Module 2:

Missing Data in Clinical Trials: Prevention and Estimands

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

(with acknowledgements to James Neaton, Ph.D., U of Minnesota)

Seattle Institute for Statistics in Clinical Research
July 25, 2016

1

Where Am I Going?

Organization of the Course

2

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.

3

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

4

Process

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

5

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

6

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    - Completion of protocol, stop intervention, consent withdrawn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

7

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 MCAR, MAR, MNAR
    - About assumptions of analytic models

8

Missing Data: Ideal

"Just say no."

(Nancy Reagan)

9

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

10

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

11

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

12

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

13

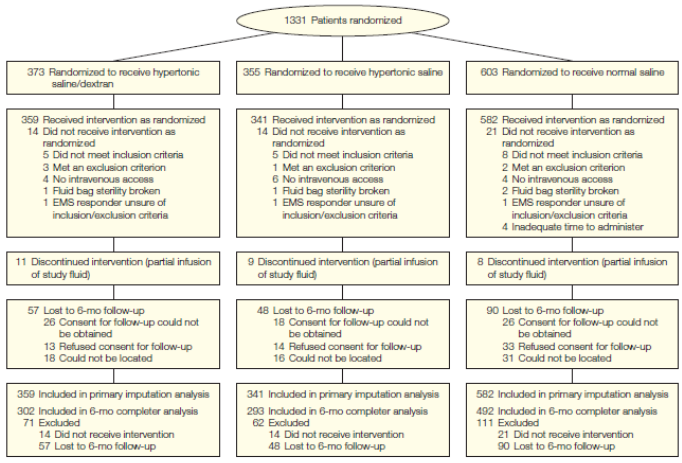
Example: Hypertonic Resuscitation in TBI

- 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

14

Example: Hypertonic Resuscitation in TBI

Figure 1. Trial Enrollment



Imputation analysis for 6-month neurologic outcome included all patients who received the intervention. Final completer analysis outcome data included all patients who received the intervention, defined as having the fluid connected to the intravenous line regardless of how much fluid was administered, and who completed 6-month follow-up. EMS indicates emergency medical system.

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?

16

Example: Rivaroxaban in ACS

- Proposed indication (from CRDAC meeting documents)
  - Reduce the risk of thrombotic cardiovascular events in patients with ACS (STEMI, NSTEMI or UA) in combination with ASA alone or with ASA plus a thienopyridine (clopidogrel or ticlopidine).
  - Rivaroxaban has been shown to reduce the risk of a combined endpoint of CV death, MI or stroke. The difference between treatments was driven by CV death and MI.
- Problem:
  - In the intial CRDAC meeting, 8.6 % of subjects (5.9% of P-Y) missing data in ITT analysis
  - With additional follow-up, 3.2% of subjects, (2.4% of P-Y) still missing data for ITT analysis
  - How to interpret results of the single registrational trial which showed difference of AR of about 2%, p value about .01 - .03?[17]

Point Meriting Extra Emphasis

- Many issues with missing data result from
  - poor choice of the primary endpoint, in combination with
  - the investigators' poor understanding of the scientific burden of proof
- Consequently, the major focus of this course is how to prevent missing data through
  - choosing the relevant primary outcome,
  - designing the study to facilitate proper data collection, and
  - educating the investigators on the importance of avoiding missing data.

18

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

19

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

20

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?

21

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?

22

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?

23

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?

24

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

25

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

26

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)

27

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?

28

Example: Antifibrinolytics in ChemoTX

- Patients undergoing chemotherapy for cancer or stem cell transplantation 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
- Major issue: Some patients will die of other causes
  - Underlying disease (cancer)
  - Infections during ablation of bone marrow
  - Graft versus host disease
- How do we record bleeding incidence in such patients?

29

Example: True Scientific Dilemmas

- Sometimes hard to score worst case
  - Death in a HTN study
- Sometimes measurement on patient becomes 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

30

Missing Data: Real Life

"Missing data happens"

(Bumper Sticker- rough translation)

31

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?

32

Missing Data: Sad Facts of Life

"Bloodsuckers hide beneath my bed"

"Eyepennies",
Mark Linkous (Sparklehorse)

33

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.)

34

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

35

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

36

Mechanisms for Missing Data

- Owing to (improper) definition of estimand
  - Competing risks, etc.
- Only three broad categories of “true” missing data
  - 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

37

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

38

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

39

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)

40

Overall Goal

- "Drug discovery"
  - More generally
    - a therapy / preventive strategy or diagnostic / prognostic procedure
    - for some disease
    - in some population of patients
- A series of experiments to establish
  - Safety of investigations / dose
  - Safety of therapy
  - Measures of efficacy
    - Treatment, population, and outcomes
  - Confirmation of efficacy
  - Confirmation of effectiveness

41

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?

42

The Enemy

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

- Maria von Trapp

(as quoted by Rodgers and Hammerstein)

43

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"

44

What Do We Want from Health Care?

- Prevention of disease
- Diagnosis of conditions
- Prognosis from disease
- Treatment of disease

- Who, what, when, where, why, how?

45

Why Might We Want to be Treated?

- **Who** do we mean by “we”?
- Personalized medicine vs (sub)populations
  - Fixed effects
    - Demographics
    - Behaviors
    - Environmental exposures
    - Genetics
    - Disease
  - Random effects

46

Why Might We Want to be Treated?

- **What** do we mean by "treatment"?
  - Drug(s), device(s), behaviors vs sequence vs combination
  - Formulation(s)
  - Dose vs dosing strategy (reductions, escalations)
- **Where / how** do we mean by "treatment"?
  - Administration (oral, topical, IV, IM, water supply)
- **When** do we mean by "treatment"?
  - Starting, stopping, frequency
  - Drug holidays
- **Why** do we want "treatment"?
  - Clinical measurement
  - Timeframe

47

Scientific Method

- We will use well-designed experiments to investigate candidate treatments
- Randomization is the key method whereby we find the causal effects of the intervention
  - "Causal" at the time of first application of the intervention
  - Ensuing treatments are potential mediators of effect
- We must do "per-randomization" analyses to ensure control of unmeasured confounding
  - Every randomized subject must have an outcome
  - *Any* missing data means that the analysis will not truly be according to randomization

48

Science: 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

49

Evidence Based Medicine

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

50

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

51

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

52

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

53

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

54

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

55

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?

56

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

57

Statistical Refinement of Hypotheses

- The group receiving the treatment will tend to 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)

58

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

59

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.

60

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

61

Basic Science

"Knowledge is good"

- Emil Faber
  Founder, Faber College

62

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

63

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

64

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)

65

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

66

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

67

Scientific Judgment in Burden of Proof

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

- Virginia Gildersleeve?

68

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?

69

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

70

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

71

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

72

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

73

Treatment: Points Meriting Extra Emphasis

- In a population receiving the same treatement strategy, different patients may at different times receive
  - Different drugs: prophylaxis, rescue, etc.
  - Different formulations: oral, IV, etc.
  - Different doses: titration, escalation, reductions due to AEs
  - Different frequencies: delays, drug holidays due to AEs
  - Different duration: response, lack of response
- It is important that both investigators and patients recognize that such differences are all anticipated variations within the same treatment strategy
- Looking ahead: None of these differences should affect our assessment of patient outcomes

74

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)

75

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

76

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

77

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?

78

Bottom Line

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

-Aretha Franklin, "Think"

79

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.)

80

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

81

"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

82

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

83

Choice of Estimands

- Ultimately, we need to know how to use the treatment
  - Who should we try it in?
  - What benefit might we expect?
  - What adverse effects might we expect / watch out for?
  - When should we stop the treatment?

84

Safety from RCT: Multiple Levels

- Safety of conducting RCTs
  - Phase I dose finding studies
- Safety in some ideal (enriched) population
  - Phase II or phase III efficacy studies
- Safety in a more general population
  - Phase III effectiveness studies
    - Safety in patients starting drug
    - Safety in patients taking drug chronically

85

Efficacy from RCT: Multiple Levels

- Preliminary efficacy in some ideal (enriched) population
  - Phase II or phase III efficacy studies
  - Often an intermediate / surrogate marker
- Efficacy / effectiveness in a more general population
  - Benefit vs preventing progression
  - Initial effects vs delayed effects
  - Initial effects vs waning effects
  - Phase III effectiveness studies
    - Averaged over all patients starting drug
    - Averaged only over patients taking drug chronically

86

Role of Adherence

- Many problems with missing data can be attributed to lack of adherence to the protocol
  - Competing risks due to toxicity, lack of response, early response, concomitant disease
- It is thus useful to consider how we might choose estimands by classifying levels of adherence

87

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?

88

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

89

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~~

90

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

91

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

92

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~~

93

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

94

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

95

Impact of Estimand on RCT

- The patients who are "relevant" differ according to the estimand of interest
- The primary goal should be to devise an experiment that only randomizes patients who are relevant to the estimand
  - This is often difficult
  - It may mean using more than one RCT, answering different aspects of the safety/effectiveness profile in different studies

96

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.

97

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

98

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

99

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

100

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

101

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

102

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

103

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

104

Possible RCT Estimand #2

- Average improvement for tolerators / compliers / "responders"
  - 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 ("responders")
  - are poor compliers

105

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

106

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
  - (Generally I am not)

107

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
  - (Generally I am not)

108

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

109

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

110

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

111

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

112

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

113

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

114

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

115

Event-Driven vs Visit-Driven Outcomes

- Event-driven: survival, hospitalization
- Visit-driven data: SBP, lab tests, QoL
- The more clinically important events are most often easier to collect than longitudinal measurements made at prescribed times
  - Longitudinal events need to be made on precise schedules
  - Visits and measurements are often cumbersome
  - On the other hand, events can often be collected retrospectively without substantial loss of accuracy (e.g., date of death).
- (NB: While time to event analyses can handle censoring, it must be noninformative censoring)

116

Methods for Visit Driven Outcomes

- Pre-define windows for scheduled visits
  - Plan for how variations from schedules will be classified
  - Use wide, contiguous follow-up visit windows
  - Define how time variation within a window will be used
- Try to define alternative methods for subjects missing a visit
- Definitely pre-specify how “interim” missing data might be imputed from previous and later measurements
- (Methods for “monotonic” missing data will typically be handled using methods described later.

117

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

118

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.

119

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?

120

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"

121

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)

122

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?

123

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
      - But could argue that we are just testing immediate vs delayed therapy with "proven" drug
  - At review scrutinize incidence of cross-in
    - Sensitivity analyses to time-varying cross-in

124

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)

125

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

126

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
- Issues
  - What is “washout” period for safety, efficacy?

127

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?

128

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

129

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)

130

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

131

Prevention of Missing Data

General Comments

Where am I going?

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

Emulating their techniques is highly recommended.

132

Primary Prevention of Missing Data

- 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 follow-up, communications and negotiations with patients
- Fully inform patients of trial requirements and importance of full participation during consent process.

133
Neaton

Primary Prevention of Missing Data

- 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.
- **Most importantly: Insist on the highest standards.**

134
Neaton

Secondary Prevention of Missing Data

- After trouble begins, act quickly
- Telephone contacts and home visits.
- Partial data collection (reduce demands of participation).
- Use central registries for vital status.

135
Neaton

Points Meriting Extra Emphasis

- 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

136
Neaton

Prevention of Missing Data

Subject Specific Issues

Where am I going?

Some methods of prevention should focus on the subjects

Of course, making certain the investigative staff know what to tell the subjects is a key part.

137

Prevention of Loss of Subjects' Data

- Minimize loss to follow-up
  - 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.
  - Include appropriate wording in Informed Consent document
    - "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"

138

Neaton

Prevention of Loss of Subjects' Data

- Minimize withdrawal of consent
  - Avoid burdensome studies
- Major predictors of missing data are related to trial design
  - 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
- Patient specific factors are less important
  - Younger age, smoking, socioeconomic status, …

139

Neaton

Training of Staff

- Staff are usually motivated to collect all necessary data, but they need to understand why the data is important
- Training of data collection staff should include
  - How the data contributes to RCT results
  - How to communicate these goals to patients
- During conduct of study
  - Make clear how the performance on data collection will be assessed
  - Provide tools for self-monitoring
  - Provide frequent feedback
  - Have plans for remedial training / quality improvement

140

"Dropout" Terminology is Problematic

- Just as with standard medical care
  - A particular treatment may be discontinued
  - However, that does not mean all medical care is stopped.
- All investigators and staff should avoid equating "stopping study drug" as meaning "study dropout"
- Instead use
  - "treatment with study drug has ended" vs
  - "patient has withdrawn consent" vs
  - "patient has been lost to follow-up"

141

Withdrawal of Consent

- Only the subject (or legally authorized representative) can withdraw consent
  - And there are gradations by which consent is withdrawn
  - Stopping study drug is not the same as withdrawal of consent
  - Modifying study visits is not the same as withdrawal of consent
- In all cases, an investigator decision to stop a study drug should not be regarded as withdrawal of consent
  - Instead, the patient should continue to monitored according to protocol (perhaps as modified for patient safety)

142

Withdrawal of Consent: Regulatory

- 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."
- Typically describe how withdrawal will affect
  - Medical care from treating HCP *(it should not)*
  - Availability of drugs *(it might)*
  - Monitoring for and treatment of possible adverse reactions

143

Additional Informed Consent Wording

- In order to maximize subject understanding and participation:
  - "The clinical trial specifies the typical time that a patient will receive the study drug. However, at some point during your participation in the clinical trial, either you or your doctor may decide that further treatment with the drug is either no longer necessary or no longer advisable. Should you and/or your doctor regard that such is the case, other treatment options will be discussed with you. However, it is still important to continue the follow-up visits as scheduled, in order to reliably answer the study question regarding any benefit of the treatment."

144

FDA Guidance on FDA-Regulated Trials

- When a subject withdraws, data collected up to time of withdrawal cannot be removed from database
- Investigators may ask subjects 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) may be used
- OHRP guidelines are also consistent with the above
  - "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."

145
Neaton

Guidance to Site Staff on Withdrawal

- Document reasons for withdrawal in medical chart and on CRF
- Discuss partial data collection with participant
- Advise participant that they always can re-consent
- Don't give up! Discourage use of terms like "off-study"
  - The DCCT only had "inactive" patients, never "lost to follow-up"

146
Neaton

CRF Documenting Withdrawal

- 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

147

Patient Withdrawal of Consent Form

- I have decided I know 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)

148

Neaton

Prevention of Missing Data

RCT Design Specific Issues

Where am I going?

A great many of the strategies used to prevent missing data relate to the design of the RCT itself

149

Prevention of Loss of Subjects' Data

- Major predictors of missing data are related to trial design
  - 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

150

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
- Incentives for participation
  - Continued access to drugs, reimburse expenses

151
Neaton

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

152
Neaton

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.

153
Neaton

Payments to Sites for Performance

- 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

154
Neaton

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

155
Neaton

Considerations for Open Label 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)

156
Neaton

Reporting Missing Data

RCT Design Specific Issues

Where am I going?

After all possible strategies for preventing missing data have been implemented, the validity of the scientific report will depend on the complete report of missingness

The RCT design should pre-specify how this will be done

157

Acceptable Rates of Missing Data

- The trial protocol should pre-specify rates of missing data that would still allow valid scientific and statistical conclusions
  - DSMB might recommend termination if these rates are exceeded
- Consider trial objectives and state missing data targets in protocol
  - Set performance goals considering:
    - Results from similar trials
    - Sensitivity analyses
- Consider possibility of missing data in sample size estimation
  - BUT: this is properly executed only by considering attenuated treatment effects and increased variability
  - It is inappropriate to just inflate the sample size and then ignore the presence of missing data

158

When is Missing Data a Problem?

- Many levels of concern leading to different opinions
  - 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.
- Fact is: There is nothing in your data that can reliably answer that question

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.

159
Neaton

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.

160
Neaton

Defining Lost-to-Follow-up

- A trial participant for whom the outcome of interest is not known – MISSING DATA- can vary over the course of the study
  - 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
- However, many "lost" patients can be found if you do not give up hope
  - Extraordinary measures are sometimes used with great success to re-contact patients

161
Neaton

Final Trial Reports

- Explicit reporting of length of *potential* follow-up and rates of missing data
  - Note: "potential follow-up" is the censoring distribution, not just the period under observation
- CONSORT diagrams
- Whenever possible, figures and tables should make clear any missing data or loss of follow-up

162

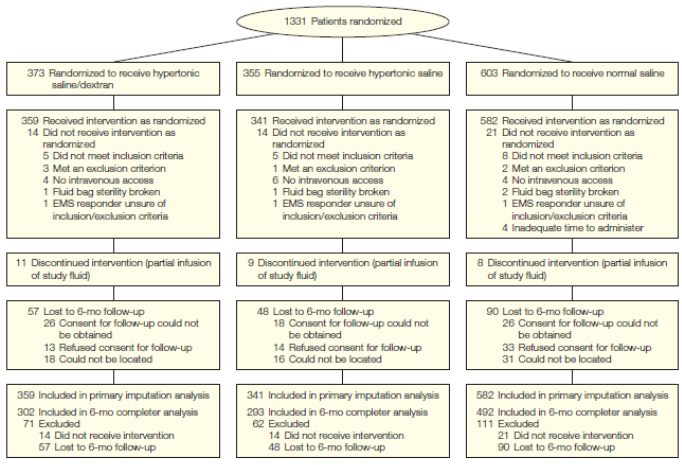
CONSORT 2010 Checklist: Follow-up

- 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.
- (Note: In time to event analyses, not all censoring mechanisms are equal. It is important to try to differentiate
  - "administrative censoring" by pre-specified calendar time
  - competing risks from mortality, etc.
  - withdrawal of consent (perhaps due to AEs, lack of response)
  - loss to follow-up)

163

Example: Hypertonic Resuscitation in TBI

Figure 1. Trial Enrollment



Imputation analysis for 6-month neurologic outcome included all patients who received the intervention. Final completer analysis outcome data included all patients who received the intervention, defined as having the fluid connected to the intravenous line regardless of how much fluid was administered, and who completed 6-month follow-up. EMS indicates emergency medical system.

Example: Methotrexate in PBC

- *Combes, et al, Hepatology, 2005*

Forty-one patients on the MTX arm and 47 patients on the placebo arm discontinued taking study drug prior to the end of the interventional phase and prior to experiencing the primary endpoint of liver transplant or death.

By the seventh year postrandomization, approximately one third of patients in both arms discontinued study treatment with no statistically significant differences between the treatment arms. Table 2 presents the numbers of patients discontinuing treatment early for each of several categories of reasons for early termination.

The overwhelming majority of patients who discontinued their study drug were still followed for occurrence of the study endpoints. Only 11 patients prematurely withdrew consent for follow-up of transplant-free survival status: 3 in the MTX arm and 8 in the placebo arm. The cumulative proportion withdrawing from the study in this manner was 1%, 3%, and 4.5% at 1, 2, and 6 years

Table 2. Reasons Provided for Discontinuing Study Treatment (Mtx or Placebo) Prior to the End of the Interventional Phase and Prior to Experiencing the Primary Endpoint of Liver Transplantation or Death

	MTX (n = 132)	Placebo (n = 133)
Signs of bone marrow suppression	4	5
Signs/symptoms of gastrointestinal toxicity	5	1
Signs/symptoms of respiratory adverse effects	2	4
Cancer	9	5
Other indications for MTX	0	2
Pregnancy	0	1
PBC progression	5	5
Other medical conditions	3	8
Other signs/symptoms (AEs)	4	2
Study burden	9	14
	41	47

165

Summary

- Missing data can be prevented – that needs to be the focus.
- Need to better educate investigators on the importance of complete follow-up.
- Insist on high standards.

166

Neaton

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.

167

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.

168

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

169

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.

170

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

171

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.

172

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

173

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.

174

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.

175

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

176

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

177

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)

178

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

179

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

180

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
    - Investigators, study coordinators, data management

181

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

182

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.

183

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.

184

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.

185

The Problem Persists

- Even in the presence of a well chosen estimand, a well designed study, and diligent follow-up of patients, missing data will likely occur
- We thus still need to consider how we will handle missing data to best address the chosen estimand

186

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

187

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

188

(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
  - BOCF: Assume first observation is *exactly* equal to missing data
  - Difficult to justify scientifically or statistically
    - Single imputation inappropriately presumes no variability

189

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

190

Methods of Analyzing Data

A Simple Statement of the Problem

in a Simple Example

Where am I going?

In order to better understand the various methods of analysis, it is useful to consider the basics of the problem with missing data.

We thus first consider the comparison of proportions in a simple RCT

191

A Simple Setting: Data, Hypotheses

- Consider a stratified comparison of proportions

Treatment $t = 0, 1$; Stratum $s = 1, \ldots, S,$; Individual $i = 1, \ldots, n_{ts}$

Outcome $Y_{tsi} \sim B(1, p_{ts})$

$$\theta = \sum_{s=1}^{S} w_s (p_{1s} - p_{0s}) \qquad \sum_{s=1}^{S} w_s = 1$$

$$H_0 : \theta = 0$$

192

A Simple Setting: Statistics

- Consider a stratified comparison of proportions

$$\hat{\theta} = \sum_{s=1}^{S} w_s \left(\hat{p}_{1s} - \hat{p}_{0s}\right) = \sum_{s=1}^{S} w_s \left(\bar{Y}_{1s\bullet} - \bar{Y}_{0s\bullet}\right)$$

$$\sim N\left(\theta, \sum_{s=1}^{S} w_s \left(\frac{p_{1s}(1-p_{1s})}{n_{1s}} + \frac{p_{0s}(1-p_{0s})}{n_{0s}}\right)\right)$$

Under stronger null $H_0^{(S)} : \forall s = 1, \ldots, S : p_{1s} = p_{0s} = p_{\bullet s}$

$$Z = \frac{\hat{\theta}}{\sqrt{\sum_{s=1}^{S} w_s \hat{p}_{\bullet s}(1-\hat{p}_{\bullet s})\left(\frac{1}{n_{1s}} + \frac{1}{n_{0s}}\right)}} \overset{H_0^{(S)}}{\sim} N(0,1); \qquad \hat{p}_{\bullet s} = \frac{n_{1s}\hat{p}_{1s} + n_{0s}\hat{p}_{0s}}{n_{1s} + n_{0s}}$$

193

A Simple Setting: Efficiency Weights

- Under the strong assumption that under the alternative there is a constant treatment effect across strata, every choice of weights is unbiased
  - Gauss-Markov thm ➔ weights proportional to inverse variance

$$w_s \propto \left(\frac{p_{1s}(1-p_{1s})}{n_{1s}} + \frac{p_{0s}(1-p_{0s})}{n_{0s}}\right)^{-1} = \frac{n_{1s}n_{0s}}{n_{0s}p_{1s}(1-p_{1s}) + n_{1s}p_{0s}(1-p_{0s})}$$

Under stronger null $H_0^{(S)} : \forall s = 1, \ldots, S : p_{1s} = p_{0s} = p_{\bullet s}$

$$w_s \propto \frac{n_{1s}n_{0s}}{n_{0s}+n_{1s}}\left(\frac{1}{p_{\bullet s}(1-p_{\bullet s})}\right) \Rightarrow \hat{w}_s \propto \frac{n_{1s}n_{0s}}{n_{0s}+n_{1s}}\left(\frac{1}{\hat{p}_{\bullet s}(1-\hat{p}_{\bullet s})}\right)$$

To avoid unstable estimates, might use

$$\hat{w}_s \propto \frac{n_{1s}n_{0s}}{n_{0s}+n_{1s}}$$

194

A Simple Setting: Importance Weights

- More plausible: Treatment effect may vary by strata
  - Importance weights would consider relative size of strata in target population and clinical impact of changes in outcome across strata
  - Typically, we lack direct information on these and instead use relative size of strata in our sample

$$\hat{w}_s \propto n_{0s} + n_{1s}$$

Under blocked r : 1 randomization :

$$\text{Efficiency } \hat{w}_s \propto \frac{n_{0s} n_{1s}}{n_{0s} + n_{1s}} = \frac{n_{0s} r n_{0s}}{n_{0s} + r n_{0s}} = \frac{r}{1+r} n_{0s} \propto n_{0s}$$

$$\text{Importance } \hat{w}_s \propto n_{0s} + n_{1s} = n_{0s} + r n_{0s} = (1+r) n_{0s} \propto n_{0s}$$

195

A Simple Setting: Missing Data

- Consider a stratified comparison of proportions with some data potentially missing

Treatment $t = 0, 1$; Stratum $s = 1, \ldots, S,$; Individual $i = 1, \ldots, n_{ts}$

$$\text{Outcome } Y_{tsi} \sim B(1, p_{ts})$$
$$\text{Indic Msng } M_{tsi} \sim B(1, \pi_{ts}) \qquad N_{ts}^{(M)} = \sum_{i=1}^{n_{ts}} M_{tsi} \qquad N_{ts}^{(O)} = n_{ts} - N_{ts}^{(M)}$$
$$\text{Aux Data } V_{tsi} \sim G_{ts}$$

$$\theta = \sum_{s=1}^{S} w_s (p_{1s} - p_{0s}) \qquad \sum_{s=1}^{S} w_s = 1$$

$$H_0 : \theta = 0$$

196

Comments on Probability Models

- With missing data, we need to consider joint distribution of responses Y_i , X_i and indicators of missingness M_i^Y , M_i^X
- Definition of probability models might be based on guesses about missing data mechanisms
- Perhaps a latent subpopulation is destined to have missing data
  - E.g., subjects prone to adverse effects, early cure, nonresponse
  - Indicator of missing data is truly indicator of latent subpopulation
  - Response distribution is "naturally" defined conditional on latent subpopulation
- Perhaps value of (future?) response "causes" missing data
  - E.g., subjects prone to drop out if *observed* blood pressure is high
  - Missing data indicator is "naturally" defined conditional on (observed or unobserved) value of response

197

"Latent Subpopulations": Conditional Dist

- We model the distribution of outcomes when conditioned on whether the value is or is not observed

$$Y_{tsi} \mid M_{tsi} = 0 \quad \sim \quad B\left(1,\ p_{ts}^{(O)}\right) \qquad \hat{p}_{ts}^{(O)} = \frac{1}{N_{ts}^{(O)}} \sum_{j=1}^{n_{ts}} \left(1 - M_{tsi}\right) Y_{tsi}$$

$$Y_{tsi} \mid M_{tsi} = 1 \quad \sim \quad B\left(1,\ p_{ts}^{(M)}\right) \qquad \hat{p}_{ts}^{(M)} = \frac{1}{N_{ts}^{(M)}} \sum_{j=1}^{n_{ts}} M_{tsi}\ Y_{tsi}$$

Correspondence with unconditional distributions :

$$p_{ts} = \left(1 - \pi_{ts}\right) p_{ts}^{(O)} + \pi_{ts}\ p_{ts}^{(M)}$$

198

"Latent Subpopulations": Hypotheses

- Correspondence with parameterization of treatment effect

$$\theta = \sum_{s=1}^{S} w_s \left(p_{1s} - p_{0s} \right)$$

$$= \sum_{s=1}^{S} w_s \left\{ \left[\left(1-\pi_{1s}\right) p_{1s}^{(O)} + \pi_{1s}\ p_{1s}^{(M)} \right] - \left[\left(1-\pi_{0s}\right) p_{0s}^{(O)} + \pi_{0s}\ p_{0s}^{(M)} \right] \right\}$$

$$= \sum_{s=1}^{S} w_s \left\{ \left(p_{1s}^{(O)} - p_{0s}^{(O)} \right) + \left[\pi_{1s} \left(p_{1s}^{(M)} - p_{1s}^{(O)} \right) - \pi_{0s} \left(p_{0s}^{(M)} - p_{0s}^{(O)} \right) \right] \right\}$$

Estimable : π_{ts} ; $p_{ts}^{(O)}$

Conjectured : $\Delta_{ts} = p_{ts}^{(M)} - p_{ts}^{(O)}$

$$\theta = \sum_{s=1}^{S} w_s \left\{ \left(p_{1s}^{(O)} - p_{0s}^{(O)} \right) + \left[\pi_{1s}\ \Delta_{1s} - \pi_{0s}\ \Delta_{0s} \right] \right\}$$

199

Ignorable Missing Data

- In this simple setting, we can explicitly state the conditions under which an analysis based only on observed data will be unbiased for the same estimand that would have been targeted in the absence of missing data

Ignorable with respect to bias (providing weights are unchanged) :

$$\sum_{s=1}^{S} w_s \left[\pi_{1s}\ \Delta_{1s} - \pi_{0s}\ \Delta_{0s} \right] = 0 \quad \Rightarrow \quad \sum_{s=1}^{S} w_s \left(p_{1s}^{(O)} - p_{0s}^{(O)} \right) = \theta$$

- More generally, we will want to consider
  - How the distribution of outcomes differ among cases with observed or unobserved values in each stratum
  - How the probability of missingness differs across strata
  - How the weights assigned in standard analyses might differ when analyses are restricted to cases with observed data
    - (Regression analyses use efficiency weights presuming correctly modeled effects)

200

Missing Completely at Random (MCAR)

- Indicator of missingness is independent of all other variables

Missing Completely at Random (MCAR):

$$\forall t = 0,1; \quad s = 1,\ldots,S: \quad \pi_{ts} = \pi \quad \& \quad \Delta_{ts} = 0$$

$$\sum_{s=1}^{S} w_s \left\{ \left(p_{1s}^{(O)} - p_{0s}^{(O)} \right) + \left[\pi_{1s}\ \Delta_{1s} - \pi_{0s}\ \Delta_{0s} \right] \right\} = \sum_{s=1}^{S} w_s \left(p_{1s}^{(O)} - p_{0s}^{(O)} \right)$$

By MCAR: $N_{1s}^{(O)} \sim B\left(n_{1s}, 1-\pi\right)$ $N_{0s}^{(O)} \sim B\left(n_{0s}, 1-\pi\right)$

In absence of missing data: $\hat{w}_s \propto n_{1s} + n_{0s}$

In presence of missing data: $\hat{w}_{ss}^{(O)} \propto N_{1s}^{(O)} + N_{0s}^{(O)}$

Hence:

$$\hat{\theta}^{(O)} = \sum_{s=1}^{S} \hat{w}_s^{(O)} \left(\hat{p}_{1s}^{(O)} - \hat{p}_{0s}^{(O)} \right) \rightarrow_p \theta$$

201

Missing at Random (MAR): Example 1

- Indicator of missingness is independent of any unobserved data
  - Many different scenarios, only a few of which are ignorable

Missing at Random (MAR): Example #1

$$\forall t = 0,1; \quad s = 1,\ldots,S: \quad \Delta_{ts} = 0 \quad \text{but} \quad \exists t, t', s, s' \ni \pi_{ts} \neq \pi_{t's'}$$

$$\sum_{s=1}^{S} w_s \left\{ \left(p_{1s}^{(O)} - p_{0s}^{(O)} \right) + \left[\pi_{1s}\ \Delta_{1s} - \pi_{0s}\ \Delta_{0s} \right] \right\} = \sum_{s=1}^{S} w_s \left(p_{1s}^{(O)} - p_{0s}^{(O)} \right)$$

By MAR Example #1: $N_{1s}^{(O)} \sim B\left(n_{1s}, 1-\pi_{1s}\right)$ $N_{0s}^{(O)} \sim B\left(n_{0s}, 1-\pi_{0s}\right)$

In absence of missing data: $\hat{w}_s \propto n_{1s} + n_{0s}$

In presence of missing data: $\hat{w}_{ss}^{(O)} \propto N_{1s}^{(O)} + N_{0s}^{(O)}$

Hence:

$$\hat{\theta}^{(O)} = \sum_{s=1}^{S} \hat{w}_s^{(O)} \left(\hat{p}_{1s}^{(O)} - \hat{p}_{0s}^{(O)} \right) \rightarrow_p \theta^{(O)} \neq \theta \quad \text{unless} \quad \forall s, s' \in \{1,\ldots S\}: \left(p_{1s} - p_{0s} \right) = \left(p_{1s'} - p_{0s'} \right)$$

202

Inverse Probability Weighting (IPW)

- In MAR Example #1, it is straightforward to ensure that an analysis that adjusts for missing data has the same estimand as would be targeted in the absence of missing data

Inverse Probability Weighting (IPW) Ex #1:

$$N_{1s}^{(O)} \sim B(n_{1s}, 1-\pi_{1s}) \qquad N_{0s}^{(O)} \sim B(n_{0s}, 1-\pi_{0s})$$

In absence of missing data: $\hat{w}_s \propto n_{1s} + n_{0s}$

In presence of missing data: $\hat{w}_{ss}^{(O)} \propto N_{1s}^{(O)} + N_{0s}^{(O)}$

Hence:

$$\hat{\theta}^{(O^*)} = \sum_{s=1}^{S} \hat{w}_s^{(O)} \left(\frac{\hat{p}_{1s}^{(O)}}{1-\pi_{1s}} - \frac{\hat{p}_{0s}^{(O)}}{1-\pi_{0s}} \right) \rightarrow_p \theta$$

Providing the sample sizes do not lead to unstable weights

$$\hat{\theta}^{(O^*)} = \sum_{s=1}^{S} \hat{w}_s^{(O)} \left(\frac{\hat{p}_{1s}^{(O)}}{1-\hat{\pi}_{1s}} - \frac{\hat{p}_{0s}^{(O)}}{1-\hat{\pi}_{0s}} \right) \rightarrow_p \theta$$

203

Missing at Random (MAR): Example 2

- Indicator of missingness is independent of any unobserved data
  - Consider a scenario where missingness depends on auxiliary variable (perhaps an intermediate measure of outcome)
  - IPW can be used with any consistent estimates of probability of observed data

Missing at Random (MAR): Example #2

$\exists t, t', s, s' \ni \pi_{ts} \neq \pi_{t's'}; \qquad \exists t, s \ni \Delta_{ts} \neq 0;$

But: $E[Y_{tsi} \mid M_{tsi} = 1, V_{tsi} = v] = E[Y_{tsi} \mid M_{tsi} = 0, V_{tsi} = v] = p_{tsv}^*$

with $\sum_v p_{tsv}^* \Pr[V_{tsi} = v] = p_{ts}$

Inverse Probability Weighting (IPW) Ex #2:

$$\hat{\theta}^{(O^*)} = \sum_{s=1}^{S} \hat{w}_s^{(O)} \left[\sum_v \left(\frac{\sum_{i=1}^{n_{ts}} Y_{1si}(1-M_{1si}) 1_{[V_{1si}=v]}}{\hat{\Pr}(M_{1si}=0 \mid V_{1si}=v)} - \frac{\sum_{i=1}^{n_{ts}} Y_{0si}(1-M_{0si}) 1_{[V_{0si}=v]}}{\hat{\Pr}(M_{0si}=0 \mid V_{0si}=v)} \right) \right] \rightarrow_p \theta$$

204

Missing at Not at Random (MNAR)

- Indicator of missingness is potentially dependent on some unobserved data

Missing Not at Random (MNAR):

$$\exists t, s, v \ni \mathrm{E}\left[Y_{tsi} \mid M_{tsi} = 1, V_{tsi} = v\right] \neq \mathrm{E}\left[Y_{tsi} \mid M_{tsi} = 0, V_{tsi} = v\right]$$

205

MNAR: Analysis Approaches

- Nothing in your data can distinguish MNAR from MAR
  - We will have to rely on conjectures about the value of the treatment effects among subpopulations with missing data

A Simple Conjecture $\Delta_{1s} = \Delta_{1\bullet}\ ; \Delta_{0s} = \Delta_{0\bullet}$

$$\hat{\theta} = \sum_{s=1}^{S} w_s \left\{ \left(\hat{p}_{1s}^{(O)} - \hat{p}_{0s}^{(O)} \right) + \left[\hat{\pi}_{1s}\ \Delta_{1s} - \hat{\pi}_{0s}\ \Delta_{0s} \right] \right\}$$

- More generally
  - Consider joint distribution of (Y_{tsi}, M_{tsi})
  - Perhaps conjectures about Δ_{ts} should be influenced by
    - Observed prevalences of missing data
    - Beliefs about mechanisms of missing data
    - Beliefs about tendencies toward skewed outcomes

206

Imputation

- So far we have considered phrasing the problem of missing data in terms of distributional parameters pertaining to the cases with unobserved outcomes
- An alternative approach is to try to predict what the unobserved data was, and then analyze the augmented data set
  - Single imputation: Do this once
  - Multiple imputation: Create several augmented data sets and perform a "meta-analysis" across those imputations
- Predictive models might use
  - Parametric models based on conjectured distributions
  - Empirical models from resampling (and perturbing) the observed values

207

Parametric Imputation

- In our simple setting of binary outcomes, parametric imputation is the most straightforward
  - (perturbing binary data does not really make sense any other way)

Conjecture $\Delta_{ts} = \Delta_{t\bullet}$

$$Y_{tsi} \mid M_{tsi} = 1 \quad \sim \quad B\left(1,\ p_{ts}^{(O)} + \Delta_{t\bullet}\right)$$

208

Multiple Imputation: General Approach

- Estimate values for the proportion observed data in each arm, stratum, and (perhaps) auxiliary strata $\hat{p}^{*}_{tsv}$
- Presume values for the difference in probabilities in observed and unobserved data Δ_{tsv}
- For *k = 1, …, K* imputations,
  - impute data unobserved values, and
  - then perform standard analyses to find estimated treatment effects for each augmented data set

$$\hat{\theta}_{(k)}; \qquad \hat{Var}_{(k)}\left(\hat{\theta}\right)$$

  - then summary statistics across imputations

$$\hat{\theta} = \underset{k}{E}\left[\hat{\theta}_{(k)}\right] = \frac{1}{K}\sum_{1}^{K}\hat{\theta}_{(k)} \qquad \underset{k}{Var}\left(\hat{\theta}_{(k)}\right) = \frac{1}{K-1}\sum_{1}^{K}\left(\hat{\theta}_{(k)} - \hat{\theta}\right)^{2} \qquad \overline{Var}_{(\bullet)}\left(\hat{\theta}\right) = \frac{1}{K}\sum_{1}^{K}\hat{Var}_{(k)}\left(\hat{\theta}\right)$$

209

Multiple Imputation: General Approach

- Base ultimate inference on distributions estimated across imputations

$$\hat{\theta} = \underset{k}{E}\left[\hat{\theta}_{(k)}\right]$$

$$\hat{Var}\left(\hat{\theta}\right) = \underset{k}{E}\left[\hat{Var}_{(k)}\left(\hat{\theta}\right)\right] + \left(\frac{K+1}{K}\right)\underset{k}{Var}\left(\hat{\theta}_{(k)}\right)$$

$$Z = \frac{\hat{\theta} - \theta_{0}}{\sqrt{\hat{Var}\left(\hat{\theta}\right)}} \overset{H_{0}}{\sim} N\left(0,\ 1\right)$$

210

Comments on Multiple Imputation

- Multiple imputation has the advantage of automatically weighting analyses across strata and/or regression covariates exactly as they would have been weighted if the imputed data were the true outcomes
- Multiple imputation allows relatively “transparent” implementation of the assumptions about mechanisms of missing data
- Multiple imputation is easily applied in MCAR, MAR, and MNAR models
  - There are still multiple ways we can parameterize the impuation schemes

211

Methods of Analyzing Data

More Complicated Models

Where am I going?

In order to better understand the various methods of analysis, it is useful to consider the basics of the problem with missing data.

We now consider other simple analytic models

212

Proportional Hazards Models

- The example of stratified proportions generalizes easily to the analysis of time to event data with either weighted logrank statistics or the proportional hazards models
- Estimated hazards are the proportion of deaths among the risk set
- Estimating equations are weighted averages of the difference in hazards

$$U(\theta_0) = \sum_t \left(d_{1t} - d_{\bullet t} \frac{n_{1t}}{n_{\bullet t}} \right) = \sum_t \frac{n_{1t} n_{0t}}{n_{1t} + n_{0t}} \left(\hat{\lambda}_{1t} - \hat{\lambda}_{0t} \right)$$

213

Two Sample Means: Complete Data

- Consider a two sample comparison of means
  - A more complicated model owing to higher dimensional parameter space

Data :

Treatment arm : $Y_i \overset{iid}{\sim} (\mu, \sigma^2); \quad i = 1, \ldots, n_Y$

Control arm : $X_i \overset{iid}{\sim} (\nu, \tau^2); \quad i = 1, \ldots, n_X$

Sample means, variances : $\bar{Y}, \hat{\sigma}^2, \bar{X}, \hat{\tau}^2$

Hypothesis Test :

Null hypothesis : $H_0 : \mu = \nu$

Test statistic : $Z = \dfrac{\bar{Y} - \bar{X}}{\sqrt{\dfrac{\hat{\sigma}^2}{n_Y} + \dfrac{\hat{\tau}^2}{n_X}}} \overset{H_0}{\sim} N(0, 1)$

214

Latent Subpopulations: Notation

- Notation with indicators of missing data as latent subpopulation

Treatment arm : $Y_i \overset{iid}{\sim} (\mu,\ \sigma^2);\quad i = 1,\ldots,n_Y$

Indicator of missing : $M_i^Y = 1_{[Y_i\ msng]};\ N_{YM} = \sum M_i^Y;\ N_{YO} = n_Y - N_{YM}$

Observed data : $Y_i \mid M_i^Y = 0 \overset{iid}{\sim} (\mu_O,\ \sigma_O^2)$

Missing data : $Y_i \mid M_i^Y = 1 \overset{iid}{\sim} (\mu_M,\ \sigma_M^2)$

Control arm : $X_i \overset{iid}{\sim} (\nu,\ \tau^2);\quad i = 1,\ldots,m$

Indicator of missing : $M_i^X = 1_{[X_i\ msng]};\ N_{XM} = \sum M_i^X;\ N_{XO} = n_X - N_{XM}$

Observed data : $X_i \mid M_i^X = 0 \overset{iid}{\sim} (\nu_O,\ \tau_O^2)$

Missing data : $X_i \mid M_i^X = 1 \overset{iid}{\sim} (\nu_M,\ \tau_M^2)$

215

Latent Subpopulations: Models

- Correspondence with complete data models

Treatment arm :

Marginal distns : $Y_i \overset{iid}{\sim} (\mu,\ \sigma^2);\quad M_i^Y \overset{iid}{\sim} B(1,\ \pi_Y);\quad i = 1,\ldots,n_Y$

Conditional distns : $Y_i \mid M_i^Y = 1 \overset{iid}{\sim} (\mu_M,\ \sigma_M^2);\quad Y_i \mid M_i^Y = 0 \overset{iid}{\sim} (\mu_O,\ \sigma_O^2)$

$$\Rightarrow \mu = \mu_O + \pi_Y(\mu_M - \mu_O);\quad \sigma^2 = \sigma_O^2\left[1 - \pi_Y + \pi_Y\frac{\sigma_M^2}{\sigma_O^2} + \pi_Y(1-\pi_Y)\left(\frac{\mu_M - \mu_O}{\sigma_O}\right)^2\right]$$

Control arm :

Marginal distns : $X_i \overset{iid}{\sim} (\nu,\ \tau^2);\quad M_i^X \overset{iid}{\sim} B(1,\ \pi_X);\quad i = 1,\ldots,n_X$

Conditional distns : $Y_i \mid M_i^Y = 1 \overset{iid}{\sim} (\nu_M,\ \tau_M^2);\quad Y_i \mid M_i^Y = 0 \overset{iid}{\sim} (\nu_O,\ \tau_O^2)$

$$\Rightarrow \nu = \nu_O + \pi_X(\nu_M - \nu_O);\quad \tau^2 = \tau_O^2\left[1 - \pi_X + \pi_X\frac{\tau_M^2}{\tau_O^2} + \pi_X(1-\pi_X)\left(\frac{\nu_M - \nu_O}{\tau_O}\right)^2\right]$$

216

Latent Subpopulations: Hypotheses

- Hypotheses tested in presence of missing data

Complete data : $H_0 : \mu = \nu$

Missing data : $H_0 : \mu_O + \pi_Y(\mu_M - \mu_O) = \nu_O + \pi_X(\nu_M - \nu_O)$

Estimable from observed data : μ_O ; π_Y ; ν_O ; π_X

Assumed based on untestable assumptions : $(\mu_M - \mu_O)$; $(\nu_M - \nu_O)$

Alternative formulations :

$$H_0 : \mu_O - \nu_O + \pi_Y(\mu_M - \mu_O) - \pi_X(\nu_M - \nu_O) = 0$$

$$H_0 : \mu_O(1-\pi_Y) - \nu_O(1-\pi_X) + \pi_Y\mu_M - \pi_X\nu_M = 0$$

217

Latent Subpopulations: Statistics

- Summary statistics and asymptotic distributions

Treatment arm : $Y_i \overset{iid}{\sim} (\mu, \sigma^2)$; $i = 1, \ldots, n_Y$

$$\bar{Y}_O = \frac{1}{N_{YO}} \sum_{i=1}^{n} (1 - M_i^Y) Y_i \qquad \bar{Y}_M = \frac{1}{N_{YM}} \sum_{i=1}^{n} M_i^Y Y_i$$

$$\bar{Y}_O \mid \vec{M}^Y \dot{\sim} N\left(\mu_O, \frac{\sigma_O^2}{N_{YO}}\right) \qquad \bar{Y}_M \mid \vec{M}^Y \dot{\sim} N\left(\mu_M, \frac{\sigma_M^2}{N_{YM}}\right)$$

$$\hat{\sigma}_O^2 = \frac{1}{N_{YO} - 1} \sum_{i=1}^{n} (1 - M_i^Y)(Y_i - \bar{Y}_O)^2 \qquad \hat{\sigma}_M^2 = \frac{1}{N_{YM} - 1} \sum_{i=1}^{n} M_i^Y (Y_i - \bar{Y}_M)^2$$

$$\hat{\sigma}_O^2 \to_p \sigma_O^2 \qquad \hat{\sigma}_M^2 \to_p \sigma_M^2$$

218

Latent Subpopulations: Statistics

- Summary statistics and asymptotic distributions

Control arm : $X_i \overset{iid}{\sim} (\nu, \tau^2); \quad i = 1, \ldots, n_X$

$$\bar{X}_O = \frac{1}{N_{XO}} \sum_{i=1}^{n} (1 - M_i^X) X_i \qquad \bar{X}_M = \frac{1}{N_{XM}} \sum_{i=1}^{n} M_i^X X_i$$

$$\bar{X}_O \mid \vec{M}^X \overset{\sim}{} N\left(\nu_O, \frac{\tau_O^2}{N_{XO}}\right) \qquad \bar{X}_M \mid \vec{M}^X \overset{\sim}{} N\left(\nu_M, \frac{\tau_M^2}{N_{XM}}\right)$$

$$\hat{\tau}_O^2 = \frac{1}{N_{XO} - 1} \sum_{i=1}^{n} (1 - M_i^X)(X_i - \bar{X}_O)^2 \qquad \hat{\tau}_M^2 = \frac{1}{N_{XM} - 1} \sum_{i=1}^{n} M_i^X (X_i - \bar{X}_M)^2$$

$$\hat{\tau}_O^2 \to_p \tau_O^2 \qquad \hat{\tau}_M^2 \to_p \tau_M^2$$

219

Latent Subpopulations: Statistics

- Correspondence with overall statistics

Treatment arm : $Y_i \overset{iid}{\sim} (\mu, \sigma^2); \quad i = 1, \ldots, n$

$$\bar{Y} = \frac{N_{YO}\bar{Y}_O + N_{YM}\bar{Y}_M}{n_Y} \qquad \hat{\sigma}^2 = \frac{(N_{YO} - 1)\hat{\sigma}_O^2 + (N_{YM} - 1)\hat{\sigma}_M^2}{n_Y - 1}$$

Control arm : $X_i \overset{iid}{\sim} (\nu, \tau^2); \quad i = 1, \ldots, n_X$

$$\bar{X} = \frac{N_{XO}\bar{X}_O + N_{XM}\bar{X}_M}{n_X} \qquad \hat{\tau}^2 = \frac{(N_{XO} - 1)\hat{\tau}_O^2 + (N_{XM} - 1)\hat{\tau}_M^2}{n_X - 1}$$

$$Z = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{\hat{\sigma}^2}{n_Y} + \frac{\hat{\tau}^2}{n_X}}} = \frac{\frac{N_{YO}\bar{Y}_O + N_{YM}\bar{Y}_M}{n_Y} - \frac{N_{XO}\bar{X}_O + N_{XM}\bar{X}_M}{n_X}}{\sqrt{\frac{(N_{YO} - 1)\hat{\sigma}_O^2 + (N_{YM} - 1)\hat{\sigma}_M^2}{n_Y(n_Y - 1)} + \frac{(N_{XO} - 1)\hat{\tau}_O^2 + (N_{XM} - 1)\hat{\tau}_M^2}{n_X(n_X - 1)}}}$$

220

Latent Subpopulations: Conditional

- Conditioning on missing data patterns

$$H_0 : \mu_O - \nu_O + \pi_Y(\mu_M - \mu_O) - \pi_X(\nu_M - \nu_O) = 0$$

$$\bar{Y}_O \mid \vec{M}^Y \sim N\left(\mu_O, \frac{\sigma_O^2}{N_{YO}}\right) \qquad \bar{X}_O \mid \vec{M}^X \sim N\left(\nu_O, \frac{\tau_O^2}{N_{XO}}\right)$$

$$Z_{\vec{M}} = \frac{\bar{Y}_O - \bar{X}_O + \hat{\pi}_Y(\mu_M - \mu_O) - \hat{\pi}_X(\nu_M - \nu_O)}{\sqrt{\frac{n_Y \hat{\sigma}_O^2}{(1-\hat{\pi}_Y)} + \frac{n_X \hat{\tau}_O^2}{(1-\hat{\pi}_X)}}}$$

$$Z_{\vec{M}} \mid \vec{M}^Y, \vec{M}^X \sim N\left(\frac{\mu_O - \nu_O + \hat{\pi}_Y(\mu_M - \mu_O) - \hat{\pi}_X(\nu_M - \nu_O)}{\sqrt{\frac{n_Y \sigma_O^2}{(1-\hat{\pi}_Y)} + \frac{n_X \tau_O^2}{(1-\hat{\pi}_X)}}}, 1\right)$$

221

Latent Subpopulations: Unconditional

- The unconditional distribution of the test statistic can be found by integrating over the sampling distribution of the probability of being missing

$$H_0 : \mu_O - \nu_O + \pi_Y(\mu_M - \mu_O) - \pi_X(\nu_M - \nu_O) = 0$$

$$Z_{\vec{M}} \mid \vec{M}^Y, \vec{M}^X \sim N\left(\frac{\mu_O - \nu_O + \hat{\pi}_Y(\mu_M - \mu_O) - \hat{\pi}_X(\nu_M - \nu_O)}{\sqrt{\frac{n_Y \sigma_O^2}{(1-\hat{\pi}_Y)} + \frac{n_X \tau_O^2}{(1-\hat{\pi}_X)}}}, 1\right)$$

$$n_Y \hat{\pi}_Y \sim B(n_Y, \pi_Y) \qquad n_X \hat{\pi}_X \sim B(n_X, \pi_X)$$

$$\Pr(Z < z \mid H_0) = \sum_i \sum_j \Pr(Z < z \mid H_0; \vec{M}^Y, \vec{M}^X) \Pr(\hat{\pi}_Y = i) \Pr(\hat{\pi}_X = j)$$

222

Latent Subpopulations: Comments

- From the previous:
  - We have restated our hypotheses in terms of
    - Estimable means, SD among the nonmissing
    - Estimable probability of being in the subpopulation
    - Nonestimable means SD among the subpopulation missing data
- Difficult issues:
  - (From a frequentist standpoint): The null hypothesis ultimately depends on 6 parameters (4 estimable, 2 presumed) and the distribution of *Z* on another 4 (2 estimable, 2 presumed)
  - (From a Bayesian standpoint): Is there a correlation between the probability of being in the subpopulation and the means, SD among the subpopulations with missing data?

223

Alternative Approach: Consider Statistics

- We can alternatively regard problem is only one of guessing what the test statistic would have been with complete data

Available statistics : $\bar{Y}_O$; $\hat{\sigma}^2_O$; $\bar{X}_O$; $\hat{\tau}^2_O$

"Guessed" statistics : $\bar{Y}_M$; $\hat{\sigma}^2_M$; $\bar{X}_M$; $\hat{\tau}^2_M$

Derived summaries :

$$\bar{Y} = \frac{N_{YO}\bar{Y}_O + N_{YM}\bar{Y}_M}{n_Y} \qquad \bar{X} = \frac{N_{XO}\bar{X}_O + N_{XM}\bar{X}_M}{n_X}$$

$$\hat{\sigma}^2 = \frac{(N_{YO}-1)\hat{\sigma}^2_O + (N_{YM}-1)\hat{\sigma}^2_M}{(n_Y-1)} \qquad \hat{\tau}^2 = \frac{(N_{XO}-1)\hat{\tau}^2_O + (N_{XM}-1)\hat{\tau}^2_M}{(n_X-1)}$$

$$Z = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{\hat{\sigma}^2}{n_Y} + \frac{\hat{\tau}^2}{n_X}}}$$

224

Multiple Imputation: General Approach

- Estimate values for the first two moments of the distributions for the observed data $\bar{Y}_O ; \hat{\sigma}^2_O ; \bar{X}_O ; \hat{\tau}^2_O$
- Presume values for the first two moments of the distributions for missing data $\mu_M , \sigma^2_M , \nu_M , \tau^2_M$
- For *k = 1, …, K* imputations,
  - impute data and/or sufficient statistics for the missing data, and
    $\bar{Y}_{M(k)} ; \hat{\sigma}^2_{M(k)} ; \bar{X}_{M(k)} ; \hat{\tau}^2_{M(k)}$
  - then derived estimates and standard errors
    $\hat{\theta}_{(k)} = \bar{Y}_{(k)} - \bar{X}_{(k)} ; \qquad \hat{Var}_{(k)}(\hat{\theta})$
  - then summary statistics across imputations

$$\hat{\theta} = \underset{k}{E}\left[\hat{\theta}_{(k)}\right] = \frac{1}{K}\sum_{1}^{K}\hat{\theta}_{(k)} \qquad \underset{k}{Var}\left(\hat{\theta}_{(k)}\right) = \frac{1}{K-1}\sum_{1}^{K}\left(\hat{\theta}_{(k)} - \hat{\theta}\right)^2 \qquad \overline{Var}_{(\bullet)}\left(\hat{\theta}\right) = \frac{1}{K}\sum_{1}^{K}\hat{Var}_{(k)}\left(\hat{\theta}\right)$$

225

Multiple Imputation: General Approach

- Base ultimate inference on distributions estimated across imputations

$$\hat{\theta} = \underset{k}{E}\left[\hat{\theta}_{(k)}\right]$$

$$\hat{Var}\left(\hat{\theta}\right) = \underset{k}{E}\left[\hat{Var}_{(k)}\left(\hat{\theta}\right)\right] + \left(\frac{K+1}{K}\right)\underset{k}{Var}\left(\hat{\theta}_{(k)}\right)$$

$$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{\hat{Var}\left(\hat{\theta}\right)}} \overset{H_0}{\sim} N(0,\ 1)$$

226

Imputation Models for Data

- Use presumed distributions to impute data that was not observed

Treatment arm : $Y_i \overset{iid}{\sim} (\mu, \sigma^2); \quad i = 1, \ldots, n_Y$

Observed data : $Y_{O\,i} \overset{iid}{\sim} (\mu_O, \sigma_O^2)$

Missing data : $Y_{M\,i} \overset{iid}{\sim} (\mu_M, \sigma_M^2)$

Imputation model : $Y_{M\,i} \overset{iid}{\sim} \sigma_M \left(\frac{Y_{O\,i'} - \mu_O}{\sigma_O} \right) + \mu_M$

Control arm : $X_i \overset{iid}{\sim} (\nu, \tau^2); \quad i = 1, \ldots, m$

Observed data : $X_{O\,i} \overset{iid}{\sim} (\nu_O, \tau_O^2)$

Missing data : $X_{M\,i} \overset{iid}{\sim} (\nu_M, \tau_M^2)$

Imputation model : $X_{M\,i} \overset{iid}{\sim} \tau_M \left(\frac{X_{O\,i'} - \nu_O}{\tau_O} \right) + \nu_M$

227

Comments on Imputation Model

- I suggested a semi-parametric imputation model that could be based on resampling the observed data
- This should not really matter too much, because all that is truly important is the presumption of the first two moments
  - To the extent that heavy tails lead to more variable estimates of SDs in small samples, this might better capture that variability
- We could also have used a parametric imputation model to compute sufficient statistics
  - Parametric based on approximate normality of sample means
    - Use sample third, fourth moments in observed data to parallel the semi-parametric model
  - Parametric model is more problematic for sample variances in small samples

228

Imputation Models for Sufficient Statistics

- By the central limit theorem, so long as the fourth moments exist

For third and fourth central moments ω and ξ, respectively

$$\begin{pmatrix} \bar{Y}_M \\ \hat{\sigma}^2_M \end{pmatrix} \dot{\sim} N_2 \left(\begin{pmatrix} \mu_M \\ \sigma^2_M \end{pmatrix}, \begin{bmatrix} \frac{\sigma^2_M}{N_{YM}} & \frac{\omega_{YM}}{N_{YM}} \\ \frac{\omega_{YM}}{N_{YM}} & \frac{\xi_{YM} - \sigma^4_M}{N_{YM}} \end{bmatrix} \right)$$

$$\begin{pmatrix} \bar{X}_M \\ \hat{\tau}^2_M \end{pmatrix} \dot{\sim} N_2 \left(\begin{pmatrix} \nu_M \\ \tau^2_M \end{pmatrix}, \begin{bmatrix} \frac{\tau^2_M}{N_{XM}} & \frac{\omega_{XM}}{N_{XM}} \\ \frac{\omega_{XM}}{N_{XM}} & \frac{\xi_{XM} - \tau^4_M}{N_{XM}} \end{bmatrix} \right)$$

229

Sensitivity Analyses

- All of the above discussion considered a single set of presumed distributions of response among those with missing data
- However, all of those presumptions are untestable
- The only solution is to cover a range of plausible presumptions and
  - Display the range of presumed values that confer “statistical significance”
  - Identify the “tipping point” at which statistical significance is lost
- For a low dimensionality of the unknown parameters, we can consider countour plots
  - Point estimates
  - Interval estimates
  - P values

230

More Complicated Settings

- So far we have considered the two-sample setting
- How do we extend this to adjusted (stratified) statistics?
- How do we extend this to repeated measurements?
- How do we extend this to a setting in which we want to presume different "effects" of missingness in different strata?
  - For example, how might we presume different means in subjects who dropped out due to "lack of response" than those who dropped out due to "inconvenience of study"?

231

Methods of Analyzing Data

Advanced Statistical Methods

Where am I going?

Various methods have been described in the statistical literature

We consider the relative merits of the various approaches

232

General Setting

- We consider the longitudinal setting common to most RCT
- Types of variables (both observed and possibly missing)
  - Baseline (pre-randomization) variables
    - Includes treatment assignment
  - Measurements of auxiliary variables (post-randomization)
    - Measures of adherence
    - Measures of safety: adverse events
    - Measures of intermediate and secondary outcomes
  - Measurements of primary clinical outcomes (post-randomization)
    - Might have repeated measurements over time
- Indicators of missing measurements
  - Potentially one indicator for every other planned measurement
  - Intermittent or monotonic missingness

233

Notation

- Observed and missing data represent an alternative partition of the random vectors *Y* and *V*

Example : $\vec{Y}$ $\quad \vec{M}_Y = \vec{m} = (0,0,0,1,0,0,1,1,1)$ $\quad \vec{Y}^*_{\vec{m}} = \begin{pmatrix} \vec{Y}_{obs} \\ \vec{Y}_{mis} \end{pmatrix}$

$$\vec{Y}_{obs} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \vec{Y} \qquad \vec{Y}_{mis} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \vec{Y}$$

234

Notation

- We find it useful to characterize for each individual the random variables that are observed and those that have missing values
- ***X*** = baseline covariates (including treatment)
- ***M*** = indicator of missingness
  - ***M*** = ***(M_Y, M_V)*** for each component of ***Y***, ***V***
- ***Y*** = outcome variables (possibly repeated measurements)
  - ***Y_m**** = ***(Y_{obs}, Y_{mis})*** denote observed and missing data, respectively
- ***V*** = auxiliary variables (generally post-randomization)
  - ***V_m*= (V_{obs}, V_{mis})*** for observed, missing auxiliary data, resp
  - (These are of interest as an aid in the missing data analysis)

235

Missing Completely at Random (MCAR)

- Missingness is unrelated to any aspect of the data values

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

- Equivalently: The data distribution is completely unrelated to whether the data is or is not observed

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

$$\Pr\left[\vec{M}_{Yi} = \vec{m}_{Yi}, \vec{M}_{Vi} = \vec{m}_{Vi} \mid \vec{Y}_i = \vec{y}, \vec{V}_i = \vec{v}_i, \vec{X}_i = \vec{x}_i\right] = \Pr\left[\vec{M}_{Yi} = \vec{m}_{Yi}, \vec{M}_{Vi} = \vec{m}_{Vi}\right]$$

$$\Pr\left[\vec{Y}_i = \vec{y}, \vec{V}_i = \vec{v}_i, \vec{X}_i = \vec{x}_i \mid \vec{M}_{Yi} = \vec{m}_{Yi}, \vec{M}_{Vi} = \vec{m}_{Vi}\right] = \Pr\left[\vec{Y}_i = \vec{y}, \vec{V}_i = \vec{v}_i, \vec{X}_i = \vec{x}_i\right]$$

236

Missing at Random (MAR)

- Missingness can depend on observed data, but not on unobserved values

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

- Equivalently: The distribution of the unobserved values can depend on the observed data, but not on being missing

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

For an arbitrary $\vec{m}$ used to define partitions $\vec{Y}^*_{\vec{m}}, \vec{V}^*_{\vec{m}}$

$$\Pr\left[\vec{M} = \vec{m} \mid \vec{Y}^*_{\vec{m}} = (\vec{y}_{obs}, \vec{y}_{mis}), \vec{V}^*_{\vec{m}} = (\vec{v}_{obs}, \vec{v}_{mis}), \vec{X}_i = \vec{x}_i\right] =$$
$$\Pr\left[\vec{M} = \vec{m} \mid \vec{Y}_{obs} = \vec{y}_{obs}, \vec{V}_{obs} = \vec{v}_{obs}, \vec{X}_i = \vec{x}_i\right]$$
$$\Pr\left[\vec{Y}_{mis} = \vec{y}_{mis}, \vec{V}_{mis} = \vec{v}_{mis} \mid \vec{M} = \vec{m}, \vec{Y}_{obs} = \vec{y}_{obs}, \vec{V}_{obs} = \vec{v}_{obs}, \vec{X}_i = \vec{x}_i\right] =$$
$$\Pr\left[\vec{Y}_{mis} = \vec{y}_{mis}, \vec{V}_{mis} = \vec{v}_{mis} \mid \vec{Y}_{obs} = \vec{y}_{obs}, \vec{V}_{obs} = \vec{v}_{obs}, \vec{X}_i = \vec{x}_i\right]$$

237

Missing Not at Random (MNAR)

- Conditional on all observed data, the distribution of the unobserved values is different than the distribution of the observed values

For an arbitrary $\vec{m}$ used to define partitions $\vec{Y}^*_{\vec{m}}, \vec{V}^*_{\vec{m}}$

$$\exists \vec{m}' \neq \vec{m}: \quad \Pr\left[\vec{Y}_{mis} = \vec{y}_{mis}, \vec{V}_{mis} = \vec{v}_{mis} \mid \vec{M} = \vec{m}, \vec{Y}_{obs} = \vec{y}_{obs}, \vec{V}_{obs} = \vec{v}_{obs}, \vec{X}_i = \vec{x}_i\right] \neq$$
$$\Pr\left[\vec{Y}_{mis} = \vec{y}_{mis}, \vec{V}_{mis} = \vec{v}_{mis} \mid \vec{M} = \vec{m}', \vec{Y}_{obs} = \vec{y}_{obs}, \vec{V}_{obs} = \vec{v}_{obs}, \vec{X}_i = \vec{x}_i\right]$$

238

Methods: CC

- Complete case analysis
  - Ignore cases with missing data
- Only appropriately unbiased for ignorable missingness
  - MCAR and some MAR
    - (MAR okay when pre-specified analysis is conditional on appropriate covariates)
  - When missingness appreciable, substantial inefficiency
- Use of complete case analysis does correspond to an assumption about the missing data mechanism
  - That assumption is poorly characterized, however

239

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 observed auxiliary variables $\boldsymbol{V_{obs}}$
  - E.g., logistic regression models
- 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

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

240

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

241

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
  - More difficult to parameterize MNAR

242

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

243

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

244

Methods: LOCF / BOCF Imputation

- Common single imputation methods
  - LOCF: Assume last observation is *exactly* equal to missing data
  - BOCF: Assume first observation is *exactly* equal to missing data
- Either of these are MNAR models

Last Observation Carried Forward (LOCF): a MNAR conjecture

$$\forall t,s: \quad M_{tsi}=1 \;\Rightarrow\; Y_{tsi}=V_{tsi}$$

- Difficult to justify scientifically
- Difficult to justify single imputation statistically
  - A MAR model could have been specified

MAR conditional on interim observation V

$$\forall t,s: \quad \mathrm{E}\left[Y_{tsi} \mid M_{tsi}=1, V_{tsi}=v\right] = \mathrm{E}\left[Y_{tsi} \mid M_{tsi}=0, V_{tsi}=v\right] = p^{*}_{tsv}$$

$$\text{with} \quad \sum_{v} p^{*}_{tsv} \Pr\left[V_{tsi}=v\right] = p_{ts}$$

245

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

246

Methods: Multiple Imputation

- General approach
  - Analysis model for full response data

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

  - Imputation model fit to observed response data

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

  - 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 across imputed datasets

247

Methods: Multiple Imputation

- Advantages
  - Can 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

248

MAR, MNAR Methods

- Selection models

$$[Y_{obs}, Y_{mis}, M \mid X] = [M \mid Y_{obs}, Y_{mis}, X] \times [Y_{obs}, Y_{mis} \mid X]$$
$$\overset{MAR}{=} [M \mid Y_{obs}, X] \times [Y_{obs}, Y_{mis} \mid X]$$

- Pattern mixture models

$$[Y_{obs}, Y_{mis}, M \mid X] = [Y_{obs}, Y_{mis} \mid M, X] \times [M \mid X]$$
$$= [Y_{mis} \mid Y_{obs}, M, X] \times [Y_{obs} \mid M, X] \times [M \mid X]$$
$$\overset{MAR}{=} [Y_{mis} \mid Y_{obs}, X] \times [Y_{obs} \mid M, X] \times [M \mid X]$$

249

MAR, MNAR Methods

- Selection models
  - Both parametric and semiparametric forms
  - Structural assumptions placed 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

250

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

251

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.

252

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.

253

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.

254

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.

255

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.

256

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

257

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

258

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.

259

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

260

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

261

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

262

Reporting Analyses

- It will be key to devote much attention to descriptive statistics that might
  - Indicate the magnitude of data that had to be imputed
  - Support the reasoning behind particular sensitivity analyses

- Descriptive statistics overall and by missingness patterns
  - Baseline (pre-randomization)
  - Treatment administered: dose delays, reductions, holidays, d/c
  - Disposition of patient: d/c drug, partial d/c assessments, withdraw consent
    - Perhaps time to entering each of these Markov “states” by important predictors
  - Any interim measures of primary, secondary, safety outcomes

263

Methods of Analyzing Data

An Example

Where am I going?

We consider a simple (simplistic?) approach that can be used to explore sensitivity to MAR assumptions

We have investigated the robustness to semi-parametric assumptions used in the sensitivity analysis

264

Example: Basic Approach

- Consider the analysis we would do with complete data
- Derive a (semi)parametric model to impute data under MAR
  - Multiple imputation to obtain inference
- Create MNAR model by couching MAR model in a larger family
  - Additional parameters model the departures from MAR
  - Parameters specific to each treatment group
- By MNAR assumption, there is nothing in the data that can estimate the additional parameters that model MNAR
  - Conduct a series of multiple imputation analyses conditional on assumed values for the additional MNAR parameters
- Find the "tipping point": the values of the MNAR parameters that substantially change inference relative to MAR model
  - Must account for "burden of proof": pivotal RCT, noninferiority, etc
  - Secondarily assess reasonableness of that tipping point

265

Example: Time to Event Analysis

- Setting of time to event examined first, because
  - The typical analysis method with noninformative censoring (complete data in a sense) is relatively standard
    - Unadjusted: logrank test
    - Adjusted: proportional hazards regression
  - There are no nuisance parameters
    - (With means of continuous data, we will have to also consider the variability of measurements)
- Mechanisms for missingness
  - Administrative censoring from times of accrual and data analysis
    - MAR that is handled well by KM
  - Potentially informative censoring due to loss of follow-up
    - (Competing risks could be handled providing consistent with the estimand of greatest interest)

266

Example: Logrank Test

- Estimating equation from score function of partial likelihood

$$U(\theta_0) = \sum_t \left(d_{1t} - d_{\bullet t} \frac{n_{1t}}{n_{\bullet t}} \right) = \sum_t \frac{n_{1t} n_{0t}}{n_{1t} + n_{0t}} \left(\hat{\lambda}_{1t} - \hat{\lambda}_{0t} \right)$$

- Under the strong null hypothesis (no treatment effect on any aspect of the distribution), PH holds for the treatment parameter
- Under the weak null hypothesis we are examining some sort of weighted time average of the hazard ratio, and presuming that average HR is 1
  - The weights will depend both on the underlying survival distribution and the censoring distribution
  - But with only administrative censoring, we will accept that

267

Example: Approach

- We use a pattern mixture model to reproduce an analysis that would only have administrative censoring
- The accrual time and data analysis time is known for all subjects
- We will ultimately impute the minimum of a survival time and the administrative censoring time

268

Example: Pattern Mixture Model

$$[Y_{obs}, Y_{mis}, M \mid X] = [Y_{obs}, Y_{mis} \mid M, X] \times [M \mid X]$$
$$= [Y_{mis} \mid Y_{obs}, M, X] \times [Y_{obs} \mid M, X] \times [M \mid X]$$
$$\overset{MAR}{=} [Y_{mis} \mid Y_{obs}, X] \times [Y_{obs} \mid M, X] \times [M \mid X]$$

$[M \mid X]$ distribution of missingness within each treatment arm

$[Y_{obs} \mid M, X]$ estimated by hazard among subjects who are at most administratively censored within each treatment arm

$[Y_{mis} \mid Y_{obs}, X]$ estimated by proportionally increased / decreased hazard after time of potentially informative censoring separately for each treatment arm

269

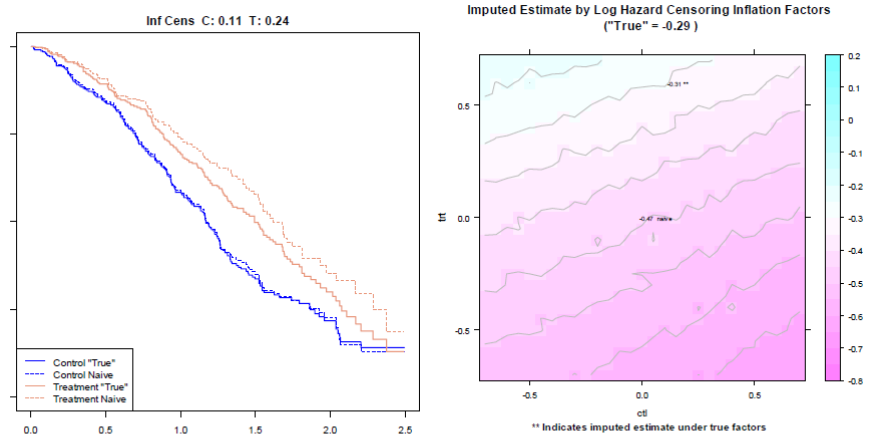
Example: Summary

- 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)
  - Multiply mpute administratively censored data
  - 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
  - HR estimates, CI, p values

(Reference: MS Thesis, Eric Meier – www.RCTdesign.org)

270

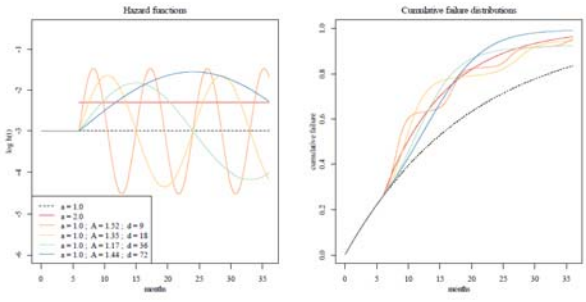
Example: Contour Plots



	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: Impact of PH Assumption

- This simplistic model presumes all potentially informative censoring shares common constant HR within treatment arms
- Is modeling an average effect adequate?
  - Various more complicated models that have same average
  - Consider hazard functions of varying shape after potentially informative censoring



272

Figure 14. Sinusoidal perturbations equivalent to $\sigma_t = 2.0$

Example: Impact of PH Assumption

- Generally reasonable (though slightly low) coverage probability across a wide variety of scenarios

	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

273

Extension to Other Settings

- Adjusted time to event analyses
  - Using estimated hazards from PH regression in imputation relatively straightforward
- Binary outcomes
  - Model treatment arm (and baseline covariate) specific MNAR odds ratios
  - Impact of departures from common OR needs to be explored
    - Mean-variance relationship may have greater impact, though PH regression can be viewed as stratified Mantel-Haenszel, so may generalize
- Means of continuous (longitudinal) data
  - Must account for two MNAR parameters on each treatment arm
    - Difference in means
    - Difference in standard deviations

274

Difference in Means

- For treatment arm *k*

$$\text{Complete data model}: \quad \bar{Y}_{k,all} \sim \left(\mu_{k,all}, \sigma^2_{k,all}\right)$$
$$\text{Observed data model}: \quad \bar{Y}_{k,obs} \sim \left(\mu_{k,obs}, \sigma^2_{k,obs}\right)$$
$$\text{Missing data model}: \quad \bar{Y}_{k,all} \sim \left(\mu_{k,mis}, \sigma^2_{k,mis}\right)$$

$$\mu_{k,mis} = \mu_{k,obs} + \delta_k \qquad \sigma^2_{k,mis} = \sigma^2_{k,obs}\omega_k$$
$$\mu_{k,all} = \pi_k \delta_k + \mu_{k,obs}$$
$$\sigma^2_{k,all} = \left(1 - \pi_k + \pi_k \omega_k\right)\sigma^2_{k,obs} + \pi_k\left(1 - \pi_k\right)\delta_k$$

275

Regression

- When adjusting for covariates or analyzing longitudinal data, we can still regard that the score function is a sum with some average difference between observed and missing data
- After computing impact of various hypothesized means, variances alternative models of covariate dependence can be explored with respect to the average mean, variance they correspond to

276

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, because the intended audience is still highly skeptical

277

Really Bottom Line

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

278

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.

279

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.

280

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.

281