We Don't Know What We Don't Know

Judging Data Fitness for Use in Submissions

Premise:

- **Are data "fit for use"? It depends upon how we want to use them.**
- We must know the rules used to define, capture, database, clean, report and analyze them, including (in)consistencies in business rules among sites, studies, and submissions. Today we don't.
- Data fit for use within a study may be unfit when aggregated with other studies.
- Rules are documented but are often unavailable to those using the data, especially later

What We Don't Know Gets Bigger

Subject

Subject 1: No pregnant males | Subject 2: AEs for all Con Meds | Subject 3: Start date before stop date | Subject A: Lab outliers confirmed | Subject B: All expected data present | Subject C: No gaps in dosing records | Subject I: AE Action Taken ≈ Subject Disposition | Subject II: All events occurred between IC & end of study | Subject III: Race Other mapped if nec | Subject a: AE End Date consistent w/ AE Outcome | Subject b: Courses of Con Meds don't overlap | Subject c: Stop date & Ongoing consistent

Subjects are generally internally consistent and reasonably consistent within a site (Data Management cleans for this)

Rules in Data Management Plan: not accessible later

Site

Site 1: Coached subjects in completing survey | Site A: Didn't coach subjects in completing survey | Site I: Coached subjects in completing survey | Site a: Didn't coach subjects in completing survey

Some cross-site checking is done but mostly by statistician identifying possible center effects after database lock

Protocol violation; may/not be identified; would go in study report; not accessed after submission

Study

Study 1A: Sites told to record diagnoses, not symptoms

Study Ia: Sites told to record symptoms, not diagnoses

Some cross-study checking for efficacy and safety patterns but ones like these may not be identified

Rules in Monitoring Guidelines; not accessible later

Submission

Submission 1AIa: Capture AEs/SAEs beginning at Informed Consent

Submission 2BIIb: Capture AE/SAEs beginning at Day 1

Cross-submission checking very rarely happens unless, e.g., safety issue arises; searches typically narrow

Rules in Protocol; not typically accessed later

Warehouse

Submission 1AIa2BIIb: Extra variability in survey results; extra variability in diagnoses (e.g., flu) vs. symptoms (e.g., fever, nausea); can't assess prevalence or incidence of AEs during baseline

And We Don't Know It!

Kit Howard
Owner & Principal
Kestrel Consultants, Inc.
Kit@KestrelConsultants.com
Data Standards and Quality Expert
Co-Leader, CDISC Med Device Team
CDISC SDS Governance Board
CDISC Trainer
SCDM Editorial Board

How to Know:

1. Reviewers: ask for the rules - Data Management Plans, Monitoring Guidelines, Programming Assumptions, Analysis Plans, & more robust Reviewer Guidelines.
2. Sponsors: design data rules on a program rather than study level. Convey them to outsourcing partners.

Longer term...

3. Develop a schema for rules to travel with the data, perhaps becoming part of metadata.