Biost 536 / Epi 536
Categorical Data Analysis in Epidemiology

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

Lecture 13:
Causal Modeling;
Missing Data

December 4, 2014

1

Lecture Outline

- Causal Estimands
- Analysis Methods
- Propensity Scores
- Missing Data

2

Causal Estimands

3

Common Questions

- Clustering
  - Individuals
  - Variables
- Associations between “risk factors” and “events”
  - “Markers” of incident events
  - “Causes” of incident events
- Prediction of events
  - Estimation of tendencies (means, medians, …)
  - Estimation of ranges (necessary tolerance, …)
  - Classification (diagnosis, prognosis)

4

My Claim

- Ultimate interest is almost always understanding cause and effect relationships
- Clustering is something we do at the very beginning of investigations
- "Associative" analysis for markers is merely recognition that we cannot infer cause and effect from observational data
  - Often a logical step on the way to investigating cause and effect
  - Sometimes all that is feasible
- "Predictive" models are most often stop-gap solutions that will work in the short term
  - If we understood the cause and effect, we would use that information

5

Causation

- A term that is at best imprecisely defined
- When considering a single “risk factor” we might consider whether it is
  - “necessary” for the event to occur
  - “sufficient” for the event to occur
- Most often, it is neither
  - The incidence of the event will depend on other variables as well
    - At least with respect to our current level of understanding
- Nonetheless, we try to refine what we mean within a context that allows us to define three broad levels of estimands

6

Three Levels of Estimands

- Effect of exposure on the population
- Effect of exposure on some subpopulation
- Effect of exposure within an individual

7

Causal Effect: Definition

- We consider a counterfactual setting in which we could know outcomes for each individual when exposed and unexposed

u Index (covariate vector) for unit of treatment

$Y_E(u)$ Outcome for unit u in presence of exposure

$Y_{\bar{E}}(u)$ Outcome for unit u in absence of exposure

Causal effect of exposure for unit u

$$CE(u) = Y_E(u) - Y_{\bar{E}}(u)$$

8

Causal Effect: Comments

- The effect within an individual
- If we consider the outcome to be deterministic, we presume that *u* captures all pertinent variables that identify the sampling unit and its determinants of outcome
  - Essentially a random effect measuring all pertinent covariates at the time of the “treatment” (exposure)
- Of course, it is not possible to observe the outcome for both the exposed and unexposed setting for any unit
  - Time (at the very least) marches on
- And it is probably not really of interest to condition on every single covariate
  - We most often find it useful to average over covariates that we do not yet understand enough to measure

9

Indiv Average Causal Effect: Definition

- We consider a counterfactual setting in which we could know a distribution of outcomes for each individual when exposed and unexposed that may depend on time-varying covariates

u Index (time invariant covariate vector) for unit of treatment

$\vec{V}$ Time - varying covariate vector for unit of treatment

$Y_E\left(u \mid \vec{V}\right)$ Outcome for unit u in presence of exposure

$Y_{\bar{E}}\left(u \mid \vec{V}\right)$ Outcome for unit u in absence of exposure

Average causal effect of exposure for unit u

$$ACE(u) = E_{\vec{V}|u}\left[Y_E\left(u \mid \vec{V}\right) - Y_{\bar{E}}\left(u \mid \vec{V}\right)\right]$$

10

Indiv Average Causal Effect: Comments

- The average effect within an individual
  - Averaged over any effect modification by *V*
- We now model the possibility that there might be a distribution of settings in which the "treatment" (exposure) would be administered to an individual
- For any arbitrary auxiliary covariate vector *V*, it is impossible to measure the outcome in both the exposed and unexposed settings
- However, depending on how *V* is defined, we may through repeated experiments on *u* estimate *ACE(u)*
  - We must include time itself as a covariate that we do not want to hold constant
  - We must also presume no time trend that would confound or modify our estimate of the desired estimand

11

Average Causal Effect: Definition

- We average the individual average causal effect across the population

u Index (time invariant covariate vector) for unit of treatment

$\vec{V}$ Time - varying covariate vector for unit of treatment

$Y_E\left(u \mid \vec{V}\right)$ Outcome for unit u in presence of exposure

$Y_{\bar{E}}\left(u \mid \vec{V}\right)$ Outcome for unit u in absence of exposure

Average causal effect of exposure

$$ACE = E_U\left\{E_{\vec{V}|u}\left[Y_E\left(u \mid \vec{V}\right) - Y_{\bar{E}}\left(u \mid \vec{V}\right)\right]\right\}$$

12

Average Causal Effect: Comments

- The average individual effect within some population
  - May or may not consider variable effect within an individual
- Again, depending on how *V* is defined, we may through repeated experiments on *u* estimate *ACE(u)*
  - Because we are using expectation, we can consider varying exposure randomly across individuals

13

Population Average Causal Effect: Defn

- We consider a counterfactual setting in which we could know outcomes for a population of individuals when exposed and unexposed

u Index for unit of treatment having some defined distribution within a population of interest

$Y_E\left(u \mid \vec{V}\right)$ Outcome for unit u in presence of exposure

$Y_{\bar{E}}\left(u \mid \vec{V}\right)$ Outcome for unit u in absence of exposure

Population average causal effect of exposure

$$PACE = E_U\left[Y_E(u)\right] - E_U\left[Y_{\bar{E}}(u)\right]$$

14

PACE: Comments

- This measures the average effect in the population
- Note that because we are using expectation, we expect the *ACE* and *PACE* to be equal

Average causal effect of exposure

$$ACE = E_U\left[Y_E(u) - Y_{\bar{E}}(u)\right]$$

Population average causal effect of exposure

$$PACE = E_U\left[Y_E(u)\right] - E_U\left[Y_{\bar{E}}(u)\right]$$

15

Subpopulations

- We can of course define an estimand similar to *PACE* for each of several subpopulations
  - We can consider causal effect in each subgroup separately, or
  - We can consider some average of the separate subgroup causal effects

Population average causal effect of exposure in subgroup

$$PACE(\vec{x}) = E_U[Y_E(u)] - E_U[Y_{\bar{E}}(u)]$$

Average subpopulation causal effect of exposure in subgroup

$$ASCE(\vec{x}) = \sum_{\vec{x}} w_{\vec{x}} PACE(\vec{x})$$

16

Modifications to Causal Estimands

- In the previous measures, we used means and differences of means to characterize the causal effect
- With binary outcomes, we have also considered
  - Ratios of means, which are equivalent to differences of log means
  - Odds ratios, which are equivalent to differences of log odds

17

Modified CE, ACE, PACE

- We consider alternatives to the expectation
  - Note that transformations of binary exposures may not make sense
  - And not that *mACE* and *mPACE* may not be equal

$f(\ )$ is some alternative transformation (e.g., log or logit)

Causal effect of exposure for unit u

$$m\text{CE}(u) = f(Y_E(u)) - f(Y_{\bar{E}}(u))$$

Average causal effect of exposure for unit u

$$mACE = E_U\left[f(Y_E(u)) - f(Y_{\bar{E}}(u))\right]$$

Population average causal effect of exposure for unit u

$$mPACE = f(E_U[Y_E(u)]) - f(E_U[Y_{\bar{E}}(u)])$$

18

Analysis Methods

Impact on Analysis Models

- When considering the analysis models used for each of the estimands, we must consider the role of:

- Matching in design

- Variable adjustment
  - Effect modification, confounding, precision

- Measure of association
  - Risk difference, risk ratio, odds ratio

20

Available Tools

- Regression models
  - Fixed effects
  - Random effects
- Matched analyses
  - Fixed effects
  - Random effects
- Stratified analyses using weighted averages

21

Population Estimands: RR and RD

- In absence of confounding we would treat as a two sample problem
- In presence of confounding we can
  - Estimate effects within strata
  - Combine strata according to some “standard” population
- RD and RR are collapsible
  - Hence, adjusting for covariates should give an estimate of a common RD or RR across all covariate defined groups
  - In the presences of effect modification, however, we still need to use weighted combinations to recreate the population

22

Population Estimands: OR

- In absence of confounding we would treat as a two sample problem
- In presence of confounding we can
  - Estimate effects within strata
  - Combine strata according to some “standard” population
    - Even with a common OR in all subgroups, we will obtain different standardized OR depending upon the population used.
- OR is not collapsible
  - We must recognize that adjusted / matched analyses are not estimating the same OR as desired for mPACE
    - Hence either logistic or conditional logistic regression would have to use weighted averages
    - In case-control studies, we may not have the information about the distribution of subjects in the entire population

23

Example

- Hypothetical case-control study: Original data

	Cases			Controls			OR
Age	**Exp**	**Unexp**	**P(stratum)**	**Exp**	**Unexp**	**P(stratum)**	
50-55	77	9	0.04	395	33	0.21	**1.40**
55-60	102	28	0.06	418	76	0.25	**1.51**
60-65	121	73	0.10	277	113	0.19	**1.48**
65-70	119	161	0.14	147	133	0.14	**1.50**
70-75	98	292	0.19	64	130	0.10	**1.47**
75-80	64	430	0.25	24	106	0.06	**1.52**
80-85	27	401	0.21	8	78	0.04	**1.52**
Total	608	1394		1333	669		**4.57**

24

Example

- Hypothetical case-control study: Standardization 1

	Cases			Controls			OR
Age	**Exp**	**Unexp**	**P(stratum)**	**Exp**	**Unexp**	**P(stratum)**	
50-55	18	2	0.01	18	2	0.01	**1.40**
55-60	31	9	0.02	34	6	0.02	**1.51**
60-65	37	23	0.03	43	17	0.03	**1.48**
65-70	51	69	0.06	63	57	0.06	**1.50**
70-75	60	180	0.12	79	161	0.12	**1.47**
75-80	65	436	0.25	92	408	0.25	**1.52**
80-85	63	938	0.50	93	908	0.50	**1.52**
Total	326	1656		423	1559		**1.38**

25

Example

- Hypothetical case-control study: Standardization 2

	Cases			Controls			OR
Age	**Exp**	**Unexp**	**P(stratum)**	**Exp**	**Unexp**	**P(stratum)**	
50-55	256	30	0.14	264	22	0.14	**1.40**
55-60	224	62	0.14	242	44	0.14	**1.51**
60-65	178	108	0.14	203	83	0.14	**1.48**
65-70	122	164	0.14	150	136	0.14	**1.50**
70-75	72	214	0.14	94	192	0.14	**1.47**
75-80	37	249	0.14	53	233	0.14	**1.52**
80-85	18	268	0.14	27	259	0.14	**1.52**
Total	907	1095		1033	969		**1.29**

26

Final Comments

- It should be apparent that we can never truly estimate a “within individual” effect
- Sometimes we can get closer than others
  - Matching, adjusting for random effects
- It should also be apparent that we rarely truly estimate a “population” effect
  - We almost always have restrictions placed on our sampling scheme
- The major point that should be made is:
  - When using OR, recognize that variations in the adjustment variables and the way they are modeled can mean that precision variables will change our estimand

27

Propensity Scores

Confounding: Definition

- The association between a predictor of interest and the response variable is confounded by a third variable if

  – The third variable is associated with the response
    - causally (in truth)
    - in groups that are homogeneous with respect to the predictor of interest, and
    - not in the causal pathway of interest,

    AND

    - The third variable is associated with the predictor of interest in the sample.

29

Interpretation of Adjusted Parameters

- Intercept
  - Corresponds to a population with all modeled covariates equal to zero
  - Most often outside range of data; quite often impossible; very rarely of interest by itself
- Slope
  - A comparison between groups differing by 1 unit in corresponding covariate, but agreeing on all other modeled covariates
    - Sometimes impossible to use this definition when modeling interactions or complex curves
- Adjustment thus handles confounding by modeling the association between the confounder and the response

30

Propensity Scores

- Motivation: Handle confounding by modeling the association between the confounders and the POI
  - An analogy to RCT: Randomization ensures independence
  - We model the propensity for individuals to receive the “treatment” (or have the exposure)

- In RCT all subjects have same probability of assignment to arms

- In an observational study, subjects with different covariate values might have different “randomization ratios” (propensity scores)
  - But we assume that all subjects with similar covariates have the same propensity to exposure

31

General Methods

- Fit a model using
  - Exposure as response variable
  - All possible covariates as potential predictors
- Analysis models
  - Logistic regression
  - Other predictive models for a probability
- Model building for propensity score can be highly adaptive
- Take estimated probability of “assignment” to exposure as an adjustment predictor in model of outcome vs POI
  - Usually use very coarse grouping of propensity scores

32

Impact

- We can show that performing analyses adjusted for similar propensity scores guarantees that within each stratum the distribution of covariates is equivalent for exposed and unexposed
  - Hence, after adjustment for propensity scores all covariates are precision variables

- HOWEVER: Above is true providing we have correctly modeled all covariates that affect the propensity to treatment
  - All covariates
  - Highly flexible model
    - Overfitting is not an issue from bias standpoint
    - But we are just estimating the propensity, so need to worry somewhat about variability

33

Advantages / Disadvantages

- Relative to adjustment for covariates
  - We end up just adjusting for one propensity score
  - May better handle sparse data
    - We just have to consider about matched propensity, the matched distribution of covariates is then addressed
  - May better address population level inference
    - But there will be some adjustment for the most dominant variables in the propensity score
  - Exposure must be binary
  - OR coverage probabilities may be poor

34

Evaluation

- Cepeda et al. (2003) studied performance of the propensity score incohort studies with moderately rare events. Comparing covariate adjustment for all confounders in logistic regression to dummy variable adjustment for quintiles of the propensity score they found:
- In the settings they considered, when there were seven or fewer events per confounder, propensity score adjustment yielded OR estimates with less bias and more precision than ordinary confounder adjustment.
- With more events per confounder, ORs based on propensity score adjustment were more biased.

35

Evaluation (cont)

- Estimates based on covariate adjustment were biased by mis-specification of the model for association between a confounder and disease, but not biased by mis-specificationof the model for the association between confounder and exposure in propensity score adjustment.
- Tests based on propensity score adjustment had greater power than tests based on covariate adjustment. (Test level wasn't evaluated.)
- Estimates based on logistic regression covariate adjustment had much worse precision than estimates based on propensity score adjustment when there were fewer than eight outcome events per confounder.

36

Evaluation (cont)

- Confidence intervals based on propensity score adjustment had much better coverage probability when there were fewer than eight observations per confounder; logistic regression covariate adjustment had much better coverage probability when there were eight or more outcome events per confounder.
- With enough events (8 per confounder) covariate adjustment was better.

37

My View

- Other attempts at evaluation also appear in the literature, with similarly mixed reviews
- Propensity scores appear attractive, but do not really seem to work well enough, if you ask me

38

Missing Data

39

Missing Data: Ideal

"Just say no."

(Nancy Reagan)

40

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

41

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
- Problems in clinical trials
  - Poorly defined treatment indications
  - Poor training of investigators
  - True scientific dilemmas exist, but they are in the minority
- Problems in observational studies
  - Poorly defined protocol in designed studies
  - Convenience samples
    - Indication bias for available data

42

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
- Study investigators can
  - Be lax in contacting patients, scheduling visits
  - Be lax in data collection, data management
  - Encourage patients to withdraw inappropriately

43

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 possible patients
  - In observational studies: What is time 0?

44

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
  - Ethics: Effect of smoking on lung function in children
  - Physiology: Effect of REM sleep deprivation on cardiovascular parameters

45

Missing Data: Real Life

"Missing data happens"

(Bumper Sticker- rough translation)

46

Analyses 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?

47

Missing Data: Sad Facts of Life

"Bloodsuckers hide beneath my bed"

"Eyepennies",
Mark Linkous (Sparklehorse)

48

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.)

49

Problem by Role of Data

- Eligibility data
  - "Misclassification"
  - Affects generalizability
- 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

50

Mechanisms for Missing Data

- Owing to (improper) definition of estimand
  - Competing risks, etc.
- Broad categories of “true” missing data
  - Biased sampling of subjects
  - Withdrawal of consent
  - Loss to follow-up
  - Sloppy data collection (repurposing data?)
- With withdrawal of consent and loss to follow-up need to consider
  - Competing risks, indication bias
- With sloppy data collection need to consider biases

51

Scientific Estimands in RCT

- 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?

52

Impact of Estimand on Design / Analysis

- The patients who are “relevant” differ according to the estimand of interest
- The primary goal should be to devise an experiment that only includes patients who are relevant to the estimand
  - This is often difficult
  - It may mean using more than one study, answering different aspects of the scientific profile in different studies
- Science is adversarial
  - When have we demonstrated safety, efficacy, effectiveness to meet reasonable doubt?

53

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

54

MCAR

- 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

55

MAR

- 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

56

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

57

MNAR

- 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

58

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

59

Bottom Line

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

-Aretha Franklin, "Think"

60

Most Important Strategies

1. Prevention
2. Prevention
3. Prevention
4. Planning for missing data
   - Anticipating where it will occur
   - Developing an analysis approach for a primary analysis
   - Developing a strategy for sensitivity analyses

61

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)

62

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

63

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
- Study 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

64

(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 (MNAR)
  - 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

65

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

66

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

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

67

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

68

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

69

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

70

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

71

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

72

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

73

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

74

MAR, MNAR Methods

- Joint distribution of observed, unobserved, and indicator of missingess based on covariates
- Selection models
  - Model distribution of all outcomes based on covariates
  - Model probability of missing based on covariates and outcome
  - If MAR: Probability of being missing depends only on observed data
- Pattern mixture models
  - Model distribution of missing observations based on observed data, indicator of missingness, covariates
  - Model distribution of observed outcomes based on indicator of missingess, covariates
  - Model distribution of missingness based on covariates
  - If MAR: Distribution of missing observations depends only on observed data

75

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

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

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

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

76

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

77

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

78

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

79

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

80

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

81

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

82

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

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)

84

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

85

Example: Pattern Mixture Model

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

$[M \mid X]$ 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

86

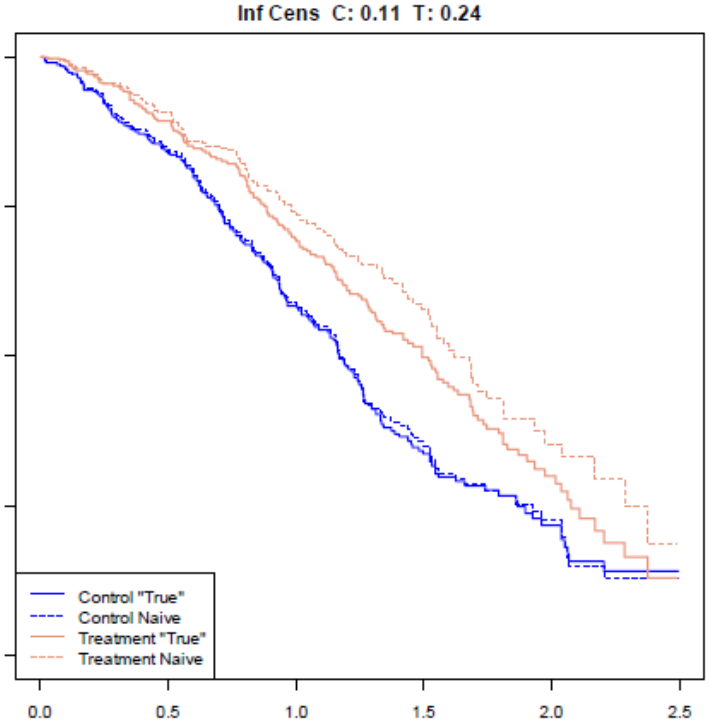
Example: Summary

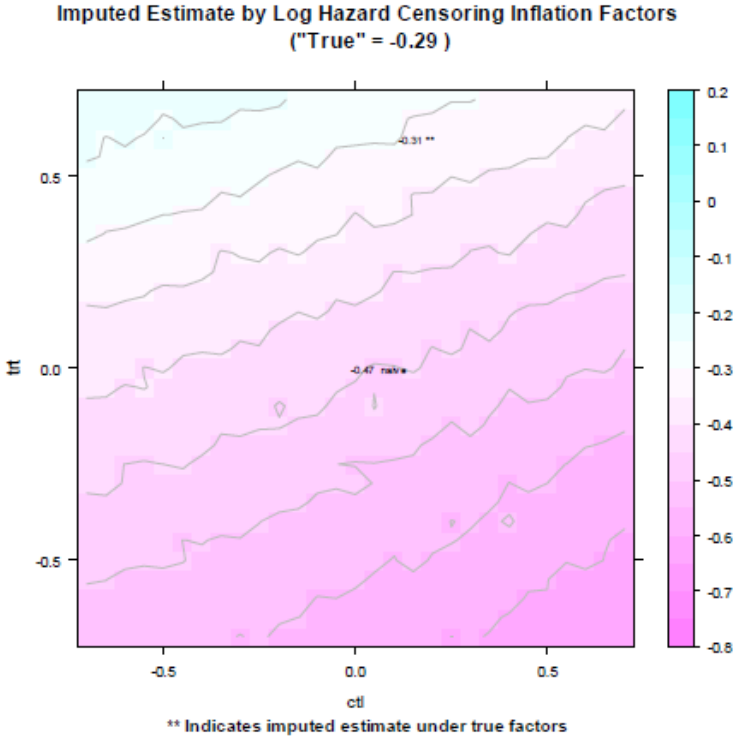
- Time to event analysis from RCT with
  - Administrative censoring
  - Potentially informative censoring
- Primary analysis: A standard KM or PH analysis (MAR)
  - Assumes imputation of missing data from all subjects still at risk
- Explore sensitivity to change in hazard at time of informative censoring (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

87

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

Example: Contour Plots

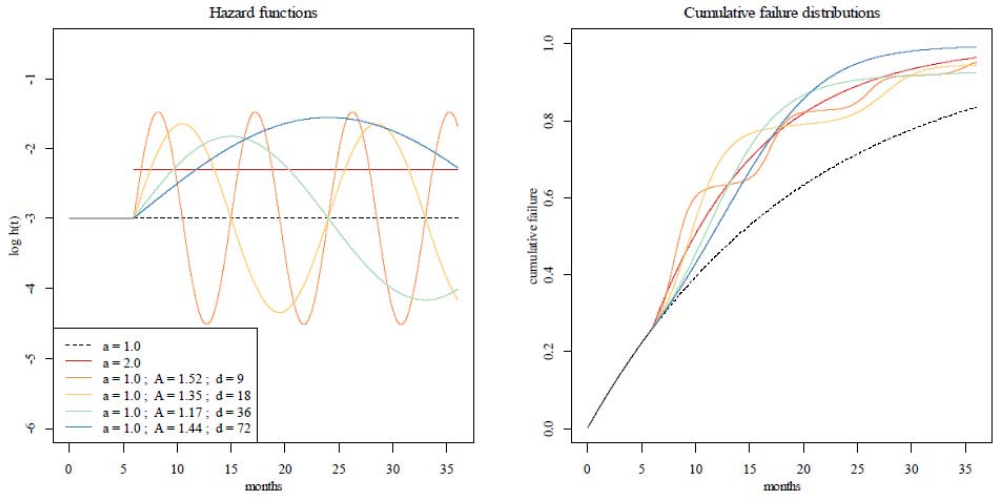




	survival	informative censoring	log hazard inflation factor
control:	Weibull(B=2.0, k=1.5)	Exponential(B=6)	0.1
treatment:	Weibull(B=2.6, k=1.5)	Exponential(B=4)	0.6

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



89

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

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

90

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

91

Difference in Means

- For treatment arm *k*

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

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

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

$$\mu_{k,mis} = \mu_{k,obs} + \delta_k \qquad \sigma^2_{k,mis} = \sigma^2_{k,obs}\omega_k$$

$$\mu_{k,all} = \pi_k\delta_k + \mu_{k,obs}$$

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

92

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

93