

A Roadmap for Developing Study Endpoints in Real-World Settings

August 28, 2020

Table of Contents

Background	1
What is an Endpoint?	2
Translating Lessons Learned to the Real-World Setting	3
Building a Real-World Endpoint	4
Concept of Interest	4
Outcome	4
Endpoint.....	5
Arriving at a Real-World Endpoint.....	9
Validating a Real-World Endpoint	10
Concept Validation.....	10
Tool Validation	10
Opportunities to Improve the Development of Real-World Endpoints.....	13
Conclusion.....	15
Appendix A. Workshop Participants.....	16
Appendix B. Glossary	20
Appendix C. Types of Endpoints	23
Appendix D. Regulatory Programs, Pathways, and Initiatives that Support Development of Real-World Endpoints	25
Appendix E. Iterative Process of Endpoint Development	27
Appendix F. Examples of Stakeholder Efforts that Support the Development of Real-World Endpoints	28
References.....	30

About the Duke-Margolis Center for Health Policy

The Robert J. Margolis, MD, Center for Health Policy at Duke University is directed by Mark McClellan, MD, PhD, and brings together expertise from the Washington, DC, policy community, Duke University, and Duke Health to address the most pressing issues in health policy. The mission of Duke-Margolis is to improve health and the value of health care through practical, innovative, and evidence-based policy solutions. Duke-Margolis catalyzes Duke University's leading capabilities, including interdisciplinary academic research and capacity for education and engagement, to inform policy making and implementation for better health and health care.

For more information, visit healthpolicy.duke.edu.

Acknowledgments

The Duke-Margolis Center would like to thank several individuals for their contributions to this white paper. This paper's development was guided by the collaboration of the working group members listed on page iv. We are grateful for the working group's extensive subject matter expertise and thoughtful feedback throughout the paper's development. We are also appreciative of the broader Real-World Evidence Collaborative Advisory Group for their guidance. We would like to thank those who participated in the Duke-Margolis private workshops, "Establishing Guideposts for Developing Real-World Endpoints" held September 16, 2019 and February 26, 2020 (a list of attendees can be found in [Appendix A](#)). In addition, we are grateful to Patricia Green from the Center for her communications guidance. We would like to thank Celeste Ferguson from Duke University for her support in developing graphics for this white paper. Finally, we acknowledge Rachael Lussos from Graham Associates for her editorial assistance.

Any opinions expressed in this paper are solely those of the authors and do not represent the views or policies of any other organizations external to Duke-Margolis. Funding for this work is made possible through the generosity of the Margolis Family Foundation, which provides core resources for the Center, as well as a combination of financial and in-kind contributions from Real-World Evidence Collaborative members, including AbbVie; Amgen; Eli Lilly and Company; Pfizer; Genentech, a member of the Roche Group; GlaxoSmithKline; Merck; Novartis; Teva; and UCB. For more information on the Real-World Evidence Collaborative, visit <https://healthpolicy.duke.edu/projects/real-world-evidence-collaborative>.

Disclosures

Mark B. McClellan, MD, PhD, is an independent board member on the boards of Alignment Healthcare, Cigna, Johnson & Johnson, and Seer; co-chairs the Guiding Committee for the Health Care Payment Learning and Action Network; and receives fees for serving as an advisor for Arsenal, Blackstone Life Sciences, and MITRE.

WHITE PAPER

A Roadmap for Developing Study Endpoints in Real-World Settings

DUKE-MARGOLIS AUTHORS

Kerra Mercon

Nirosha Mahendraratnam

Joy Eckert

Christina Silcox

Morgan Romine

Kathryn Lallinger

Adam Kroetsch

Huzyfa Fazili

Marta Wosińska

Mark McClellan

WORKING GROUP

Jeff Allen

Friends of Cancer Research

Aylin Altan

OptumLabs

Shrujal Baxi

Flatiron Health

Marc Berger

ISPOR

Andrew Bevan

Evidera

Ariel Bourla

Flatiron Health

Susan Colilla

Teva Pharmaceuticals

Gracy Crane

F. Hoffmann- La Roche

Cynthia de Luise

Pfizer

Shannon Ferrante

GlaxoSmithKline

Luca Foschini

Evidation Health

Marni Hall

IQVIA

Rohini Hernandez

Amgen

Stacy Holdsworth

Eli Lilly and Company

Kristen Johnson

Novartis

Linda Kalilani

UCB

Lee Kallenbach

Veradigm Health

Eric Klein

Eli Lilly and Company

Laura Koontz

Flatiron Health

Erem Latif

Evidera

Grazyna Lieberman

Genentech

Renato Lopes

Duke University

Nicole Mahoney

Flatiron Health

Annie McNeill

AbbVie

Linda Nelsen

GlaxoSmithKline

Josephine Norquist

Merck & Co.

Elisabeth Oehrlein

National Health Council

Sally Okun

UnitedHealth Group R&D

Lucinda Orsini

ISPOR

Bryce Reeve

Duke University

Matthew Reynolds

IQVIA

Frank Rockhold

Duke University

Jason Simeone

Evidera

Rachel Sobel

ISPE

United Biosource

Corporation

Mark Stewart

Friends of Cancer Research

David Thompson

Syneos Health

Eileen Mack Thorley

PatientsLikeMe

Aracelis Torres

Flatiron Health

Emese Toth

UCB

Stuart Turner

Novartis

Melissa Van Dyke

GlaxoSmithKline

Vince Willey

HealthCore

Wei Zhou

Merck & Co.

ADVISORY GROUP

Aylin Altan
OptumLabs

Marc Berger
ISPOR

Barbara Bierer
The Multi-Regional Clinical
Trials Center of Brigham and
Women's Hospital and
Harvard

Cathy Critchlow
Amgen

William Crown
The Heller School for Social
Policy and Management at
Brandeis University

Riad Dirani
Teva Pharmaceuticals

Nancy Dreyer
IQVIA

Andrew Emmet
Pfizer

Carlos Garner
Eli Lilly and Company

John Graham
GlaxoSmithKline

Solomon Iyasu
Merck & Co.

Ryan Kilpatrick
AbbVie

Lisa LaVange
UNC Gillings School of Global
Public Health

Jacqueline Law
Genentech

Christina Mack
ISPE Representative
IQVIA

Sally Okun
UnitedHealth Group R&D

Eileen Mack Thorley
PatientsLikeMe

Nicole Mahoney
Flatiron Health

Brian Mayhew
Novartis

David Miller
UCB

Bray Patrick-Lake
Evidation Health

Eleanor Perfetto
National Health Council

Richard Platt
Harvard Medical School

Stephanie Reisinger
Veradigm Health

Debra Schaumberg
Evidera

David Thompson
Syneos Health

Richard Willke
ISPOR

Marcus Wilson
HealthCore

Observer

Amanda Wagner-Gee
National Academies of
Sciences,
Engineering, and Medicine

EXECUTIVE SUMMARY

With growing interest in using real-world data (RWD) and real-world evidence (RWE) to support regulatory decision-making, stakeholders are considering how to develop robust real-world study endpoints to evaluate medical product effectiveness when fit-for-use data and valid methods are available. Despite extensive literature and guidance for developing *clinical trial* endpoints, few resources support real-world endpoint development. Some principles can be carried over from the clinical trial setting, but differences in patient populations, care settings, and data collection in the real-world setting result in unique considerations for endpoint development. Additionally, studies conducted in the real-world setting have the potential to capture outcomes that are more relevant to patients than outcomes captured in clinical trials.

This paper explores how key differences in study settings influence a researcher’s considerations for developing study endpoints in the real world. First, because stakeholders involved in the real-world endpoint development process have multidisciplinary backgrounds, this paper details the current landscape of endpoint development, provides standardized definitions of key concepts, and introduces existing frameworks. Second, this paper presents a roadmap for endpoint development, beginning with selection of a concept of interest and study outcome that reflect the research question. Within this roadmap, the paper details how real-world settings impact selection of a concept of interest, outcome, and endpoint components, raising challenges for researchers to consider when developing real-world endpoints. Third, this paper addresses key considerations for the validation of real-world endpoints. Finally, this paper examines opportunities to enhance the use of real-world endpoints through stakeholder collaboration.

How This Paper Was Developed

This paper is informed by a literature review, two private workshops on “Establishing Guideposts for Developing Real-World Endpoints” (September 16, 2019 and February 26, 2020), and the expert opinion of the Duke-Margolis RWE Collaborative RWE Endpoints Working Group. During the workshops, stakeholder experts representing sponsors, academic research groups, data vendors, providers, and patient networks discussed the current and evolving landscape around endpoint development in the real-world setting. This work builds on Duke-Margolis’s recommendations published in: 1) [Adding Real-World Evidence to a Totality of Evidence Approach for Evaluating Marketed Product Effectiveness \(2019\)](#), 2) [Need for Non-Interventional Studies Using Secondary Data to Generate Real-World Evidence for Regulatory Decision Making, and Demonstrating Their Credibility \(2019\)](#), 3) [Determining Real-World Data’s Fitness for Use and the Role of Reliability \(2019\)](#), 4) [Characterizing RWD Quality and Relevancy for Regulatory Purposes \(2018\)](#), and 5) [A Framework for Regulatory Use of Real-World Evidence \(2017\)](#).

Background

Stakeholders are eager to increase the use of real-world data (RWD)—“data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources”—throughout the life-cycle of drug development, approval, and access.^{1*} In particular, stakeholders want to analyze RWD to generate real-world evidence (RWE) about the use, benefits, and risks of medical products and then make that RWE actionable by health care decision makers.¹ FDA is exploring the use of RWD and RWE for regulatory decision-making, per Congressional mandates in the 21st Century Cures Act and 6th Prescription Drug User Fee Act (PDUFA). The December 2018 Framework for FDA’s Real-World Evidence Program is an important step in this exploratory process.¹

Integral to improving the acceptability of real-world studies by regulatory decision-makers is study quality, including data fitness for use and the ability of the methods to support valid causal inference, as well as the regulatory and clinical contexts.²⁻⁶ One key step toward generating regulatory-grade RWE is developing robust and relevant endpoints that can address a research question about a medical product’s safety or effectiveness in the real-world setting: real-world endpoints.

The Value of Real-World Data and Real-World Evidence

RWE studies can complement evidence from randomized controlled trials (RCTs) and contribute to a robust evidence package to support regulatory decision-making. There is a well-established history of the FDA using RWE to support labeling changes related to safety; however, RWE studies might also be useful in labeling changes related to effectiveness. RWD is often collected by providers as part of clinical practice throughout the health system. Therefore, RWD can support analyses that better represent the broader impact of a medical product, including routine clinical care and self-care. RWD can also continuously capture the evolving standard of care, whereas RCTs capture information during a specified timeline. Drawing from RWD, RWE studies often have broader inclusion criteria than traditional RCTs, which might provide insight into the impact of a drug on patients who were not represented in the RCT. RWE studies might also capture outcomes that are more relevant to prescribers and patients. RWE might be generated more efficiently and with fewer resources, increasing the availability of information that might not otherwise be generated.

* A glossary of relevant terms can be found in [Appendix B](#).

What is an Endpoint?

As defined in the FDA-NIH Biomarker Working Group’s Biomarkers, Endpoints, and other Tools (BEST) glossary, an endpoint is “a precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question.”⁷ Endpoints are characterized by the type of research question they aim to answer, the outcomes they capture, and how they are used in the study design (Table 1). Effectiveness endpoints answer research questions that intend to demonstrate that an intervention or exposure results in a clinical benefit, defined as “a positive effect on how an individual feels, functions, or survives.”⁷ Endpoint types are characterized by the manner in which the outcome (or outcomes) are captured. Endpoints are also grouped within the statistical hierarchy,[†] as defined by the study design.⁸ Endpoints can also be classified according to whether they are novel compared to commonly accepted endpoints. (For more information on the types of endpoints, including a discussion on endpoint novelty, see [Appendix C](#)).

Table 1. Endpoint types are categorized by how outcomes are captured and by their position in the statistical hierarchy.

	ENDPOINT TYPE	DEFINITION
Outcomes Captured	Single-measure	Single variable that reflects a single outcome of interest
	Composite	Combination of “clinical outcomes into a single variable” ⁸
	Multi-component	Combination of components or domains to create a single score according to specified rules ⁸
	Intermediate	“Clinical outcome that can be measured earlier than an effect on irreversible morbidity or mortality (IMM) and that is considered reasonably likely to predict the medical product’s effect on IMM or other clinical benefit” ⁷
	Surrogate	“Substitute for a direct measure of how a patient feels, functions, or survives” ⁷
	ENDPOINT POSITIONING	DEFINITION
Statistical Hierarchy	Primary	“Establish the effectiveness, and/or safety features, of the drug in order to support regulatory action” ⁸
	Secondary	“To demonstrate additional effects after success on the primary endpoint” ⁸
	Exploratory	“Include clinically important events that are expected to occur too infrequently to show a treatment effect or endpoints that for other reasons are thought to be less likely to show an effect but are included to explore new hypotheses” ⁸

Researchers balance the tradeoffs of an ideal real-world endpoint with practical considerations, such as the feasibility and relevancy of the endpoint. For example, a composite endpoint may best answer a research question, but if capturing multiple outcomes in the RWD source is not feasible or presents a significant burden for providers, a different endpoint might be considered.

Characterizing the endpoint by the outcomes it captures and the endpoint’s position in the statistical hierarchy is necessary for determining the appropriate statistical analyses. Positioning in the statistical hierarchy can also impact the regulatory acceptability of the endpoint. For example, secondary and exploratory endpoints might be less likely to inform a product’s label.

[†] The statistical hierarchy refers to a grouping of endpoints by clinical importance, expected frequency of the event, and anticipated drug effects.

Translating Lessons Learned to the Real-World Setting

Developing real-world endpoints is challenging due to the lack of adequate literature, standardized best practices, and regulatory guidance that address the differences in endpoint development between the clinical trial and real-world settings. Differences in data collection practices, patient populations, and care patterns in the real-world setting might require certain endpoint components that a clinical trial for the same disease or condition might not use. The uncertainty introduced by these differences may also require analytical and study design approaches distinct from the approaches used in clinical trials.

Although literature on developing real-world endpoints is limited, many lessons can be learned from clinical trial endpoint development, which has been detailed extensively for decades across peer-reviewed publications, multi-stakeholder standards-setting bodies, and international collaborative efforts. FDA itself has outlined many key considerations for clinical trial endpoint development in at least four cornerstone guidance documents:

- *Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics*
- *Multiple Endpoints in Clinical Trials*
- *Expedited Programs for Serious Conditions – Drugs and Biologics*
- *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*

FDA does not state a preference for type of endpoint chosen to demonstrate effectiveness. However, the *Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products* guidance notes that “the most straightforward and readily interpreted endpoints are those that directly measure clinical benefit or are validated surrogate endpoints shown to predict clinical benefit.”⁹

Throughout its endpoint guidances, FDA also references the estimand framework: a structured framework on developing a regulatory-grade research question to determine if an intervention or exposure results in a clinical benefit.^{10,11} FDA provides feedback on endpoint development through a variety of mechanisms summarized in [Appendix D](#).

Endpoint development is framed by the clinical and regulatory contexts surrounding the research question. Clinical context includes the understanding of the disease, treatment alternatives, therapy, patient perspective, and provider perspective.^{6,12} Important regulatory context factors include the intended purpose of the endpoint (including labeling), the available regulatory review and approval pathways, and the relevant information and evidence from any previous regulatory decisions for the given disease or condition. For example, endpoints used previously to support a regulatory approval may have greater acceptability to support labeling changes for other medical products. It is important to note that regulatory acceptability is based on the evaluation of clinical studies through the totality of evidence approach, where the evidence base to support effectiveness is consistently growing and evolving and real-world studies are often not evaluated in isolation.⁶

⁹The estimand framework is detailed in *ICH E9(R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials Harmonised Guideline*. More on how the estimand framework relates to endpoint development, including a case study, can be found in the Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 4: *Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making* (PFDD Discussion Document 4).

Building a Real-World Endpoint

Figure 1 depicts a new roadmap for developing a real-world endpoint. This roadmap is to be applied to research questions studied using real-world data in conjunction with tools such as the estimand framework or target trial approach.^{10,13}

The real-world endpoint may be similar to the commonly accepted endpoint used to support clinical trials for the same disease or condition. However, even “standardized” clinical trial endpoints often differ in definition across trials. For example, major adverse cardiovascular events (MACE) is a commonly used composite endpoint to assess cardiac outcomes; however, definitions of MACE differ across clinical trials.¹⁴ Therefore, clearly defining the endpoint through the selection of the concept of interest, outcome, and endpoint components is vital for any clinical trial or real-world study.

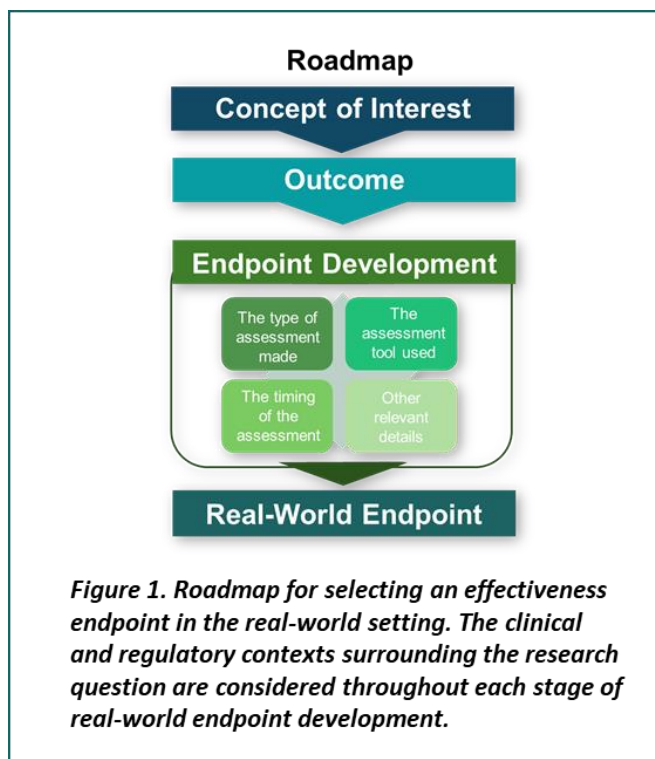


Figure 1. Roadmap for selecting an effectiveness endpoint in the real-world setting. The clinical and regulatory contexts surrounding the research question are considered throughout each stage of real-world endpoint development.

Concept of Interest

The concept of interest (COI) is the “aspect of an individual’s clinical, biological, physical, or functional state, or experience that the outcome assessment is intended to capture or reflect.”⁷ For each disease or condition, a variety of COIs (e.g., functional status, mental health) are applicable.^{15,16} The COI depends on the research question and the clinical benefit of interest. The COI can also be informed by patient input, the natural history of the disease, the aspect of the disease modified through a study, or the targeted labeling.¹⁷

The COI is likely consistent regardless of study setting. However, if a different COI is easier or more available to measure in the real-world setting (e.g., clinical vs. physical) or more clinically relevant, that COI may be used instead. Choice of COI may also depend on the purpose of the study: to inform regulatory decision-making, payer decision-making, or the standard of care in clinical practice.

Outcome

After the COI is chosen, an outcome can be selected. An outcome is a “measurable characteristic that is influenced or affected by an individuals’ baseline state or an intervention as in a clinical trial or other exposure.”⁷ Most clinical studies that support regulatory decision-making examine clinical outcomes (e.g., change in blood pressure, occurrence of stroke) or humanistic outcomes (e.g., leg mobility, health-related quality of life). In contrast, economic outcomes (e.g., cost per hospital stay day, incremental cost effectiveness ratio) related to medical products may support payer and health system decision-making.¹⁸

In many cases, the outcome measured in a study is the same regardless of setting because the outcome chosen for the clinical trial was carefully selected based on the disease definition and the impact of treatment on the disease. However, the outcome may change if the measurement of treatment benefit is not captured in the real world in the same way as clinical trials. For example, cancer progression is measured using RECIST for clinical trials, but may be monitored by radiographic images or tumor markers in the real world. Furthermore, researchers may consider whether there are outcomes more routinely captured in RWD that might better reflect the COI and whether the outcome is associated with an event is likely to be medically attended.

Endpoint

An endpoint is developed to measure the outcome. An endpoint is made of four components:

- 1) Type of assessment made
- 2) Assessment tool used
- 3) Timing of the assessment
- 4) Other relevant details.⁷

Each component is selected to reflect the COI and address the research question. The sequence in which each component is selected, and subsequent iteration, depends on the clinical and regulatory contexts.

The following sections define each of the four components and discuss specific considerations for choosing each component in the real-world setting.

1

Type of Assessment

The type of assessment refers to the three types of outcome assessments to evaluate clinical benefit: survival, clinical outcome assessments (COAs), and biomarkers. Survival often has a “well-defined means for determination.”¹⁹ Generally, COAs measure symptoms, and biomarkers measure a patient’s physiological state.

If the research question is the same for both the clinical trial and real-world settings, the type of assessment may be the same. The type of assessment may change if there is a better way of measuring the clinical benefit in the real-world setting (e.g., using an electronic PRO [ePRO] rather than a ClinRO to capture patient experience). Availability of the assessment tool in the real world may also impact the decision (e.g., a PerfO [e.g., spirometry] may be used to evaluate COPD exacerbations in a clinical trial, but symptoms captured through ePROs may be used in the real world).

Types of Outcome Assessments

1. **Survival:** Duration of survival.
2. **Clinical outcome assessments (COAs):** Measurements of how patients feel and function, influenced by the judgement of a person (respondent).^{7,19} The four types of COAs are clinician-reported outcomes (ClinROs), patient-reported outcomes (PROs), observer-reported outcomes (ObsROs), and performance outcomes (PerfOs).⁷
3. **Biomarkers:** Measurements of “normal biologic processes, pathogenic processes, or responses to an exposure or intervention” that serve as an objective, indirect patient assessment (e.g., protein levels in a blood sample).^{7,19} Biomarkers are often used in surrogate endpoints.

The assessment tool is chosen to measure the outcome assessment. Traditionally, tools to measure COAs have included paper or phone questionnaires, while biomarkers have been measured through molecular, histologic, radiographic, and physiologic tools.⁷ Many tools used in clinical trials may not be practical, cost-effective, or relevant for use in the real-world setting. For example, frequent use of MRIs to measure an outcome is likely not possible as part of routine clinical care. Although some tools may be used in both clinical trial and real-world settings, real-world tools should be chosen in accordance with relevance to patient care, regardless of whether the real-world tool is closely related to the commonly accepted tool. Secondary use data algorithms and digital measurement tools are two types of tools often used in real-world studies.

Secondary Use Data Algorithms

In the real-world setting, tools to measure outcomes may rely on primary data capture, as is typical for clinical trials, or secondary use data. Common secondary use data sources include electronic health records (EHRs), insurance claims, patient-generated health data, laboratory values, or genetic, biometric, or diagnostic reports. For secondary use data, an outcome might not be routinely collected or reported to the data source. Whether the outcome (or any variable) is captured in the dataset depends on whether the primary purpose of that data source has a systemic reason to report the outcome. For claims data, a code associated with the outcome is required for billing, whereas an EHR relies on clinical observation and reporting of an outcome. If the outcome is not included within a dataset, a researcher may be able to extract key variables from raw data (when available and accessible) or link the research dataset with another data source with the relevant outcome information. Alternatively, researchers can use an algorithm to extract the outcome, extract a variable selected as a “proxy” for the outcome, or derive the outcome based on available data from one or more sources.

Developing an algorithm to address a research question is a multistage process. First, the researcher must determine if a commonly accepted standard for assessment of the outcome exists. If no commonly accepted standard exists or the standard is not accessible, the researcher must determine whether a clinically objective measurement exists. Some outcomes, such as lupus flares, do not have commonly accepted standards for assessment or clinically objective measurements. Developing an algorithm to assess lupus flares is therefore more difficult than for diseases or conditions with clinically objective measurements (e.g., blood pressure as a biomarker for hypertension).^{20,21}

Because many sources of RWD are not collected specifically for research use, researchers must address the reliability of the data, including how to interpret data gaps. In most cases, the data is not truly “missing” but rather has not been documented. For example, data may not be present in an EHR because the clinician did not feel the test was necessary, the test was not accessible, or the results of the test were not recorded in the EHR. Another limitation of developing an algorithm for EHR data is that the data usually reflects interactions with a particular clinician or health care system and is not representative of the patient’s entire health care experience. In claims data, challenges exist with the coding systems. Because multiple coding systems (e.g. ICD, WHO) have multiple versions, researchers must understand which coding system was used when the algorithm was developed. Researchers also must account for miscoding in claims data.²² Additionally, the recorded diagnosis may be uncertain. For inpatient settings, ICD-10 guidelines state that “If the diagnosis documented at the time of discharge is qualified as ‘probable,’ ‘suspected,’ ‘likely,’ ‘questionable,’ ‘possible,’ or ‘still to be ruled out,’ ‘compatible with,’ ‘consistent with,’ or other similar terms indicating uncertainty, code the condition as

if it existed or was established.”²³ This practice may make it difficult for researchers to determine if the diagnosis was the true diagnosis or a probable diagnosis.

Multiple RWD sources may be used for algorithm development, and these sources may be discordant. As such, an algorithm derived from claims data will likely differ from an algorithm derived from EHR data. Depending on the sources of the data, some endpoint types may be more feasible to use than others. For example, composite and multi-component endpoints may be difficult to obtain in claims data if a patient’s comorbidities are not consistently coded upon hospital and clinician office visits.

Digital Measurement Tools

Digital measurement tools are increasingly used to measure COAs or biomarkers in both clinical trials and real-world studies.^{24,25} Digital measurement tools refer to both devices used in clinical care and patient-generated health data collected through mobile health technologies and consumer devices.²⁵ Digital measurement tools may be electronic versions of traditional tools (e.g., paper questionnaires) or tools that measure an outcome in a different way than the traditional tool.²⁴ For example, ePROs can be captured through digital questionnaires, potentially administered through apps or texts sent to patients, or captured in the EHR. Digital questionnaires can also be used to capture ClinROs or ObsROs. PerFOs are typically measured digitally through active sensors as a patient knowingly performs a task.²⁵

Digital biomarkers (i.e., “objective, quantifiable, physiological, and behavioral measures that are collected by means of digital devices that are portable, wearable, implantable, or ingestible”) may be collected through active or passive sensor data.^{24,26} For example, a continuous glucose monitor is passive sensor data, collected at pre-programmed intervals or contexts without patient involvement. A standard glucose monitor that requires the user to initiate a finger prick is active sensor data.²⁵

In the real-world setting, digital measurement tools can capture a more complete picture of the outcome than other tools. Because digital measurement tools have the capability to capture continuous data over long periods or allow for frequent discrete data capture, more information can be gathered about patient experience, including a patient’s symptoms and physiological being during their daily life.^{24,26} For example, continuous glucose monitors collect an uninterrupted stream of longitudinal data, while standard glucose monitors are used at specific time points.

The types and uses of digital measurement tools may differ in the real-world setting from the clinical trial setting. For example, consumer-grade devices are more common in the real-world setting than in the clinical trial setting. With consumer-grade devices, the researcher does not have control over when firmware or software updates are applied or what those changes do to the resulting data. Furthermore, patients and caregivers often do not receive training on how to use these tools correctly. RWD from digital measurement tools may also be biased based on the characteristics of the population who own and use the tool, how frequently each patient uses the tool, how well the patient uses the tool according to the manufacturer specifications, the feasibility of using the tool given connectivity requirements (e.g., smartphone), and the affordability of the tool. During analysis, researchers might have to consider the motivation for use of the tool and how to address selection bias.

⁵ It is important to acknowledge that use of a digital tool does not classify the endpoint or study as “real-world.”

In both the clinical trial and real-world study settings, the timing of assessments must be clinically relevant, incorporate a baseline measurement, and reflect when changes in outcomes are expected to occur due to disease or treatment.¹¹ In the clinical trial setting, measurements often follow a strict assessment frequency schedule. When deviations from the pre-specified assessment schedule occur due to missed or unexpected site visits, they are often accounted for through pre-specified windowing.

In the real-world setting, patient behavior, environmental characteristics (e.g., insurance design), appointment scheduling, start and end of treatment, and length of follow-up can all impact the timing of assessments. Because patient care in the real-world setting is not standardized, the timing of assessments between and within patients will vary, posing study design and analysis challenges. Assessments that are administered too frequently may contribute to patient or administrator fatigue or inadvertently alter patient care.¹¹ In contrast, less frequent assessments may result in data that has not been recorded for the time point of interest (e.g., COA administered outside of the acceptable recall period). The timing of the outcome assessment in the real-world setting must limit interference in patient care, to avoid impacting the quality of care or the generalizability. For example, more frequent assessments may bias the analysis because the information is recorded at faster rates than is common in routine clinical practice.

The importance of these timing variations depends on the research question and purpose of the endpoint. For example, if using RWD as a comparator arm in a single-arm trial, timing variations may greatly impact analysis and interpretation. For many sources of RWD, the assessment frequency cannot be pre-specified, but the timing of collection can be pre-specified (e.g., forming a retrospective cohort in a claims database). Ultimately, researchers must consider if variable timing is acceptable to assess the outcome for the given research question or whether to use a different tool.

Researchers must understand how the motivation for the timing of assessment impacts interpretation of the measurement. Daily measurements to monitor disease progression reflect a different outcome than single or infrequent measurements for diagnosis during clinic visits. For example, frequent spirometry may assess overall lung function, while spirometry performed in response to patient concern about worsening lung function may assess exacerbations. The characteristics of the cohort may also impact timing variations. For example, a patient receiving treatment intravenously may have more office visits than patients taking oral tablets. The timing of assessments may dictate whether the measurement is performed in a clinic or within a patient's home, which may affect the precision of the measurements.

The other relevant details that may impact endpoint development largely depend on the research question. To define the endpoint, researchers must choose the causal contrast measure or statistic (e.g., time to event, time to deterioration, percent of responders, change in score) that appropriately measures the change in the outcome. Researchers must choose the causal contrast measure or statistic in accordance with the COI, outcome, study design, and population of interest. For example, measuring

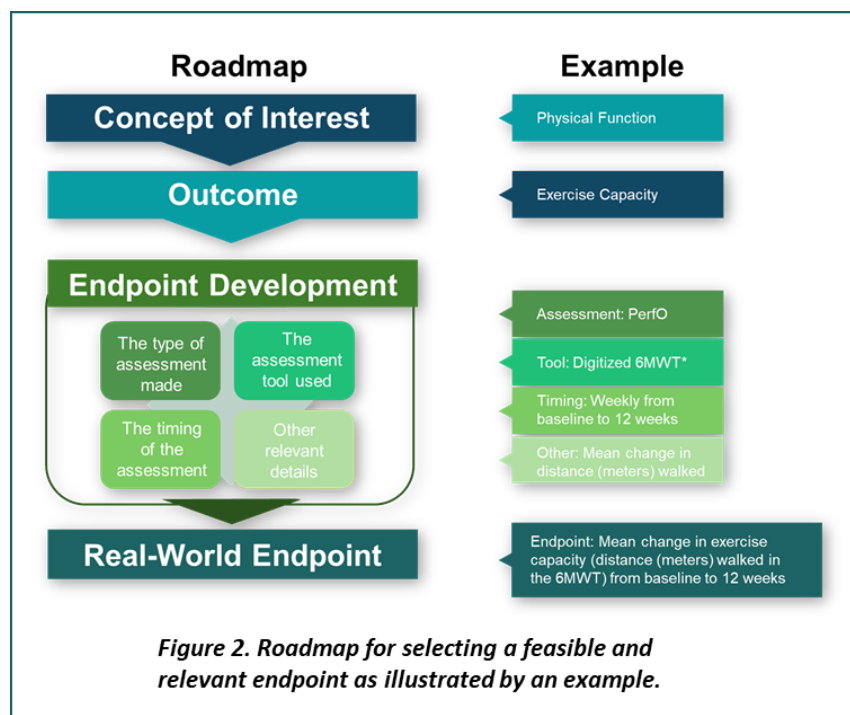
the percentage of patients with asthma-free days may be sufficient for patients with mild asthma, but measuring reduction in asthma symptom severity may be more appropriate for patients with severe asthma.

When developing the protocol for answering a research question, researchers must pre-specify their study design and statistical analysis plan as well as discuss any potential challenges. In some instances, researchers may use a different endpoint to mitigate specific challenges. Potential challenges may include conflicting data, intercurrent events, unmeasured confounding, and data gaps.^{4,5} Researchers also want to ensure that the variables required for statistical adjustments have been captured. As is the case for clinical trials, researchers must consider how the measure of center and variability for the summary statistic impacts endpoint interpretation.

Arriving at a Real-World Endpoint

When designing a real-world study to answer a research question, choosing the COI, outcome, and endpoint can be an iterative process. As many outcomes can reflect a single COI, multiple outcome assessments can be used to evaluate a single outcome. For example, a study with an outcome of acute myocardial infarction may use a combination of symptoms of ischemia, as measured through a COA, and the rise or fall of cardiac troponin, a biomarker.²⁷ Multiple endpoints may be defined to address a research question, and these endpoints may reflect multiple outcomes and COIs. For a visual explanation and full example, see [Appendix E](#).

Figure 2 presents examples for each endpoint component for how a researcher might assess the impact of a drug on patients with heart failure. To address this question, the COI is physical function and the outcome is exercise capacity. The type of assessment is a PerfO, the tool is a digitized 6 Minute Walk Test (6MWT), and the timing is weekly from baseline to study completion at 12 weeks. Other relevant details include the causal contrast measure: mean change in distance (meters) walked.



Validating a Real-World Endpoint

A real-world endpoint must undergo validation to demonstrate its ability to elucidate the treatment effect. Validation is “a process to establish that the performance of a test, tool, or instrument is acceptable for its intended purpose” and has different elements based on the outcome assessment and the context of use (COU).^{7,28} Endpoint validation requires both validation of the concept and of the tool component of the endpoint, discussed below.

Researchers should approach validation as a continual process. Multiple studies may be needed to establish a body of evidence to support the validity and reliability of a tool. A tool may have strong evidence for validity and reliability in one population but may need additional steps to ensure that the tool is fit for use to assess the COI in the target population. Regulatory acceptability of the endpoint depends on how well the body of evidence demonstrates that the tool is appropriate for the COU. **

Concept Validation

Concept validation is demonstrating that the outcome and the intended purpose of the tool reflect the COI. Each type of outcome assessment and tool may capture different components of the COI. Researchers must ensure that the tool actually captures the component of the COI that is reflected in the outcome. Considering the clinical context and conceptualizing the clinical benefit (i.e., identifying the COU) are critical for determining if the outcome reflects the COI. Whether the outcome reflects the COI is often setting-agnostic.

Tool Validation

Tool validation determines whether the measurement tool meets its intended purpose by accurately and adequately capturing the outcome. As part of the validation process, different measurement properties determine whether the tool is appropriate for its COU. The names and definitions of these measurement properties differ depending on the type of tool and whether the tool measures a biomarker or COA, but the ultimate goal of the measurement properties is the same. The measurement properties seek to demonstrate that the tool adequately measures the COI, maintains logical relationships with related measures, and has consistent and reproducible measurements. The measurement properties also assess the relevant performance characteristics for the tool and, if applicable, the associated sensors (Table 2). Demonstrating that the tool measures the COI is the most

Context of Use (COU)

COU is the “statement that fully and clearly describes the way a medical product development tool is to be used and the medical product development-related purpose of the use.”⁷ For COAs, additional factors that help define COU include “disease definition (e.g., disease subtype, disease severity, history of previous treatment), target population (e.g., demographics, culture and language), clinical practice and trial setting (e.g., inpatient, outpatient, controlled/uncontrolled trial), and endpoint positioning (e.g., primary, co-primary, secondary, exploratory).”⁹ For biomarkers, the COU includes the BEST biomarker category* and the biomarker’s intended use and is written in a standard form.^{7,28} The COU depends on the study design, study setting, and the data collected (e.g., primary or secondary use data).

* The BEST biomarker categories include: diagnostic, monitoring, predictive, prognostic, pharmacodynamic/response, safety, and susceptibility/risk.

** For more information on COU, reference: PFDD Discussion Document 3, PFDD Discussion Document 4, and Biomarker Qualification: Evidentiary Framework Guidance.

important aspect for validation and should be demonstrated before other measurement properties.¹⁷ The measurement properties for COAs and biomarkers are applicable regardless of study setting and generally can be applied to digital measurement tools and secondary use data algorithms.

Table 2. Measurement properties that researchers *might* demonstrate as part of the validation process for each type of tool. ** 17,24,28,29

	SENSOR PERFORMANCE MEETS TECHNICAL SPECIFICATIONS (Verification)	TOOL MEASURES COI (Content Validity, Clinical Validation)	TOOL HAS ACCEPTABLE PERFORMANCE CHARACTERISTICS (Analytical Validation, Ability to Detect Change)	TOOL HAS CONSISTENT AND REPRODUCIBLE MEASUREMENTS (Reliability)	TOOL MAINTAINS LOGICAL RELATIONSHIPS WITH RELATED MEASURES (Construct Validity)
Traditional COAs		✓	✓	✓	✓
Traditional Biomarkers		✓	✓	✓	✓
Algorithms		✓	✓	✓	✓
Digital Measurement Tools for COAs	✓	✓	✓	✓	✓
Digital Biomarkers	✓	✓	✓	✓	✓

** The commonly used terms to describe each attribute are in parentheses.

The amount of validation testing required must be evaluated on a case-by-case basis. However, extensive testing is likely to be required if the tool is very different than previous tools used to support regulatory approvals. The types of validation needed are based on the measurement properties that are logical to demonstrate for the tool and research question (sometimes referred to as “common sense” validation).

Clinical Outcome Assessments

For COAs, researchers must demonstrate these measurement properties: the tool measures the COI (i.e., content validity), the tool maintains logical relationships with related constructs and distinguishes between groups known to be different (i.e., construct validity), the measurements are consistent and reproducible (i.e., reliability), and the tool has acceptable performance characteristics (i.e., ability to detect change).^{17,30††} Demonstrating that the tool has the ability to detect change is the most difficult performance characteristic to define. Other measurement properties can be demonstrated with data from a single assessment (i.e., cross-sectional study). Conversely, demonstrating the ability to detect change requires longitudinal data collection under circumstances when the effect of disease or treatment is “known.” Funding for longitudinal psychometric studies can be difficult to obtain. Beyond the key measurement properties, researchers might consider other factors such as the interpretability of score, availability of language translations, accessibility for low literacy populations, and evidence of mode equivalence when using a COA.^{17,31,32††}

Biomarkers

For biomarkers, researchers must demonstrate these measurement properties: the tool measures the COI (i.e., clinical validation), the tool has acceptable performance characteristics (i.e., analytical

†† Additional information on the demonstration of measurement properties for COAs can be found PFDD Discussion Document 3 and *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims Guidance to Industry*.

†† More on these aspects can be found in the PFDD Discussion Documents and Guidances and in resources provided by the Critical Path Institute PRO Consortium.

validation), and the results are consistent and reproducible (i.e., reliability).²⁸ Researchers may also assess whether the biomarkers maintain logical relationships with related concepts and can differentiate between groups known to be different (i.e., construct validity). Additionally, the usefulness and feasibility of the tool are essential to consider. Usefulness includes usability factors, such as if the tool has features that are easy to use, and utility, such as whether the tool has features that are required to measure the outcome.³³ Feasibility of using a tool is based on the disease or condition and COU, among other factors.

Traditionally, many validation studies for COAs have been conducted in the real-world setting. Furthermore, the *Biomarker Qualification: Evidentiary Framework Guidance for Industry and FDA Staff* states that while “the strongest level of evidence to support the association of a biomarker with an outcome of interest comes from prospective studies that are specifically designed and powered to assess the association...data from studies conducted for other purposes are used to support biomarkers qualification...[C]linical trial data is not critical for all COUs.”²⁸ However, validation of measurement properties is just the first step in demonstrating that a tool may be used as part of an endpoint.

Secondary Use Data Algorithms

Often algorithms that identify outcomes in secondary use data are validated by comparing key measurement properties such as sensitivity, specificity, negative predictive value, positive predictive value, accuracy, and internal reliability to a commonly accepted standard. Examples of commonly accepted standards include manual chart review, adjudication, clinician confirmation, patient confirmation, or comparison to established national databases (e.g., National Death Index).³⁴⁻³⁷ The most appropriate standard for comparison depends on the research question. Depending on the method of abstraction, additional measurement properties may be required. For example, inter- and intra-reliability may be necessary to assess the consistency across human abstractors.³⁸ The implications of suboptimal performance must be understood when interpreting the study results.

Typically, algorithms are validated within a single dataset. To validate an algorithm across datasets, researchers must consider additional factors, such as differences in practice patterns, populations, and health care systems.

Digital Measurement Tools

Digital measurement tools share the same measurement properties as traditional COAs or biomarkers, respectively.^{24,29} However, digital measurement tools must also be assessed to ensure that the sensor performance meets technical specifications (i.e., verification) when applicable.^{24,29} These measurement properties are applicable regardless of whether the digital measurement tool is an electronic version of a traditional tool or is measuring the outcome in a new way. Challenges arise when no objective measurement or commonly accepted standard exists for comparison to ensure that the tool has acceptable performance characteristics. In this instance, performance may be based on the ability of the tool to measure the COI and its ability to produce consistent measurements. The underlying IT system and algorithms must be stable to prevent issues such as the unintentional loss or replication of data. Feasibility,⁵⁵ usability, and utility of the tool and user interface are also important to consider because patients use the tools outside the clinical trial setting and compliance depends on how well the tool is integrated into the daily lives of patients.³⁹

⁵⁵ CTTI has developed an Interactive Database of Feasibility Studies of different digital health tools.

Opportunities to Improve the Development of Real-World Endpoints

Endpoints vary in their meaningfulness across multiple decision makers, and understanding this variation is necessary to support efficient drug development and patient access. For example, selecting outcomes that are truly meaningful to patients, but can also demonstrate a clinical benefit in a study can be challenging. As a specific example, hemoglobin A1c is commonly used for determining long-term risk of diabetes complications and as a measure to assess the effectiveness of therapies, but glucose time in range may be more meaningful to patients and provide more detailed insights for providers to manage care.^{40,41} Stakeholders must continue to engage with patients to effectively incorporate the patient experience into drug development, including capture of the most relevant data at the point of care. Stakeholder initiatives such as mCODE and OneSource are underway to support this effort by standardizing capture of key data elements to improve use of EHR data for clinical research.^{***42,43}

The lack of endpoint standardization across a therapeutic area creates difficulties for regulators comparing therapeutics with the same indication across pivotal trials to determine the meaningfulness of results. Understanding the anticipated clinical benefit for each therapy is critical especially when there are small effect sizes that are susceptible to variation. Comparing studies is further complicated when researchers use a novel endpoint, especially when no commonly accepted standard exists for comparison. Further clarity is needed on how regulators assess novel endpoints as part of an evidence package.

The challenges in assessing the meaningfulness of an endpoint and creating potential standards for endpoint development illustrates a real need for additional pre-competitive collaboration among sponsors. In this space, sponsors may be able to harmonize real-world endpoints across some therapeutic areas. Sponsors also may be able to collaborate on some aspects of real-world endpoint validation to begin to build the body of evidence, understanding that some later aspects of validation are case-specific and potentially proprietary. This increased openness about real-world endpoint development can decrease potentially duplicative work and overall cost burden, as sponsors may be able to leverage existing tools. Additionally, this openness could make it easier to understand patient experiences with different treatments for the same disease or condition.

Pre-competitive collaborations could also be established between data aggregators and health care technology companies to harmonize on endpoint development with secondary use data. For example, Friends of Cancer Research Pilot Project 2.0: Establishing the Utility of Real-World Endpoints assessed real-world oncology outcomes in ten different RWD sources.⁴⁴ Data companies might work together to standardize criteria to assess the fitness of use of a dataset for capturing or deriving outcomes.

Collaborations among researchers can also support initiatives to improve outcome identification. For example, within the Sentinel system, researchers are developing computational phenotypes so that data using the same common data models can use a standardized algorithm to identify outcomes.⁴⁵⁻⁴⁹

There may also be a need for multidisciplinary collaboration across the entire health care system, including developers, sponsors, patients, researchers, practitioners, and payers. This multidisciplinary

***See [Appendix F](#) for additional ongoing stakeholder efforts that address uncertainties around real-world endpoints.

collaboration might improve the capture, standardization, and quality of RWD to develop harmonized real-world endpoints. These harmonized endpoints can not only inform regulatory decision-making but also increase payer confidence in the potential outcomes associated with a product's use.

Endpoint harmonization can also benefit from an FDA published library of validated real-world endpoints that supported a regulatory approval. Publishing a list of validated endpoints—in addition to FDA's lists of qualified COAs and biomarkers and surrogate endpoints with associated COUs—may also improve the use and acceptability of novel endpoints. Because endpoint acceptability changes over time, stakeholders should be aware of endpoints that were recently validated for a COU to support approvals across a therapeutic area.

Currently, sponsors get feedback from FDA as part of review pathways, where sponsors can discuss the study protocol and statistical analysis plan, as appropriate.^{†††} Sponsors are encouraged to discuss with FDA the use of COAs and biomarkers that have not been used to support regulatory approval in the past. However, limited opportunities exist for tool developers that are not sponsors to get similar feedback from FDA. While non-sponsors can use FDA's current qualification programs, these programs can be complex and lengthy. Exploring additional opportunities for gaining feedback from FDA on real-world endpoint development would be helpful. To improve collaborative discussions with FDA on real-world endpoints, stakeholders should document research practices to facilitate regulatory understanding and demonstrate reproducibility.

A specified process for feedback from FDA could be especially useful for real-world endpoint validation. For example, consensus multi-stakeholder recommendations, such as the Clinical Trials Transformation Initiative's (CTTI) published steps for developing and validating endpoints that use mobile technology could inform thinking on digital measurement tool validation and support this feedback process.⁵⁰⁻⁵³ Sponsors, developers, and researchers need additional clarity on how to demonstrate validation of digital measurement tools, especially as the software and algorithms undergo frequent updates. Sponsors, developers, and researchers also need clarification on whether digital measurement tools can or should undergo qualification and which qualification program to use (i.e., biomarker, COA, or other). Guidances across the Centers at FDA on the development and validation of real-world endpoints can help clarify these points for stakeholders and improve the quality of submitted RWE.

To increase the acceptability of real-world endpoints, sponsors should consider engaging in real-world endpoint pilot projects. (Notably, in a March 2020 Funding Opportunity Announcement, the Agency highlighted determining and evaluating endpoints using RWD as a priority project area.⁵⁴) These pilot projects allow sponsors to gain feedback from FDA in exchange for sharing examples of a study endpoint used in the real-world setting to demonstrate the effectiveness of a marketed medical product. These pilot projects could also offer an opportunity for data companies and other stakeholders to become more involved in the endpoint development process and engage with FDA at an earlier stage. Because the real-world endpoint development process is iterative, stakeholders must be flexible and maintain a willingness to adapt as needed based on feedback from FDA.

^{†††} See [Appendix D](#) for other mechanisms in which FDA provides feedback on endpoint development, such as the COA Compendium and lists of Qualified Clinical Outcome Assessments and Biomarkers.

Conclusion

To evaluate the effectiveness of a therapy in the real world, a robust research question must be developed to conceptualize the treatment effect. Then, an endpoint is designed to address the research question by reflecting the clinical benefit, through the identification of a COI and outcome that inform the selection endpoint components. Selecting the endpoint components is an iterative process, because it can be challenging to develop an endpoint that not only captures the treatment effect but is also relevant to stakeholders and feasible in a real-world setting. Validation of measurement properties is required to demonstrate that the endpoint can capture the treatment effect, while consideration of the clinical and regulatory contexts can assist in ensuring that the endpoint is relevant and feasible in the real-world setting. The endpoint development process is complex but can be simplified by a greater understanding of the considerations for endpoint development in the real world coupled with multi-stakeholder collaboration in the pre-competitive setting to determine the most relevant outcomes to capture (and the process for doing so) for specific therapeutic areas.

APPENDIX A. WORKSHOP PARTICIPANTS

Establishing Guideposts for Developing Real-World Endpoints

Duke-Robert J. Margolis, MD, Center for Health Policy
1201 Pennsylvania Ave, NW, Suite 500 • Washington, DC 20004
September 16, 2019

Participant List

Jeff Allen

President & CEO, Friends of
Cancer Research

Aylin Altan

Senior Vice President of
Research, OptumLabs

Shrujal Baxi

Medical Director, Flatiron
Health

Susan Colilla

Director, Epidemiology &
RWE, GHEOR, Teva
Pharmaceuticals

Gracy Crane

Senior Principal Data
Scientist, RWD Policy,
F.Hoffman-La Roche

Kourtney Davis

Senior Director and Head,
Therapy Area Matrix, Global
R&D Epidemiology, Janssen

Susan Ellenberg

Professor of Biostatistics,
Medical Ethics and Health
Policy, Perelman School of
Medicine, University of
Pennsylvania

Rachael Fleurence

Executive Director, NESTcc

Jessica Franklin

Assistant Professor of
Medicine, Division of
Pharmacoepidemiology &
Pharmacoeconomics,
Brigham and Women's
Hospital and Harvard
Medical School

Marni Hall

Vice President, Clinical
Evidence and Regulatory
Science, Real World
Solutions, IQVIA

Lee Kallenbach

Senior Clinical Consultant,
Veradigm Life Sciences,
Allscripts

Lindsay Kehoe

Project Manager, Clinical
Trials Transformation
Initiative

Eric Klein

Senior Director, GPORWE—
Oncology, Eli Lilly &
Company

Laura Koontz

Director, Regulatory Policy,
Flatiron Health

Harlan Krumholz

Harold H. Hines Jr. Professor
of Medicine, Yale University,
and Director, Yale New
Haven Hospital Center for
Outcomes Research and
Evaluation

Erem Latif

Director, Patient
Engagement, Evidera

Gracie Lieberman

Senior Director of Regulatory
Policy, Genentech

Christie Mangir

Project Manager, Provider
Education, Association for
Community Cancer Centers

Anna McCollister-Slipp

Principal, Four Lights
Consulting

Annie McNeill

Senior Director,
Epidemiology, AbbVie, Inc.

Linda Nelsen

Senior Director and Head,
Value Evidence and
Outcomes – Patient
Centered Outcomes, R&D
Chief Medical Office,
GlaxoSmithKline

Elisabeth Oehrlein

Senior Director, Research and Programs, National Health Council

Sally Okun

Vice President, Policy & Ethics, PatientsLikeMe

Ernesto Ramirez

Senior Data Scientist, Evidation

Matthew Reynolds

Vice President, Real World Evidence, Real World Analytics & Solutions, IQVIA

Frank Rockhold

Professor of Biostatistics and Bioinformatics, Duke Clinical Research Institute, Duke University Medical Center

Amy Rudolph

Head of HE&OR and Medical Access, Novartis Pharmaceuticals

Sebastian Schneeweiss

Chief, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, and Professor of Medicine, Harvard Medical School

Mark Stewart

Vice President, Science Policy, Friends of Cancer Research

David Thompson

Senior Vice President, Real World and Late Phase, Syneos Health

Emese Toth

Real World Evidence Innovation Lead, UCB

Sean Tunis

Principal, Rubix Health, and Senior Strategic Advisor, Center for Medical Technology Policy

Vince Willey

Principal Scientist, HealthCore

Wei Zhou

Executive Director, Oncology Lead, Department of Pharmacoepidemiology, Merck Center for Observational and Real-World Evidence

Duke-Robert J. Margolis, MD, Center for Health Policy

Joy Eckert

Adam Kroetsch

Nirosha Mahendraratnam Lederer

Kerra Mercon

Gillian Sanders Schmidler

Rachel Sherman

Christina Silcox

Establishing Guideposts for Developing Real-World Endpoints

Duke-Robert J. Margolis, MD, Center for Health Policy
1201 Pennsylvania Ave, NW, Suite 500 • Washington, DC 20004
February 26, 2020

Participant List**Jeff Allen**

President & CEO, Friends of
Cancer Research

Aylin Altan

Senior Vice President of
Research, OptumLabs

Marc Berger

Special Advisor for Real
World Evidence, ISPOR

Ariel Bourla

Medical Director, Oncology,
Flatiron Health

Susan Colilla

Director, Epidemiology &
RWE, GHEOR,
Teva Pharmaceuticals

Gracy Crane

Senior Principal Data
Scientist, RWD Policy,
F.Hoffman-La Roche

Cynthia de Luise

Senior
Director/Epidemiologist,
Safety Surveillance Research,
Pfizer, Inc.

Luca Foschini

Co-founder and Chief Data
Scientist,
Evidation Health

Marni Hall

Vice President, Clinical
Evidence,
Head US Regulatory Science
and Strategy, IQVIA

Rohini Hernandez

Observational Research
Director, Amgen

Lee Kallenbach

Senior Clinical Consultant,
Veradigm Life Sciences,
Allscripts

Eric Klein

Senior Director, GPORWE—
Oncology,
Eli Lilly & Company

Erem Latif

Director, Patient
Engagement, Evidera

Gracie Lieberman

Senior Director of Regulatory
Policy, Genentech

Renato Lopes

Professor of Medicine, Duke
Clinical Research Institute,
Duke University Medical
Center

Nicole Mahoney

Senior Director, Regulatory
Policy, Flatiron Health

Paivi Miskala

Senior Principal Scientist,
Outcomes Research,
Merck Research Laboratories

Linda Nelsen

Senior Director and Head,
Value Evidence and
Outcomes – Patient
Centered Outcomes, R&D
Chief Medical Office,
GlaxoSmithKline

Sally Okun

Policy & Ethics, UnitedHealth
Group R&D

Lucinda Orsini

Associate Chief Science
Officer, ISPOR

Bryce Reeve

Director of Center for Health
Measurement, Duke
University School of
Medicine

Matthew Reynolds

Vice President, Real World
Evidence, IQVIA

Frank Rockhold

Professor of Biostatistics and
Bioinformatics, Duke Clinical
Research Institute, Duke
University Medical Center

Amy Rudolph

Global Head, Evidence &
Launch Excellence, Novartis
Pharmaceuticals

Rachel Sobel

ISPE
United Biosource
Corporation

Duke-Robert J. Margolis, MD, Center for Health Policy

Adam Aten

Joy Eckert

Mark Stewart

Vice President, Science
Policy,
Friends of Cancer Research

Eileen Mack Thorley

Senior Research Scientist,
PatientsLikeMe

Mira Gill

Kerra Mercon

Morgan Romine

Emese Toth

Real World Evidence
Innovation Lead, UCB

Vince Willey

Principal Scientist,
HealthCore

Christina Silcox

Marta Wosińska

APPENDIX B. GLOSSARY

- **Ability to detect change** – “Evidence that a [test, tool, or] instrument can identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept.”³⁰
- **Analytical validation** – “A process to establish that the performance characteristics of a test, tool, or instrument are acceptable in terms of its sensitivity, specificity, accuracy, precision, and other relevant performance characteristics using a specified technical protocol (which may include specimen collection, handling and storage procedures).”⁷
- **Biomarker** – “A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions. Molecular, histologic, radiographic, or physiologic characteristics are types of biomarkers. A biomarker is not an assessment of how an individual feels, functions, or survives. Categories of biomarkers include:
 - Susceptibility/risk
 - Diagnostic
 - Monitoring
 - Prognostic
 - Predictive
 - Pharmacodynamic/response
 - Safety.”⁷
- **Clinical (treatment) benefit** – “A positive clinically meaningful effect of an intervention, i.e., a positive effect on how an individual feels, functions, or survives.”⁷
- **Clinical outcome** – “Medical events that occur as a result of a disease or treatment.”⁵⁵
- **Clinical outcome assessment (COA)** – “Assessment of a clinical outcome can be made through report by a clinician, a patient, a non-clinician observer or through a performance-based assessment. There are four types of COAs.”⁷
 - *Clinician-reported outcome* – “A measurement based on a report that comes from a trained health-care professional after observation of a patient’s health condition.”⁷
 - *Observer-reported outcome* – “A measurement based on a report of observable signs, events or behaviors related to a patient’s health condition by someone other than the patient or a health professional.”⁷
 - *Patient-reported outcome* – “Measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient’s health condition without amendment or interpretation of the patient’s response by a clinician or anyone else.”⁷
 - *Performance outcome* – “A measurement based on standardized task(s) actively undertaken by a patient according to a set of instructions. A PerfO assessment may be administered by an appropriately trained individual or completed by the patient independently.”⁷
- **Clinical validation** – “A process to establish that the test, tool, or instrument acceptably identifies, measures, or predicts the concept of interest.”⁷
- **Composite endpoint** – An endpoint that is a combination of “clinical outcomes into a single variable.” The endpoint is “defined as the occurrence or realization in a patient of any one of the specified components.”⁸

- **Concept of Interest*** – “The aspect of an individual’s clinical, biological, physical, or functional state, or experience that the assessment is intended to capture (or reflect).”⁷
- **Construct validation** – “A process to establish, using quantitative methods, the extent to which the relationships among items, domains, and concepts of a clinical outcome assessment conform to *a priori* hypotheses concerning logical relationships that should exist with other measures or characteristics of patients and patient groups.”⁷
- **Content validation** – “Extent to which the COA measures the concept of interest including evidence that the items and domains are appropriate and comprehensive relative to its intended measurement concept(s), population, and use.”¹⁷
- **Context of use** – “A statement that fully and clearly describes the way the medical product development tool is to be used and the medical product development-related purpose of the use.”⁷
- **Economic outcome** – “Direct, indirect, and intangible costs compared with the consequences of medical treatment alternatives.”⁵⁵
- **Endpoint** – “A precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question. A precise definition of an endpoint typically specifies the type of assessments made, the timing of those assessments, the assessment tools used, and possibly other details, as applicable, such as how multiple assessments within an individual are to be combined.”⁷
- **Estimand** – “A precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarises at a population-level what the outcomes would be in the same patients under different treatment conditions being compared.”¹⁰
- **Exploratory endpoint** – An endpoint that “may include clinically important events that are expected to occur too infrequently to show a treatment effect or an endpoint that for other reasons are thought to be less likely to show an effect but are included to explore new hypotheses”.⁸
- **Humanistic outcome** – “Consequences of disease or treatment on patient functional status or quality of life.”⁵⁵
- **Intercurrent events** – “Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest.”¹⁰
- **Intermediate clinical endpoint** – “An endpoint measuring a clinical outcome that can be measured earlier than an effect on irreversible morbidity or mortality (IMM) and that is considered reasonably likely to predict the medical product’s effect on IMM or other clinical benefit.”⁷
- **Measurement** – “The obtained value using a test, tool, or instrument.”⁷
- **Novel endpoint** – An endpoint that has not been used before in a real-world study for the specified context of use.
- **Outcome** – “The measurable characteristic that is influenced or affected by an individual’s baseline state or an intervention as in a clinical trial or other exposure.”⁷
- **Primary data** – Data collected to answer a research question.
- **Primary endpoint** – An endpoint that “consists of the outcome or outcomes (based on the drug’s expected effects) that establish the effectiveness, and/or safety features, of the drug in order to support regulatory action.”⁸

* This term has been changed from “concept,” as listed in the BEST glossary, to “concept of interest” to align with the terminology in this paper.

- **Qualification** – “A conclusion, based on a formal regulatory process, that within the stated context of use, a medical product development tool can be relied upon to have a specific interpretation and application in medical product development and regulatory review.”⁷
- **Reliability** – “Ability to yield consistent, reproducible estimates of true treatment effect.”³⁰
- **Secondary use data** – Data used to answer additional research questions other than the purpose for which it was originally collected.
- **Secondary endpoint** – An endpoint that “may be selected to demonstrate additional effects after success on the primary endpoint” and “may also provide evidence that a particular mechanism underlies a demonstrated clinical effect.”⁸
- **Single-measure endpoint** – A single variable that reflects a single outcome of interest.
- **Surrogate endpoint** – “An endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself, but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. From a U.S. regulatory standpoint, surrogate endpoints and potential surrogate endpoints can be characterized by the level of clinical validation:
 - validated surrogate endpoint
 - reasonably likely surrogate endpoint
 - candidate surrogate endpoint.”⁷
- **Test, tool, or instrument** – “An assessment system comprising three essential components: 1) materials for measurement; 2) an assay or method or procedure for obtaining the measurement; and 3) method and/or criteria for interpreting those measurements.”⁷
- **Validated surrogate endpoint** – “An endpoint supported by a clear mechanistic rationale and clinical data providing strong evidence that an effect on the surrogate endpoint predicts a specific clinical benefit. A validated surrogate endpoint can be used to support marketing approval of a medical or tobacco product in a defined context without the need for additional studies to demonstrate the clinical benefit directly. Although the term has been used in a conceptually broader way, from a U.S. regulatory standpoint, a validated surrogate endpoint almost always refers to a biomarker.”⁷
- **Validation** – “A process to establish that the performance of a test, tool, or instrument is acceptable for its intended purpose.”⁷
- **Verification** – “Assessment of sensor accuracy (which describes the agreement between the measurement made by a single sensor vs. a ground truth), precision (which describes the agreement between multiple measurements made by a single sensor back-to-back), consistency (which describes the agreement between multiple measurements made by a single sensor over longer time periods), and/or uniformity which describes the agreement across measurements made by multiple sensors simultaneously).”²⁴

APPENDIX C. TYPES OF ENDPOINTS

This appendix defines the following endpoints: single-measure, composite, multi-component, intermediate clinical, surrogate, and novel.

A **single-measure endpoint** is a single variable that reflects a single outcome of interest.

For a **composite endpoint**, multiple clinical outcomes are combined into a single variable and a single statistical test is performed.⁸ One commonly used composite endpoint is major adverse cardiac events (MACE). MACE does not have a single definition, but since the mid-1990s, researchers have used MACE to include a range of cardiovascular-related adverse effects.⁵⁶ More information about composite endpoints can be found in FDA's *Multiple Endpoints in Clinical Trials Guidance for Industry*.⁸

Multi-component endpoints can include a combination of components or domains to create a single score according to specified rules.⁸ Multi-component endpoints can be further classified as categorical, continuous, event-time endpoints, or other more complex endpoints.⁸ For COAs with multiple domains, within-patient combinations of all domain scores are used to calculate a single overall rating.¹⁷ Limited correlation among the multiple endpoints can negatively affect study power.⁸ The American College of Rheumatology 20/50/70 criteria (ACR20/50/70) is an example of a multi-component endpoint.⁵⁷

Intermediate clinical endpoints measure “a clinical outcome that can be measured earlier than an effect on irreversible morbidity or mortality (IMM) and that is considered reasonably likely to predict the medical product’s effect on IMM or other clinical benefit.”⁷ Intermediate clinical endpoints can be used when researchers are unsure whether a short-term significant treatment effect will remain durable over a longer period of time or when a clinical benefit is reasonably likely to predict an effect on IMM or for a particular disease.⁵⁸ In clinical trials, exercise tolerance has been utilized as an intermediate clinical endpoint for medical devices that treat heart failure.⁷

A **surrogate endpoint*** “does not measure the clinical benefit of primary interest in and of itself, but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence” and is used as a “substitute for a direct measure of how a patient feels, functions, or survives.”⁷ FDA uses surrogate endpoints, including markers such as laboratory measurements, radiographic images, or physical signs, to predict clinical benefit.⁵⁸ Surrogate endpoints can be classified by their level of validation: validated surrogate endpoints, reasonably likely surrogate endpoints, and candidate surrogate endpoints.⁷ For an accelerated approval, a surrogate endpoint that is “reasonably likely to predict clinical benefit” can be used in place of a validated surrogate endpoint.^{9,58}

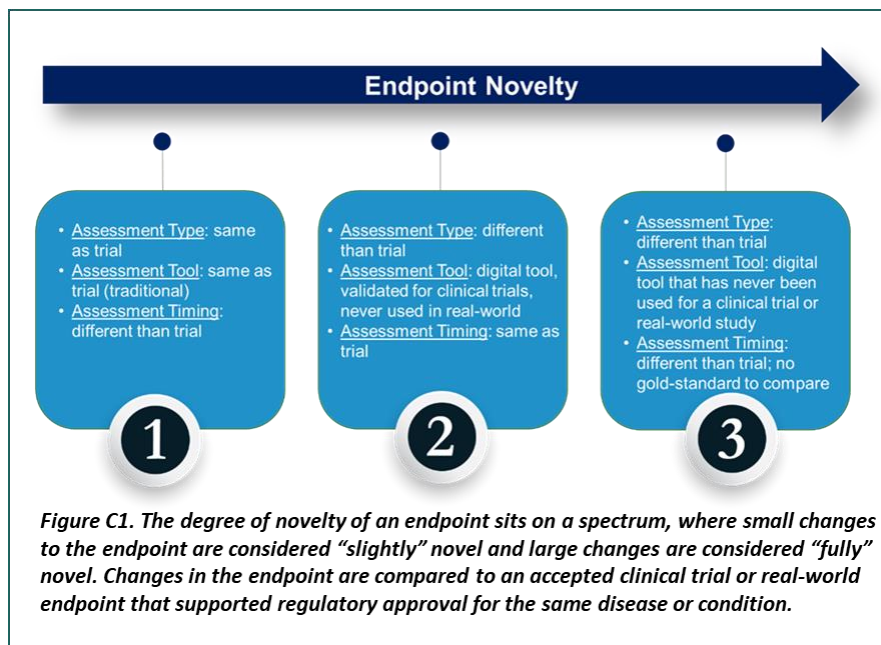
Surrogate endpoints are useful in cases where the natural course of the disease is long and an extended period of time is needed to measure any clinical benefit. For example, if investigators want to demonstrate an effect on survival or morbidity in patients with human immunodeficiency virus (HIV), lengthy, large trials are often required due to the duration of the disease course. To limit time requirements, a surrogate endpoint that has been validated to demonstrate the relationship between prolonged viral load suppression and morbidity or mortality could be used to support a traditional approval.⁵⁸

* For most surrogate endpoints, the type of assessment is a biomarker; however, it is possible for a surrogate endpoint to be based on a clinical outcome assessment or survival.

A Note on Novel Endpoints

When developing a real-world endpoint, researchers may have to consider the regulatory implications of including a novel endpoint, where the acceptability has not been previously assessed. The term “novel endpoint” is used commonly; however, significant variability exists in the definition, assignment, and interpretation of the term. This paper considers endpoint novelty on a spectrum from “slightly novel” to “fully novel.” A “slightly” novel endpoint exists when most of the endpoint components are the same as an endpoint that has been used before to support regulatory approval (e.g., the assessment type and tool are the same, but the assessment frequency has changed). On the other end of the spectrum, the majority or all of the endpoint components have changed. For example, the endpoint includes a new digital measurement tool that has never been used to support regulatory approval at a frequency at which the outcome assessment has never been measured. Many combinations of these examples exist, as well as endpoints that may fall more centrally within the spectrum (Figure C1). Novel endpoints may offer an opportunity to generate evidence on clinical concepts or outcomes that have not been studied in clinical trials. For example, if a digital measurement tool is used to measure a traditional outcome used in a trial (e.g., measuring pulse to discern risk of cardiovascular disease), the endpoint could be considered “slightly” novel, whereas if a non-traditional tool is used to measure a non-traditional outcome (e.g., tracking facial expressions to evaluate depression severity), the endpoint would be “fully” novel.^{26,59,60}

Endpoint novelty is complex and cannot be precisely defined. Endpoint novelty should be determined on a case-by case basis; however, it is important to remember that an endpoint can be novel even if it does not include a digital measurement tool. Also, the novelty of endpoints may change over time as familiarity and regulatory experience with the endpoint evolves.



APPENDIX D. REGULATORY PROGRAMS, PATHWAYS, AND INITIATIVES THAT SUPPORT DEVELOPMENT OF REAL-WORLD ENDPOINTS

This appendix describes FDA’s programs, review pathways, and initiatives that support real-world endpoint development.

Investigational New Drug (IND)/ New Drug Application (NDA)/ Biologic Licensing Application (BLA)⁶¹⁻⁶⁴

The IND, NDA, and BLA applications are used by sponsors applying to test or market a medical product. Through these review pathways, FDA can give individualized feedback to sponsors related to the specific medical product in question.

RWE Subcommittee¹

As part of CDER’s Medical Policy and Program Review Council, the RWE Subcommittee guides policy development around use of RWE. The RWE Subcommittee provides recommendations on “whether the underlying data, methods, and other study design elements are appropriate to provide support for a regulatory decision” and advises review divisions on how to evaluate RWE.¹ The Office of New Drugs can also consult with the RWE Subcommittee when reviewing an evidence package that includes RWE.

Drug Development Tool Qualification Programs⁶⁵⁻⁶⁹

Drug Development Tools (DDTs), “methods, materials, or measures that can aid in drug development and regulatory review,” can be qualified under FDA’s DDT Qualification programs.⁶⁵ DDTs may include biomarkers, COAs, and animal models used for efficacy testing. Qualification is a “determination that a DDT and its proposed COU can be relied upon to have a specific interpretation and application in drug development and regulatory review.”⁶⁵

The optional 3-step qualification program begins when the tool developer submits a letter of intent (LOI) that summarizes the proposed DDT, its relevant need, and its COU. The second step is a comprehensive review of the current literature, knowledge gaps, and data analysis plan, as well as addressing comments from FDA from the first step. The final step for qualification includes detailed descriptions of all studies, analyses, and results related to the DDT and its COU. FDA aims to complete the qualification program for a COA or biomarker within 10 months of the LOI submission. Qualification is not required for use of a COA or biomarker in a clinical study. Currently, 6 COAs and 8 biomarkers have been qualified, and many have been submitted for qualification.

Critical Path Innovation Meeting (CPIM)⁷⁰

A CPIM is a meeting to “discuss a methodology or technology proposed by the meeting requester and for CDER to provide general advice on how this methodology or technology [including novel technologies] might enhance drug development.”⁷⁰ The “meeting requester” can be a patient advocacy group, academic centers, sponsor, or even other government agencies. This meeting is considered informal and non-binding, but also non-drug-specific.

Patient Listening Session⁷¹

Hosted in partnership between FDA and the National Organization for Rare Disorders (NORD), listening sessions “are a resource for the medical product Centers to quickly engage with patients or their advocates and are one of many ways the patient community can share their experience with a disease or condition by talking directly with FDA staff.”⁷¹ Listening sessions can be FDA-led or patient-led. These

sessions enhance the regulatory decision-making process by educating the review staff on a rare disease and its effects on patients. These sessions are non-public, non-regulatory, and not specific to a medical product.

FDA Patient-Focused Drug Development Program (PFDD) ^{11,17,72-75}

In accordance with mandates from the Prescription Drug User Fee Act VI (PDUFA VI) and the 21st Century Cures Act, FDA is developing guidance documents for integrating patient experience data into the drug development process. FDA held a series of four public workshops, with accompanying discussion documents, to inform guidance development. Currently, four discussion documents and two guidance documents have been released that focus on collecting information from patients, identifying the concepts most important to patients, developing fit-for-purpose COAs, and incorporating COAs into endpoints.

Clinical Outcome Assessment (COA) Compendium⁷⁶

The COA Compendium is a tool that provides information on specific COAs that have been used previously in clinical trials to support labeling changes and COAs qualified under the Drug Development Tool Qualification Program. The COA compendium lists the disease or condition, concept that the COA is evaluating, type of COA, tool used to evaluate the COA, COU, drug name, approval date, and the qualification link (if applicable).

CDER Pilot Grant Program: Standard Core Clinical Outcome Assessments (COAs) and their Related Endpoints⁷⁷

Part of FDA's PFDD program, this recently developed grant program supports the "development of publicly available core set(s) of Clinical Outcome Assessments (COAs) and their related endpoints for specific disease indications."⁷⁷ Currently, the program has approved three awards to groups investigating improvement in migraine outcomes, acute pain therapies in young children, and various chronic conditions.

FDA's MyStudies Application^{78,79}

The MyStudies App was developed by FDA as a resource to link EHR data to electronically collected patient-reported outcome data to support clinical studies and registry development. Some of the capabilities of the MyStudies App include the ability to take informed consent, determine if patients meet the inclusion criteria, and remind patients to complete a questionnaire. The code for the app is open source, so stakeholders can customize the app to their specific needs.

APPENDIX E. ITERATIVE PROCESS OF ENDPOINT DEVELOPMENT

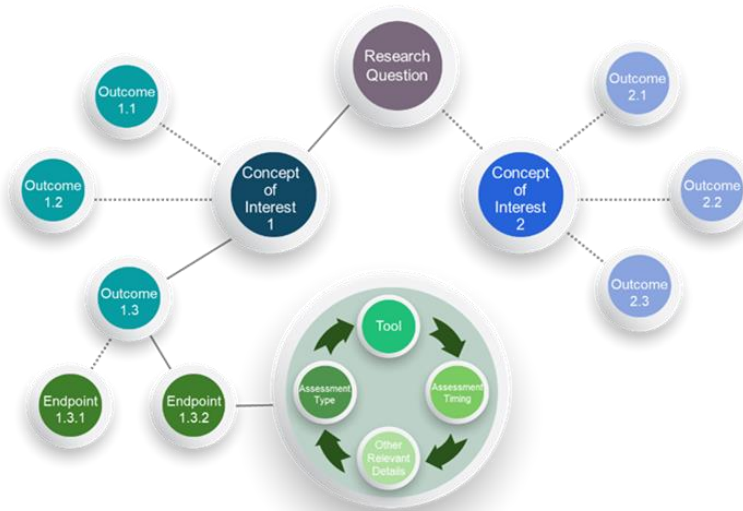


Figure E1. Demonstrates the complex relationship between the research question and building the endpoint, with emphasis on the iterative nature of endpoint development. For a given research question, multiple concepts of interest could be chosen. For each concept of interest, many outcomes may reflect the concept of interest. Furthermore, many endpoints (with varying components) can be developed to measure the outcome.



Figure E2. An example is used to illustrate the complex relationship between the research question and the endpoint definition.

APPENDIX F. EXAMPLES OF STAKEHOLDER EFFORTS THAT SUPPORT THE DEVELOPMENT OF REAL-WORLD ENDPOINTS

The Clinical Trials Transformation Initiative (CTTI) ⁵⁰⁻⁵³

In response to the growing interest in digital health technologies, CTTI released detailed recommendations for developing novel endpoints that incorporate digital health technologies for use in clinical trials. To support these recommendations for novel endpoint development, CTTI also developed a flowchart of steps and a tool with suggested approaches for applying these steps. To provide additional context, CTTI applied the recommendations and flowchart steps in four use cases.

Friends of Cancer Research⁴⁴

Friends of Cancer Research's real-world evidence project operationalizes the collection and use of RWD to support regulatory decision-making. Pilot Project 1.0, Operationalizing and Validating Real-World Evidence, focused on standardizing data collection and developing a framework for the validation of real-world endpoints for non-small cell lung cancer. Ten healthcare research organizations collaborated for Pilot 2.0: Establishing the Utility of Real-World Endpoints to characterize the patient population with advanced non-small cell lung cancer, evaluate the ability of several real-world endpoints to measure the treatment effect, and assess the performance of the selected real-world endpoints. Next steps include examining differences in treatments among specific patient populations using RWD.

Digital Medicine Society (DiMe)^{80,81}

To increase the use of endpoints that incorporate digital technologies, DiMe created an open-access and crowdsourced Digital Endpoint Library that compiles use of digital endpoints in clinical trials. The library includes information on the technology type, the medical condition, the endpoint, manufacturer notes, specific measurements, and the clinical trial information. Additionally, DiMe is developing recommendations for the verification and validation of digital tools to enable better integration into clinical trials.

The Patient-Reported Outcome (PRO) Consortium⁸²

In a partnership with the Critical Path Institute, FDA, and pharmaceutical industry representatives, the PRO Consortium aims "to establish and maintain a collaborative framework with appropriate stakeholders for the qualification of PRO measures and other clinical outcome assessment (COA) tools that will be publicly available for use in clinical trials when COA-based endpoints are used to support product labeling claims."⁸² The Consortium brings together experts to contribute to COA development across several therapeutic areas.

Flatiron Mortality Endpoint³⁴

Survival is well-defined but can be difficult to ascertain in the real world because the status is not clear. The date of death and the cause of death are typically not recorded for secondary research use. Ascertainment of mortality data is especially difficult when using claims or EHR data due to loss of follow-up, errors in mortality data collection systems, or lack of appropriate clinical workflows to capture mortality data. Although public data on mortality exists in the U.S. through the US Social Security Death Index (SSDI) and National Death Index (NDI), these datasets have issues with linkability, timeliness, and restricted access. Researchers have supplemented mortality information from EHR data with commercial death datasets that include credit card, insurance, or obituary information. It is important to note that challenges exist with point of care data entry as each database may capture different variables or capture the variables in a different manner. In one RWE study, researchers from

Flatiron and Genentech gathered data from EHR structured data from the Flatiron Health database, abstractions from unstructured EHR data, commercial death data, and publicly available U.S. mortality data from SSDI. For validation, the researchers matched data from the study cohort to NDI data to calculate the sensitivity, specificity, positive predictive value, negative predictive value, and death date agreement of each of the datasets to understand the contribution to the composite.

REFERENCES

1. U.S. Food and Drug Administration. *Framework For FDA's Real-World Evidence Program*. US Department of Health & Human Services; December 2018.
2. Berger M, Daniel G, Frank K, Hernandez A, McClellan M, Okun S. A Framework for Regulatory Use of Real-World Evidence: Duke-Margolis Center for Health Policy; 2017.
3. Daniel G, McClellan M, Silcox C, Romine M, Bryan J, Frank K. Characterizing RWD Quality and Relevancy for Regulatory Purposes: Duke-Margolis Center for Health Policy; 2018.
4. Mahendraratnam N, Silcox C, Mercon K, et al. Determining Real-World Data's Fitness for Use and the Role of Reliability: Duke-Margolis Center for Health Policy; 2019.
5. Mahendraratnam N, Eckert J, Mercon K, et al. Using Secondary Data to Generate Real-World Evidence for Regulatory Decision Making, and Demonstrating Their Credibility: Duke-Margolis Center for Health Policy; 2019.
6. Mahendraratnam N, Mercon K, Eckert J, et al. Adding Real-World Evidence to a Totality of Evidence Approach for Evaluating Marketed Product Effectiveness: Duke-Margolis Center for Health Policy; 2019.
7. FDA-NIH Biomarker Working Group. BEST (Biomarkers, Endpoints, and other Tools) Resource-Glossary. Published 2016. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK338448/#IX-B>. Accessed August 27, 2020.
8. U.S. Food and Drug Administration. *Multiple Endpoints in Clinical Trials Guidance for Industry*: U.S. Department of Health & Human Services; 2017.
9. U.S. Food and Drug Administration. *Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products Guidance to Industry*: U.S. Department of Health & Human Services; 2019.
10. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1); 2019.
11. U.S. Food and Drug Administration. Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 4: Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making: U.S. Department of Health & Human Services; 2019.
12. Sherman RE, Davies KM, Robb MA, Hunter NL, Califf RM. Accelerating development of scientific evidence for medical products within the existing US regulatory framework. *Nature Reviews Drug Discovery*. 2017;16:297.
13. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American journal of epidemiology*. 2016;183(8):758-764.
14. Kip KE, Hollabaugh K, Marroquin OC, Williams DO. The Problem With Composite End Points in Cardiovascular Studies: The Story of Major Adverse Cardiac Events and Percutaneous Coronary Intervention. *Journal of the American College of Cardiology*. 2008;51(7):701-707.
15. Wilson IB, Cleary PD. Linking Clinical Variables With Health-Related Quality of Life: A Conceptual Model of Patient Outcomes. *JAMA*. 1995;273(1):59-65.
16. Available PROMIS® Measures for Adults. Available at: https://www.healthmeasures.net/index.php?option=com_content&view=category&layout=blog&id=113&Itemid=808. Accessed August 27, 2020.

17. U.S. Food and Drug Administration. Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 3: Select, Develop or Modify Fit-for-Purpose Clinical Outcome Assessments: U.S. Department of Health & Human Services; 2018.
18. Outcomes Research. AMCP. 2020. Available at: https://www.amcp.org/sites/default/files/2020-03/Outcomes%20Research%20-%202020%20Updates_New.pptx. Accessed August 27, 2020.
19. Walton MK, Powers JH, 3rd, Hobart J, et al. Clinical Outcome Assessments: Conceptual Foundation- Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for Outcomes Research Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2015;18(6):741-752.
20. Siu A. Screening for High Blood Pressure in Adults: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine* 2015;163:778-786.
21. FDA-NIH Biomarker Working Group. BEST (Biomarkers, Endpoints, and other Tools) Resource- Diagnostic Biomarker. Published 2016. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK402285/>. Accessed August 27, 2020.
22. De Lusignan S, Khunti K, Belsey J, et al. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. 2010;27(2):203-209.
23. Centers for Disease Control and Prevention. ICD-10-CM Official Guidelines for Coding and Reporting FY 2020. Available at: https://www.cdc.gov/nchs/data/icd/10cmguidelines-FY2020_final.pdf. Accessed August 27, 2020.
24. Coravos A, Goldsack JC, Karlin DR, et al. Digital Medicine: A Primer on Measurement. *Digital Biomarkers*. 2019;3(2):31-71.
25. Daniel G, Colvin H, Silcox C, McClellan M, Sharma I, J B. *Mobilizing mHealth Innovation for Real-World Evidence Generation*. Duke-Margolis Center for Health Policy;2017.
26. Babrak LM, Menetski J, Rebhan M, et al. Traditional and Digital Biomarkers: Two Worlds Apart? *Digital Biomarkers*. 2019;3(2):92-102.
27. Chan D, Ng LL. Biomarkers in acute myocardial infarction. *BMC Med*. 2010;8:34-34.
28. U.S. Food and Drug Administration. Biomarker Qualification: Evidentiary Framework Guidance for Industry and FDA Staff: U.S. Department of Health & Human Services; 2018.
29. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *npj Digital Medicine*. 2020;3(1):55.
30. U.S. Food and Drug Administration. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims Guidance to Industry: U.S. Department of Health & Human Services; 2009.
31. Withycombe JS, McFatrigh M, Pinheiro L, et al. The association of age, literacy, and race on completing patient-reported outcome measures in pediatric oncology. *Quality of Life Research*. 2019;28(7):1793-1801.
32. O'Donohoe P, Muehlhausen W, Byrom B. Ensuring Equivalence of Electronic and Paper Administration of Patient-Reported Outcome Measures. Presented at: Critical Path Institute PRO Consortium; 2016. Available at: <https://c-path.org/ensuring-equivalence-of-electronic-and-paper-administration-of-patient-reported-outcome-measures-2/>. Accessed August 27, 2020.

33. Coravos A, Doerr M, Goldsack J, et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. *npj Digital Medicine*. 2020;3(1):37.
34. Curtis MD, Griffith SD, Tucker M, et al. Development and Validation of a High-Quality Composite Real-World Mortality Endpoint. *Health services research*. 2018;53(6):4460-4476.
35. Chen X, Agiro A, Martin AS, Lucas AM, Haynes K. Chart validation of an algorithm for identifying hereditary progressive muscular dystrophy in healthcare claims. *BMC Med Res Methodol*. 2019;19(1):174.
36. Nissen F, Quint JK, Morales DR, Douglas JJ. How to validate a diagnosis recorded in electronic health records. *Breathe*. 2019;15(1):64-68.
37. Brennan JM, Wruck L, Pencina MJ, et al. Claims-based cardiovascular outcome identification for clinical research: Results from 7 large randomized cardiovascular clinical trials. *Am Heart J*. 2019;218:110-122.
38. Griffith SD, Miksad RA, Calkins G, et al. Characterizing the Feasibility and Performance of Real-World Tumor Progression End Points and Their Association With Overall Survival in a Large Advanced Non-Small-Cell Lung Cancer Data Set. *JCO Clin Cancer Inform*. 2019;3:1-13.
39. Clinical Trials Transformation Initiative. CTTI Feasibility Studies Database. Available at: <https://feasibility-studies.ctti-clinicaltrials.org/>. Accessed August 27, 2020.
40. Battelino T, Danne T, Bergenstal RM, et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. *Diabetes care*. 2019;42(8):1593-1603.
41. Runge AS, Kennedy L, Brown AS, et al. Does Time-in-Range Matter? Perspectives From People With Diabetes on the Success of Current Therapies and the Drivers of Improved Outcomes. *Clinical Diabetes*. 2018;36(2):112-119.
42. American Society of Clinical Oncology. mCODE™: Minimal Common Oncology Data Elements. Available at: <https://mcodeinitiative.org/>. Accessed August 27, 2020.
43. Rocca M, Asare A, Esserman L, Dubman S, Gordon G. *Source Data Capture from EHRs: Using Standardized Clinical Research Data*. 2019.
44. Friends of Cancer Research. Real-World Evidence. Available at: <https://www.focr.org/rwe>. Accessed August 27, 2020.
45. Sentinel Coordinating Center. Health Outcome of Interest Validations and Literature Reviews. Available at: <https://www.sentinelinitiative.org/sentinel/surveillance-tools/validations-lit-review>. Accessed August 27, 2020.
46. Sentinel Coordinating Center. Health Outcome Algorithm Inventory. Available at: <https://www.sentinelinitiative.org/sentinel/surveillance-tools/validations-lit-review/health-outcome-algorithm-inventory>. Published 2016. Accessed August 27, 2020.
47. Kirby JC, Speltz P, Rasmussen LV, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *J Am Med Inform Assoc*. 2016;23(6):1046-1052.
48. Richesson RL, Smerek MM, Blake Cameron C. A Framework to Support the Sharing and Reuse of Computable Phenotype Definitions Across Health Care Delivery and Clinical Research Applications. *EGEMS (Wash DC)*. 2016;4(3):1232-1232.
49. Duke-Robert J. Margolis, MD, Center for Health Policy. Improving the Efficiency of Outcome Validation in the Sentinel System. 2018.

50. Clinical Trials Transformation Initiative. Project: Novel Endpoints. Available at: <https://www.ctti-clinicaltrials.org/projects/novel-endpoints>. Published 2017. Accessed August 27, 2020.
51. Clinical Trials Transformation Initiative. Steps for Novel Endpoint Development, with Suggested Approaches and Considerations; 2017.
52. Clinical Trials Transformation Initiative. Flowchart of Steps for Novel Endpoint Development; 2017.
53. Clinical Trials Transformation Initiative. CTTI Recommendations: Developing Novel Endpoints Generated by Mobile Technology for Use in Clinical Trials; 2017.
54. U.S. Food and Drug Administration. RFA-FD-20-030: Exploring the use of Real-World Data to Generate Real-World Evidence in Regulatory Decision-Making (U01) Clinical Trials Optional. Available at: <https://grants.nih.gov/grants/guide/rfa-files/RFA-FD-20-030.html>. Published 2020. Accessed August 27, 2020.
55. Gunter MJ. The Role of the ECHO Model in Outcomes Research and Clinical Practice Improvement. *The American Journal of Managed Care* 1999;5.
56. Poudel I, Tejpal C, Rashid H, Jahan N. Major Adverse Cardiovascular Events: An Inevitable Outcome of ST-elevation myocardial infarction? A Literature Review. *Cureus*. 2019;11(7):e5280-e5280.
57. Felson DT and Pincus T. American College of Rheumatology 20/50/70 criteria (ACR20/50/70). Mapi Research Trust. Available at: <https://eprovide.mapi-trust.org/instruments/american-college-of-rheumatology-20-50-70-criteria>. Accessed August 27, 2020.
58. U.S. Food and Drug Administration. Expedited Programs for Serious Conditions – Drugs and Biologics: U.S. Department of Health & Human Services; 2014.
59. Perret-Guillaume C, Joly L, Benetos A. Heart Rate as a Risk Factor for Cardiovascular Disease. *Progress in Cardiovascular Diseases*. 2009;52(1):6-10.
60. Girard JM, Cohn JF, Mahoor MH, Mavadati S, Rosenwald DP. Social Risk and Depression: Evidence from Manual and Automatic Facial Expression Analysis. *Proc Int Conf Autom Face Gesture Recognit*. 2013:1-8.
61. U.S. Food and Drug Administration. New Drug Application. Available at: <https://www.fda.gov/drugs/types-applications/new-drug-application-nda>. Accessed August 27, 2020.
62. U.S. Food and Drug Administration. Investigational New Drug (IND) Application. Available at: <https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application>. Accessed August 27, 2020.
63. U.S. Food and Drug Administration. Therapeutic Biologics Applications (BLA). Available at: <https://www.fda.gov/drugs/types-applications/therapeutic-biologics-applications-bla>. Accessed August 27, 2020.
64. U.S. Food and Drug Administration. Biologics License Applications (BLA) Process (CBER). Available at: <https://www.fda.gov/vaccines-blood-biologics/development-approval-process-cber/biologics-license-applications-bla-process-cber>. Accessed August 27, 2020.
65. U.S. Food and Drug Administration. Qualification Process for Drug Development Tools: Guidance for Industry and FDA Staff: U.S. Department of Health & Human Services; 2019.
66. U.S. Food and Drug Administration. Qualified Clinical Outcome Assessments (COA). Available at: <https://www.fda.gov/drugs/development-approval-process-drugs/qualified-clinical-outcome-assessments-coa>. Accessed August 27, 2020.

67. U.S. Food and Drug Administration. List of Qualified Biomarkers. Available at: <https://www.fda.gov/drugs/cder-biomarker-qualification-program/list-qualified-biomarkers>. Accessed August 27, 2020.
68. U.S. Food and Drug Administration. Clinical Outcome Assessment (COA) Qualification Program. Available at: <https://www.fda.gov/drugs/drug-development-tool-qualification-programs/clinical-outcome-assessment-coa-qualification-program>. Accessed August 27, 2020.
69. U.S. Food and Drug Administration. CDER Biomarker Qualification Program. Available at: <https://www.fda.gov/drugs/drug-development-tool-qualification-programs/cder-biomarker-qualification-program>. Accessed August 27, 2020.
70. Office of Translational Sciences. Critical Path Innovation Meetings Policy and Procedures. Center for Drug Evaluation and Research (CDER); 2014.
71. U.S. Food and Drug Administration. Patient Listening Sessions. Available at: <https://www.fda.gov/patients/learn-about-fda-patient-engagement/patient-listening-sessions>. Accessed August 27, 2020.
72. U.S. Food and Drug Administration. FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient's Voice in Medical Product Development and Regulatory Decision Making. Available at: <https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical>. Accessed August 27, 2020.
73. U.S. Food and Drug Administration. Plan for Issuance of Patient-Focused Drug Development Guidance Under 21st Century Cures Act Title III Section 3002; 2017.
74. U.S. Food and Drug Administration. Patient-Focused Drug Development: Collecting Comprehensive and Representative Input Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders: U.S. Department of Health & Human Services; 2018.
75. U.S. Food and Drug Administration. Patient-Focused Drug Development: Methods to Identify What Is Important to Patients Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders: U.S. Department of Health & Human Services; 2019.
76. U.S. Food and Drug Administration. Clinical Outcome Assessment (COA) Compendium. Available at: <https://www.fda.gov/media/130138/download>. Published 2019. Accessed August 27, 2020.
77. U.S. Food and Drug Administration. CDER Pilot Grant Program: Standard Core Clinical Outcome Assessments (COAs) and their Related Endpoints. U.S. Department of Health and Human Services. Available at: <https://www.fda.gov/drugs/development-approval-process-drugs/cder-pilot-grant-program-standard-core-clinical-outcome-assessments-coas-and-their-related-endpoints>. Published 2019. Accessed August 27, 2020.
78. U.S. Food and Drug Administration. Mobile App Quick Overview for Research. Available at: <https://www.fda.gov/media/119834/download>. Accessed August 27, 2020.
79. U.S. Food and Drug Administration. FDA's MyStudies Application (App). Available at: <https://www.fda.gov/drugs/science-and-research-drugs/fdas-mystudies-application-app>. Accessed August 27, 2020.
80. Digital Medicine Society (DiMe). DiME: Active Research Projects. Available at: <https://www.dimesociety.org/index.php/research/dime-research-projects>. Accessed August 27, 2020.

81. Digital Medicine Society (DiMe). DiMe's Library of Digital Endpoints. Available at: <https://www.dimesociety.org/index.php/knowledge-center/library-of-digital-endpoints>. Accessed August 27, 2020.
82. Critical Path Institute. PRO Consortium. Available at: <https://c-path.org/programs/proc/>. Accessed August 27, 2020.