FDA-NIH TERMINOLOGY FOR CLINICAL RESEARCH

Glossary of Terms and Definitions

FDA-NIH Clinical Research Working Group

Glossary Terms and Definitions

Introduction:

The U.S. Food and Drug Administration and the National Institutes of Health (NIH) Joint Leadership Council convened a working group to bring clarity to terms that are inconsistently used within the scientific community. Working together with the goals of improving communication and scientific understanding, the two agencies developed a glossary with definitions of terms used in clinical research. To fulfill their charge, the FDA-NIH Clinical Research Working Group (CRWG) developed definitions for 37 clinical research terms related to innovative clinical study designs, including studies using real-world data (RWD) to generate real-world evidence (RWE), that support scientific, clinical, and/or regulatory decision-making. In addition, 8 terms are included for reference and FDA-NIH are not seeking comment on these 8 terms. The membership of the FDA-NIH CRWG consisted of statisticians, epidemiologists, pharmacologists, clinicians, biomedical engineers, and policy experts from both agencies.

This document is intended to facilitate communication within the clinical research community, by helping establish a common vocabulary to more uniformly characterize clinical research, including innovative trial designs and studies using RWD to generate RWE. In turn, the community may be better situated to evaluate potential strengths and weaknesses of individual studies, and to convey innovative and other aspects of clinical research in a meaningful way to funders, reviewers, and other interested parties.

Terms in the glossary are grouped into the following two areas:

1. Terms for Comment: These terms include approaches to clinical study design, methodology, and interpretation and description of research results. Where an existing definition is used or adapted, the source(s) are listed.
2. Terms for Reference: These are terms with well-established definitions (e.g., well-established in the scientific community) included for completeness and are not newly defined with this project. They are included as Appendix A to add context for the terms the FDA-NIH CRWG have defined under the “Terms for Comment”.

Glossary Terms and Definitions

Terms for Comment

A

Administrative Claims Data
The information obtained from claims that health care providers submit to insurers to receive payment for treatments and other interventions. Claims data use standardized medical coding systems (nomenclatures), such as the World Health Organization International Classification of Diseases Coding (ICD-CM) to identify diagnoses, National Drug Code (NDC) to identify drugs, and Current Procedural Terminology (CPT®) to identify procedures.

C

Causal Effect
A measure of difference in outcome that would be expected in individuals subjected to an exposure of interest compared to the expected outcome if those same individuals were subjected to a specified alternative exposure (including no exposure).

Source: Adapted from Musci R.J. & Stuart, E. (2019). Ensuring causal, not casual, inference. *Prevention Science*, 20, 452–456, https://doi.org/10.1007/s11121-018-0971-9.

Causal Inference
The process of evaluation, estimation, and attribution of a causal effect.

Cluster Randomized Trial
A trial in which randomization is at a group level (for instance, by community, health care facility or medical provider) rather than an individual level.

Source: Adapted from the Secretary's Advisory Committee on Human Research Protections, Recommendations on Regulatory Issues in Cluster Studies. (2014). Available at: Attachment C: Recommendations on Regulatory Issues in Cluster | HHS.gov.

Common Data Element
A standardized, precisely defined variable or question that is paired with a set of specific allowable values or responses, that are used systematically across different sites, studies, or clinical trials to ensure consistent data collection and/or analysis.

Source: Adapted from the NIH CDE Repository.

Glossary Terms and Definitions

Common Data Model (CDM)
Comprehensive framework that includes definitions, specifications, and operational rules for data to be presented and used in a common manner to enable interoperability.

Source: Adapted from Duke Margolis Center for Health Policy, *Characterizing RWD Quality and Relevancy for Regulatory Purposes*. Available at https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf.

Completeness of Capture
The extent to which a data source includes a complete representation of the exposures, outcomes, and covariates needed for the proposed analysis. Incomplete capture may be due to variables that were not recorded or variables that include some missing values.

Computable Phenotype
A clinical condition or characteristic that can be ascertained using a computerized query to Electronic Health Record (EHR) system, administrative claims database, or clinical data repository using a defined set of data elements and logical expressions. Computable phenotype definitions provide the specifications for identifying populations likely to have the conditions or characteristics of interest.

Source: Adapted from FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

Confounding
Systematic error in estimation of the measure of the effect of a medical product on an outcome due to another factor that is associated with both the exposure and the outcome and not through the causal pathway between exposure and outcome.

Source: Adapted from Porta M. et al, A Dictionary of Epidemiology. (2014). United Kingdom: Oxford University Press.

Continuity of Coverage
The period of time over which an individual is enrolled in a health care system (provider, pharmacy, insurer, or other) and for which data on provided or reimbursed healthcare services and treatments are captured in that system.

Source: Adapted from FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

Glossary Terms and Definitions

D

Data Curation
Processing of source data (unstructured and/or structured data) into a dataset suitable for analyses. The curation process involves the application of standards for the exchange, integration, sharing, and retrieval of source data, often from various sources. For example, the application of standard medical diagnostic codes to adverse events, disease staging, the progression of disease, and other medical and clinical concepts.

Source: Adapted from FDA final guidance *CVM GFI #266 Use of Real-World Data and Real-World Evidence to Support Effectiveness of New Animal Drugs (October 2021)*.

Data Harmonization
The process of combining data from different sources and reorganizing it according to a single schema so that data are compatible and comparable. Data are combined by either identifying equivalent data elements between the sources or by applying specific transformations between the elements to derive a common data element.

Source: Adapted from the National Institute of Environmental Health Sciences, Common Language Glossary from the Environmental Health Language Collaborative.

Data Imputation
A process using statistical techniques to estimate missing data values, facilitating subsequent analyses.

Source: Adapted from Food and Agricultural Organization of the United Nations, Statistical Standard Series Imputation Version 2.0. Available at: https://www.fao.org/3/cb9339en/cb9339en.pdf.

Data Lake
A controlled, centralized environment that stores structured and unstructured data in its native form and provides infrastructure for organizing large volumes of diverse data from multiple sources.

Source: Adapted from Amazon AWS.

Data Transformation
The process of converting data from one format or structure into another format or structure. It is a process of data extraction and conversion or normalization in construction of analytic datasets.

Source: Adapted from FDA final guidance *CVM GFI #266 Use of Real-World Data and Real-World Evidence to Support Effectiveness of New Animal Drugs (October 2021)*.

Glossary Terms and Definitions

Data Warehouse
A controlled, centralized environment that stores structured data in a processed form for analysis and provides infrastructure for data access by multiple applications.

Source: Adapted from FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

Distributed Data Network
A network in which data from multiple sites are transformed into a single common data model with the ability to execute a query without substantial modifications on multiple datasets.

Source: Adapted from FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

E

Electronic Health Record
An individual patient record contained within an electronic system. A typical individual record may include a patient medical history, diagnoses, treatment plans, immunizations, allergies, imaging, pharmacy orders, laboratory values, and test results.

Source: Adapted from FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

I

Immortal Time
A span of time in the observation or follow-up period of a cohort during which the outcome under study could not have occurred, due to the cohort design and/or exposure definition.

Source: Adapted from Suissa, S. (2008). Immortal time bias in pharmacoepidemiology, *American Journal of Epidemiology*, 167(4),492–499, https://doi.org/10.1093/aje/kwm324.

Glossary Terms and Definitions

Information Bias
Systematic error in estimation of an association or other parameter of interest arising from measurement error in the data. For categorical variables, measurement error is usually called classification error or misclassification.

Source: Adapted from Daniel, G., Silcox, C., Bryan, J., McClellan, M., Romine, M., & Frank, K. (2018). Characterizing RWD Quality and Relevancy for Regulatory Purposes. Duke Margolis Center for Health Policy, https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf and Rothman KJ, Greenland S, Lash TL, Lippincott Williams & Wilkins. (2008). Modern Epidemiology.

Interventional Study
A study involving participants (e.g., healthy individuals or individuals with a disease or condition of interest) whose exposure or interaction with a medical product is assigned according to a study protocol to evaluate the effect on health outcomes or product performance.

Source: Adapted from FDA draft guidance Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products (December 2021). When final, this guidance will represent FDA's current thinking on these topics.

M

Missing Data
Data that would have been used in the study analysis but were not observed, collected, or accessible. This refers to information that is intended to be collected but is absent and information that is not intended to be collected and is therefore absent. There may be special considerations regarding real world data sources (e.g., electronic health records or claims); such data are generally not collected for primary research purposes and therefore may not have systematic data capture to answer a research question.

Source: Adapted from FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

N

N of 1 Trial
A clinical trial evaluating an intervention or multiple interventions in a single participant according to a protocol in which there is either switching between intervention(s) and control or a planned comparison between an intervention and a natural history.

Glossary Terms and Definitions

Non-interventional (observational) study
A type of study in which individuals are not assigned to a medical product according to a protocol.

Source: Adapted from FDA final guidance Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices (August 2017) and FDA draft guidance Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products (December 2021). When final, this guidance will represent FDA's current thinking on these topics.

O

Observational Study, Retrospective
A study that identifies the population and determines the exposure/treatment from data collected before the initiation of the study. The variables and outcomes of interest are determined at the time the study is designed.

Source: Adapted from the Framework for FDA's Real-World Evidence Program (December 2018) and FDA final guidance Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices (August 2017).

P

Pragmatic Clinical Trial
A clinical trial designed to efficiently inform decision-making on the benefits, burdens, and risks of health interventions in representative populations by including pragmatic elements (see definition below) that 1) are partially or fully integrated into routine clinical practice and/or 2) that streamline trial design and conduct.

Source: Adapted from Califf R.M., Sugarman J. (2015). Exploring the ethical and regulatory issues in pragmatic clinical trials. *Clinical Trials*, (5), 436-441, https://journals.sagepub.com/doi/10.1177/1740774515598334.

Pragmatic Elements
Design features that can be integrated into a clinical trial, including but not limited to, one or more of the following elements: broad eligibility criteria, simplified recruitment and follow-up, flexibility in delivery of the intervention (e.g., community settings), flexibility in assessment frequency, and measurement of outcomes relevant to the population.

Propensity Score
The conditional probability of assignment to a particular treatment given a set (e.g., vector) of observed covariates.

Source: Adapted from Rosenbaum, P.R., Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70(1), 41–55, https://doi.org/10.1093/biomet/70.1.41.

R

Registry
An organized system that collects clinical and other data in a standardized format for a population defined by a particular disease, condition, or exposure.

Source: Adapted from FDA final guidance Real-World Data: Assessing Registries To Support Regulatory Decision-Making for Drug and Biological Products (December 2023) and FDA final guidance Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices (August 2017).

Residual Confounding
Confounding (see definition) that remains after adjusting for measured confounders.

Source: Adapted from Porta, M. A dictionary of epidemiology (6th ed.), https://www.oxfordreference.com/display/10.1093/acref/9780199976720.001.0001/acref-9780199976720-e-1649.

S

Selection Bias
Systematic error in estimation of an association or other parameter that occurs from factors that influence study participation [or eligibility].

Source: Adapted from Rothman KJ, Greenland S, Lash TL, Lippincott Williams & Wilkins. (2008). Modern Epidemiology.

Sequential, Multiple Assignment, Randomized Trial (SMART)
A trial designed to evaluate a collection of interventions guided by a sequence of decision rules that specifies when and how the type and/or intensity of an intervention should be modified depending on the patient's past or present characteristics and/or ongoing clinical state or performance (e.g., response, adherence) to optimize clinically important outcomes. In such a trial, patients move along multiple stages and are randomly assigned to one of several suitable intervention options at each stage.

Source: Adapted from FDA final guidance Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (December 2020).

Glossary Terms and Definitions

Stepped Wedge Cluster Randomized Trial
In a stepped wedge group-randomized trial, also called a stepped wedge cluster randomized trial, groups or clusters are randomized to sequences that direct them to switch from a control to the intervention at predetermined time points in a sequential, staggered fashion until all clusters receive the intervention.

Source: Adapted from Home | Research Methods Resources (nih.gov).

Synthetic Data
Data that have been created artificially (e.g., through statistical modeling, computer simulation) so that new values and/or data elements are generated. Generally, synthetic data are intended to represent the structure, properties and relationships seen in actual patient data, except that they do not contain any real or specific information about individuals.

T

Target Trial Emulation
A framework for designing and analyzing an observational study based on conceptualizing a target randomized trial to answer a scientific question and designing the observational study to mimic the trial estimand(s) (including specification of population eligibility criteria, treatment strategies and assignment procedures, outcomes, handling of intercurrent events, and follow-up period).

Time-Related Bias
Systematic error in estimation of an association or other parameter of interest due to misclassification or exclusion of person-time attributed to the treatment, intervention, or exposure. Examples include protopathic bias, latency time bias, immortal time bias, time-window bias, depletion-of-susceptibles, immeasurable time bias, and other such biases.

U

Umbrella Trial
Trial designed to evaluate multiple medical products in separate sub-studies concurrently for a single disease or condition.

Source: Adapted from FDA draft guidance Master Protocols for Drug and Biological Product Development | FDA (December 2023). When final, this guidance will represent FDA's current thinking on these topics.

Glossary Terms and Definitions

Appendix A

Terms for Reference – These terms are included because they provide context for terms under "Terms for Comment". We are not requesting comment on these terms.

Adaptive Design
A clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulating data from subjects in the trial.

Source: FDA final guidance Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry (December 2019).

Basket Trial
Trial designed to evaluate a medical product for different diseases, conditions, or disease subtypes.

Source: FDA draft guidance Master Protocols for Drug and Biological Product Development | FDA (December 2023). When final, this guidance will represent FDA's current thinking on these topics.

Conceptual Definition
Explains a study construct (e.g., exposure, outcomes, covariates) or feature in general or qualitative terms.

Source: FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

Master Protocol
A protocol designed with multiple sub-studies, which may have different objectives and involve coordinated efforts to evaluate one or more medical products in one or more diseases or conditions within the overall study structure.

Source: FDA draft guidance Master Protocols for Drug and Biological Product Development | FDA (December 2023). When final, this guidance will represent FDA's current thinking on these topics.

Glossary Terms and Definitions

Observational Study, Prospective
A study in which the population of interest is identified at the start of the study, and exposure/treatment and outcome data are collected from that point forward. The start of the study is defined as the time the research protocol for the specific study question was initiated.

Source: Framework for FDA's Real-World Evidence Program (December 2018).

Operational Definition
The data-specific operation or procedure a researcher followed to measure constructs in a particular study.

Source: FDA draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021). When final, this guidance will represent FDA's current thinking on these topics.

Patient Generated Health Data
Health-related data created, recorded, or gathered by or from patients, family members, or other caregivers to help address a health concern.

Source: Definition adapted from https://www.healthit.gov/topic/scientific-initiatives/pcor/patient-generated-health-data-pghd and included in FDA draft guidance Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices | FDA (December 2023). When final, this guidance will represent FDA's current thinking on these topics.

Platform Trial
Trial designed to evaluate multiple medical products for a disease or condition in an ongoing manner, with medical products entering or leaving the platform.

Source: FDA draft guidance Master Protocols for Drug and Biological Product Development | FDA (December 2023). When final, this guidance will represent FDA's current thinking on these topics.