Recommended “Good” Practices for Real World Evidence (RWE) Algorithms

ISPOR Virtual 2020 Conference

James F Murray, Ph.D.
Julie Beyrer, MTSC, MPH
Hamed Abedtash, PharmD, PhD
Kenneth Hornbuckle, PhD, DVM

Eli Lilly and Company, Inc.
Indianapolis, IN

Lilly

Abstract

- **Objectives:** RWE studies addressing the same research question using the same data and analysis methods may arrive at different results and conclusions. This variability in practice suggests bias, unmeasured confounders, and undesirable heterogeneity that may cause an inability to validate and replicate results. There is a growing body of work about the need for the quality and relevance of real world data as well as the transparency and replicability of RWE. We explore why variability may occur in a fundamental building block of RWE, the “algorithm.” An algorithm is a set of rules to be followed in calculations or problem-solving operations. RWE algorithms are comprised of real-world data, coding schemes, metadata, and logic. We address four types of algorithms: cohort and subgroup identification, exposures, other covariates, and outcomes measures. While algorithms may be transparent, the criteria used to evaluate whether they are “fit-for-purpose,” including their “operating characteristics,” are often not.
- **Methods:** We looked for standards and recommendations from regulatory bodies (i.e., Federal Drug Agency (FDA), European Medicines Agency (EMA), International Council on Harmonization (ICH)), other governmental agencies (i.e., Agency for Healthcare Research and Quality (AHRQ), Patient Centered Outcomes Research Institute (PCORI)), external agencies (i.e., Duke Margolis), payer organizations (i.e., NCQA, NQF), and professional societies (i.e., ISPOR, ISPE). We abstracted the criteria proposed by which algorithms should be evaluated as well as statistics and operating characteristics to be reported.
- **Results:** Our research yielded the following recommended criteria: Clinical importance, Data Quality, Feasibility, Interpretation of scores (Outcomes), Relevance, Reliability or reproducibility, Responsiveness, Safety (i.e., no harm to subjects), Transparency & Replicability, and Validity. Specific statistic recommendations are: Sensitivity, Specificity and Positive & Negative Predictive Values, Sample size available as well as other appropriate measures of validity and reliability. The presentation will define and map these criteria to their source with the evaluation metrics.
- **Conclusion:** Applying these criteria for algorithms allows RWE to be judged both on its transparency and quality.

Rationale and Scope

- There is a growing interest about the quality and relevance of Real World Evidence (RWE) to make it “fit for purpose” for decisions by regulators and other population decision makers.
- While there is a growing body of work about producing relevant RWE with high quality, there are issues still to be addressed. This presentation makes recommendations about a fundamental building block of RWE, which is the “algorithm.”
- The basic premise and rationale for our work is that the appropriateness, applicability and quality of RWE are directly affected by the “suitability” of the algorithms used.
- Specifically, we will address four types of algorithms: cohort identification to include cohort subgroups, covariates, exposures as the primary predictive or explanatory variables, and outcomes measures.

Company Confidential © 2017 Eli Lilly and Company

Source Data

- We drew our recommendations from a targeted review of guidance and/or methods published by regulators, quality organizations or government agencies that contained assessment criteria or characteristics for one or more of our algorithm types. Source documents are given in the appendix.
- From these sources we documented:
  - General Criteria - applicable to any or all of our selected algorithms,
  - Specific Criteria - criteria and operating characteristics applicable to specific algorithm type
- The next slide documents the general criteria that we found with their definition

Company Confidential © 2017 Eli Lilly and Company

General Recommendation for Algorithms

Category	Definition	Source
Clinical importance	RWE algorithm needs to be consistent with current clinical guidelines or with clinical expert judgment.	FDA, EMA, NCQA
Feasibility	Measures should have clear specifications for data sources & methods for data collection & reporting. The following should be addressed: clear specifications, reasonable cost, confidentiality & data availability for the measure.	FDA, NQF, NCQA
Interpretation of scores	Summary of the logic and methods used to interpret the clinical meaningfulness of clinical trial results at the individual patient level on outcomes. This requires an understanding of what is meaningful change.	FDA, NCQA
Relevance or fit for purpose	Evidence must be clinically relevant to the decision or intervention under consideration and has the potential to achieve important outcomes (i.e., either clinical, humanistic or economic) in a healthcare setting or system.	FDA,EMA, NQF, NCQA
Reliability or reproducibility	Reliability is the extent to which evidence is consistent and reproducible over time or in different datasets in multiple uses/scenarios.	FDA,EMA, NQF, NCQA, AHRQ
Responsiveness	The key concept when assessing responsiveness is the “ability to detect change” and identify differences in results over time in individuals or groups, who have changed with respect to the measurement concept.	FDA, EMA, AHRQ
Safety to participants	This is an assessment that process of data collection & evidence generation has no potential for patient harm.	FDA, EMA
Validity	Establish a clinically appropriate definition for the evidence and evidence that it provides an appropriate and correct insight to the desired decision or application at hand. Valid algorithms measure what they are intended to measure	FDA,EMA, NQF, NCQA, AHRQ

Company Confidential © 2017 Eli Lilly and Company

Understanding Operating Characteristics & Concurrent Validity

- As a general recommendation, we recommend assessing how a variable produced by any of the four RWE algorithm types performs on a standard set of descriptive statistics. Variables from algorithms can be assessed by:
  - Analyze the discrete or continuous distribution of the variable
  - Analyze the distribution of variable produced by an algorithm against concurrent variables of interest. Examples are:
    - Age
    - Disease severity
    - Gender
    - Morbidity
    - Race
    - Correlation and distribution of cohorts, exposures and covariates on relevant outcomes

Specific Criteria for Cohort Algorithms

- For cohort algorithms, the analyst needs to know the true prevalence of the condition in the target population of interest to correctly estimate the Positive Predictive Value (PPV) using the Sensitivity and Specificity of the algorithm.
- If evaluating a cohort identification algorithm against a gold standard and the gold standard does not reflect the true prevalence of the disease in the real world target population, then the positive predicted values cannot be accurately estimated directly from the data. There are adjustments to calculate the correct NPV and PPV[1].
- Estimates of true prevalence may be obtained from the literature or databases that have a representative same of the targeted population (e.g., NHANES for estimates of US prevalence).

1. Molinaro A. Diagnostic tests: how to estimate positive predictive value. *Neuro-Onco Pract*. 2015; 2(4):162-166.

Company Confidential © 2017 Eli Lilly and Company

Specific Criteria for Cohort Algorithms

- The FDA Sentinel has done significant amount of research on identifying individual with a disease as an outcome. However, the same principles and criteria apply when identifying a study cohort. The Sentinel initiative recommended a "floor" 0.7 for PPV for as the minimum level of the PPV for a disease (aka cohort) identification algorithm[1]. This criteria was recommended but not set as an immutable threshold.
- Maximizing PPV is necessary since having a significant number of false positives in the cohort makes the results more prone to bias or confounding as well as potentially has an adverse effect on the generalizability of the results.

1. Schmouk GT, Lee TA, Pickard AS, et.al. FDA Mini-Sentinel Project. Mini-Sentinel Methods — Alternative Methods For Health Outcomes of Interest Validation. FDA White Paper. August 31, 2013. Available at: https://www.sentinelinitiative.org/sites/default/files/surveillance-tools/validations-literature/Mini-Sentinel-Alternative-Methods-for-Health-Outcomes-of-Interest-Validation_0.pdf. Accessed November 30, 2018

Company Confidential © 2017 Eli Lilly and Company

Exposures and Covariates

While Exposures and Covariates are fundamentally different variables, they share criteria for assessing their suitability with Exposures having additional considerations.

- **Exposures –** An *exposure* is a known or potential variable of interest assumed to affect an outcome of interest. An exposure is usually an exogenous agent (e.g., medication or procedures) but it may also be environmental (e.g., socioeconomic factors). There may also be endogenous factors (e.g., individual demographics).

- **Covariates** - a *covariate* is a variable that is a known or potentially influential variable related to the outcome under study. A covariate may be of direct interest to the outcome or it may be a confounding or interacting variable. In conducting RWE studies the first step, after defining the cohort, is to select the exposures and outcomes. Once these primary variables are selected, the other variables in the study (whether measured or not measured) are called covariates.

Company Confidential © 2017 Eli Lilly and Company

Conceptual Considerations for Defining Covariates & Exposure Variables

These are recommendations from the Agency for Healthcare Research and Quality (AHRQ)[1].

- **Examine the relationship their relationship with key outcome variables**
  - Is it a Single or cumulative outcome or outcomes?
  - How immediate is the outcome?
  - How rapidly is the factor's effect lost?
  - Are "spillover" effects possible?
  - Is it a linear or other type of relationship?
  - **Bias between the exposure/covariate variables and the outcomes**
    - Are there any issues of bias or confounding between the variables and the outcomes?
    - Does the unit of analysis between the variables and the outcome(s) agree?
    - What are the data sources used for each? If different, what effect might this have on the conclusions one can draw?

1. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Effective Health Care Program. Available at: https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/user-guide-observational-cer-130113.pdf. Accessed November 30, 2018

Company Confidential © 2017 Eli Lilly and Company

Conceptual Considerations for Defining an Exposure Definition

- **Time Window**
  - What was the time window of measurement (calendar time and etiologic relevant time window)?
  - What is desired or the frequency of measurement by time windows within which to define and measure the variable?
- **Induction and Latent Periods (relative to the Time Window)**
  - What was the desired mode or context of delivery by which to measure the variable?
  - Were there any changes in Exposure Status
  - Is there a single exposure involved or multiple exposures?
  - What was the intensity of the exposure(s)?

Company Confidential © 2017 Eli Lilly and Company

Conceptual Considerations for Defining an Exposure Definition

- **For drugs**
  - What is the relationship between the Dosage and Dose-Response?
  - What are the PharmacoKinetics (PK) and PharmacoDynamics (PD) properties of the drug?
  - What is the recommended administration of the drug?
  - What is the recommended frequency of the drug(s)?
  - What were the actual vs. the recommended administration & frequency?

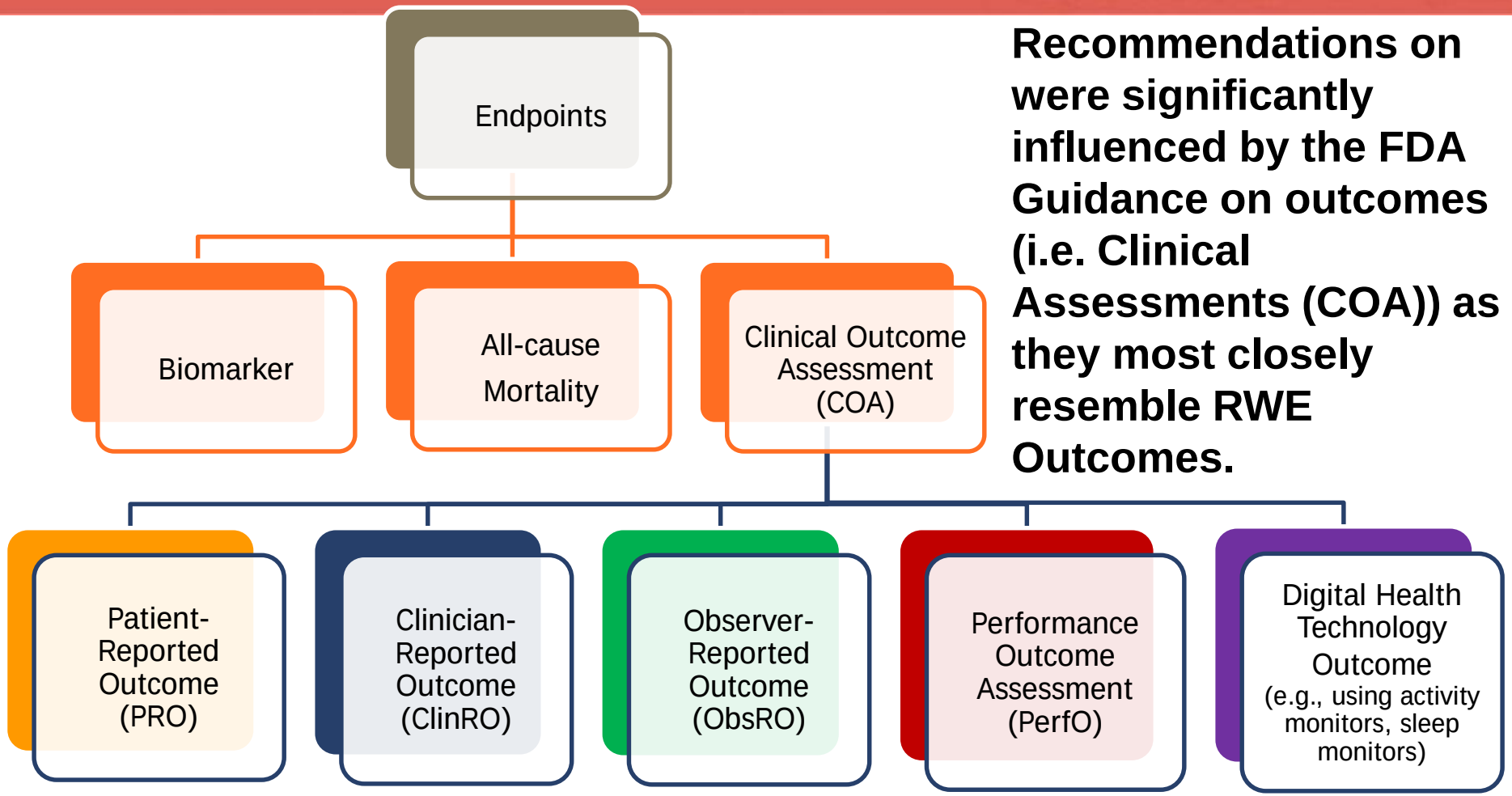
Company Confidential © 2017 Eli Lilly and Company

Conceptual Considerations for Measuring Covariates & Exposure Variables

- **Measurement Error**
  - Are there any known differences in measurement methods between reporters (e.g., data abstractors, labs, etc.)?
  - Are there any changes in the measurement method(s) over time?
  - What were the quality control procedures for the measurement method, including inter-observer or intra-observer reliability?
- **Measurement Scale**
  - The method of measurement (e.g., script fills, patient report, recall, etc.)
  - What is the quantitative representation of the exposure? variable: continuous, categorical, dichotomous measure?
  - What was the frequency of measurement?

Company Confidential © 2017 Eli Lilly and Company

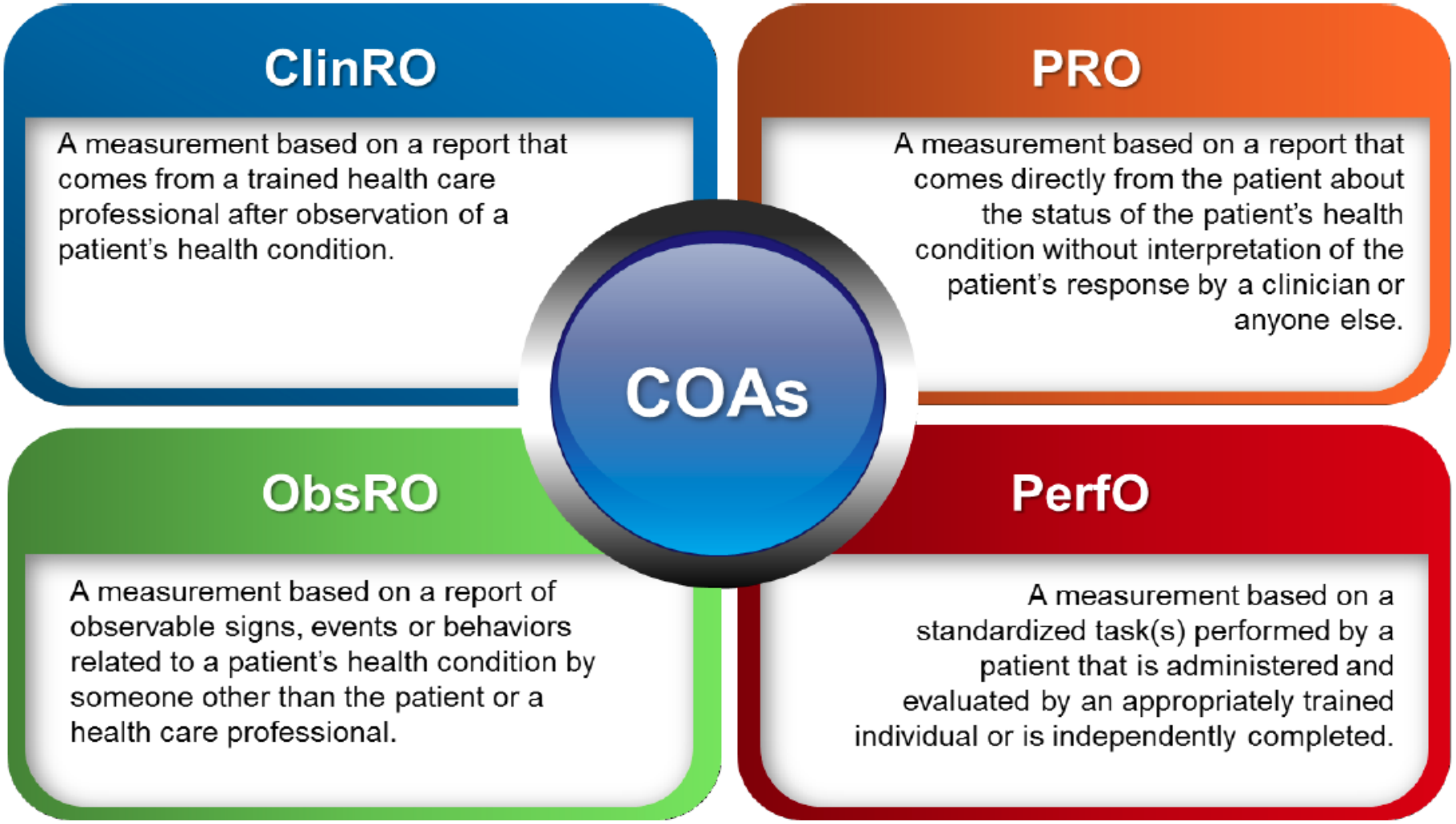
FDA Taxonomy for Endpoints*



Recommendations on were significantly influenced by the FDA Guidance on outcomes (i.e. Clinical Assessments (COA)) as they most closely resemble RWE Outcomes.

**Endpoint*: A precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question

Clinical Outcome Assessments (COAs)



*There are certain types of COAs derived from mobile health technologies (e.g., activity monitors, sleep monitors) that do not fall into one of the other types of COAs.

FDA 2018

Outcomes Recommendations - Validity

Measurement Property	Type	What Is Assessed?
Validity	Content validity	Evidence that the instrument measures the concept of interest including evidence from qualitative studies that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use. Testing other measurement properties will not replace or rectify problems with content validity.
Validity	Construct validity	Evidence that relationships among items, domains, and concepts conform to a priori hypotheses concerning logical relationships that should exist with measures of related concepts or scores produced in similar or diverse patient groups

Outcomes Recommendations - Reliability

Measurement Property	Type	What Is Assessed?
Reliability	Test-retest or intra-rater reliability	Stability of scores over time when no change is expected in the concept of interest
	Internal consistency	Extent to which items comprising an outcomes (e.g., scale) measure the same concept
	Inter-rater reliability	Agreement among responses when the COA is administered by two or more different raters

Outcomes Recommendations - Responsive & Interpretable

Measurement Property	What Is Assessed?
Responsiveness	Evidence that a COA instrument can identify differences in scores over time in individuals or groups (similar to those in the clinical trials) who have changed with respect to the measurement concept
Interpretation of Scores	Regardless of whether the primary endpoint is based on individual responses to treatment or the group response, it is usually useful to display individual responses, often using an a priori meaningful within-patient change threshold (i.e., the individual patient COA score change over a predetermined time period that should be interpreted as a treatment benefit). The meaningful within-patient change threshold used for interpretation may vary by target population or other characteristics

Company Confidential © 2017 Eli Lilly and Company

Conclusions

- In conclusion, the recommendations made here are for guidance only and are not intended as immutable decision rules.
- Decisions about the criteria and characteristics of any algorithm need to be made in the context of decision makers needs and the criteria presented here.
- While we acknowledge that some use cases may warrant more flexibility, we generally recommend that algorithms used in generating RWE meet the highest, feasible standard.

Company Confidential © 2017 Eli Lilly and Company

Definitions

Company Confidential © 2017 Eli Lilly and Company

Company Confidential © 2017 Eli Lilly and Company

Definitions for the General Criteria

- **Clinical Importance –** The algorithm needs to produce results consistent with current clinical guidelines or with clinical expert judgment..
- **Feasibility –** Specifically, the following constructs should be addressed:
  - **Logistical feasibility:** Are the required data available? Are the data reasonably accessible?
  - **Reasonable cost:** Does the measure impose an undue burden on scarce resources (i.e., time or money)? A measure should not impose an inappropriate burden on healthcare systems (e.g., expensive primary data collection).
  - **Clear Specifications:** Are there clear specifications for data sources and methods for data collection and reporting? Measures should have clear specifications for data sources and methods for data collection and reporting.
- **Relevance – Relevance** is assessing whether the algorithm provides important and necessary information to the conclusions drawn from the algorithm in either a decision or its application.
- **Generalizability and representativeness**: Assess the ability of the algorithm to produce findings that address the question posed for a specific decision or application (e.g., matches the target population or the general population of interest)..

Definitions for the General Criteria

- **Potential for Improvement -**Example of where meaningful differences can occur are: Demonstration of quality problems, data demonstrating considerable variation, or overall less-than-optimal performance in the quality of care across providers; and/or disparities in care across population groups
- **Reliability –** Reliability is the extent to which the data results are consistent and reproducible over time or in different datasets in multiple uses/scenarios. There are many specific quantitative measures of reliability that are suited to specific characteristics of the algorithm being tested.
- **Responsiveness –** The key concept when assessing responsiveness is the "ability to detect change" and identify differences in results over time in individuals or groups (similar to those in the clinical trials), who have changed with respect to the measurement concept.
- **Safety considerations –** The major concern in the external guidance is that any measure or algorithm has no potential for patient harm during its collection
- **Validity -** Valid algorithms measure what they are intended to measure. There are specific measures of validity documented for outcomes

Data Sources and Documentation

Company Confidential © 2017 Eli Lilly and Company

Company Confidential © 2017 Eli Lilly and Company

Data Sources used to form recommendations (1)

While there has been no definitive guidance from regulatory agencies on the standards and criteria for RWD or RWE, it is unlikely that the regulatory agencies will significantly change or overturn existing guidance. Therefore, there exists a significant amount current information that can give insight on what the forthcoming guidance on RWE and RWD may require as the necessary standards and characteristics of RWE and RWD. The following are standards and recommendations from the Food and Drug Administration (FDA) and EMA were reviewed as these organizations set or influence the standards for both RWE and RWD.

Notable Regulatory Guidance discussing Real World Data

- FDA 2013 – Pharmacoepidemiology Guidance – "This guidance describes best practices pertaining to conducting and reporting on pharmacoepidemiologic safety studies that use electronic healthcare data, which include administrative claims data and electronic medical record (EMR) data." (1)
- EMA 2015 – Scientific Guidance on Post-Authorization Efficacy Studies (PAES) – "This guidance is intended to provide scientific guidance for marketing authorization holder (MAHs) and for Competent Authorities on PAES in the context of EU regulatory decision-making with regard to: the general need for such studies, general methodological considerations, specific situations and study conduct." (2)
- EMA 2016, 2017 – ENCePP guide on methodological standards in pharmacoepidemiology – "This Guide aims to offer a dynamic and publicly available web resource for methodological guidance in pharmacoepidemiology. It provides links to selected published articles and guidelines that illustrate important principles of pharmacoepidemiological research." (3)
- EMA (2102-2017) - "Guideline on good pharmacovigilance practices (GVP) - Module VIII – Post-authorization safety studies (Rev 3)." A post-authorization safety study (PASS) is defined in DIR Art 1(15) as any study relating to an authorized medicinal product conducted with the aim of identifying, characterizing or quantifying a safety hazard, confirming the safety profile of the medicinal product, or of measuring the effectiveness of risk management measures. (4)

1. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. FDA Guidance for Industry and FDA Staff. Silver Spring, MD. May 2013. https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf
2. European Medicines Agency. Scientific guidance on post-authorisation efficacy studies. 2015. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/11/WC500196379.pdf
3. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. ENCePP guide on methodological standards in pharmacoepidemiology. Revision 5. EMA/95098/2010 Rev. 5. July 2016. Available at: http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml.
4. EMA (2102-2017) - "Guideline on good pharmacovigilance practices (GVP) - Module VIII http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129137.pdf (Accessed April 23, 2018)

Data Sources used to form recommendations (2)

Notable Regulatory Guidance discussing Real World Data

- **FDA 2017 - Use of Real-World Evidence for Medical Devices – "**FDA is issuing this guidance to clarify how we evaluate real-world data to determine whether they are sufficient for generating the types of real-world evidence that can be used in FDA regulatory decision making for medical devices." (5)
- **FDA 2018 – FDA Data Standards Strategy – "**The purpose of the CBER-CDER Data Standards Strategy is to reinforce the ongoing commitment to the development, implementation, and maintenance of a comprehensive data standards program that will facilitate the pre- and post-market regulatory review process so that safe and effective medical products are available to patients." (6)

Regulatory Guidance applicable to RWE

- **ICH E9–Statistical Principles for Clinical Trials** – EMA has not issued any specific guidance on RWE. However, the FDA has cited this guidance for any outcomes submitted to the agency. This guidance is written primarily to harmonize the principles of statistical methodology applied to clinical trials for marketing applications submitted in Europe, Japan and the United States. (7)
- **FDA 2009 - PRO Guidance** – This guidance describes how the FDA reviews and evaluates existing, modified or newly created patient-reported outcome (PRO) instruments used to support claims in approved medical product labeling. Since its release, the recommendations have been applied to any Clinical Outcomes Assessments submitted to the agency. [8]

5. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. FDA Guidance Draft. July 27, 2016. 1-28. https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.
6. Data Standards Strategy: FY2018-FY2022. Center for Biologics Evaluation and Research (CBER) Center for Drug Evaluation and Research (CDER).https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM595163.pdf
7. ICH Harmonised Tripartite Guideline - Statistical Principles For Clinical Trials—E9. February 5, 1998. Available at: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf**.** Accessed November 30, 2018.
8. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Guidance for Industry. December 2009. Available at: https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed November 30, 2018.

Data Sources used to form recommendations (3)

Regulatory Guidance applicable to RWE

- **FDA 2013–Pharmacoepidemiology Guidance** – "This guidance describes best practices pertaining to conducting and reporting on pharmacoepidemiologic safety studies that use electronic healthcare data, which include administrative claims data and EMR data. The guidance includes recommendations for documenting the design, analysis and results of pharmacoepidemiologic safety studies to optimize FDA's review [1]."system. By applying validated algorithms for adverse events/health outcomes of interest (HOI) to individual patient data, new or existing cases of such HOIs can be detected. (9)
- **FDA Mini-Sentinel Project 2013** – The FDA Mini-Sentinel project is a prototype for the Sentinel initiative that is conceptualized as a nation-wide medical product safety surveillance system. By applying validated algorithms for adverse events/health outcomes of interest (HOI) to individual patient data, new or existing cases of such HOIs can be detected. (10)
- **FDA Healthcare Economic Information Guidance** - 2018 – This guidance provides answers to common questions regarding firms' communication of HCEI regarding their prescription drugs and medical devices to payers, formulary committees or other similar entities with knowledge and expertise in the area of healthcare economic analysis (collectively referred to as payers). (11)
- **FDA 2019 – RWE Framework** - FDA's RWE Program will be multifaceted. This document outlines the framework FDA plans to use to implement its RWE Program. (12)

9. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. Guidance for Industry and FDA Staff. Silver Spring (MD). May 2013. Available at: https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf. Accessed November 30, 2018.Data Standards Strategy: FY2018-FY2022. Center for Biologics Evaluation and Research (CBER) Center for Drug Evaluation and Research (CDER).
10. Schmouk GT, Lee TA, Pickard AS, et.al. FDA Mini-Sentinel Project. Mini-Sentinel Methods — Alternative Methods For Health Outcomes of Interest Validation. FDA White Paper. August 31, 2013. Available at: https://www.sentinelinitiative.org/sites/default/files/surveillance-tools/validations-literature/Mini-Sentinel-Alternative-Methods-for-Health-Outcomes-of-Interest-Validation_0.pdf. Accessed November 30, 2018.
11. Drug and Device Manufacturer Communications With Payors, Formulary Committees, and Similar Entities — Questions and Answers. Guidance for Industry and Review Staff. June 2018. Available at: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm537347.pdf. Accessed November 30, 2018
12. **Framework for FDA's Real-World Evidence Program** December 2018 https://www.fda.gov/media/120060/download (Accessed November 22, 2019)

Data Sources used to form recommendations (4)

Standards used by payer organizations for the development and validation of quality measures - Although quality measures are not labeled as 'RWE algorithms' per se, quality measures often include RWE algorithms in defining "outcomes" and "process" measures. Since payers are an important set of stakeholder and decision makers, the standards for cohorts, factors of interest and outcomes measures used as or within quality measures by payers were also investigated:

- **National Quality Forum (NQF)** – NQF measures and standards serve as a critically important foundation for initiatives to enhance healthcare value, make patient care safer, and achieve better outcomes. The Department of Health and Human Services (CMS) relies on the guidance of NQF's Measure Applications Partnership (MAP) standards and criteria. (13)
- **National Committee for Quality Assurance (NCQA) –** NCQA is a private not-for-profit organization dedicated to improving healthcare quality. NCQA has been a central figure in driving improvement throughout the US healthcare system, helping to elevate the issue of healthcare quality to the top of the national agenda. The **Healthcare Effectiveness Data and Information Set (HEDIS)** is one of the most widely used sets of healthcare performance measure in the United States. The **NCQA measurement development process** has expanded the size and scope of HEDIS to include measures for physicians, PPOs and other organizations. It should be noted that all NCQA HEDIS measures have NQF endorsement and have were adopted by CMS and all US payer organizations. The specific standards and requirements adopted by the work stream are described in the HEDIS manual Volume 1., (14,15)
- **Agency for Healthcare Research and Quality (AHRQ) –** AHRQ is the lead Federal agency charged with improving the safety and quality of America's healthcare system. AHRQ develops the knowledge, tools and data needed to improve the healthcare system and help Americans, healthcare professionals and policymakers make informed health decisions. **AHRQ generates measures and data used by providers and policymakers. The AHRQ guide was used for** "Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide" to extract general and specific criteria and recommendations for RWE algorithms (16)

13. Measure Evaluation Criteria. Available at: https://www.qualityforum.org/Publications/2013/10/Review_and_Update_of_Guidance_for_Evaluating_Evidence_and_Measure_Testing_-_Technical_Report.aspx. Accessed November 30, 2018.
14. HEDIS Measures and Technical Resources. Available at: http://www.ncqa.org/hedis-quality-measurement/hedis-measures. Accessed November 30, 2018.
15. HEDIS 2018 – Volume 1, National Committee for Quality Assurance, September 2017.
16. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Effective Health Care Program. Available at: https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/user-guide-observational-cer-130113.pdf. Accessed November 30, 2018

Data Sources used to form recommendations (5)

Standards used by payer organizations for Comparative Effectiveness

Patient Centered Outcomes Research Institute (PCORI) **–** PCORI was founded to improve the quality and relevance of evidence available to help patients, caregivers, clinicians, employers, insurers and policy makers make better-informed health decisions. To do this, PCORI works with those healthcare stakeholders to identify critical research questions and answer them through comparative clinical effectiveness research, or CER, focusing on outcomes important to patients. PCORI has established methodology standards (17)

17. Portfolio of Funded Projects. PCORI. Available at: https://www.pcori.org/research-results/about-our-research/research-methodology/pcori-methodology-standards. Accessed November 30, 2018.