Data Coordination and Management in Cardiovascular Disease

C. Michael Gibson, M.S., M.D.

Harvard Medical School

Chairman, PERFUSE Study Group

Founder and Chairman, WikiDoc & WikiPatient, The World's Open Source Textbook of Medicine Viewed 896 Million Times A Year

Data Coordination and Management Overview

- Goals of a Data Coordinating Center
- Study Design and Importance of Sample Size
- Role of Substudies
- Data Collection and Case Report Forms
- Active Trial Management

Goals of a Data Coordinating Center

- Assist in Study Design and Exploration of Study Objectives in varying scenarios
- Process, Clean, and Organize data
- Perform primary and secondary analyses of data

The ultimate goal is to be able to synthesize data ACROSS studies and to develop a library of data that is easily accessible and easy to query

Academic Research Organization (ARO)

- Non profit/academic
  - Usually associated with an academic medical center
  - Benefit: Able to analyze data and publish scientific manuscripts from the trial database in a timely and cost efficient manner
  - Examples include:
    - TIMI (Thrombolysis In Myocardial Infarction)
    - HCRI (Harvard Clinical Research Institute)
    - DCRI (Duke Clinical Research Institute)
    - C5 (Cleveland Clinical Coordinating Center)

The Spectrum of Clinical Trials

Registry

Case Control study

Single Center Randomized study

Multi-Center Randomized study

Retrospective

Prospective

Weak

Strong

Study objectives

- Each clinical trial must have a primary question the study seeks to address
  - The question the study is most interested in answering
  - used for the primary sample size calculation
  - should be framed in the form of a hypothesis
- Primary question as well as all secondary questions should be clearly defined and stated in advance

Study objectives - Example

- Demonstrate an improvement in epicardial patency with emergency room-based eptifibatide administration vs cath lab-based eptifibatide administration among ST elevation MI patients

Study objectives - Specifying hypotheses

- Hypotheses for a two-sided test to demonstrate a difference between interventions

$H_0: S_T = S_C$

Patency rate ***not*** higher with ER-based administration of eptifibatide vs cath lab based administration

$H_A: S_T \neq S_C$

Patency rate higher with ER-based administration of eptifibatide vs cath lab based administration

- Want to reject the null hypothesis of no difference

REJECT: Patency rate ***not*** higher with ER-based administration of eptifibatide vs cath lab based administration

Study objectives - Secondary questions

- Major secondary questions, like the primary question, should be stated in advance
- May be related to the primary question (e.g. cardiovascular death in a study of mortality)
- Sample size calculations should also be considered
- May be related to a *subgroup* of patients

Study objectives - Subgroups

- Results from subgroup analyses should be considered with caution
  - If enough statistical tests are done, some will be significant by chance (Type I Error)
  - Number of patients in a subgroup may be too small to show any difference even if one truly exists (Type II Error)
  - Looking for consistency with overall trial results

Sample Size Considerations

- Sample size needed to show a statistically robust difference in treatments

- Sample size usually based on primary endpoint, although can be based on secondary endpoint

Sample Size Considerations

- Sample size estimate based on three factors:
  - Estimated event rate in control arm (generally based on historical data)
  - Expected treatment difference
  - Acceptable error
    - alpha error – p-value
    - Beta error – Power

Example

- The study has an 80% odds of detecting a 20% treatment effect if it really exists ($p<0.05$)
  - Power 80%
  - 20% treatment effect
  - 2 sided test with $p<0.05$

The Spectrum of Clinical Trials

Registry

Case Control study

Single Center Randomized study

Multi-Center Randomized study

Retrospective

Prospective

Weak

Strong

Basic study designs

- *Randomized controlled clinical studies* are the standards against which all other studies are compared

- Randomization assigns patients to either the intervention group or control group “with the same probability”

Basic study designs - Randomized controls

- Advantages of randomizing treatment assignment
  - eliminates selection biases
  - produces comparable groups with respect to known (and unknown) risk factors
  - increases validity of statistical tests

Basic study designs – Non-randomized controls

- Patients are assigned to one of two groups, but not in a random fashion
- Patients are assigned *concurrently*
  - e.g., First patient in ER with MI treated with PCI, second patient in ER treated with lytic+PCI
- Advantages: easier to convince patients and investigators to participate
- Disadvantages: potential of ending up with groups that are not comparable

Basic study designs - Historical controls

- New intervention is studied in all patients prospectively
- Results are compared to the outcome from a previous study of comparable patients
- Historical controls are non-randomized, non-concurrent

Basic study designs - Historical controls

- Arguments for historical controls
  - all patients receive the “new” intervention
  - greater participation from investigators, patients
  - shorter studies

Basic study designs - Historical controls

- Concerns when using historical controls
  - accuracy and completeness when collected
  - open to bias
  - changes in patient population or patient management over time

- *A historical control study is no substitute for a randomized control clinical trial*

Data Collection Tools: Case Report Forms

- Case report forms should be designed to balance the need for parsimony and ease of data acquisition at the clinical site, with the scientific goal of obtaining a comprehensive and exhaustive data set

Case Report Form Design Requirements

- Capture the pre-specified endpoints of the trial
- Capture both expected and unexpected events
- Limit data collection to those items essential to the study's goals and that are practical to gather

- Capture the specific nuances associated with specific medications:
  - Thienopyridine
    - which thienopyridine
    - at what dose (for load and maintenance)
    - timing of dose (pre or post-PCI, how long in advance)
  - Devices
    - which stent (type, DES or bare metal)
    - what was the sequence of devices used (balloon or direct stent)
    - what segment were they used in

Case Report Form Design Requirements

- Questions must be clear to the user
- Careful attention to the “flow” of forms and questions
  - logical progression
  - concise instructions on each form questions directly relate to the protocol
- Instructions are aimed to assist the user
- Limit the amount of free text, use tickboxes - utilize a narrative summary form

Data Management

- Many programs available with different costs and skills needed
  - Expensive:
    - ClinTrials
    - Velos
  - Inexpensive:
    - Access
    - Excel

Data Cleaning & Double Data Entry

Cleaning algorithm detects that first entry does not equal second entry

Entry 1	Entry 2		
1	1	1	
2	2	2	
3	3	3	
4	4	4	
5	50	5	Adjudicated by CMG on 1/26/05

Cleaning algorithm detects that value lies outside of range, up to 5 in this case

Electronic paper trail created

Security of Data: Access to Data

- Release of data may have impact on market valuation, therefore must be kept secure
- Data coordinating center is secure
- Minimize transmission of data over the internet
- Critical data on one PC (not one network or multiple PCs), password protected
- Data PCs not connected to internet, cannot be “hacked into”

Database Management: The Academic Perspective

- Goals:
  - Conduct trials
  - Perform substudies
    - Subgroup analyses of treatment effect
    - Pathophysiology, hypothesis generating
  - Plan future trials
    - Anticipated event rates
    - Sample size estimates
    - Subgroup analyses