Society for Clinical Trials 33rd Annual Meeting

Workshop P9

The Prevention and Treatment of Missing Data in Clinical Trials

Sunday, May 20, 2012
1:00 PM - 5:00 PM
Tuttle South

Society for Clinical Trials 2012: 33rd Annual Meeting

Workshop 9

The Prevention and Treatment of Missing Data in Clinical Trials

At the request of the U.S. Food and Drug Administration, the National Academy of Sciences convened the Panel on the Handling of Missing Data in Clinical Trials to prepare a report that would make recommendations that could be used to aid in the FDA's eventual development of a Guidance for Industry on that topic. This half day workshop presents an overview of the findings and recommendations of the resultant report from the perspective of two clinical trialist members of the NAS panel. The workshop will follow the basic organization of the NAS report, though it will place greatest emphasis on aspects of trial design and trial conduct that can be used to minimize issues arising from missing data. However, because trial protocols must also describe how any missing data will be handled at the end of the study, methods for analysis of clinical trial results will be discussed at a conceptual level. We will focus more on the common features of such analyses, than on the technical details of particular analytic methods. To that end, the target audience for this workshop includes biostatisticians and epidemiologists involved in the design of clinical trials, as well as study coordinators and CRAs involved in the conduct of the studies.

We first review settings in which missing data commonly arise and pose difficult problems in the analysis and interpretation of clinical trial results, as a basis for discussing aspects of clinical trial design that could minimize or even eliminate the most troublesome missing data. In particular we focus on aspects of clinical trial design that relate to appropriate definition of primary endpoints, anticipating problems that might arise when patients drop off study drug due to adverse events, lack of efficacy, or competing risks such as newly developed contraindications to therapy or deaths from other causes. We further consider alternative trial designs that would facilitate randomized comparisons among patients who can adhere to protocol defined treatment strategies.

We then consider aspects of trial conduct that will promote the collection and analysis of complete data on all randomized subjects. Proper attention should be paid to informing both investigators and participants of the scientific importance of complete data collection. We describe ways in which the Study Protocol, the Manual of Operations, and the Case Report Forms can facilitate the investigators' understanding of and adherence to the actions that must be taken to minimize missing data, as well as discussing the impact that careful subject education (including the Informed Consent documents) can have on preserving the scientific and statistical relevance of clinical trial results.

Major recommendations of the Panel also included the need for lead investigators to anticipate missing data and to plan for appropriate methods for the statistical analysis of the clinical trial results. We briefly discuss the need for easily understood and clearly described methods based on reasonable assumptions about the mechanisms giving rise to missing data and assumptions about the likely impact that missingness would have on conclusions drawn from the RCT. We give a broad, non-technical overview of some of the approaches that might be used for the primary analysis of the trial results. Then, owing to the impossibility of ever knowing that assumptions about missing data mechanisms are valid, we conclude with an overview of general criteria that should be met by sensitivity analyses that explore the potential impact of the assumptions about missing data.

Faculty: Scott S. Emerson, University of Washington
James D. Neaton, University of Minnesota

Workshop Organizer: Rick Chappell, University of Wisconsin

The Prevention and Treatment of Missing Data in Clinical Trials

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

James D. Neaton, Ph.D.
Professor of Biostatistics
University of Minnesota

Society for Clinical Trials

May 20, 2012

1

Background

Panel on Handling Missing Data in Clinical Trial

Where am I going?

The FDA commissioned the National Academy of Sciences to convene a panel to

- gather expert opinion and
- make recommendations

pursuant to the FDA's eventual development of a Guidance for Industry on how to address the pervasive problem of missing data in RCT.

It is of interest to consider the types of input we received.

2

Oversight Committee

- Experts in missing data methodology and clinical trial methodology

Roderick Little, Chair	
Ralph D'Agostino	Susan Murphy
Kay Dickersin	James Neaton
Scott Emerson	Andrea Rotnizky
John Farrar	Daniel Scharfstein
Constantine Frangakis	Weichung (Joe) Shih
Joseph Hogan	Jay Siegel
Geert Molenberghs	Hal Stern

3

Process

- Gathering of information: Workshops
  - FDA
  - Industry
  - Academia
- Preparation of report
- Outside review of report

4

Outline of Report

- Ch 1: Introduction and Background
  - RCT setting, randomization, regulatory setting
- Ch 2: Trial Designs to Reduce Missing Data
  - Estimands, alternative study designs, continued data collection
- Ch 3: Trial Strategies to Reduce Missing Data
  - Actions at design, actions during conduct, targets
- Ch 4: Drawing Inferences from Incomplete Data
  - Missing data probability models, analytic methods
- Ch 5: Principles and Methods of Sensitivity Analyses
- Ch 6: Summary and Recomendations

5

Workshops: What I Learned

- Mission 0a: Consolidation of Clinical Trial Terminology
  - Safety, efficacy, effectiveness
    - What is the estimand?
  - Definition of treatment
    - Treatment versus strategy
  - Study design
    - Standard cohort, placebo vs active run-in
  - Timeframe for primary endpoint
    - Event time, study time, calendar time
  - Multiple endpoints
    - Composite vs co-primary vs primary & secondary
  - Study termination
    - Completion of protocol, stop intervention, consent withdrawn
  - Analysis populations
    - ITT, mITT, per-protocol, safety

6

Workshops: What I Learned

- Mission 0b: Consolidation of Missing Data Terminology
  - Mechanisms generating missing data
    - Toxicity, efficacy (or lack), no longer relevant
    - Sloppy data capture, loss to follow-up, withdrawn consent
  - Statistical definition of missing data mechanisms
    - MCAR, MAR, MNAR
  - Statistical impact of missing data mechanisms
    - Ignorable/non-ignorable
  - Statistical methods
    - Direct imputation (LOCF, BOCF), MMRM, MI, pattern mixture, weighting
  - Types of sensitivity analyses
    - About assumptions of analytic models
    - About assumptions of MCAR, MAR, MNAR

7

Missing Data: Ideal

"Just say no."

(Nancy Reagan)

8

Common Problems (Report)

- Missing data due to discontinuation of treatment
  - Adverse events vs lack of efficacy vs efficacy
  - Specified by protocol vs perception of subjects or investigators
    - Relevance of data *vis a vis* health status, rescue therapies
- Outcomes undefined or unmeasurable for some patients
  - Counterfactual estimands (e.g., QoL after death)
  - Competing risks (e.g., renal function after transplant)
- Missing data because of attrition in the course of the study
  - Missed visits, loss to follow-up, withdrawal of consent
- Missing data in composite outcomes
- Missing data due to death

9

Regulatory Setting

- Need to establish
  - Safety, efficacy, effectiveness
  - Short vs long term effects, dose response
  - Subpopulations, concomitant treatments
- Clinical trials
  - Science: Basic science vs clinical science
  - Statistics: Magnitude of effect vs strength of evidence
  - Game theory: “Intent to cheat” analyses
    - Need for prespecification of endpoints, analyses
- Attempts to use a single trial to address all goals often leads to missing data

10

Primary Findings (SSE)

- From the viewpoint of a statistician scientist:
  - Always: define testable hypotheses relevant to question
  - Build necessary evidence from multiple studies as indicated
- Most difficult problems with missing data in clinical trials are due to poorly defined indications being tested
  - Disease, population, treatment, and/or outcome
- The second major cause is poor training of investigators
  - Poor understanding of true clinical question that needs to be addressed and regulatory environment
  - Leads to terminating data collection early
- True scientific dilemmas exist, but they are in the minority
  - Economic dilemmas are more often the problem

11

Common Problems: “Data Issues”

- Sometimes the problem is one of adherence to the protocol
- Patients can
  - Refuse individual measurements
  - Miss visits
  - Discontinue treatments
  - Move away
  - Withdraw consent
- RCT investigators can
  - Be lax in contacting patients, scheduling visits
  - Be lax in data collection, data management
  - Encourage patients to withdraw inappropriately

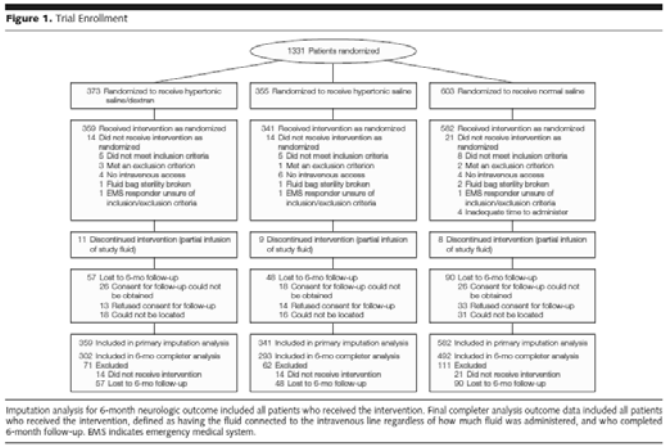
12

Example: Hypertonic Resuscitation in TBI

- RCT conducted in prehospital setting
  - Exception from informed consent in emergency setting
- Patients with low level of consciousness randomized by EMS
  - Notification of inclusion after regain consciousness
  - Notification of right to withdraw
- Primary endpoint: Glascow Outcome Score – Extended at 6m
- Issues with missing data: Difficult follow-up
  - (Deaths)
  - Withdrawn consent
  - Variable adherence to timing of follow-up
  - Loss to follow-up

13

Example: Hypertonic Resuscitation in TBI



Example: Hypertonic Resuscitation in TBI

- How to handle missing data?
- Withdrawn consent not missing completely at random
  - Only surviving patients could withdraw
  - Patients in extended rehabilitation unlikely to move away
- Nonadherence to monitoring schedule
  - Follow-up at 5 mos instead of 6 mos?
  - Follow-up at 12 mos instead of 6 mos?

15

Common Problems: “Scientific Issues”

- Sometimes the problem is the definition of the question
- In their usual clinical course, patients can
  - Need ancillary therapies to control AEs, etc.
  - Develop contraindications to treatments
  - Need to advance to other therapies
  - Die
- There is a need to define outcomes such that they apply to all randomized patients

16

Example: Second Line Therapy NSCLC

- TAX317
  - Non-small cell lung cancer
  - Patients who have “failed” first line therapy
  - Docetaxel 75, 100 mg/m^2 vs best supportive care (BSC)
    - 100 mg/m^2 arm dropped at interim analysis
- Secondary endpoint of overall survival (OS)
  - Median: 7.5 mos DOC75 vs 4.6 BSC
  - HR: 0.484, p = .004 (adjusted)
- Above analysis censored subjects at the time they advanced to other therapies

17

Example: Second Line Therapy NSCLC

- Pemetrexed as second line in NSCLC
- Noninferiority trial compared to docetaxel
- Patients who progress on pemetrexed may cross-over to docetaxel
  - Ethics: They have not yet been tried on approved therapy
- How should we analyze this data for OS?

18

Example: Everolimus in NET

- Neuroendocrine tumors
  - Pancreatic neuroendocrine tumors
  - Carcinoid
- Trial design
  - Primary endpoint: PFS by central radiology
  - Randomized, double blind, placebo controlled
  - Treatment: Randomized intervention until investigator determined progression
- Placebo group crosses over to open-label everolimus
  - How to analyze PFS when discordant views on progression?
  - How to analyze OS in presence of cross-over?

19

Example: Chronic Renal Disease

- Effect of treatment on glomerular filtration rate
- Primary endpoint: GFR at 6 months
- Some patients progress to dialysis
  - Does this preclude measurement of the endpoint?
- Some patients progress to renal transplant
  - How about now?

20

Example: Radiation Oncology

- Local irradiation of tumors
- Primary outcome: Tumor recurrence
- Some patients die of distant disease
  - How do we handle measurements of local recurrence?

21

Example: Hypothermia in AMI

- Patients presenting with acute myocardial infarction receive hypothermia
- Primary endpoint: Size of infarct measured by SPECT at 30 days
  - Deaths will be imputed to have same infarct size as worst observed
    - Still a problem even if using ranks and assigning worst rank
- Secondary endpoint: Size of infarct ignoring deaths

22

Example: Uveitis

- Patients initially treated with steroids
- Randomized to receive anti-inflammatory drug
  - Taper of steroids
- Primary endpoint: Visual acuity at 24 weeks
  - Patients who do not successfully taper steroids are ignored
    - Data collection stops prior to 24 weeks

23

Example: Chronic Pain

- Study design
  - Patients randomized to experimental treatment or placebo
  - Patients often recruited after being on some therapy chronically
- High rates of dropout
  - Potential toxicities to new therapy
  - Potential lack of efficacy to placebo
- Actions on progression
  - Return to prior therapy
  - Use of more potent analgesia (e.g., morphine)

24

Example: Quality of Life

- Improvement in quality of life in cancer treatments
- Primary endpoint: Average QoL over 12 months
  - QoL measured using validated instrument q 6 weeks
    - Questionnaires with specific subdomains
    - Functional tests (e.g., 6 minute walk)
  - How to handle deaths?

25

Example: Antifibrinolytics in ChemoTX

- Patients undergoing chemotherapy for cancer often experience increased risk of bleeding due to low platelets
- Hypothesize that platelets are being used up due to repeated dissolving of clots
- Consider prophylaxis with antifibrinolytics to decrease rates of serious bleeding in first 30 days of chemotherapy
- Issues
  - Some patients will die of their underlying disease
  - How do we record bleeding incidence in such patients?

26

Example: True Scientific Dilemmas

- Sometimes hard to score worst case
  - Death in a HTN study
- Sometimes measurement on patient truly irrelevant
  - Liver function in patients awaiting liver transplant
  - HTN in preeclampsia preceding delivery
- Some populations are notoriously difficult
  - Psychiatric patients, drug users, homeless, …
- AND: Some questions cannot be answered with a RCT
  - Ethics: Effect of smoking on lung function in children
  - Physiology: Effect of REM sleep deprivation on cardiovascular parameters

27

Missing Data: Real Life

"Missing data happens"

(Bumper Sticker- rough translation)

28

Analyses of RCT with Missing Data

- Methods should use the best scientific information we have available
  - As simple and straightforward as possible
  - But certainly not overly simplistic
- HOWEVER: nothing in your data can tell you whether missing data is ignorable or nonignorable
  - You just have to deal with what you worry about
  - At the time of study design, plans should be made
    - Do the best that you can to prevent it!
    - Sensitivity analyses? Imputation? Ignore?

29

Missing Data: Sad Facts of Life

"Bloodsuckers hide beneath my bed"

"Eyepennies",
Mark Linkous (Sparklehorse)

30

Missing Data

Preliminary Terminology

Where am I going?
We need to define
- Mechanisms by which missing data occur
- Statistical classification of missing data mechanisms
- Statistical impact of missing data

(I will later provide more detailed terminology.)

31

Roles of Data in RCT

- Eligibility variable
- Precision variables
- Treatment indicator
- Ancillary treatments
- Adverse events
  - While still actively taking treatment
  - During active follow-up
  - While on different treatments (rescue) / trials
- Interim measures of outcomes
- Outcomes

32

Problem by Role of Data

- Eligibility data
  - Affects generalizability
  - Especially a problem in "modified intent to treat analyses" (mITT)
    - mITT: Restricted based on variables defined prior to randomization
- Ancillary treatments
  - Truly an outcome, but of interest as effect modifier
- Efficacy / effectiveness outcomes (longitudinal)
  - Major focus of methods has been on partial follow-up
    - "Monotone" missing data: Once missing, always missing thereafter
- Safety outcomes (longitudinal)
  - May be of interest in wider population than efficacy population
  - Time frame of interest may differ from the efficacy endpoint

33

Mechanisms for Missing Data

- Owing to (improper) definition of estimand
  - Competing risks, etc.
- Only three broad categories
  - Withdrawal of consent
  - Loss to follow-up
  - Sloppy data collection
- With withdrawal of consent and loss to follow-up need to consider
  - Toxicity profiles
  - Efficacy or lack thereof
- With sloppy data collection need to consider biases

34

Statistical Classification of Missing Data

- Missing completely at random (MCAR)
  - The indicator of missingness does not depend upon any measured data
    - Sometimes confused with ignorability
- Missing at random (MAR)
  - Within groups defined by some observed data, the data is missing completely at random
  - Information about missing data can be borrowed from data that is available
- Missing not at random (MNAR)
  - Even after conditioning on all observed data, the subjects missing data would have outcomes distributed differently than those for subjects with observed data

35

Statistical Impact of Missing Data

- Ignorable
  - Weak: Analyzing complete cases in the planned analyses provides unbiased estimates of the desired estimand
    - MCAR
    - MAR if we were going to adjust anyway
  - Strong: Just as precisely?
- Nonignorable
  - Failure to account for missingness results in biased estimation of the desired estimand

36

Overview of Clinical Trial Design

Science and Statistics

Where am I going?

In the real world, clinical trial design must consider

- scientific theory
- statistical theory
- logistical issues
- game theory

I make an argument (plea?) for clinical trial design to consider science first, then statistics

(Game theory is a necessary evil)

37

Clinical Trials

- Experimentation in human volunteers
- Investigates a new treatment/preventive agent
  - Safety:
    - Are there adverse effects that clearly outweigh any potential benefit?
  - Efficacy:
    - Can the treatment alter the disease process in a beneficial way?
  - Effectiveness:
    - Would adoption of the treatment as a standard affect morbidity / mortality in the population?

38

The Enemy

"Let's start at the very beginning, a very good place to start…"

- Maria von Trapp

(as quoted by Rodgers and Hammerstein)

39

First

- Where do we want to be?
  - Describe some innovative experiment?
  - Find a use for some proprietary drug / biologic / device?
    - "Obtain a significant p value"
  - Find a new treatment that improves health of some individuals
    - "Efficacy"
  - Find a new treatment that improves health of the population
    - "Effectiveness"

40

Treatment "Indication"

- Disease
  - Therapy: Putative cause vs signs / symptoms
    - May involve method of diagnosis, response to therapies
  - Prevention / Diagnosis: Risk classification
- Population
  - Therapy: Restrict by risk of AEs or actual prior experience
  - Prevention / Diagnosis: Restrict by contraindications
- Treatment or treatment strategy
  - Formulation, administration, dose, frequency, duration, ancillary therapies
- Outcome
  - Clinical vs surrogate; timeframe; method of measurement

41

Evidence Based Medicine

- Decisions about treatments should consider PICO
  - Patient (population)
  - Intervention
  - Comparators
  - Outcome
- There is a need for estimates of safety, effect

42

Safety

- Multiple levels of concern
- Safety of conducting RCTs
  - Phase I dose finding studies
- Safety in the ideal population
  - Phase II or phase III efficacy studies
- Safety in the general population
  - Phase III effectiveness studies
  - Vulnerable populations
  - Concomitant renal, liver disease
  - Expansion of indication to patients with little benefit
  - Changes in behavior associated with adoption
  - Rare but serious adverse events

43

Efficacy

- An efficacious treatment has demonstrated an ability to beneficially modify
  - An endpoint thought to be an indicator of good clinical outcome
  - In some subset of patients
  - Under some conditions that are at least marginally relevant

44

Effectiveness

- An effective treatment will, upon adoption, improve the average health of the population
- N.B.: Effectiveness is a very hard thing to demonstrate in a RCT, but there are gradations

45

Efficacy: A Moving Target

- Definition of efficacy can vary widely according to choice of endpoint and magnitude of importance
  - Basic science
    - Does treatment have any effect on the pathway
  - Clinical science
    - Does treatment have a sufficiently large effect on a clinically relevant endpoint

46

Effectiveness: A Moving Target

- A treatment is “effective” if its introduction improves health in the population
- A treatment can be both efficacious and ineffective depending on factors of clinical trials
  - Target population
  - Control treatment
  - Intervention
  - Measurement of outcome(s)
  - Summary measure of outcome distribution

47

Ultimate Goal

- Medical science in general, and the FDA in particular, is rightly concerned with the process by which new treatments are adopted
- Randomized clinical trials are the mainstay of this process
- Obviously, effectiveness is our eventual goal
  - There are many ways to get there, however
    - Study safety, efficacy separately
    - Study bottom-line “effectiveness”
- Sometimes scientific / clinical judgment holds sway
  - Can results of RCT be safely generalized to other settings?

48

RCT Tools

- The key tools for a well conducted RCT are all part of the scientific method
  - Interventional experiment
    - Ensures proper definition of indication
  - Well defined study protocol
    - Avoids multiple comparisons
  - Randomized assignment
    - Ensures comparability of treatment arms (on average)
  - Unbiased ascertainment of results

49

Statistical Refinement of Hypotheses

- The group receiving the treatment will <u>tend to</u> have outcome measurements that are

higher than, / lower than, or / about the same as

an absolute standard, or / measurements in an otherwise comparable group (that did not receive the treatment)

50

Points Meriting Repeated Emphasis

- Randomization is our friend…
  - If we randomize, we do not (on average) need to worry about differences between the treatment groups with respect to factors present at time of randomization
  - Any difference in outcomes can be attributed to treatment
  - Again, recognize that treatment can lead to differential use of other ancillary treatments, however
- But like all friends, we must treat it with respect.
  - We must analyze our data in groups defined at the time of randomization
  - Discarding or missing data on randomized subjects may lead to bias
  - It certainly leads to diminished scientific credibility

51

Comment on “Intent to Treat”

- I view this term problematic
  - Originally, it was coined to describe the estimand associated with a “by-randomization” analysis when the target population is everyone who would ultimately be started on an effective therapy
  - The term is widely abused
- “By-randomization” is the true goal
  - The RCT may not be considering an intention to treat, e.g.,
    - Randomized withdrawal among tolerators
    - Randomized withdrawal among responders
    - Restricted eligibility criteria
    - Restricted ancillary therapies
    - etc.

52

Clinical Trial Design

- Finding an approach that best addresses the often competing goals: Science, Ethics, Efficiency
  - Basic scientists: focus on mechanisms
  - Clinical scientists: focus on overall patient health
  - Ethical: focus on patients on trial, future patients
  - Economic: focus on profits and/or costs
  - Governmental: focus on safety of public: treatment safety, efficacy, marketing claims
  - Statistical: focus on questions answered precisely
  - Operational: focus on feasibility of mounting trial

53

Basic Science

"Knowledge is good"

- Emil Faber
  Founder, Faber College

54

Clinical Science

- Goal tends to be more bottom line
- What can improve the health of a patient?
- Considers the entire sequence of treatments administered to a patient

55

Regulatory Agencies

- Considers treatment costs / benefits
  - Safety
  - Efficacy
- Considers public health
  - Effectiveness
- Ultimately a governmental setting
  - Approval of introduction of drugs, biologics, devices, diagnostics
  - Oversight of marketing claims
  - Responding to political (economic) pressures

56

Carrying Coals to Newcastle

- Wiley Act (1906)
  - Labeling
- Food, Drug, and Cosmetics Act of 1938
  - Safety
- Kefauver – Harris Amendment (1962)
  - Efficacy / effectiveness
    - " [If] there is a lack of substantial evidence that the drug will have the effect ... shall issue an order refusing to approve the application. "
    - "...The term 'substantial evidence' means evidence consisting of **adequate and well-controlled investigations, including clinical investigations**, by experts qualified by scientific training"
- FDA Amendments Act (2007)
  - Registration of RCTs, Pediatrics, Risk Evaluation and Mitigation Strategies (REMS)

57

Medical Devices

- Medical Devices Regulation Act of 1976
  - Class I: General controls for lowest risk
  - Class II: Special controls for medium risk - 510(k)
  - Class III: Pre marketing approval (PMA) for highest risk
    - "…**valid scientific evidence** for the purpose of determining the safety or effectiveness of a particular device … adequate to support a determination that there is reasonable assurance that the device is safe and effective for its conditions of use…"
    - "Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, **from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness**…"
- Safe Medical Devices Act of 1990
  - Tightened requirements for Class 3 devices

58

Body of Evidence

- Ultimately, regulatory approval will be based on
  - Current scientific knowledge and beliefs
  - Historical studies, both observational and RCT
  - Preclinical evidence: *in vitro* and animal studies
  - Preliminary studies: Phase 1, 2
  - Registrational confirmatory trials
- Strength of evidence
  - Rigorous evidence from adequate and well-controlled RCT
  - Scientific and clinical judgment generalizing those results to
    - Related diseases and more general populations
    - Variations in treatment strategies
    - Impact on long term outcomes

59

Scientific Judgment in Burden of Proof

"Keep an open mind, but not so open that your brains fall out."

- Virginia Gildersleeve?

60

Scientific Judgment in Burden of Proof

- We cannot answer every question with a RCT
- We always have to take some leap of faith
  - But we should try to keep it to a hop
- Science is adversarial
  - When have we demonstrated safety, efficacy, effectiveness to meet reasonable doubt?

61

Ultimate Goals

- The ultimate goals of basic science, clinical science, and regulatory agencies are somewhat different
- These goals are manifest in all aspects of a treatment indication

62

Definition of Disease

- Basic science: Ideal
  - Defined by cause of disease
- Clinical science: Moving target
  - Putative cause vs constellation of symptoms, signs
  - Sometimes reflects response to prior therapy
    - Second line chemotherapy, MRSA, etc.
  - Ultimately refined according to effective therapies
- Regulatory
  - Reproducible definition

63

Definition of Population

- Basic science
  - Restrictions based on contraindications of treatment
- Clinical science
  - Considers perceived risks
  - Considers alternative therapies
- Regulatory
  - Legal criteria for efficacy
  - Legal criteria for safety of public
  - Reproducibility
    - E.g., diagnostic criteria for genetics

64

Definition of Treatment

- Basic science
  - Effect of precisely defined formulation, dose, administration, frequency, duration, concomitant treatments
- Clinical science: Treatment strategy encompassing some of
  - Modifications of dose, frequency, etc.
  - Prophylactic or concomitant control of adverse treatment effects
  - Rescue and follow-on therapies
- Regulatory
  - Safety margins
  - Reproducibility of treatment definitions

65

Definition of Outcomes

- Basic science
  - Intermediate endpoints along causal pathway
- Clinical endpoint
  - Measurable (ethically) for every subject (e.g., anticipate deaths)
  - Long term: clinical benefit
  - Short term: until next treatment decision point
- Regulatory
  - Concordance with public health benefit
  - Concordance with clinical practice
    - Perceived clinical goal (e.g., HTN)

66

Impact on RCT Design

- Ultimately, regulators must approve a specific indication
- However, in the process of gathering evidence in support of approval, different RCT may be actually testing different indications
  - Integrating these results will often come down to scientific and clinical judgment
- But we want each RCT to rigorously answer the question it was designed to answer

67

Recommendation # 1

- A RCT protocol should explicitly define
  - Objective(s) for the trial
  - Primary outcome(s)
  - How, when, on whom the outcome is measured
  - The measures of intervention effects that reflect the causal estimands of primary interest
- The measures of intervention effects must
  - be meaningful for all study subjects, and
  - be estimable with minimal assumptions
- The protocol should therefore address the potential impact and treatment of missing data

68

Current Focus: Adequate, Well-Controlled RCT

- How does the potential for missing data alter the experimental strategy (series of studies) to establish effectiveness?
- How can we minimize the occurrence of missing data in RCT?
- How does the presence of missing data in a RCT change the analysis strategies?
- How can we assess the potential impact that missing data (and the prespecified methods for dealing with it) has on our confidence in the RCT results?

69

Bottom Line

"You better think (think)
about what you're
trying to do…"

-Aretha Franklin, "Think"

70

Estimands

Scientific Goals

Where am I going?

Given my claim that most of the truly difficult missing data problems are from poorly defined primary endpoints, it is useful to consider the quantities that we wish we knew.

(Later, we will consider the processes by which we might be able to estimate these in a RCT.)

71

Scientific Questions of Interest

- We can first consider the types of questions we might want to know the answer to
  - Safety & efficacy vs effectiveness
    - Single endpoints, multiple endpoints, composite endpoints
  - Population defined by treatment compliance
- Conceptually, these can be discussed in a single arm interventional trial

72

"Competing Risks"

- Incidence of one event precludes observation of another
  - Time to event analyses: Cause specific mortality
  - All analyses: Withdrawal of consent, loss to follow-up
  - Depending on estimand: Noncompliance, death
- Possible solutions
  - Most important endpoint
    - E.g., overall survival
  - Composite endpoints
    - Progression free survival
    - Quality adjusted life years
    - Major cardiovascular adverse events (MACE)
    - Ventilator free days alive during first 28 days

73

Composite Endpoints: Issues

- Clinical relevance of time to first event
- Need to avoid combining endpoints of markedly different clinical importance
  - Death
  - Progression
  - Termination of study drug
- Composites involving invasive procedures require special consideration
  - E.g., liver biopsies may be too risky in some patients
- Regulatory issue
  - How to write an indication for a nonstandard composite endpoint

74

Scientific Estimands

- Efficacy of treatment
  1. What is impact among patients who follow protocol?
  2. What is impact among patients who could follow protocol?
  3. What is impact among patients who start treatment?
- Safety of treatment
  1. What is impact among patients who follow protocol?
  2. What is impact among patients who could follow protocol?
  3. What is impact among patients who start treatment?
- Effectiveness of treatment
  - What is impact among patients who would knowingly start treatment?

75

Scientific Efficacy Estimand #1

- What is impact among patients who follow protocol?
  - No matter what: An interesting basic science question
  - Clinically may be used to explore mechanism of action
- Patients who do not follow protocol are irrelevant
  - Patients who do not follow directions
  - Patients who have intolerable adverse reactions
    - Perhaps "intolerable" only because uncertain of efficacy, or
    - Perhaps leading to serious consequences with continued therapy
  - Patients with real or perceived lack of efficacy
    - Early clinical course is discouraging, or
    - Definitive progression to serious condition prior to primary endpoint
  - Development of contraindication to treatment (e.g., pregnancy)
  - Patients with early evidence of cure

76

Scientific Efficacy Estimand #2

- What is impact among patients who could follow protocol?
  - No matter what: A relevant basic science question
  - Depending on safety: Possibly relevant to clinical science
  - Requires estimating outcomes among noncompliant patients
- Some patients who do not follow protocol are irrelevant
  - ~~Patients who do not follow directions~~
  - Patients who have intolerable adverse reactions
    - ~~Perhaps "intolerable" only because uncertain of efficacy, or~~
    - Leading to serious consequences with continued therapy
  - Patients with real ~~or perceived~~ lack of efficacy
    - ~~Early clinical course is discouraging, or~~
    - Definitive progression to serious condition prior to primary endpoint
  - Development of contraindication to treatment (e.g., pregnancy)
  - ~~Patients with early evidence of cure~~

77

Scientific Efficacy Estimand #3

- What is impact among patients who start protocol?
  - No matter what: A relevant basic science question
  - Highly relevant to clinical science
    - But does presume no change in behavior after knowing efficacy
  - No need to estimate outcomes among noncompliant patients
- All patients' data is relevant
  - Hence need to collect efficacy data (in an unbiased fashion) following stopping therapy

78

Scientific Safety Estimand #1

- What is impact among patients who follow protocol?
  - No matter what: An interesting basic science question
  - Clinically, may be used to estimate dose response
- Patients who do not follow protocol are irrelevant
  - Patients who do not follow directions
  - Patients who have intolerable adverse reactions
    - Perhaps "intolerable" only because uncertain of efficacy, or
    - Perhaps leading to serious consequences with continued therapy
  - Patients with real or perceived lack of efficacy
    - Early clinical course is discouraging, or
    - Definitive progression to serious condition prior to primary endpoint
  - Development of contraindication to treatment (e.g., pregnancy)
  - Patients with early evidence of cure

79

Scientific Safety Estimand #2

- What is impact among patients who could follow protocol?
  - No matter what: A relevant basic science question
  - If answerable: Definitely relevant to clinical science
  - Requires estimating outcomes among noncompliant patients
- Some patients who do not follow protocol are irrelevant
  - ~~Patients who do not follow directions~~
  - Patients who have intolerable adverse reactions
    - ~~Perhaps "intolerable" only because uncertain of efficacy, or~~
    - Leading to serious consequences with continued therapy
  - Patients with real ~~or perceived~~ lack of efficacy
    - ~~Early clinical course is discouraging, or~~
    - Definitive progression to serious condition prior to primary endpoint
  - Development of contraindication to treatment (e.g., pregnancy)
  - ~~Patients with early evidence of cure~~

80

Scientific Safety Estimand #3

- What is impact among patients who start protocol?
  - No matter what: A relevant basic science question
  - Highly relevant to clinical science
    - But does presume no change in behavior after knowing efficacy
  - No need to estimate outcomes among noncompliant patients
- All patients' data is relevant
  - Hence need to collect safety data (in an unbiased fashion) following stopping therapy

81

Scientific Effectiveness Estimand

- What is impact among patients who would knowingly start treatment?
  - Ideally considers benefit / cost tradeoffs through a therapeutic index
  - No matter what: A relevant basic science question
  - Highly relevant to clinical and public health science
    - But does presume no change in behavior after knowing efficacy
  - No need to estimate outcomes among noncompliant patients
- All patients' data is relevant
  - Hence need to collect all data (in an unbiased fashion) following stopping therapy

82

Estimands

RCT Goals

Where am I going?

We now consider the processes by which we accurately and precisely we might be able to estimate the scientific estimands of interest with a rigorous RCT.

83

Added Issues in RCT

- RCT are meant to allow the causal effect of the treatment
  - We truly might be interested in within patient effects
  - But these are never truly measurable in the same place, time
  - We thus consider differences between populations who, through randomization, are otherwise comparable
- As we try to quantify the "Scientific Estimands" we face the problem that missing data might not be on comparable subjects
  - We generally do not randomize patients to missingness
- Whenever possible we want an analysis based on randomization

84

MCAR in RCT

- Missing completely at random (MCAR)
  - The indicator of missingness does not depend upon any measured data
  - If MCAR, then ignorable
    - Precision might be gained by special analysis, however
- Possible mechanisms
  - By design
    - Measurements made on random subset of subjects
  - By accident
    - Clerical data loss
    - Meteors killing subjects
- MCAR should be rare by accident
  - Can prove missingness is not MCAR, but can not prove MCAR

85

MAR in RCT

- Missing at random (MAR)
  - Within groups defined by some observed data, the data is missing completely at random
  - MAR based on pre-randomization variables might be ignorable
- Possible mechanisms
  - Administrative censoring in longitudinal and time to event data
    - Missingness depends solely on date of accrual
    - No time trends in patient characteristics
  - Selected subsampling (e.g., case-cohort studies)
  - Withdrawal of consent or loss to follow-up?
    - Adverse effects, efficacy or lack of efficacy,etc.
    - Possibly differential across arms in incidence and reasons
- Can not use your data to differentiate MAR from MNAR

86

MAR Motivating Example: KM

- Administrative censoring in time to event analysis
  - Subjects accrued to study and followed until time of analysis
  - (Presume no time trends in study accrual)
- Subjects with missing data on time of event
  - "Redistribute to the right"
  - We can borrow information from other subjects in the risk set at time of censoring
  - Under noninformative censoring, a censored subject is equally likely to behave like any of the subjects who were still at risk at not censored at that time

87

KM: Imputed Data

- KM estimate is in some sense "imputing" the missing data
- We "impute" a censored observation by substituting any of the survival times from others still at risk at the censoring time
  - Each person at risk is equally likely to be used in the imputation
  - We can thus simulate repeated RCT, substituting a randomly selected individual from the risk set for the censored individual
  - We then average the results of the simulated RCTs
- Note that in the case of KM, we can use a formula to perform the multiple imputation

88

MNAR in RCT

- Missing not at random (MNAR)
  - Even after conditioning on all observed data, the subjects' missing data would have outcomes distributed differently than those for subjects with observed data
- Possible mechanisms (there are zillions)
  - A sudden change in health status
    - is not reflected in any of the scheduled clinic visits / measurements
    - causes a patient to be lost to follow-up or withdraw consent
  - Protopathic signs cause study withdrawal
    - Adverse events are associated with impending events
  - Depending on the estimand, e.g., cause specific mortality
    - Competing risks share a common frailty or tend toward mutual exclusivity

89

Possible RCT Estimand #1

- Average improvement for those initially prescribed drug
  - Corresponds to randomized "intent-to-treat" analysis
- Data on all patients is relevant up to the time of the protocol defined primary endpoint
- Unless there is a problem with measurement safety, there should be no missing data from the definition of the estimand

90

Possible RCT Estimand #2

- Average improvement for tolerators / compliers
  - An efficacy outcome
  - Safety would need to be assessed in another way
- This could be assessed in an RCT using randomized withdrawal or an experimental treatment run-in followed by washout
- Such would eliminate subjects who
  - cannot tolerate due to AEs
  - cannot tolerate due to perception of lack of efficacy
  - are poor compliers

91

Possible RCT Estimand #3

- Average improvement if everyone tolerated
  - This is not directly observable in all cases
  - Requires some sort of modeling of subjects stopping study treatment
    - Models based on MAR, MNAR – unlikely to be MCAR
- This could be partially assessed in a RCT with extraordinary incentive
  - Perhaps would handle mild toxicity and mild lack of efficacy
  - Could not be addressed for all cases of stopping study drug
    - Need to avoid coercive incentives

92

Possible RCT Estimand #4

- Average AUC improvement during adherence
  - Measure efficacy outcome only while adherent
  - Integrate area under the curve
  - Does not require efficacy data following stopping treatment
- Incorporates adherence as the timeframe of interest, with both longer adherence and better efficacy reflected in the magnitude of the effect
  - Depending on similarity of efficacy and safety outcomes, might in some sense equate two treatments
    - one having low dropout, with mild efficacy benefit
    - one having high dropout, but high efficacy benefit
- This can be addressed in a RCT, providing comfortable with the composite adherence-efficacy endpoint

93

Possible RCT Estimand #5

- Average improvement during adherence
  - Incorporates adherence as the timeframe of interest, but length of adherence is averaged out
  - No need for efficacy data after stopping treatment
- This approach would equate two treatments in which
  - one has high efficacy during a short phase of tolerability
  - other has high efficacy during a long period of tolerability
- This can be addressed in an RCT if comfortable with the timeframe of measurement

94

Assessing Effectiveness

- For all but the first estimand, safety must be assessed separately
- Need to consider safety in the general population, including non-tolerators
  - Short- and long-term AEs from short term exposure
  - Harm from delay of starting efficacious treatment

95

Study Design Issues

Minimizing Missing Data

Where am I going?

There are several design issues that can be used to facilitate the quantification of the various estimands while maintaining analyses based on randomization

96

Possible Methods

- Study design
  - Randomized withdrawal for long term effects
- Eligibility
  - Run-ins / enrichment for tolerators, compliers
  - Patients having “unmet need”
- Treatment
  - Add-on therapies
  - Flexible dose, titration
  - Rescue therapies
- Outcomes
  - Appropriate choice of estimand
  - Reduce follow-up times

97

Run-In Periods/ Enrichment

- Randomization only after successful completion of a trial “run-in” period
- Placebo run-in
  - Minimize patients with poor compliance behaviors
  - Ensure ability to make measurements
- Experimental treatment run-in
  - Ensure tolerability
  - Enrich “response”
    - “Clearly identifying the target population”? (In general I think not)
- Appropriate for efficacy studies: May not fully generalize

98

Flexible Dose (Titration) Studies

- Indications based on a treatment strategy, rather than a narrowly defined dose, frequency, duration
  - Incorporate dose reductions, drug holidays, etc.
  - (An aside: Avoid temptation to attribute toxicities always to other treatments)
- Often more closely mimics clinical practice
- Regulatory issues
  - Eventual product labeling needs to reflect the conditions used in testing
    - A problem that has been solved: insulin, chemotherapy, asthma

99

Treatments Added to SOC

- Avoids switching patients from effective therapies
- When added to standard of care, may decrease dropout due to lack of efficacy
  - But: not possible with noninferiority studies or when mere equivalence to existing therapies is anticipated
- Still need to avoid “cross-in” to experimental therapy

100

Treatments Added to SOC: My View

- Most experimental therapies do not pan out
- In cancer especially, I am frequently told that there is an ethical imperative to allow cross-in of the placebo group to the experimental therapy
- On average, the trials I have seen do not support this view
- In any case, decisions to cross-in should be based on clinical, not subclinical endpoints

101

Reduce Follow-up Period

- When clinical relevance is maintained, shorter follow-up periods will often reduce problems with
  - adverse events and
  - lack of efficacy
- Shorter duration is quite often adequate to assess safety in subjects who will only take the drug on a trial basis
- Additional studies will be necessary, however, to study
  - long term effects on efficacy
  - long term effects on safety
- These additional studies might consider randomized withdrawal for longer time follow-up

102

Estimands

Examples of Choice of Estimands

Where am I going?

We can examine a few examples to illustrate the sorts of considerations that might go into choosing to establish the effectiveness of a particular therapy.

103

Example: Everolimus in NET

- Neuroendocrine tumors
  - Pancreatic neuroendocrine tumors
  - Carcinoid
- Trial design
  - Primary endpoint: PFS by central radiology
  - Randomized, double blind, placebo controlled
  - Treatment: Randomized intervention until investigator determined progression
- Placebo group crosses over to open-label everolimus
  - How to analyze PFS when discordant views on progression?
  - How to analyze OS in presence of cross-over?

104

Everolimus: Relevant Estimands

- Setting: Blinded, add-on therapy
  - Chemotherapies highly toxic ➔ safety a major concern
    - Hoping to show equivalence ➔ infinite sample size needed
  - Time is of the essence
    - Starting one therapy is generally precluding another therapy
  - Usual clinical course
    - Intolerable adverse events lead to change of treatment
    - Clinical progression leads to change in treatment
      - I am not a fan of subclinical PFS, but in any case change therapy after clinical progression
      - Experimental treatment is unproven, so "cross-in" unneeded
    - Any impact of treatment on ancillary care is thus "standard"

105

Everolimus: Relevant Estimands

- Missing data
  - If we want OS on randomized therapy, possibly have MNAR due to any association between PFS (determined by investigator) and OS
    - But, maybe association is quite weak: Treating symptom, not disease
  - If we merely want PFS on randomized therapy, possibly have MNAR due to any association between PFS by investigator and PFS by central review
    - Likely an association, but extreme variability
  - If we want effectiveness of OS, need to recognize the "cross-in" can attenuate the effect either way
    - Need to avoid "cross-in"
    - cf: Laromustine (VNP40101M, Cloretazine, Onrigin)

106

Example: Second Line Therapy NSCLC

- Pemetrexed as second line in NSCLC
- Noninferiority trial compared to docetaxel
- Patients who progress on pemetrexed may cross-over to docetaxel
  - Ethics: They have not yet been tried on approved therapy
- How should we analyze this data for OS?

107

Pemetrexed: Relevant Estimands

- Setting
  - Noninferiority trial on efficacy
    - Presumably hoping for unspecified advantages on safety
    - But: cross-over to control might make arms comparable
  - Time is of the essence
    - Starting one therapy is generally precluding another therapy
    - But: comparator is an approved therapy ➔ cross-in option
- Missing data
  - If estimand is outcome on randomized therapy, then data likely MNAR after cross-over to control
    - Possible attenuation, but highly unlikely that progression is good
    - Best belief: time is of the essence in cancer treatment
  - At review scrutinize incidence of cross-in
    - Sensitivity analyses to time-varying cross-in

108

Example: Chronic Pain

- Study design
  - Patients randomized to experimental treatment or placebo
  - Patients often recruited after being on some therapy chronically
- High rates of dropout
  - Potential toxicities to new therapy
  - Potential lack of efficacy to placebo
- Actions on progression
  - Return to prior therapy
  - Use of more potent analgesia (e.g., morphine)

109

Chronic Pain: Relevant Estimands

- Setting
  - Chronic use of pain medications
    - Time not of essence in starting therapy
    - Multiple efficacious treatments of clinical value
  - Patients often characterize reasons for dropout differentially by treatment arm
  - Rescue therapies often known to be effective
- Missing data
  - Dropout for lack of efficacy: Likely MNAR
  - Dropout for adverse events:
    - Perhaps MAR if very different receptors for efficacy, toxicity
    - Perhaps MNAR if common pathways

110

Chronic Pain: Relevant Estimands

- My choice of design / estimand
  - RCT in all eligible
    - Safety among all starting therapy
    - Tolerability measuring compliance
    - Efficacy among compliers
  - RCT using randomized withdrawal in tolerators
    - Longer term safety, efficacy

111

Example: Antifibrinolytics in ChemoTX

- Patients undergoing chemotherapy for cancer often experience increased risk of bleeding due to low platelets
- Hypothesize that platelets are being used up due to repeated dissolving of clots
- Consider prophylaxis with antifibrinolytics to decrease rates of serious bleeding in first 30 days of chemotherapy
- Issues
  - Some patients will die of their underlying disease
  - How do we record bleeding incidence in such patients?

112

Possible Estimands: WRONG

- Effect of treatment among patients who survive
  - Eliminate any patient who dies within 30 days from analysis
  - Inflate sample size by 11% to account for anticipated 10% deaths

$$N_{analyze} = 0.9 \times N_{accrue} \quad \Rightarrow \quad N_{accrue} = \frac{N_{analyze}}{0.9}$$

- This conditions on a post-randomization variable
  - We are not assured of comparability of treatment groups if treatment affects death
  - Sample size inflation merely increased precision of a potentially biased observation

113

Possible Estimands: MAR

- Effect of treatment if all would survive
  - Assume missing at random
  - Censor subjects who die
  - Imputation assumes that with improved treatment of underlying disease, their clinical course re bleeding would be the same as any of the patients in study that lived past the time of death
  - Inflate sample size to account for censoring using event driven analyses (complicated, but manageable for KM)
    - If all survived N= 465 ➔ With competing risk, N=486
- Caveats
  - Possibly reasonable given extensive experience of treatment in somewhat related indications
  - Will need to plan for sensitivity analyses of MAR assumption
    - (more later)

114

Possible Estimands: Composite

- Effect of treatment on disease free survival
  - Include death in the definition of the primary endpoint
  - Inflate sample size by attenuating treatment effect to account for the belief that treatment does not affect mortality
    - Using MAR model, believe reduction of bleeding 57% to 42.8%
      - 486 subjects
    - With competing risk, reduce incidence of event 59.2% to 45.6%
      - 506 subjects (4% increase from MAR, 9% from survivors only)
- Caveats
  - May equate death with relatively minor, treatable bleeding
  - Efforts to score deaths as “most severe” may shift estimand more toward overall survival
  - Still need to evaluate overall survival, as treatment could increase rate of death after initial bleeding

115

Prevention of Missing Data

Recommendations

Where am I going?

There have been many clinical trialists who are successful at avoiding missing data.

Emulating their techniques is highly recommended.

116

Recommendation # 2

- Investigators, sponsors and regulators should design RCT with the goal of maximizing the number of participants who are maintained on the protocol-specified intervention until the outcome data are collected.

117

My View

- Some of the report seems to suggest that shorter time periods for observation should be chosen to avoid missing data
- This should only be done when scientific relevance is maintained
  - And too many investigators only consider the timeframe for their perceived mechanism of action on a surrogate endpoint
  - Safety and longterm effects are very important

118

Recommendation # 3

- Trial sponsors should
  - continue to collect information on key outcomes on participants who discontinue their protocol-defined intervention in the course of the study, and
  - record and use this information in the analysis.
- Exceptions to this should only be made when a compelling cost-benefit analysis argues otherwise.

119

My View

- Many site investigators have to be educated on the relevance of such data
  - I believe they should be tested on their comprehension of
    - clinical relevance
    - regulatory setting
- "Compelling cost/benefit" is most appropriately measured in patient safety
  - E.g., liver biopsy might not be wise if there are no continuing concerns about liver damage from the treatment

120

Recommendation # 4

- The trial design team should consider whether participants who discontinue the protocol intervention should have access to and be encouraged to use specific alternative treatments.
- Any such “follow-on” treatments should be specified in the study protocol.

121

My View

- These follow-on therapies can still cause problems with some estimands
  - Efficacy vs effectiveness
  - Rescue therapies consistent with eventual use vs artificial setting of RCT
- But we should maintain as close to effectiveness studies as practical
  - Proscriptive therapies not in keeping with standard of care should usually be avoided

122

Recommendation # 5

- Data collection and information about all relevant treatments and key covariates should be recorded for all initial study participants, whether or not participants received the intervention specified in the protocol.

123

Recommendation # 6

- Study sponsors should explicitly anticipate potential problems of missing data.
- In particular, the protocol should contain a section that addresses missing data issues, including
  - the anticipated amount of missing data,
  - the steps taken in trial design to limit the impact of missing data, and
  - the steps taken in trial conduct to monitor the incidence of missing data.

124

Preliminaries

- Identification of burden of proof
  - Safety
  - Efficacy
  - Effectiveness
- Selection of primary endpoint
  - Timeframe (event time or calendar time)
  - Composite endpoints (when safety issues critical for combination vs handled separately)
- Potential mechanisms of missing data
  - Issues with estimand (hopefully minimized)
  - Withdrawal of consent (what scientific factors lead to it)
  - Loss to follow up (what scientific factors lead to it)
  - Sloppy data

125

Design Time Issues

- Protocol definitions
  - Specific aims (estimand)
  - Precise definition of primary, secondary endpoints
  - Primary analysis by randomization (ITT)
  - Planning for missing data
    - Potential scientific mechanisms
    - Prevention strategies (protocol wording, investigator training, subject education, informed consent)
    - Resulting statistical classification: MCAR, MAR, MNAR
    - When will study be stopped
    - How it will be handled analytically

126

Planned Analyses

- Descriptive statistics to describe missing data patterns
- Results that would be compatible with presumed mechanisms
- Description of models to be used for sensitivity analyses
  - MAR to MNAR
  - Inclusion of covariates
  - Modeling of covariates
- Primary analyses
  - Available measurements that will be used
  - How they will be modeled
  - The statistical model (MMRM, MI, pattern mixture ?but never single imputation?)
  - Standards for inference (frequentist, Bayesian)

127

Conduct and Analysis

- Conduct time issues
  - Monitoring of success
- Analysis time issues
  - ?blinding of analysts
- Interpretation time issues
  - Were missing data patterns consistent with hypotheses
  - Robustness of sensitivity analyses

128

History of Prior Success

- Successful conduct of RCT with minimal missing data has been documented in the literature
  - DCCT
  - HIVNET 012
- Such success does not happen by accident, however
- There are many RCT implementation strategies that will help minimize the problems

129

Strategies

- Minimize patient burden
  - Minimize number of visits, and make them pleasant experiences
  - Collect only the necessary information
  - Use user-friendly CRFs
  - Use direct data capture
  - Use relatively large time window for ascertainment
- Provide incentives for continued participation
  - Access to health care for participants
  - Adequate reimbursement for investigators
- Use experienced investigators and provide good training
  - Particularly important to educate on need for continued data collection

130

Strategies

- Informed consent
  - Tell participant either patient or physician may decide to modify or stop the investigational treatment
  - Inform participant of scientific importance of continued data collection
  - Choices for stopping study drug
    - Stop drug, continue visits
    - Stop drug and visits, continue medical records surveillance
    - Withdraw consent completely
- Close monitoring of data completion and data quality
  - Terminate sites with poor performance
  - DSMB might terminate study if overall poor performance

131

Recommendation # 7

- Informed consent documents should
  - emphasize the importance of collecting outcome data from individuals who choose to discontinue treatment during the study, and
  - encourage participants to provide this information whether or not they completed the anticipated course of study treatment.

132

Recommendation # 8

- All trial protocols should recognize the importance of minimizing the amount of missing data.
- All trial protocols should set a minimum rate of completeness for the primary outcome(s) based on what has been achievable in similar past trials.

133

The Prevention and Treatment of Missing Data in Clinical Trials

Society for Clinical Trials
May 20, 2012

Jim Neaton

Chapter 3: Trial Strategies to Reduce the Frequency of Missing Data

National Research Council (2010). Panel on Handling Missing Data in Clinical Trials. National Academies Press.

Panel Report

- 18 recommendations
  - 8 on prevention through design and conduct
  - 7 on methods, including sensitivity analysis
  - 3 on data sharing, training and future research

Outline of Presentation

- Some general considerations
- Actions for design and data management teams
- Actions for investigators and site personnel
- Targets for acceptable rates of missing data
- Reporting missing data
- Panel recommendations

It Is Easier to Collect Event-Driven as Compared to Visit-Driven Outcomes

- **Event-driven:**
  - Survival
  - Hospitalization
  - Adverse events
- **Visit-driven data:**
  - Blood pressure
  - Blood test
  - Procedures:
    - Bone-mineral density (radiology center)
    - Exercise stress test

Loss-to-Follow-up Rates in Selected Trials with Mortality Outcome

	Total Enrolled	Unknown Vital Status (%)
MRFIT	12,866	0.2
HDFP	10,940	0.6
AMIS	4,524	0.2
ELITE II	3,152	0.1
EPHESUS	6,642	0.3

Range of follow-up: 1.5 to 7 years.

Loss-to-Follow-up Rates in Selected Trials with Morbidity/Mortality Outcome

	Total Enrolled	Unknown Primary Endpoint (%)
Phidisa II	1,771	4.9
ESPRIT	4,111	6.1
HEAAL	3,846	2.5
POET-COPD	7,384	5.7

Range of follow-up: 1 to 6 years.

Loss-to-Follow-up Rates in Selected Trials with Visit-Driven Outcomes

	Total Enrolled	Unknown Primary Endpoint (%)
CATT	1,185	6.8
HPTN 052	1,771	9.9
TOMHS	902	10.9 (0.5)
Turner's Synd.	149	38.9 (8.1)

Range of follow-up: 1 to 7.2 years.

Acronyms and References

- MRFIT: Multiple Risk Factor Intervention Trial (JAMA 1982)
- HDFP: Hypertension Detection and Follow-up Program (JAMA 1979)
- AMIS: Aspirin Myocardial Infarction Study (JAMA 1980)
- ELITE II: Evaluation of Losartan on Mortality (Lancet 2000)
- EPHESUS: Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study (N Engl J Med 2003)
- Phdisa II: HIV Treatment Trial (J Infect Dis 2010)
- ESPRIT: Evaluation of Subcutaneous Proleukin in a Randomized International Trial (N Engl J Med 2009)
- HEAAL: Heart Failure Endpoint Evaluation of Angiotensin II Antagonist Losartan (Lancet 2009)
- POET-COPD: Prevention of Exacerbations with Tiotropium in COPD (N Engl J Med 2011)
- CATT: Comparison of Age-Related Macular Degeneration Treatments Trials (N Engl J Med 2011)
- HPTN 052: HIV Prevention Trials Network (N Engl J Med 2011)
- TOMHS: Treatment of Mild Hypertension Study (JAMA 1993)
- Turner's Syndrome: (N Engl J Med 2011)

Recommendations for Visit Driven Outcomes

- Use wide, abutting follow-up visit windows.
- Make contingency plans to collect data for those not attending data collection visits.
- Make plans to provide treatment if data collection visit is missed.

Minimize missing data and maximize adherence by planning ahead.

Most Staff Responsible for Data Collection Want To Do It Correctly

- **Recommendations:**
  - –Prior to study implementation as part of training:
    - Make sure they understand why the proposed data collection plan is important – they need to communicate this clearly to patients.
    - Give them tools to facilitate self-monitoring.
  - –During study implementation:
    - Give them regular feedback on performance, including comparative performance statistics.
    - Have a plan for performance improvement.

"Analysis" and "Treatment" Dropouts Are Not the Same

- While "analysis" dropout usually implies "treatment" dropout, reasons for these types of dropout vary
- Reasons for "analysis" dropouts:
  - – Withdrawal of consent by patient or legally authorized representative
  - – Patient moves away or cannot be contacted

There are Many Misconceptions About Withdrawal of Consent

- It is <u>not</u> the investigator's decision; it is the patient's (at least, these should be differentiated)
- Treatment discontinuation ≠ withdrawal of consent
- Unwillingness to attend follow-up visits ≠ withdrawal of consent

Withdrawal of Consent

- **Elements of the consent form: 45 CFR 46.116 (b)(4):**

"The consequences of a subject's decision to withdraw from the research and orderly termination of participation by the subject."

Sample Language: HIV Trial

WHAT IF YOU DON'T WANT TO BE IN THE STUDY ANY LONGER?

If you enroll in this study, you may decide to stop participating at any time. Withdrawing from this study will not affect the benefits of your regular medical care. However, if you are receiving HIV medicines provided by the study, you will not continue to be given HIV medicines through the study after you withdraw. Your doctor or nurse will help you find another way to get HIV medicines.

Withdrawal Language Should Be Balanced by Language Like This

If you or your doctor decide it is best not to take the study drugs, other treatment options will be discussed with you. You will continue to be scheduled for follow-up visits every 4 months until the study ends. Your continued participation is very important in order to reliably answer the study question.

FDA Guidance on Data Retention When Subjects Withdraw from FDA-Regulated Trials

- When a subject withdraws, data collected up to time of withdrawal cannot be removed from database
- Investigator may ask subject if they wish to provide additional data collection following discontinuation of intervention
- Additional data collection following discontinuation of intervention requires consent
- Following withdrawal, medical and other confidential records cannot be used but public records (e.g., survival status) may be used

Good Clinical Practice Program and Office of Chief Counsel, October 2008.

Office for Human Research Protections (OHRP) Interpretations of Guidance Do Not Conflict with FDA Guidance - 1

"OHRP recommends that when a subject decides to withdraw from a clinical trial, the investigator conducting the clinical trial ask the subject to clarify whether the subject wishes to withdraw from all components of the trial or only from the primary interventional component."

Guidance on Withdrawal of Subjects from: Data Retention and Other Related Issues. September 2010.

Office for Human Research Protections (OHRP) Interpretations of Guidance Do Not Conflict with FDA Guidance - 2

"OHRP recommends that investigators and IRBs consider whether and how the withdrawal of a subject from a research study should be documented".

- is it a decision by subject or investigator?
- all components of research or just intervention?

Withdrawal of Consent in Two Large HIV Trials

	SMART (33 Countries)	ESPRIT (25 Countries)
USA	65 / 2989 (2.2%)	16 / 949 (1.7%)
Germany	5 / 215 (2.3%)	6 / 266 (2.3%)
Canada	4 / 102 (3.9%)	5 / 141 (3.5%)
All Others	14 / 2166 (0.6%)	16 / 2755 (0.6%)
Total	88 / 5472 (1.6%)	43 / 4111 (1.1%)
Median follow-up (months)	36	84

"Analysis" Dropout Due to Moving Away

- **This can be minimized**
  - Collect contact information at entry and regularly update
  - Establish procedures to transfer care to another site
  - In some cases, it may be possible to use national laboratories for blood measurements.

Sample Consent Language: HIV Trial

WHAT IF YOU MOVE?

If you move or transfer your medical care to another doctor, the study staff would like to continue to collect information about your health. If you give permission, your study doctor or nurse will contact your new doctor…When you move, you will be asked to sign a "Release of Medical Information" form

Important Predictors of Missing Data Are Not Patient Related

- Research has usually focused on patient characteristics:
  - Younger age
  - Smokers
  - Socioeconomic status – employment, stable housing, education
- Helpful to a degree, but these are not the major risk factors

More Important Factors

- Clinical research site (whether source of primary care, stability, experience, number of patients enrolled, staff turnover)
- Trial design (e.g., event- versus visit-driven data)
- Trial conduct conditions
- Quality assurance procedures

Etiology of Missing Data

Trial protocol

↓

Selection and training of investigators

↓

Informed consent

↓

Trial conduct conditions

↓

Monitoring against realistic, but high, performance goals

Etiology Suggests the Following for Primary Prevention

- Write protocol with minimization of losses in mind (do not overburden patients and staff).
  - Avoid complicated and cumbersome record keeping.
  - Make it easy to obtain prescriptions.
  - Choose easily ascertainable endpoints.
- Select sites in a convenient location for patients with demonstrated record of excellent follow-up.
- Train study staff on the importance of excellent follow-up (minimizing missing data).
- Fully inform patients of trial requirements and importance of full participation during consent process.

Primary Prevention (cont.)

- Collect contact information at entry.
- Adopt a flexible appointment schedule.
- Remind patients about appointments and follow-up immediately after missed appointments.
- Minimize waiting time during appointments.
- Provide reports to staff to monitor follow-up completeness.
- Insist on the highest standards.

After Trouble Begins – Secondary Prevention

- Telephone contacts and home visits.
- Partial data collection (reduce demands of participation).
- Use central registries for vital status.

Key Points So Far

- It is possible to design and conduct trials, even long-term trials, with a minimal amount of missing data.
- Missing data are less likely with event-driven outcomes
- Site staff will do it right if motivated, trained and provided feedback
- With patient consent, there are no regulatory impediments to collection of data after discontinuation of intervention
- The etiology of missing data suggests several practical steps for prevention

Outline

- Some general considerations
- Actions for design and data management teams:
  - Limit participant burden
  - Increase incentives for participation
  - Select investigators with a good track record
  - Train investigators
  - Use payment schedules that reward excellent follow-up
  - Monitor data collection

Limiting Participant Burden

- Focus on essential data items
- Consider subsamples for secondary outcomes (e.g., lower grade adverse events)
- Make it easy to stay in the study
  - wide windows for data collection visits
  - evening/weekend appointments
  - home visits, if necessary, for essential data

Increase Incentives for Participation

- **Phase 3 trial of interleukin -2:**
  - Consent form: "If the research finds that IL-2 is safe and effective for HIV patients, and you and your doctor decide that you want to take it, the company supplying IL-2 during the study will provide it until it is approved for use for HIV infection. After the approval…"
  - Extension protocol written prior to study closure

Selection of Clinical Sites

- **Location, convenience for patients, stability**
- **Track record: retention as well as recruitment.**
  - treatment and analysis dropouts in past trials
  - data queries
- **PI motivation and commitment to research question**
- **Availability of trial coordinator/manager with appreciation of local QA**

Training Considerations

- Plan for initial and refresher/remedial training.
- Choose trainers who understand the goals of the study, who can discuss study design, and foster a team mentality.
- Provide rationale for study and motivate the importance of a high quality data in addition to study procedures.
- Describe how to use self-monitoring tools.

Payment Schedules that Emphasize Excellent Follow-up

- Modest up-front payment to sites for training and protocol IRB approvals.
- Quarterly payments for case-report forms completed – no follow-up, no money.
- Final payment for end of study visit at which patient status is verified

Monitoring Data Collection

- Provide appointment schedules following randomization
- Provide visit reminders (in advance of window opening and last chance before closing)
- Provide easily accessible web-based reports on follow-up summary statistics as well as for individual patients
- Use of on-site visits for training and checking for missed events
- Discussion of importance of excellent follow-up at investigator meetings; rewards to investigators for follow-up
- Assist site develop local QA procedures

Visit Reminder Report

The TRIAD Study
Visit Reminder Report
Patients with Visit Windows Opening Between
1 Apr 2012 and 30 Apr 2012

Site 622

Participant ID	Rand. Date	Study	Month	Target Date	Visit Window
32-0109-7	03JUN2010	TRIAD	24	03JUN2012	04APR12-03AUG12
32-0110-5	02JUN2010	TRIAD	24	02JUN2012	03APR12-02AUG12
		Substudy	24	02JUN2012	03APR12-02AUG12
32-0111-3	29JUN2010	TRIAD	24	29JUN2012	30APR12-29AUG12
		Substudy	24	29JUN2012	30APR12-29AUG12
32-0112-1	29JUN2010	TRIAD	24	29JUN2012	30APR12-29AUG12

Last Chance Visit Reminder Report

The TRIAD Study
Visit Closing Report
Participants with Visit Windows Closing Between
1 Apr 2012 and 30 Apr 2012

Site 622

Participant ID	Rand. Date	Study	Month	Target Date	Visit Window
32-0110-5	02JUN2010	TRIAD	20	02FEB2012	03DEC11-02APR12
32-0111-3	29JUN2010	TRIAD	20	29FEB2012	30DEC11-29APR12

Regional Data Completeness Report

The TRIAD Study
Completeness of Visit Forms, HIV RNA, and CD4 Data
Northeast Region

	Visit CRF			CD4+			HIV RNA		
Visit	Exp.*	No.	Pct.	Exp.*	No.	Pct.	Exp.*	No.	Pct.
Baseline**	615	609	99.0	1230	1229	99.9	615	609	99.0
Month 1	578	560	96.9	578	558	96.5	578	558	96.5
Month 4	493	487	98.8	493	486	98.6	493	485	98.4
Month 8	424	417	98.3	424	414	97.6	424	412	97.2
Month 12	378	366	96.8	378	363	96.0	378	360	95.2
Month 16	319	310	97.2	319	309	96.9	319	309	96.9
Month 20	208	199	95.7	208	199	95.7	208	199	95.7
Month 24	70	69	98.6	70	69	98.6	70	69	98.6
Month 28	28	27	96.4	28	27	96.4	28	27	96.4
Month 32	1	1	100.0	1	1	100.0	1	1	100.0
Overall	3114	3045	97.8	3729	3655	98.0	3114	3029	97.3

Regional Missing Outcome Data Report

The TRIAD Study
Lost to Follow-up Summary by Country – European Region

			No data in last 8 months		Withdrawn consent		Total lost to follow-up	
Region	Country	Current pts[1]	N	%	N	%	N	%
Europe	Austria	3	0	0.0	0	0.0	0	0.0
	Belgium	80	1	1.3	1	1.3	2	2.5
	Czech Republic	7	0	0.0	0	0.0	0	0.0
	Denmark	28	0	0.0	0	0.0	0	0.0
	Estonia	3	0	0.0	0	0.0	0	0.0
	Finland	18	0	0.0	0	0.0	0	0.0
	Germany	249	2	0.8	6	2.4	8	3.2
	Luxembourg	2	0	0.0	0	0.0	0	0.0
	Norway	10	0	0.0	0	0.0	0	0.0
	Poland	64	2	3.1	0	0.0	2	3.1
	Portugal	24	0	0.0	0	0.0	0	0.0
	Spain	135	1	0.7	0	0.0	1	0.7
	Sweden	1	0	0.0	0	0.0	0	0.0
Subtotal		624	6	1.0	7	1.1	13	2.1

Special Consideration for Non-Blind Studies

- Ensure outcome assessments are similar for each treatment group (e.g., progression-free survival in oncology trials)
- Special considerations when contact schedule differs by protocol (e.g., postcards in MRFIT)

Outline

- Some general considerations
- Actions for design and data management teams
- Actions for investigators and site personnel
  - Emphasize participation for full duration of trial
  - Minimizing the burden on participants
  - Collect information on participants at risk for dropout
  - Educate participants on continued engagement
  - Study-branded gifts.
  - Make study participation enjoyable
  - Regularly update contact information

A Suggested Withdrawal of Consent Form (Patient Version)

- **I have decided I no longer want to participate in this clinical study the way it was planned...**
  - I am/I am not willing to take study medication
  - I am/I am not willing to attend study visits
  - I am/I am not willing to let you contact me by telephone or letter
  - I am/I am not willing to let you contact my family doctor to check on my progress
  - I am/I am not willing to let you use information from my medical record to check on my progress

Cleland JG et al, Eur J Heart Failure 2004
(see also the example in Report for DART trial)

Document Reasons for "Analysis" and "Treatment" Dropout

- Written statement for withdrawal
- Case report form documenting "partial" withdrawal
- Reasons for "treatment" dropout (study medication discontinuation)
  - adverse event (type of event and severity)
  - lack of efficacy
  - concomitant medication/contraindication
  - patient or physician directed

Guidance to Site Staff on Withdrawal

- Document reasons for withdrawal in medical chart and on CRF
- Discuss/negotiate partial data collection with participant
- Advise participant that they always can re-consent
- Don't give up! Discourage use of terms like "off-study"

Outline

- Some general considerations
- Actions for design and data management teams
- Actions for investigators and site personnel
- Targets for acceptable rates of missing data
- Reporting missing data
- Panel recommendations

Target for Acceptable Rates of Missing Data

- Consider trial objectives and state missing data targets in protocol
- Consider possibility of missing data in sample size estimation
- Set performance goals considering:
  - Results from similar trials
  - Sensitivity analyses

Taken from an HIV Treatment Protocol

"The primary analysis will be intention-to-treat".

Implication: Since focus is on treatment policy and ITT estimand, all patients must be followed for the primary endpoint (e.g., AIDS or death) until the end of the study.

Note: It may be very reasonable to define your estimand as a "modified intent to treat estimand" in which patients not meeting entry criteria are excluded (e.g., participants who are found to be HIV- at the time of randomization). The implications for patient follow-up are the same.

In Other Cases the Term ITT Is Not Appropriately Used

"We used a five-point scale…to grade adverse events occurring while the patient was taking study drugs and during the eight weeks after their permanent discontinuation".

"All analyses were performed according to intention to treat".

N Eng J Med 1996;335:1099-1106.

Another Example

- **"Efficacy analyses were performed for the intent-to-treat population defined as all randomized patients who received at least one dose of study drug and provided at least one post-randomization efficacy evaluation. For patients who withdrew early, the last available pain evaluation was carried forward…The primary efficacy analysis was pain reduction…from the final visit to baseline."**
- **Approximately 50% withdrawal!**

Acta Neurol Scand 2004; 110:221-231.

When is Missing Data a Problem?

- Anything but zero is bad
- If number of losses exceeds number of events, results are questionable
- < 5% is okay, but if > 20%, do not believe the results
- If > 0% and differential by group, question the results
- If > 0%, and different assumptions concerning losses yield different trial results, e.g., P<0.05 to P> 0.05.

Sackett DL, Evidence based medicine: how to practice and teach EBM, 1997.
Matts J et al. Stat Med 1997.
Schultz KF, Grimes D. Lancet 2002.

A Perspective from 35 Years Ago

- "Rigorous entry criteria are not necessary for a randomised trial, but rigorous follow-up is."
- "One excellent policy is to accept no withdrawals under any circumstances."
- "Patients who move away from the centres where they were admitted to the trial should not be allowed to disappear from the trial."
- "…our policy is to accept no reason for loss except emigration…"

Peto R, Pike MC, Armitage P, et al. Br J Cancer 1976.

A Perspective from 1.5 Years Ago

- "A preferred approach to addressing missing data is to prevent it."
- "Procedures should be in place to maximize the likelihood that outcome data will be obtained at scheduled times of evaluation for all surviving patients who have not withdrawn consent."

Fleming TR. Ann Intern Med 2011.

Sample Size: HIV Protocol

From sample size justification:

"Two percent of patients will be lost to follow-up each year. It is recognized that if the loss rate is as high as 2% per year, then estimates of treatment differences…could be severely biased…Nevertheless, this conservative adjustment to sample size was made in order to increase power because some losses are inevitable."

Increase in Sample Size to Account for Loss of Power

$$N_{NEW} = \frac{N_{OLD}}{1 - L}$$

L = fraction of patients expected to be missing outcome data

This adjustment may take care of the loss of power but not bias resulting from the missing data.

Monitoring Guidelines: HIV Protocol

"The trial may be terminated or modified…if 1 year lost-to-follow-up is > 2.5%, or projected overall 3-year lost-to-follow-up is > 10% or the absolute difference between treatment groups is more than 7.5%".

Defining Lost-to-Follow-up (Missing Outcome Data) in Event-Driven Trial

- A trial participant for whom the outcome of interest is not known – MISSING DATA
  - at the time routine reports are prepared for sites and protocol team
  - at the time of interim analyses for Data and Safety Monitoring Committee
  - for final report

Operational Definition of Missing Outcome Data

- During the trial – more than 8 months with no data (no case-report forms)
- At end of trial -- event status unknown at closing date (usually a calendar date chosen to ensure the target number of events has been achieved)

Note: Many interim losses are eventually found so sites should be encouraged to continue looking.

Outline

- Some general considerations
- Actions for design and data management teams
- Actions for investigators and site personnel
- Targets for acceptable rates of missing data
- Reporting missing data
- Panel recommendations

Reporting Missing Data in Final Trial Reports

- Quality of reporting
- Suggested text and CONSORT diagrams
  - Event-driven trials
  - Visit-driven trials
- Labeling of figures

Quality of Reporting Missing Data

- 477 parallel group trials published in 2006
- 32% included a flow diagram
- 74% reported losses (missing outcome data) for each group
- 85% reported reasons for loss

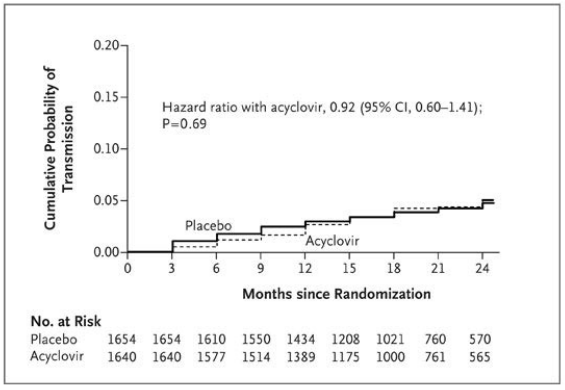
Hopewell S, et al. BMJ 2010

Reporting of Survival Data in Cancer Trials

- 125 cancer trials published in 2004
- Completeness of reporting:
  - starting point (randomization/initiation of treatment) – 78%
  - censoring – 58%
  - patients at risk – 53%
  - extent of follow-up – 57%

J Clin Oncol 2008

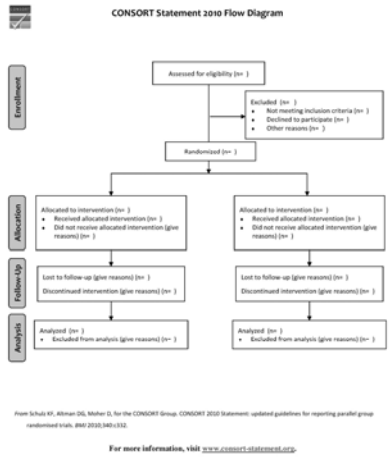
HIV Prevention Trial: Illustrative Survival Curve



N Engl J Med 2010

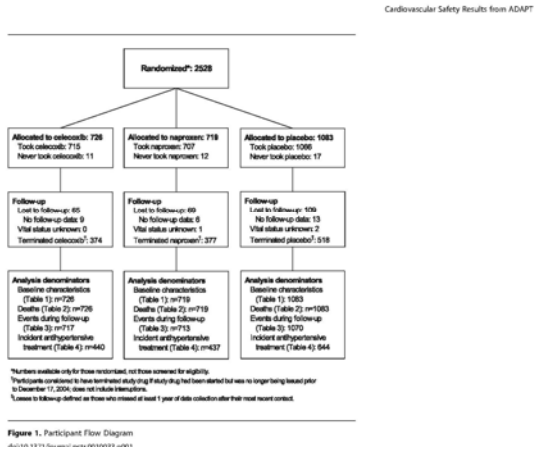
CONSORT 2010 Checklist: Follow-up

- For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome.
- For each group, losses and exclusions after randomization, together with reasons.



Flow Diagram for ADAPT Study

(PLoS Clinical Trials Nov. 2006)



Text for Morbidity/Mortality Trial: Delta

"284 participants (95 AZT, 102 AZT plus ddI, 87 AZT plus ddC had not had a full clinical assessment for at least 4 months on September 25, 1995, but the vital status of 77 (25 AZT, 28 AZT plus ddI, 24 AZT plus ddC) of these was known."

Delta Coordinating Committee, Lancet 1996.

Sample Text: Visit-Driven Primary Endpoint with Event-Driven Secondary Endpoint (TOMHS)

"Median follow-up was 4.4 years for analyses of time to death or non-fatal CVD events. Between March 1 and May 31 1992, the vital status of all but four participants was confirmed.

Attendance at follow-up visits was high, averaging 90.6% and ranging from 88.9 (enalapril group) to 93.7 (acetbutolol group) among the six treatment groups. For the 3-, 12-, 24-, 36- and 48-month follow-up visits, at which echocardiographic and ECG measurements were made, attendance averaged 92.8%. Only five participants (0.5%) never returned for a follow-up visit."

Outline

- Some general considerations
- Actions for design and data management teams
- Actions for investigators and site personnel
- Targets for acceptable rates of missing data
- Reporting missing data
- Panel recommendations

Panel Recommendation 6

Study sponsors should explicitly anticipate potential problems of missing data. In particular, the trial protocol should contain a section that addresses missing data issues, including the anticipated amount of missing data, and steps taken in trial design and trial conduct to monitor and limit the impact of missing data.

Panel Recommendation 7

Informed consent documents should emphasize the importance of collecting outcome data from individuals who choose to discontinue treatment during the study, and they should encourage participants to provide this information whether or not they complete the anticipated course of study treatment.

Panel Recommendation 8

All trial protocols should recognize the importance of minimizing the amount of missing data, and, in particular, they should set a minimum rate of completeness for the primary outcome, based on what has been achievable in similar past trials.

Summary

- Missing data can be prevented – that needs to be a major focus during the design and implementation stages of a study.
- Need to better educate investigators on the importance of complete follow-up.
- Insist on high standards.

Thank you!

Methods of Analyzing Data

From the Report

Where am I going?

Multiple methods have been described in the statistical literature, but appropriate methods have not yet seen widespread use.

Methods should be easily understood by the audience of RCT results.

137

Major Themes

- Inference from missing data is necessarily based on subjective, untestable assumptions about the distribution of missing values
- But not all such assumptions are equally reasonable
- In particular, the overly simplistic single imputation methods of last one carried forward (LOCF) and baseline carried forward (BOCF) are most often hard to justify scientifically and statistically

138

Basic Principles

- The missingness must hide a potentially useful value
- The estimand must be scientifically (clinically) relevant
- Reasons for missing data must be documented fully
- Trial designers should decide on primary assumptions about missing data mechanisms
- A statistically valid analysis under those assumptions should consider both consistency and variability of estimates
- The robustness of the conclusions to the untestable assumptions should be investigated

139

(Overly) Simplistic Methods

- Complete case analysis
  - Ignore cases with missing data
  - Only appropriately unbiased for ignorable missingness
  - Otherwise assumes some poorly characterized mechanism
  - Inflate sample size to account for missingness
    - "A more precise biased answer"
- Common single imputation methods
  - LOCF: Assume last observation is ***exactly*** equal to missing data
  - **B**OCF: Assume first observation is ***exactly*** equal to missing data
  - Difficult to justify scientifically or statistically
    - Single imputation inappropriately presumes no variability

140

Advanced Statistical Methods

- "Inverse Probability Weighting"
  - With MAR, analogous to methods used in political polling
- Modeling missing data
  - "Likelihood methods"
  - "Selection models"
  - "Pattern mixture models"
- "Multiple imputation" from prediction models
  - Borrow information from available data
    - MAR : straightforward borrowing
    - MNAR : perturb observed results
  - Sample repeatedly from prediction models to assess variability

141

Notation

- $\boldsymbol{X}$ = baseline covariates (including treatment)
- $\boldsymbol{Y}$ = outcome variables (possibly repeated measurements)
  - $\boldsymbol{Y_{obs}}$ and $\boldsymbol{Y_{mis}}$ denote observed and missing data, respectively
- $\boldsymbol{V}$ = auxiliary variables to aid in missing data analysis
  - $\boldsymbol{V_{obs}}$ and $\boldsymbol{V_{mis}}$ denote observed and missing data, respectively
- $\boldsymbol{M}$ = indicator of missingness (one for each component of $\boldsymbol{Y}$)

142

Notation

- Missing completely at random (MCAR)

$$[M \mid X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M]$$

- Missing at random (MAR)

$$[M \mid X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M \mid X, V_{obs}, Y_{obs}]$$

$$[Y_{mis}, V_{mis} \mid X, V_{obs}, Y_{obs}, M] = [Y_{mis}, V_{mis} \mid X, V_{obs}, Y_{obs}]$$

- Missing not at random (MNAR)

$$[Y_{mis}, V_{mis} \mid X, V_{obs}, Y_{obs}, M=1] \neq [Y_{mis}, V_{mis} \mid X, V_{obs}, Y_{obs}, M=0]$$

143

Methods: CC

- Complete case analysis
  - Ignore cases with missing data
- Only appropriately unbiased for ignorable missingness
  - MCAR and some MAR
  - When missingness appreciable, substantial inefficiency
- Use of complete case analysis does correspond to an assumption about the missing data mechanism
  - Though that assumption is poorly characterized

144

Methods: IPW

- Inverse probability weighting
  - Appropriate for MAR
- Estimate a model for the probability of observed response (***M = 0)*** as a function of covariates ***X*** and auxiliary variables ***V***
  - E.g., a logistic regression model
- Estimate mean ***Y*** as a weighted average of observed $\boldsymbol{Y_{obs}}$
  - Weights are inversely proportional to the probability of observed response for each corresponding ***X,V***
  - Standard errors analytically or by bootstrap

$$\hat{\mu} = \frac{1}{n}\sum_i \frac{(1-M_i)Y_i}{\hat{\Pr}(M_i = 0 \mid X_i, V_i)}$$

145

Methods: IPW

- Inverse Probability Weighting properly adjusts for bias, providing the model estimating probability of missingness is correctly specified
  - Variables and form of model
- High variability when probability of observed response is low
  - There must always be some observed Y for each auxiliary variable combination
  - The support of the missing data distribution is the same as that for the observed distribution
- Augmented IPW makes better use of incomplete cases in the presence of repeated measures
  - Doubly robust

146

Methods: IPW

- Advantages
  - Simple to implement with monotone data
  - Generally easy to understand
- Disadvantage
  - Relies on correct specification of missingness model
  - Instabilities where missingness is high

147

Methods: Likelihood Methods

- Uses a parametric model for full data distribution
  - Model for full response data and missingness

$$p(y, m \mid x; \theta, \psi) = p(y \mid x; \theta)\, p(m \mid y, x; \psi)$$

  - Integrate over all possible realizations of missing data
  - Under MAR, simplifies to involve only observed data
    - Does not depend on functional form of missingness model

$$L(\theta, \psi \mid y_{obs}, x, m) = p(m \mid y_{obs}, x, \psi)\, p(y_{obs} \mid x; \theta)$$

  - Inference under asymptotics or Bayes

148

Methods: Likelihood Methods

- Advantages
  - If missingness ignorable, then models generally easy to fit
  - Random effects models can help simplify multivariate distribution
- Disadvantages
  - Untestable parametric assumptions

149

Methods: LOCF / BOCF Imputation

- Common single imputation methods
  - LOCF: Assume last observation is ***exactly*** equal to missing data
  - **B**OCF: Assume first observation is ***exactly*** equal to missing data
- Either of these are MNAR models
- Difficult to justify scientifically
- Difficult to justify single imputation statistically

150

Methods: Multiple Imputation

- Multiple data sets are created with sampling of missing data from its predictive distributions
- Each dataset then analyzed
  - Conditional on specific imputed dataset
- Results from analyses combined in a simple way
  - Unconditional variance from
    - Variance of conditional expectations, and
    - Expectation of conditional variances

151

Methods: Multiple Imputation

- General approach
  - Analysis model for full response data

$$p(y \mid x;\theta)$$

  - Imputation model fit to observed response data

$$p(y_{obs} \mid x, v;\phi)$$

  - Generate datasets from predictive distribution

$$p(y \mid x, v) = \int p(y \mid x, v;\phi)p(\phi)\,d\phi$$

  - Analyze complete datasets
  - Combine results

152

Methods: Multiple Imputation

- Advantages
  - Can easily use auxiliary variables for imputation models that are not desired in the main analysis
  - Handle arbitrary missing data mechanisms
  - Assumptions explicit in imputation model
- Disadvantages
  - Does rely on parametric methods
  - Data models may be incompatible with imputation models
    - Auxiliary variables do not integrate out

153

MAR, MNAR Methods

- Selection models

$$\left[Y_{obs}, Y_{mis}, M \mid X\right] = \left[M \mid Y_{obs}, Y_{mis}, X\right] \times \left[Y_{obs}, Y_{mis} \mid X\right]$$

- Pattern mixture models

$$\begin{aligned}\left[Y_{obs}, Y_{mis}, M \mid X\right] &= \left[Y_{obs}, Y_{mis} \mid M, X\right] \times \left[M \mid X\right] \\ &= \left[Y_{mis} \mid Y_{obs}, M, X\right] \times \left[Y_{obs} \mid M, X\right] \times \left[M \mid X\right]\end{aligned}$$

154

MAR, MNAR Methods

- Selection models
  - Both parametric and semiparametric forms
  - Structural assumptions place on full data assumptions
- Pattern mixture models
  - Can be viewed as imputation of missing values from predictive distributions
  - Transparency of assumptions owing to models
  - Well suited to sensitivity analyses

155

MAR, MNAR Methods

- Relative advantages of selection models
  - Modeling of full data is natural approach
  - But model of nonresponse and outcome is less clear
- Relative advantage of pattern mixture models
  - Describes how observed and missing data distributions differ
    - Description through imputation methods
  - Analogy with time to event data for imputation

156

Recommendation # 9

- Statistical methods for handling missing data should be specified by clinical trial sponsors in study protocols.
- Assumptions associated with the analysis methods should be stated in a way that can be understood by clinicians.

157

Recommendation # 10

- Single imputation methods like last observation carried forward (LOCF) and baseline observation carried forward (BOCF) should not be used as the primary approach to the treatment of missing data unless the assumptions that underlie them are scientifically justified.

158

Recommendation # 11

- Parametric models in general, and random effects models in particular, should be used with caution, with all their assumptions clearly spelled out and justified.
- Models relying on parametric assumptions should be accompanied by goodness-of-fit procedures.

159

Recommendation # 12

- The primary analysis of the data from a RCT should account for the uncertainty attributable to missing data so that
  - associated significance tests have valid type I error rates, and
  - associated confidence intervals have valid coverage probabilities
- For inverse probability weighting and maximum likelihood methods, this can be accomplished by appropriate computation of standard errors using either asymptotic results or the bootstrap.
- For imputation, it is necessary to use appropriate rules for multiply imputing missing responses and combining results across imputed datasets.
  - Single imputation does not account for all sources of variability.

160

Recommendation # 13

- Weighted generalized estimating equations methods should be more widely used in settings when missing at random can be well justified and a stable weight model can be determined
  - These serve as a possibly useful alternative to parametric modeling.

161

Recommendation # 14

- When substantial missing data are anticipated, auxiliary information should be collected that is believed to be associated with reasons for missing values and with the outcomes of interest. Such can
  - allow use of a more appropriate MAR analysis, or
  - help in the conduct of sensitivity analyses
- Investigators should seriously consider following up on all or a random sample of trial dropouts who have not withdrawn consent in order to obtain
  - their reasons for dropping out of the study, and
  - relevant outcome measurements

162

Treatment Effects Among Compliers

- The report contains a brief section on CACE
  - Complier-averaged causal effects
- I have difficulty with this estimand scientifically
  - Bias due to noncomparability of per-protocol analyses
  - This bias may not be removed by adjustment for covariates owing to unmeasured confounding
    - Typically the R^2 of measured covariates is quite small

163

Recommendation # 15

- Sensitivity analyses should be part of the primary reporting of findings from clinical trials. Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting.

164

Sensitivity Analyses

- Analytic methods for missing data rely on untestable assumptions
- It is therefore of great importance to
  - Prespecify assumptions and
  - Explore the dependence of results to those assumptions

165

Types of assumptions

- Presumed mechanisms of missing data
  - MCAR, MAR, MNAR
- Analytic models involve
  - Distributional assumptions (mean, variance, parametric family)
  - Form of modeled variables (linear, dichotomized, interactions)
  - Auxiliary variables included in models
  - Analysis populations (efficacy, safety, etc.)
  - Departures from MAR or MNAR assumptions
  - Augmented data collection that could be used

166

Framework for Sensitivity Analyses

- Pattern mixture models show great flexibility for being able to model dependence on the various assumptions
  - Straightforward parameterization on differences in distributions between missing and nonmissing observations
    - Difference in means, odds ratios, etc.
- There remains much work to be done to better understand the extent to which sensitivity analyses should be conducted
  - The methods of handling missing data should not require more publications to describe than did the main clinical trial results

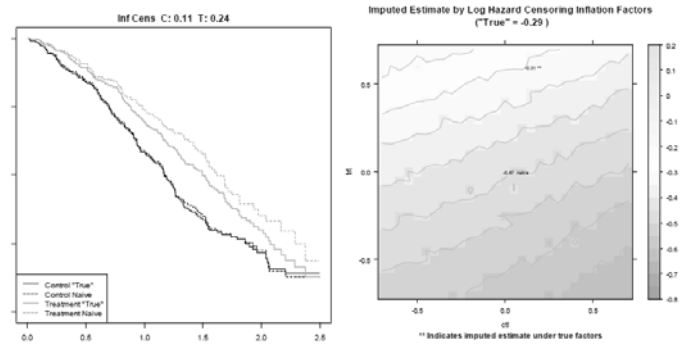
167

Example: Informative Censoring

- Time to event analysis from RCT with
  - Administrative censoring
  - Potentially informative censoring
- Primary analysis: A standard KM or PH analysis (MAR)
  - Assumes imputation of missing data from all subjects still at risk
- Explore sensitivity to change in hazard at time of informative censoring (MNAR)
  - Estimate treatment effect for each hypothesized change in hazard
- Display contour plot of inference as change in hazard varies
  - Consider bias of missing data varies by treatment group

168

Example: Contour Plots



	survival	informative censoring	log hazard inflation factor
control:	Weibull(B=2.0, k=1.5)	Exponential(B=6)	0.1
treatment:	Weibull(B=2.6, k=1.5)	Exponential(B=4)	0.6

Example: Informative Censoring

- This simplistic model presumes that all potentially informative censoring shares a common change in hazard within treatment groups
- Is modeling an average effect adequate?
  - Various more complicated models that have same average

	Estimated Treatment log(HR)								
Scenario	Mean "True"	"True" CI Coverage Rate	Mean "True" CI Width	Mean Naïve	Naïve CI Coverage Rate	Mean Naïve CI Width	Mean Imputed	Imputed CI Coverage Rate	Mean Imputed CI Width
base	-0.272	0.950	0.422	-0.392	0.834	0.480	-0.273	0.930	0.458
a	-0.276	0.961	0.422	-0.393	0.846	0.480	-0.273	0.941	0.458
b	-0.280	0.948	0.423	-0.393	0.849	0.480	-0.272	0.932	0.458
c	-0.280	0.946	0.423	-0.393	0.849	0.480	-0.272	0.932	0.458
d	-0.267	0.954	0.421	-0.392	0.826	0.480	-0.273	0.930	0.458
e	-0.278	0.951	0.423	-0.392	0.845	0.480	-0.273	0.929	0.458

Example: Informative Censoring

- Current research into readily understood methods for sensitivity analyses
- Time to event analysis from RCT with
  - Administrative censoring
  - Potentially informative censoring
- Assume change in hazard at time of informative censoring
  - Estimate treatment effect for each hypothesized change in hazard
- Display contour plot of inference

171

Final Comments

- Careful design of RCT to minimize missing data is all important
- Protocol should anticipate problems and pre-specify how they will be handled
- Sensitivity analyses should be included to quantify the possible impact of the missing data
- There is some hope that simple sensitivity analyses are possible
  - But it is not clear that they are ready for prime time

172

Really Bottom Line

"An ounce of prevention is worth a pound of cure"

173

Recommendation # 16

- The FDA and NIH should make use of their extensive clinical trial databases to carry out a program of research, both internal and external,
  - to identify common rates and causes of missing data in different domains,
  - to identify how different models perform in different settings, and
  - to use the results of such research to inform future study designs and protocols.

174

Recommendation # 17

- The FDA and the drug, device, and biologic companies that sponsor clinical trials should carry our continued training of their analysts to keep abreast of up-to-date techniques for missing data analysis.
- The FDA should also encourage continued training of their clinical reviewers to make them broadly familiar with missing data terminology and missing data methods.

175

Recommendation # 18

- The treatment of missing data in clinical trials, being a crucial issues, should have a higher priority for sponsors of statistical research such as NIH and NSF, including
  - Methods for sensitivity analyses and their resulting decisions,
  - Methods for non-monotone missing data,
  - Sample size calculations in the presence of missing data,
  - Designs for follow-up after treatment discontinuation,
  - Doable robust methods, and
  - Development of appropriate software.

176