Analyzing Sensitivity to Data Missing Not At Random (MNAR): A Framework for Design, Analysis, and Reporting

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

11th Annual Conference on Statistical Issues in Clinical Trials
Center for Clinical Epidemiology and Biostatistics (CCEB)
University of Pennsylvania
,April 18, 2018

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

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

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

Estimands

- Clinical issues:
  - The indication
    - Disease, Population, Treatment, Outcome
  - Clinical importance of distributional summary measure
  - Clinical importance of stratification

- Statistical issues:
  - Summarizing the outcome distribution
    - Mean, geometric mean, median of continuous data
    - Proportion or odds above threshold
    - (Time averaged) hazard ratio of censored data
  - Covariate adjustment (precision)
  - Per randomization analyses

Scientific Efficacy / Safety Estimands

- 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 don't follow protocol may be 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
  - Patients with competing risk that prevents measurement

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

- Sometimes, however, a counterfactual estimand is of greatest scientific interest
  - In these cases, all results are subjective

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
  - A binary endpoint in absence of competing risks
- Issues
  - Some patients will die of their underlying disease
  - How do we record bleeding incidence in such patients?

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

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

Possible Estimands: Cause Specific Hzrd

- Effect of treatment on hazard, cumulative incidence of bleeding
  - Estimate cause specific hazard in presence of competing risks
    - Fine & Gray
  - Assumes incidence of competing risks remains unchanged
    - Though does allow that competing risk might be informative

- Caveats
  - Still need to evaluate overall survival, as treatment could increase rate of death after initial bleeding
  - Untestable hypothesis about unchanging competing risks

Possible Estimands: MAR

- Effect of treatment if all would survive until risk of bleeding is gone
  - Assume missing at random
  - Censor subjects who die
    - Use KM to estimate proportion bleeding within 30 days
  - 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)

Missing Data: Ideal

"Just say no."

(Nancy Reagan)

Prevention of Missing Data

- The most important issue
  - Through diligence and hard work, experienced RCT investigators have managed to greatly reduce prevalence of missing data
  - It can be done

- Recommendations 2 – 8
  - Design, conduct, analysis
  - Investigators, subjects

- But not discussed further in this talk

Missing Data: Real Life

"Missing data happens"

(Bumper Sticker- rough translation)

Missing Data: Sad Facts of Life

"Bloodsuckers hide beneath my bed"

"Eyepennies",
Mark Linkous (Sparklehorse)

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

- But missingness is a post-randomization variable
  - Conditioning on missingness is necessarily fraught with peril

Recommendations for Analysis

- #9: Statistical methods for handling missing data should be specified by clinical trial sponsors in study protocols.
  - Assumptions should be understood by clinicians.
- #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.
- #12: The primary analysis of the data from a RCT should account for the uncertainty attributable to missing data so that type I error rates and associated CI are valid
- #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

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
  - Necessarily subjective and untestable
- 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

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

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

(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

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

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.

Types of assumptions

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

Guiding Principles

- Maintain RCT / regulatory standards
  - Per randomization analyses
  - Pre-specification of analysis methods
  - ***Try for analysis obtained if no missing data occurred***
    - Estimands, estimates, inferential methods

- Avoid unnecessary assumptions
  - Avoid (semi)parametric assumptions across treatment arms

- Minimize dimensionality of untestable assumptions
  - "Don't make up 14 numbers to arrive at 1 made up number"

- "Tipping point" analyses to identify limits of evidence

Primary Analysis

- Addresses missing data: typically MAR, but perhaps MNAR
  - E.g., presume 30 day survival following hospital discharge
- Relatively distribution-free probability model
  - Distribution-free interpretation of common analysis methods
- Preserves weighting used in absence of missing data
- Analysis method treats MAR as special case of MNAR
- All of above easily addressed using multiple imputation
  - Also relatively transparent assumptions

Framework for Sensitivity Analyses

- Separate MNAR parameters for each treatment arm
- Assume average MNAR ”effect” across modeled baseline covariates
- Tipping point analyses (using contour plots?) for p values, CI
  - Based on minimal parameterization of MNAR
- Descriptive analyses of missing data
  - Exploration of stratum specific MNAR ”effects” that correspond to average MNAR ”effects” reported in tipping point analyses

Reporting Analyses

- 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
  - Patient disposition: d/c drug, d/c assessments, withdraw consent
    - Perhaps time to d/c or withdrawal by important predictors
  - Any interim measures of primary, secondary, safety outcomes

- Provide primary analysis inference and summarize tipping points
  - Summarize how more detailed assumptions might translate to parameterizations used in sensitivity analyses

Example: Methotrexate in PBC

- *Combes, et al, Hepatology, 2005*
  - Issues with stopping study drug, clinic visits, or all follow-up

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

Example: Methotrexate in PBC

- Premature discontinuation of clinic visits
  - Missing time to subclinical progression of hyperbilirubinemia

Table 6. Actuarial (Kaplan-Meier) estimates of the cumulative probability of a patient discontinuing clinic visits prior to the end of the interventional phase of the study and prior to experiencing the primary endpoint of liver transplant or death.

Years post randomization	Probability of Early Discontinuation of Clinic Visits (95% CI)		
	MTX arm	Placebo arm	P value
1	2.3% (0.0% - 4.8%)	3.0% (0.1% - 5.9%)	0.72
2	2.3% (0.0% - 4.8%)	6.0% (1.9% - 10.0%)	0.13
3	3.1% (0.1% - 6.0%)	7.6% (2.9% - 12.0%)	0.10
4	3.1% (0.1% - 6.0%)	7.6% (2.9% - 12.0%)	0.10
5	3.9% (0.5% - 7.2%)	9.1% (4.1% - 13.9%)	0.09
6	4.8% (1.0% - 8.6%)	14.0% (7.5% - 20.0%)	0.01
7	6.0% (1.5% - 10.3%)	17.5% (10.0% - 24.4%)	0.01
8	8.3% (1.9% - 14.4%)	19.4% (11.1% - 26.9%)	0.03

- Compare those with / without premature discontinuation
  - Baseline characteristics
  - Most recent bilirubin levels
  - Trends in bilirubin levels
  - (But interested in patterns of bilirubin after discontinuation)

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

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)

Potential Methods

- Many proposals varying in
  - Analysis models: (semi)parametric vs nonparametric
  - Modeling of missingness: assumptions, predictive markers
  - Goals: estimation, inference
    - Fisher, Kanarek, *Rel and Biometry, Stat Analysis of Lifelength*,1974.
    - Lagakos, Williams, *Biometrika*, 1978.
    - Slud, Rubinstein, *Biometrika*, 1983.
    - Klein, Moeschberger, *Biometrics*, 1988.
    - Robins, Rotnitzky, *Aids Epi Meth*, 1992.
    - Robins, *Proc Biopharm, ASA*, 1993.
    - Zheng, Klein, *Biometrika*, 1995.
    - Scharfstein, Robins, Eddings, Rotnitzky, *Biometrics*, 2001.
    - Scharfstein, Robins, *Biometrika*, 2002.
    - Siannis, Copas, Lu, *Biostatistics*, 2005.
    - Zhang, Heitjan, *Clin Trials*, 2005.
    - Rotnitzky, Farall, Bergesio, Scharfstein, *JRSS Series B*, 2007.
    - Liu, Heitjan, *StatMed*, 2011.

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 typically accept that

Example: Approach

- We use a pattern mixture model to reproduce an analysis that would only have administrative censoring
  - We presume we were happy with interpretation of HR in presence of administrative censoring

- The accrual time and data analysis time is known for all subjects
  - We thus compute an administrative censoring time

- We will ultimately impute the minimum of a survival time and the administrative censoring time

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

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 impute 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

Imputation Probability Model

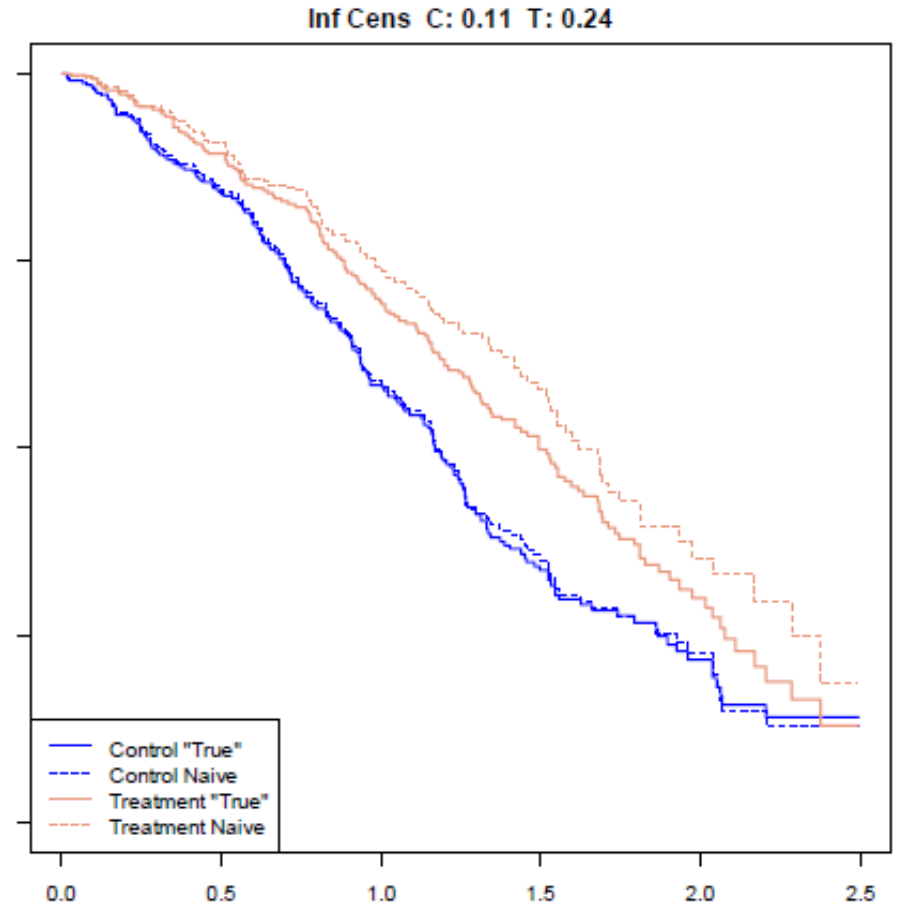
- Distinguish subjects with
  - Observed event : $(Y_{ki}, \delta_{ki} = 1);\ M_{ki}(t) \equiv 0$
  - Administrative censoring : $(Y_{ki}, \delta_{ki} = 0);\ M_{ki}(t) \equiv 0$
  - Potential informative censoring: $(Y_{ki}, \delta_{ki} = 0);\ M_{ki}(t) = \mathbf{1}_{[t>Y_{ki}]}$

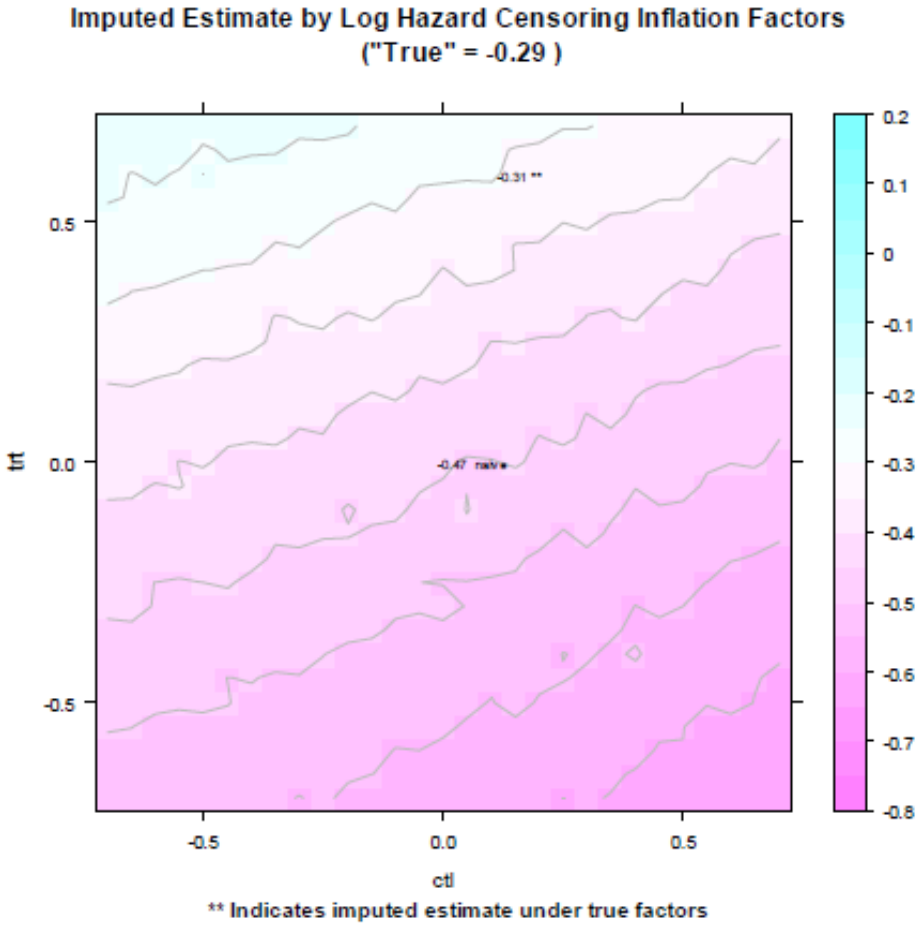
- Presume conditional hazard model for each arm using PH

$$\lambda_{ki}\left(t \mid M_{ki}(t)\right) = \lambda_{k0}(t) \times \Delta^{M_{ki}(t)}$$

- Note
  - $\lambda_{k0}(t)$ is estimable using complete data from all subjects
  - Impute administratively censored event times using presumed Δ
  - Untestable use of PH here is motivated by dimension reduction

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

Interpretation

- Primary analysis might be based on MAR: $\Delta = 0$ for both arms

- Tipping point might identify $\Delta's$ when p value or CI bounds meet regulatory burden of proof
  - E.g., for pivotal or noninferiority trials

- Adequacy of tipping point is then judged subjectively by considering for each treatment arm the patterns missingness by
  - Baseline characteristics: age, sex, concomitant disease
  - Putative reasons for dropout: early response, lack of response, AEs, general health status

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

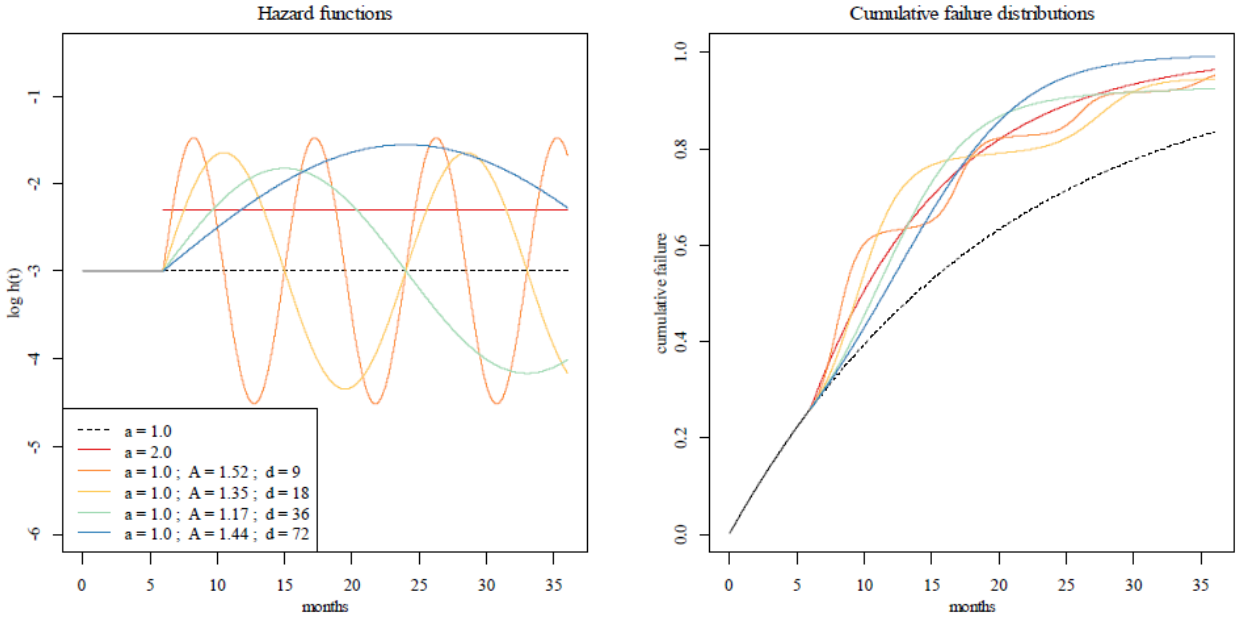


Figure 14. Sinusoidal perturbations equivalent to $\alpha_t = 2.0$

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

(Reference: 2012 MS Thesis, Eric Meier – www.RCTdesign.org)

Extension to Other Settings - 1

- Adjusted time to event analyses
  - Using estimated hazards from (possibly stratified) 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

Extension to Other Settings - 1

- Adjusted time to event analyses
  - Using estimated hazards from (possibly stratified) 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

Extension to Other Settings - 2

- Comparison of means of continuous data
  - Distribution of data to use in imputations
    - Probably not too important owing to central limit theorem
      - Ultimately only need means and SDs
    - May be able to use perturbations to resampling
    - Within strata or by residual error distribution when adjusting for covariates
  - Two MNAR parameters on each treatment arm
    - Difference of means
    - Difference (ratio) of standard deviations
    - Consider mean-variance relationships?

Difference in Means

- For treatment arm *k*

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

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

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

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

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

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

Extension to Other Settings - 3

- Longitudinal and other correlated outcome data
  - Imputation for the primary (MAR?) analysis will typically be more complicated
    - Imputations might condition on partially observed data
  - Sensitivity for MNAR is probably best conducted assuming a single average effect due to missingness over both time and clusters
  - Then, having described impact of a hypothesized average effect, explore how more complicated assumptions about time- or stratum-varying effects of missingness might correspond to an average effect

Further Dimension Reduction

- I have parameterized the effect of missingness separately for each treatment arm

- This is only important if there is no (semi)parametric model that is valid across treatment and control groups
  - Under the strong null hypothesis such separate treatment is not necessary, but under alternatives it may be more important

- If there are not large departures from a (semi)parametric model, it is likely sufficient to report contrasts across the MNAR parameters
  - Data analyst can explore the richer parameterization and only report the lower dimension tipping point if relatively constant
  - One dimension for odds or hazards; two dimensions for means

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
  - Frequentist vs Bayesian vs minimax
  - How many researchers have we convinced vs the ”average” researcher
- 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

Bottom Line

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

Really Bottom Line

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

-Aretha Franklin, "Think"