Ensuring Data Quality in Animal Health Studies

Research Perspective

David Renter, DVM, PhD
Professor, Epidemiology
Director, Center for Outcomes Research and Epidemiology
College of Veterinary Medicine, Kansas State University
drenter@vet.ksu.edu

Center for Outcomes Research & Epidemiology
CORE
Kansas State University

Kansas State University

Ensuring Data Quality in Animal Health Studies

Research Perspective

Professor --- principal investigator….blinded clinician, necropsy (tech)……data analysis

Director, Center - Outcomes Research & Epidemiology

Consultation on design, implementation, data analysis and interpretation

---- consulting, service provider

CENTER FOR OUTCOMES RESEARCH & EPIDEMIOLOGY
CORE
KANSAS STATE UNIVERSITY

Ensuring Data Quality in Animal Health Studies

- A description of all circumstances that may have affected the quality or integrity of the data.
- …description of the transformations, calculations, or operations performed on the data, a summary and analysis of the data, and a statement of the conclusions drawn from the analysis.

Code of Federal Regulations

Ensuring Data Quality in Animal Health Studies

- Communications
- Minimizing Bias and Error
- “tracking the” Data – documentation
- Variability in Methods

Center for Outcomes Research & Epidemiology
CORE
Kansas State University

Kansas State University

Communications

- Federal – Sponsor – CRO – [analyst]….
- At all stages – from design, implementation, analysis, reporting…
- [analyst] - not just at the end!
  - frame the "question"
  - ensure design, etc. result in data that "fit" intended result
  - content expertise

Communications

- Sponsor asks for input on protocol development (now more than ever)
  - Procedures, but also design, data management and analysis
- Would CVM/Sponsors enable consultants (CROs and/or analysts) to communicate directly with the CVM on the sponsors behalf?
  (e.g. to avoid miscommunication on data quality issues)

Minimizing Bias and Error

- Bias – differential with respect to treatments
  - Systematically “wrong”
  - Randomization, blinding, etc.
- Error – not differential among treatments
  - Lack of precision (inaccurate)
  - Improve precision (processes) in data collection through analysis

KANSAS STATE UNIVERSITY

Minimizing Bias and Error

- Blinding and other bias reduction techniques
- For – clinical observations, but also for data analysis/interpretation
  - Blinding for analytics – (re)coding
  - Process, protocol (including who) defined in advance

Documentation - “tracking” Data

- Observations/measurements
- Records
- Data files
- Dataset(s) for analysis
- Analysis results

For examples:

"From the original raw datasheets provided by investigators, several data entry and management steps were performed……."

"Multiple spreadsheets were provided by the investigators; a single dataset was created for analysis following reformatting of existing variables and creating relevant outcome measures (e.g. adg) based on data provided…."

Documentation - “tracking” Data

- Defining procedures and standards of operation in advance; e.g. for:
  - *Procedures* management, files, transformations
  - *Logs* for data management – file naming, scripts/logs for manipulations, calculations, coding, etc.
  - *Logs* for model(s) and output(s) captured from analysis phase

For examples:

ID#	Treatment	Weight1	Weight2	Weight3	Date1…
1	Drug A	56 g	55 g	54 g	Jan 1
2	Drug B	55 g	55 g	54 g	Jan 1
3	Drug A + B	56 g	55 g	54 g	Jan 1
4	Control	56 g	56 g	57 g	Jan 1

Etc.

Often multiple “datasets” like this
(e.g. weight, clinical score, diagnostics….)

For examples:

ID	FactorA	FactorB	Weight	Day
1	1	0	56	1
1	1	0	55	5
1	1	0	54	10
2	0	1	55	1
……				

etc

Documentation - “tracking” Data

- Raw data, dataset(s), → dataset(s) for analysis
  - Log how data files are related, changed at each step
  - Logs for data management – “changes” in format e.g.
  - Data Dictionary (defines variables, units of measure, etc.)
  - Log: Result “X” from Output file “X” from Stat. Model “X” (code) ran on Dataset “X” created
    from Record (raw data) file(s) “X”

Variability in Methods/Approaches

- Statistical power “needs”, and calculations
- Design structure (pens vs animal EU)
  - Accounting for clustering (hierarchical, temporal)
- Statistical models/analysis methods
  - Software differences, options within software platforms
  - (methods to optimize “model fit”)

General and generalized linear mixed models we
used for all analyzes. Models were fitted usi
binomial (e.g. mortality), ordinal (e.g. s
normal (e.g. ADG) distributions, m
likelihood estimation, compli g-log link,
Kenward-Roger degrees and Newton-
Raphson and Ridgi ation procedures (Proc
GLIMMIX SAS 9 effects included the
treatment , and a random effect (intercept)
....to or the design structure (lack of
ence among pens within
s).....autoregressive covariance structure to
account for repeated measures over time.....with
Tukey adjustment for multiple comparisons.....

Could Be Implemented MANY Ways!

Ensuring Data Quality in Animal Health Studies

- Communications
- Minimizing Bias and Error
- “tracking the” Data – documentation
- Variability in Methods

Center for Outcomes Research & Epidemiology
CORE
Kansas State University

Kansas State University

Ensuring Data Quality in Animal Health Studies

- Communications
- Training/Consistency/Transparency
- Improve Precision – capture -> modeling
- Documentation - Details
- Communications

CENTER FOR OUTCOMES RESEARCH & EPIDEMIOLOGY
CORE
KANSAS STATE UNIVERSITY

KANSAS STATE UNIVERSITY

Ensuring Data Quality in Animal Health Studies

www.vet.k-state.edu/core

drenter@vet.ksu.edu

Kansas State University