment and examples
4. Clarification of data collection and analysis expectations

Access to Patient-level Source Data



Data Privacy
- Records may contain identifiable patient-specific information
- Subject to local privacy laws governing the data which vary

Data Access
- Data holder may not permit 3rd party access
- Data not permitted to leave borders
- Analyses run at local sites

Data Transformation
- Data derived from unstructured data
- Lineage of RWD sources involve various activities to create analytic data set

Conversion of RWD to a supported standard

Additional guidance on how to apply Agency's recommendation to Common Data Models (CDMs)

Sponsors expected to ensure that RWD associated programming codes and algorithms are **"documented, well-annotated and complete"**

Granularity in the source data may be lost when mapping to currently supported format when using Common Data Models (CDM)

EHR/Claims guidance: Data in CDM-driven networks "rarely contain all of the source information present at the individual healthcare sites"	Conversion to CDISC SDTM resulted in: - Additional investigator burden - Reduced the amount of information in the data

Data quality assessment and examples

***Assess quality** of registry data and demonstrate **acceptable quality** of different data sources*

Data Quality Assessment

Can data quality be assessed without providing access to raw data when authorization is not feasible or appropriate?

Examples Needed

- **Risk-based database quality assurance practices** considered appropriate for registries
- **Acceptable approaches** to resolve issues of data quality for validating a common outcome of interest and applying quality standards, etc.

Clarification of data collection & analysis expectations

Document all analyses** including feasibility evaluations and exploratory analyses and **describe all accessed data and results in study protocol

Study Design Phase

- Clarify when **study design** phase **begins and ends**
- Acknowledge **scope** is **study-specific**
- Key **distinction between** conducting **feasibility analyses** to **identify fit-for-purpose data** vs testing a **study hypothesis**
- Feasibility evaluations, a pilot study, statistical analysis plan, or a data management plan should **include** how **inclusion/exclusion criteria** were **selected and** how **data** were **transformed**

Feasibility Evaluations

- **Clarification of study design phase from feasibility evaluations**
- **"Data source"** may be used by **different sponsors** and for **different** development **programs**, and over time a data source may be further **curated** by an RWD/E organization
- Clarify feasibility is focused on a **specific time period** of each study
- Submit results of feasibility assessments and exploratory analyses in a **separate documents** rather than study protocol due to length and volume of potential data sources

Pilot 2.0: Performance of Real World Overall Survival and Methodological Recommendations

Treatment effects varied from 0.80 to 1.15 across RW data sources

Heterogeneity of missing data for entry criteria and prognostic factors

- Up to 25% missing ECOG PS
- Up to 87% missing laboratory values to ascertain organ function
- Varied evidence of brain metastases and low sensitivity of ICD codes in identifying brain metastases

Mortality Se and Sp ranged by RW data source

- Poor completeness of mortality data
- Granularity of variable death dates and handling of partial complete dates varied

See Lasiter L, Tymejczyk O, Garrett-Mayer E et al. Clin Pharmacol Ther. 2022 Feb;111(2):444-454. doi: 10.1002/cpt.2443

Data Quality Assessment Recommendations

Develop template

- For quantitative evaluation of data distributions, quality, and missingness

Use quantitative approach

- To understand data availability and missingness for improved interpretation

Use careful evaluation

- By a representative team that has deep knowledge of the data curation, extraction, and provenance

Use quality indicators

- For data or consensus on problematic missingness for key covariates may inform the study design

See Rivera DR, Henk HJ, Garrett-Mayer E et al. Clin Pharmacol Ther . 2022 Jan;111(1):283-292. doi: 10.1002/cpt.2453.

Thank you!