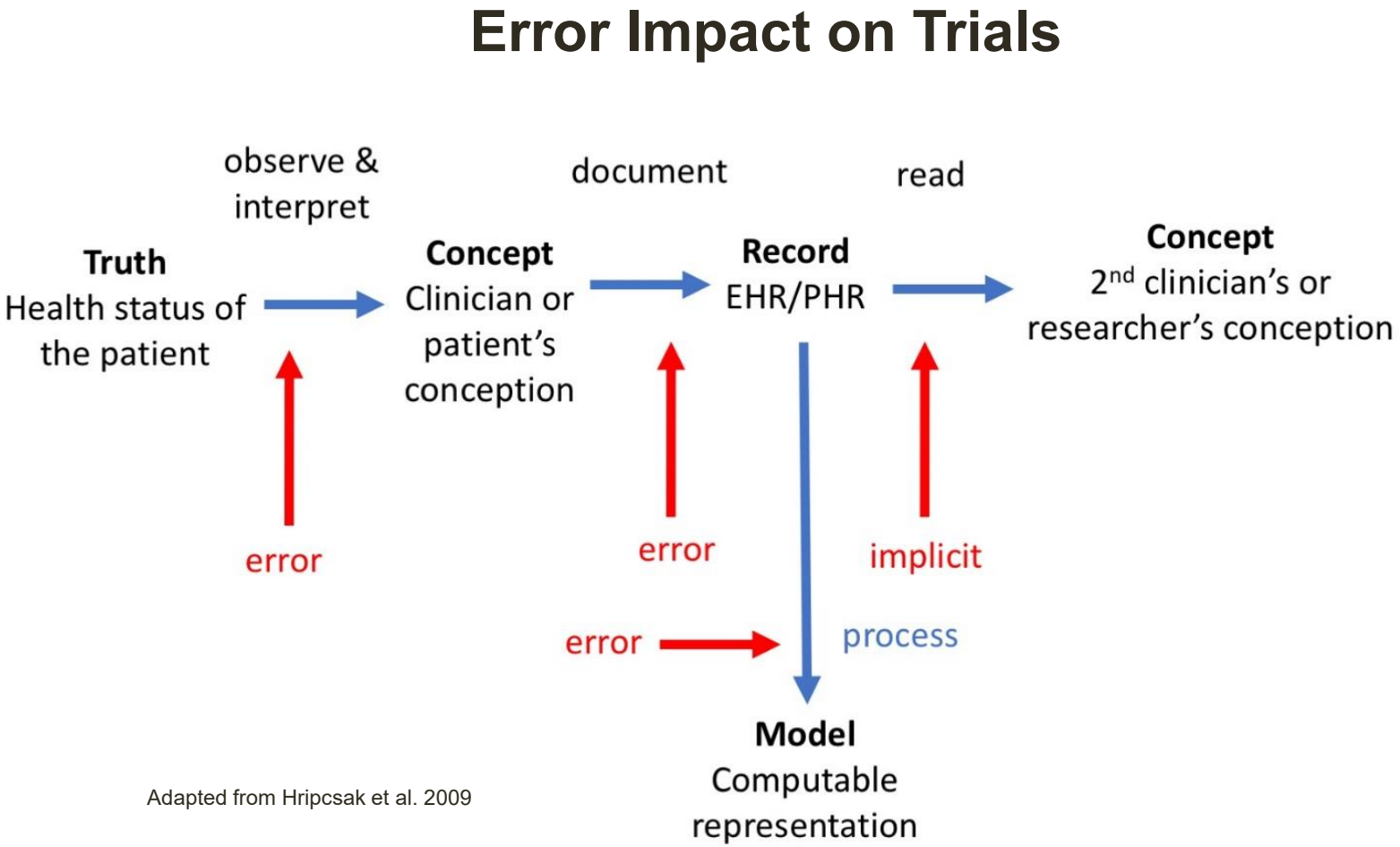
NIH Collaboratory *Rethinking Clinical Trials®*

Health Care Systems Research Collaboratory

Data Quality Assessment Recommendations

Rachel Richesson, PhD
Duke University
May 14, 2018

Data is a surrogate for clinical phenomena



Adapted from Hripcsak et al. 2009

Data Quality Assessment Recommendations

- Identify variation between populations
- Recommend formal assessment of accuracy, completeness & consistency for key data
- Data quality should be described, reported & informed by workflows

Rethinking Clinical Trials®

A Living Textbook of Pragmatic Clinical Trials

Welcome to the Living Textbook | Topic Chapters | Tools for Research | Blog | Contact Us | How to Cite

Assessing Data Quality

Created by the Collaboratory Phenotypes, Data Standards, and Data Quality Core

This white paper is a product of the Collaboratory Phenotypes, Data Standards, and Data Quality Core. Titled *Assessing Data Quality for Healthcare Systems Data Used in Clinical Research (V. 1.0)*, it provides guidance, based on the best available evidence and practice, for assessing data quality in pragmatic clinical trials (PCTs) conducted through the Collaboratory. Topics covered include an overview of data quality issues in clinical research settings, data quality assessment dimensions (completeness, accuracy, and consistency), and a series of recommendations for assessing data quality. Also included as appendices are a set of data quality definitions and review criteria, as well as a data quality assessment plan inventory. The full text of the document is available for download here.

Please note: this document opens as an Adobe PDF. If you do not have software that can open a PDF, click here to download a free version of Adobe Acrobat Reader.

https://www.nihcollaboratory.org/Products/Assessing-data-quality_V1%200.pdf

NIH Collaboratory Rethinking Clinical Trials®
Health Care Systems Research Collaboratory

Find a copy of the PCT Reporting Template in your meeting materials

Recommendations

- Key data quality dimensions should be assessed for data elements used in subject identification, outcome measures, and important covariates
  - accuracy, completeness, and consistency

- Describe formal assessments for completeness, accuracy, consistency, and impact

Recommendations

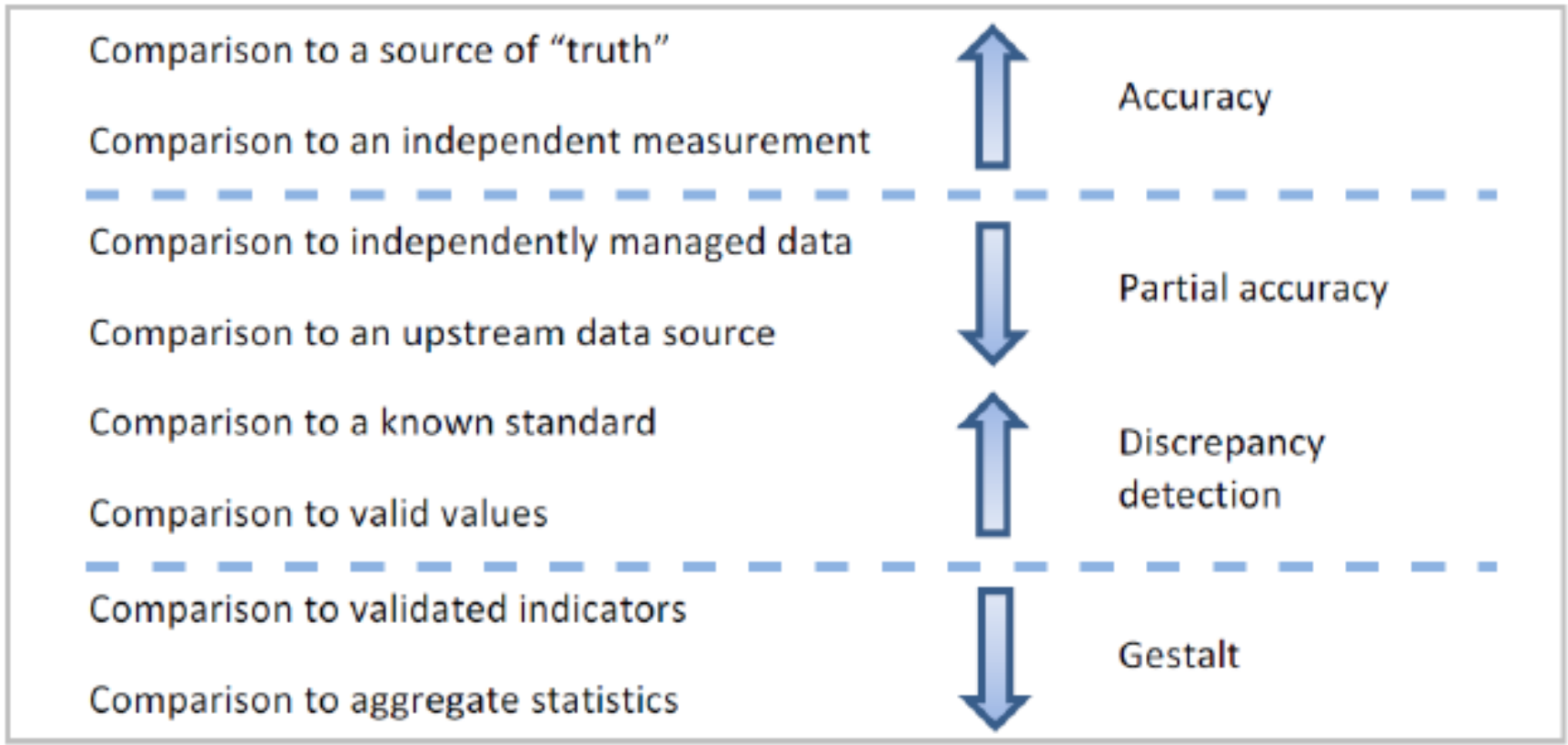
- Use of workflow and data flow diagrams to inform data quality assessment
  - “Talk though” (source, format) for each data element used in cohort identification
  - Describe all transformations from source data to final research repository
  - Are there differences in data capture, documentation, or transformation processes across sites?
  - Are there any subsets of data that may be collected or documented differently?
- Reporting data quality assessment with research results

Data Quality Dimensions Determining Fitness for Use of Research Data

Dimension	Conceptual definition	Operational examples
Completeness	Presence of the necessary data	Presence of necessary data elements, percent of missing values for a data element, percent of records with sufficient data to calculate a required variable (e.g., an outcome)
Accuracy	Closeness of agreement between a data value and the true value*	Percent of data values found to be in error based on a gold standard, percent of physically implausible values, percent of data values that do not conform to range expectations
Consistency	Relevant uniformity in data across clinical investigation sites, facilities, departments, units within a facility, providers, or other assessors	Comparable proportions of relevant diagnoses across sites, comparable proportions of documented order fulfillment (e.g., returned procedure report for ordered diagnostic tests)

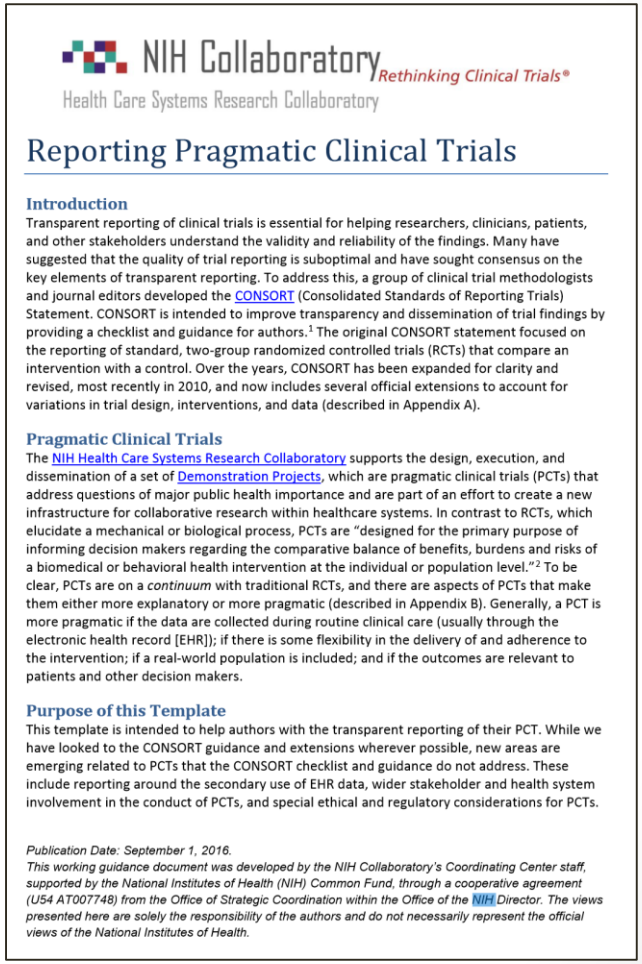
**Consistent with the International Organization for Standardization (ISO) 8000 Part 2 definition of accuracy, replaced "property value" in the ISO 8000 definition with "data value" for consistency with the language used in clinical research.*

Hierarchy of Approaches to Data Accuracy Assessment



Benefits of Data Quality Assessment Plan

- A robust data quality assessment plan can improve value of data and to detect and address data issues
- Data quality assessment results should be reported with final study results.
- Will enable readers to understand, interpret, and trust results.



NIH Collaboratory Rethinking Clinical Trials®
Health Care Systems Research Collaboratory

Reporting Pragmatic Clinical Trials

Introduction

Transparent reporting of clinical trials is essential for helping researchers, clinicians, patients, and other stakeholders understand the validity and reliability of the findings. Many have suggested that the quality of trial reporting is suboptimal and have sought consensus on the key elements of transparent reporting. To address this, a group of clinical trial methodologists and journal editors developed the CONSORT (Consolidated Standards of Reporting Trials) Statement. CONSORT is intended to improve transparency and dissemination of trial findings by providing a checklist and guidance for authors.[1] The original CONSORT statement focused on the reporting of standard, two-group randomized controlled trials (RCTs) that compare an intervention with a control. Over the years, CONSORT has been expanded for clarity and revised, most recently in 2010, and now includes several official extensions to account for variations in trial design, interventions, and data (described in Appendix A).

Pragmatic Clinical Trials

The NIH Health Care Systems Research Collaboratory supports the design, execution, and dissemination of a set of Demonstration Projects, which are pragmatic clinical trials (PCTs) that address questions of major public health importance and are part of an effort to create a new infrastructure for collaborative research within healthcare systems. In contrast to RCTs, which elucidate a mechanical or biological process, PCTs are "designed for the primary purpose of informing decision makers regarding the comparative balance of benefits, burdens and risks of a biomedical or behavioral health intervention at the individual or population level."[2] To be clear, PCTs are on a *continuum* with traditional RCTs, and there are aspects of PCTs that make them either more explanatory or more pragmatic (described in Appendix B). Generally, a PCT is more pragmatic if the data are collected during routine clinical care (usually through the electronic health record [EHR]); if there is some flexibility in the delivery of and adherence to the intervention; if a real-world population is included; and if the outcomes are relevant to patients and other decision makers.

Purpose of this Template

This template is intended to help authors with the transparent reporting of their PCT. While we have looked to the CONSORT guidance and extensions wherever possible, new areas are emerging related to PCTs that the CONSORT checklist and guidance do not address. These include reporting around the secondary use of EHR data, wider stakeholder and health system involvement in the conduct of PCTs, and special ethical and regulatory considerations for PCTs.

Publication Date: September 1, 2016.
This working guidance document was developed by the NIH Collaboratory's Coordinating Center staff, supported by the National Institutes of Health (NIH) Common Fund, through a cooperative agreement (U54 AT007748) from the Office of Strategic Coordination within the Office of the NIH Director. The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

NIH Collaboratory Rethinking Clinical Trials®
Health Care Systems Research Collaboratory