Biost 524:
Design of Medical Studies

•••••••••••••••••••••••••••••

Lecture 10:
Data Collection / Management

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

May 24, 2010

1

Lecture Outline

•••••••••••••••••••••••••••••

• Planning for Data Collection
• Data Collection
• Data Management

2

Data Collection

•••••••••••••••••••••••••••••

• Sir Josiah Stamp (my best recollection of the quote):

    – "The government are very fond of their statistics. They table them, raise them to the nth degree. But they always seem to forget that the information was collected in the first place by the village watchman, who just writes down what he damn well pleases."

3

Planning for Data Collection

•••••••••••••••••••••••••••••

By Role in Clinical Trial

Where am I going?

• We collect a wide variety of data in a RCT for a wide variety of purposes

4

Ultimate Goal

- Reporting a scientific experiment
  - Overall goal

  - Specific aims

  - Materials and Methods

  - Results

  - Conclusions

5

Ultimate Goal

- Reporting a scientific experiment
  - Overall goal

  - Specific aims

  - Materials and Methods
    - Patients, dosing, adherence to monitoring
  - Results
    - Disposition, compliance, adverse events, outcomes
  - Conclusions

6

Overall Goal / Specific Aims

- Role of data
  - Goal / aims ideally determined prior to start of study
  - BUT, the question actually answered is specific to
    - the subjects actually sampled
    - the methods actually used
    - the data actually gathered
    - the analysis actually performed
  - Generalization of results depends on all of the above

7

Materials

- Role of data
  - Eligibility criteria are usually broad
  - Need to describe the population actually sampled
  - Need to describe how the sample might differ from the ultimate target population

8

Conceptual Framework

• •

- Population of patients with disease
  - Definition of disease by cause vs signs / symptoms
- Subpopulation with disease targeted by intervention
  - I argue "disease" is really defined by treatment
- Subpopulation eligible for study accrual
  - Restricted due to general clinical trial setting
- Eligible patients from which sampled
  - Restricted due to specific clinical trial (location, time)
- Study sample
  - Restricted due to willingness to participate
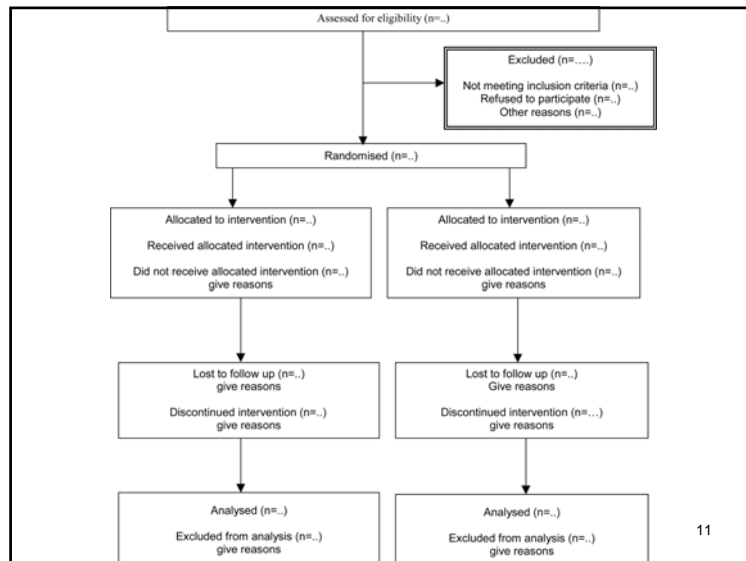- Analysis sample
  - Data collection

9

Generalizability

• •

- CONSORT: Consolidated Standards of Reporting Trials
  - Evidence based, minimum standards
  - Report flow of patients from screening to collection of primary outcomes
    - Screened
    - Enrolled
    - Randomized
    - Completed

10

11

Initial Screening Data

• •

- Source of screened patients
- Number screened
- Characteristics (may require consent)
  - Demographics
  - Disease characteristics
- Reasons for ineligibility
  - Inclusion criteria
  - Exclusion criteria
  - No participation
    - Unable to contact
    - Refused participation

12

Screening Visit(s) Data

- Consent for screening
- Contact information: Name, address, alternative contacts…
- Demographics: Sex, age, race, ethnicity…
- Disease characteristics: Duration, severity, …
- Prior and ongoing treatments
- Eligibility data
  - Inclusion criteria
  - Exclusion criteria
- Consent for randomization

13

Baseline Visit(s) Data

- Baseline data
  - Characterize patients
    - Severity of disease, concomitant disease…
  - Baseline measures of outcomes
    - Concomitant medications
    - Adverse events
    - Efficacy outcomes
      - E.g., initial SBP for reduction of HTN
      - E.g., tumor size for progression

- Note differing detail needed for screening vs baseline

14

Run-in Data

- Some clinical trials involve a run-in
  - Placebo: All patients take placebo
    - Washout vs assessing compliance
    - Patients may be blinded to existence of run-in
  - Active: All patients take experimental therapy
    - Allows randomized comparison of efficacy in patients actually taking drug
      - Randomized withdrawal of drug (among "responders"?)
      - Usually patients aware of run-in
    - Assess tolerability for AEs
    - Assess compliance

15

Randomization Data

- Documentation of eligibility
- Informed consent
- Stratification variables
- Variables needed for determination of dosing
  - Weight, BSA, renal function, severity of disease…
- Time, date of randomization
- Documentation of assigned group (blinded)
  - Cluster?
- Receipt of first treatment: time, date

16

Treatment Data: Why

- Intention to treat analysis is the standard for efficacy
  - Patients are analyzed in assigned group irrespective of their compliance
  - Compliance data is an outcome
    - Assess possible AEs
    - Assess possible mechanism for lack of effect
    - Describe realized exposure to treatment
    - Exploratory analyses for dose / response?
- Safety analyses are typically analyzed according to drug exposure
  - AEs / SAEs occurring within 28 days (?) of last dose

17

Treatment Data: What

- Initial assignment
  - Dose, administration, frequency, duration, ancillary treatments
- Protocol specified modifications
  - Dose reduction / escalation / holidays
    - Date, time
    - Reasons for change (AE, efficacy or lack of efficacy)
- Patient compliance
  - Dose, frequency, duration
  - Intermittent vs permanent change

18

Treatment Data: How

- Protocol specified modifications
  - Regularly scheduled visits
  - Interim visits
- Patient compliance
  - Patient diaries
  - Pill counts
  - Clocks on container lids
  - Biochemical measures: blood, biopsies

19

Patient Monitoring Data: Safety

- Protocol defined safety endpoints
  - Clinical events, subclinical laboratory measurements
- Adverse events
  - Review of interim AEs at regular visits
  - Undesirable clinical events that occur during the study
    - Treatment emergent: new or exacerbated
    - Classification (e.g. MEDRA), grade of severity
  - Treatment relatedness (but do not necessarily believe)
- Serious adverse events
  - Fatal, life-threatening, hospitalization or prolongation, birth defects
  - Expedited reporting if unexpected

20

Patient Monitoring Data: Efficacy

- Protocol defined efficacy endpoints
  - Clinical events
    - Create patient symptoms
  - Quality of life
  - Subclinical events
    - Signs thought to be indicative of clinical risk
    - Protocol specified monitoring schedules of
      - Patient performance (FEV, 6 minute walk, etc.)
      - Blood
      - Tissue biopsies
      - Radiology

21

Missing Data: Efficacy

- Missing data is a pervasive problem in RCT
  - Lack of training: Patients, investigators
    - Off study drug
    - Decline invasive procedures
    - Withdraw consent
  - Poor endpoint definition
    - All randomized patients must have defined outcome
      - E.g, Quality of life after death, GFR in dialysis, symptom relief with noncompliance
  - Sloppy conduct of RCT
    - Excessive loss of follow-up

22

Missing Data: Solutions

- Prevention of missing data is the only sure method
  - Increasing enrollment to "replace" missing data just attains greater precision on a biased outcome
  - Methods of modeling missing data are based on untestable assumptions
    - There is nothing in your data that can prove your assumptions are appropriate

23

Patient Monitoring Data: Compliance

- Patient adherence to measurement of outcomes
  - Clinic visits
    - Timing relative to window
  - Outcome assessments
    - Efficacy
      - Blood, tissue samples; radiology, special exams
      - Withdrawn consent for invasive procedures?
    - Adverse events
      - Periodic reports per protocol
      - Capture of interim SAEs

24

Patient Monitoring Data: Logistics

- Patient change of address
  - (sometimes schedule phone visits to maintain contact)
- Site compliance with timeliness completeness

25

End of Study

- Reason for stopping study
  - Completion per protocol
    - May be off study drug but still followed
    - Death
    - Withdrawn consent
      - Reasons
- Permission for further follow-up
  - Change of address
- Conjectured treatment assignment?

26

Sources of Data

- Subject self report
- Proxy for subject
- Clinic staff and study records
  - Standard medical care
  - Protocol specified procedures
- Medical records
- Laboratory, radiology, pathology
  - Local vs central labs
- Adjudication panels
- Public health records
  - Registries
  - National Death Index

27

Data Collection Issues

- Timeliness
  - Data collection and storage
- Completeness
  - Missing data
- Accuracy
  - Measurement methods, adjudication panels
- Precision
  - Measurement methods, adjudication panels

28

Data Collection Methods

- Forms
  - Abstracted from medical records
    - Indication bias
  - Completed by subject
  - Completed by proxy
  - Administered by study personnel
  - Completed by clinic staff, study personnel
  - Completed by adjudication panels
- Data files
  - E.g., laboratory, Medicare, National Death Index

29

Data Collection

- Development of forms
  - Administrative information
    - For follow-up, etc.
    - Often text
  - Scientific information
    - Needs to be appropriate for statistical analysis
      - Free text is difficult to analyze
      - Coding of response by person closest to the source

30

Data Collection

- Development of forms (cont.)
  - Format of forms should facilitate
    - Completion of form
      - Brief as possible
      - Make sure no portions overlooked
        » "skip patterns", two columns, back of page
      - Cover all cases (explicit "does not apply")
    - Data entry
      - Coding on form

31

Data Collection

- Issues in form development
  - Number of distinct forms
  - Guidance to the subject, clinic staff on form
    - Study specific definitions
    - Indications for study procedures, other forms
    - Convenience versus increased length
  - Manual and training for form completion
  - Forms for subject vs proxy vs administered
  - Translations
  - Pretesting: subject, staff, investigators, statistician
  - Mapping between different versions over time

32

Data Management

• Planning for data management
  – Data to be collected: What? Why?
  – Methods of collection: Who? Where? When? How?
    • Forms development
  – Methods for data storage
    • Development of database
      – Administrative data: often dynamic
      – Scientific data: usually static
  – Methods for data entry
    • Distributed versus central

33

Data Management

• Handling of data
  – Collection
  – Data entry
  – Storage of forms, primary records
  – Data verification
  – Checking for errors
  – Data reporting
  – Data analysis
  – Final database

34

Data Management: Entry

• Data entry
  – Transcription of data from forms into computerized data base

  – Personnel often low-level clerical staff
    • Little scientific knowledge

  – Minimize data entry errors
    • Screen for impossible values
    • Screen for inconsistencies within form
    • Double entry

35

Data Management

• Storage of forms, primary records

  – Subject confidentiality is a major concern

  – Must ensure limited access to confidential information

36

Data Management

- Data verification and checking for errors
  - Data entry errors

  - Data collection errors
    - Audit clinics
    - Compare study data to medical records

  - Maintaining an audit trail
    - Changing database versus making corrections in separate files

37

Data Management

- Data reporting
  - Administrative analyses
    - Accrual rates
    - Timeliness of data collection
    - Completeness of data collection

  - Baseline characteristics

  - Event rates (combined treatment groups only)

38

Data Management

- Data analysis

  - The ultimate purpose of collecting the data

  - Much easier, more generalizable if all the previous stages conducted properly

  - Complete record of all analyses should be maintained
    - date of analysis
    - version of data base and software

39

Data Management

- Bottom line

  - The key qualification you should look for in the data management personnel is attention to detail

40