Using Multiple Imputation, Survival Analysis, And Propensity Score Analysis In Cancer Data With A Large Amount Of Missing Data

Master's Thesis

Nathan Berliner [1]

[1]Department of Statistics
Rice University

11/30/2015

Outline

Outline

In an ideal world

- We would have a large dataset
  - That was obtained from an RCT
  - That would help answer a clearly defined question
  - That had all the covariates of scientific interest
  - That contained no missing data

In Reality

- RCT's are expensive and often unethical
  - We often get retrospective observational data
  - Pulled from a database or historical records
- The questions we have may not be answerable from the data on hand
  - The data obtained often doesn't support the original question in mind
- The covariates collected are out of our control
  - Since often no control of experiment, no control over what is collected
- Lots of missing data
  - Since no control over how the data is collected, we can't guarantee that everything is collected
  - This issue is seemingly omnipresent in all types of data collection

Is This a Problem?

- Without an RCT, we can't be sure if differences in treatments is due to the treatment or something else
- Omitting important factors may bias our results
- With missing data, we will be throwing away data and biasing our results

The Solution

This thesis aims to fix some of these problems

- Fill in missing data via multiple imputation
- Create meaningful analytical models via survival analysis
- Get a causal interpretation from observational data

Motivation

- This thesis is motivated by cancer survival data with moderate missingness
- We will build the theory for dealing with this sitation
- And then apply it to a cancer data set

Abstract

In this thesis, multiple imputation, survival analysis, and propensity score analysis are combined in order to answer questions about cancer data with moderate missingness. While each of these fields have been studied individually, there has been little work and analysis on using the three in trio. Starting with an incomplete dataset, we aim to impute the missing data, run survival analysis on each of the imputed datasets, and then do propensity score analysis to observe causal effects. Along the way, many theoretical and analytical decisions are made. I explain why each decision is made, and offer ample evidence for the other choices such that the interested reader may implement the methods if they so choose. I apply the methodology to a cancer survival dataset in a case study, but the methods used are general, and could be adapted for any type of data.

Outline

What is missing data

- Missing data happens when we intend to collect a piece of data but don't actually get it
- Historical approaches
  - Complete Case analysis: Throw away any record that is not complete
  - Available Case analysis: Use records so long as they are complete for the specific analysis in question

Imputation

Definition

The English verb "to impute" comes from the Latin imputo, which means to reckon, attribute, make account of, charge, ascribe. [1]

- In the 1930's, Allan, Wishart,and Yates laid framework for missing data
  - Idea: Fill in the missing value, deduct degrees of freedom to account for it
  - Issue: Dogmatic, and variance can't be estimates correctly

Multiple Imputation

Throughout the 70's and 80's Donald Rubin worked to improve on this

- Instead of imputing one value, lets impute it $m \geq 2$ times
- Draw the values from the missing datas posterior distibution given the observed data and the process that generated the missing data

This idea is called Multiple Imputation (MI) and was formalized in 1987 [2]. It is the gold standard method for missing data currently.

How does MI work?

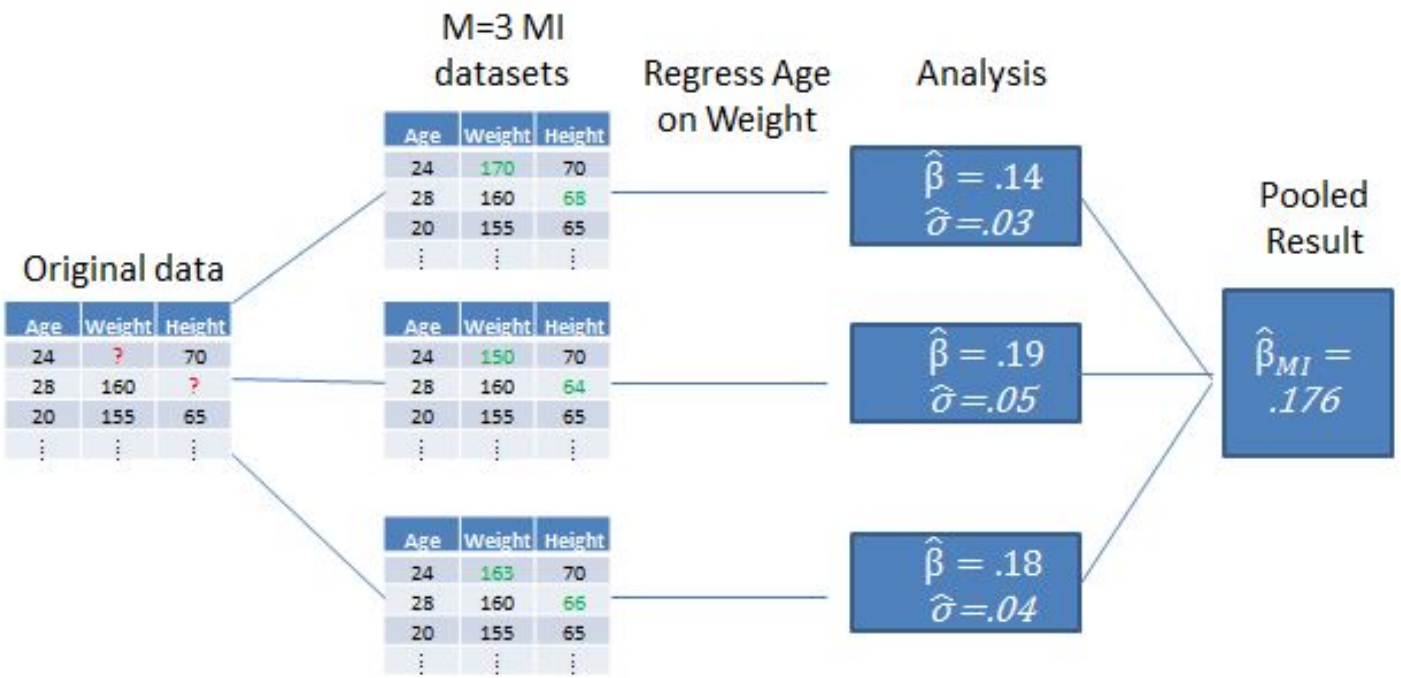


Figure: Visualization of MI data

Missingness is displayed by ?'s and the imputed data is shown as #'s. We then regress age on weight, get the results from the individual datasets, and then pool them together.

Outline

Survival Analysis

Survival Analysis

Survival analysis is a field of statistics concerned with analyzing time to event data, often in the face of censoring or truncation.

Examples:

- The survival of patients after a liver transplant in a hospital
  - Complications: study ending, patients die before study starts, subject moves away
- The time until a child learns a new task
  - Complications: refuse participation, move away, don't recall the exact time they learned, already learned the task

Kaplan-Meier Estimator

- The survival function $S(t) = P(T > t) = \int_t^\infty f(u)du$ is estimated by the nonparametric Kaplan-Meier Estimator

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

- n_i is the number of subject in the risk set at time t_i
- d_i is the number of deaths at time t_i

Log rank test

The log rank test compares two survival curves to see if from the same distribution

$$\frac{\sum_{j=1}^{J} w_j(O_{1j} - E_{1j})}{\sqrt{\sum_{j=1}^{j} w_j^2 V_j}} \sim N(0,1)$$

- Where w_j is the weight of each observation (must be ≥ 0, we will set all to be 1)
- $N_j = N_{1j} + N_{2j}$ is the number at risk at time j (composed from deaths in each group)
- $O_j = O_{1j} + O_{2j}$ is the observed number of deaths at time j (composed from the observed deaths in each group)
- $E_{1j} = \frac{O_j N_{1j}}{N_j}$
- $V_j = \frac{O_j(N_{1j}/N_j)(1-N_{1j}/N_j)(N_j-O_j)}{N_j-1}$

Cox Regression

- Hazard is the instantaneous rate of event given that you have survived until time t, given by

$$h(t) = \lim_{\Delta t \to 0+} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t}$$

- Cox regression models hazard by

$$h(t|Z) = \underbrace{h_0(t)}_{\text{time}} * \underbrace{exp(\sum_{k=1}^{p} \beta_k Z_k)}_{\text{covariates}}$$

- Where $h_0(t)$ is the baseline hazard
- Z_k is the k^{th} covariate
- β_k's are found by maximizing the partial likelihood function

The covariates act to multiply the hazard function.

Outline

Causal Analysis

Suppose we have a new drug we want to test to see how efficacious it is.

- We would like to be able to say "The drug leads to better health"
  - But need an RCT to say this
  - We only have observational data
  - Thus differences could be attributed to the drug or confounding
    - e.g more Healthier people at baseline were much more likely to take the drug

Idea: Try to balance the covariates to reduce the effects of confounding so the two groups seem identical at baseline

Counterfactual Model

- Suppose that for or every person, there are two potential outcomes
  - $Y_i(0)$ - The outcome if they had taken the control, $Z = 0$
  - $Y_i(1)$ - The outcome if they had taken the treatment, $Z = 1$
- Obviously, we only observe one. *The fundamental problem of causal inference*
- If we could observe both, then we could observe the causal effects for each person
- Estimands of interest:
  - Average Treatment Effect (ATE) $E[Y_i(1) - Y_i(0)]$. The effect of moving entire population from treated to untreated
  - Average treatment effect for the treated (ATT) $E[Y_i(1) - Y_i(0)|Z = 1]$. The average treatment effect for those actually treated
- In survival, the ATE is the difference in survival time

Propensity scores

Definition

The propensity score is the probability that the subject received the treatment given the subjects *pretreatment* covariates. It is computed using the patient's baseline (pretreatment) information [3]

- Defined as $e_i(x) = P(Z_i = 1|X_i)$
- Assume that the covariates play a role in how the subject chose treatment
- If we assume that $(Y(0), Y(1)) \perp T|X \implies (Y(0), Y(1)) \perp T|e(X)$
- Controling for propensity score will make groups seem indistinguishable
- Thus, we may treat it as if it were an RCT

Common Propensity Score Methods

- Matching: Match treatment and controls on their propensity score, calculate ATE
- Stratification: Stratify on propensity score, weight and combine ATE in each strate
- Weighting: Weight each observation by the inverse of its propensity score, and then calculate ATE

Outline

A path with many options

- There are many different options to choose
- I explain my choices but dicuss other options
- Goal: Be clear so other researchers can adapt my methodology to their problems

MI primer

- MI forms the base of this thesis
- There are lots of different ways to impute
- As long as we can impute valid imputations, we can analyze them
- Poor imputation leads to poor results (bias, variability, loss in power)

MI Notation

- Y is our whole dataset. It will have i rows and j columns. Some of the covariates in the dataset will be completely observed, and others will have missingness.
- Y_j is a specific column of Y. Y_j is composed as $Y_j = (Y_{j,obs}, Y_{j,mis})$, where
  - $Y_{j,obs}$ is the data we have observed for covariate j
  - $Y_{j,mis}$ is the missing data covariate j
- Y_{obs} is all of the data that we have observed
- Y_{mis} is all the data that we have not observed
- R is a binary matrix the same size as Y where a 1 indicates we observed the data, and 0 means it is missing
- ψ is a vector of parameters for the missing data model.
- The missing data model is given as $p(R|Y_{obs}, Y_{mis}, \psi)$
- θ is a vector of the parameters for the full model of Y

MI Concepts

- Ignorability

$$p(Y_{mis}|Y_{obs}, R) = p(Y_{mis}|Y_{obs})$$

That is, we may "ignore" the R. The probability of the data being missing does not depend on how the data is missing. Equivalently, we may write this as

$$p(Y_{mis}|Y_{obs}, R = 1) = p(Y_{mis}|Y_{obs}, R = 0)$$

- Non ignorability:

$$p(Y_{mis}|Y_{obs}, R = 1) \neq p(Y_{mis}|Y_{obs}, R = 0)$$

So we must take into account the missing data structure for imputation.

Missing data Mechanisms

Now, we may discuss the three main types of missing data mechanisms.

- MCAR: Missing completely at random:

$$P(R = 0|Y_{obs}, Y_{mis}, \psi) = P(R = 0|\psi)$$

  - The missingness in the data is not at all related to any of the data that we do or don't have
- MAR: Missing at random:

$$p(R = 0|Y_{obs}, Y_{mis}, \psi) = p(R = 0|Y_{obs}, \psi)$$

  - The missingness we have is related to something in the data
- MNAR: Missing not at random:

$$p(R = 0|Y_{obs}, Y_{mis}, \psi)$$

  does not simplify
  - and the missingness depends on data that we have as well as have not collected

Joint Modelling

- Assume ignorable MAR missing data mechanism
- Missing data imputed by sampling from a user specified distribution
- A lot of theory developed for Normal, not much else
  - Normal imputation has been shown to perform well, even under non normality [4]
- Idea: pull imputations by missing data row pattern

JM pseudocode

1. Sort the rows of Y into S missing data patterns $Y_{[s]}, s = 1, \ldots, S$.
2. Initialize $\theta^0 = (\mu^0, \Sigma^0)$ by a reasonable starting value.
3. Repeat for $t = 1, \ldots, T$:
4. Repeat for $s = 1, \ldots, S$:
5. Calculate parameters $\dot{\phi}_s = \text{SWP}(\dot{\theta}^{t-1}, s)$ by sweeping the predictors of pattern s out of $\dot{\theta}^{t-1}$.
6. Calculate p_s as the number missing data in pattern s. Calculate $o_s = p - p_s$.
7. Calculate the Choleski decomposition C_s of the $p_s \times p_s$ submatrix of $\dot{\phi}_s$ corresponding to the missing data in pattern s.
8. Draw a random vector $z \sim N(0, 1)$ of length p_s.
9. Take $\dot{\beta}_s$ as the $o_s \times p_s$ submatrix of $\dot{\phi}_s$ of regression weights.
10. Calculate imputations $\dot{Y}^t_{[s]} = Y^{\text{obs}}_{[s]}\dot{\beta}_s + C'_s z$, where $Y^{\text{obs}}_{[s]}$ is the observed data in pattern s.
11. End repeat s.
12. Draw $\dot{\theta}^t = (\dot{\mu}, \dot{\Sigma})$ from the normal inverted-Wishart distribution according to Schafer (1997, p. 184).
13. End repeat t.

JM Pros and Cons

Pros

- Fast
- Easy to derive posteriors with common distributions

Cons

- Inflexible
- Limited to known distributions
- How to deal with mixed categorical and continous missing data

Full Conditional Specification

- Assume MAR missing data mechanism
- Missing data is imputed iteratively on a variable by variable basis
- Requires no distributional assumptions
- Idea: Specify k one dimensional models to impute on the missing data columns

FCS Algorithm

1. Specify an imputation model $P(Y_j^{\text{mis}}|Y_j^{\text{obs}}, Y_{-j}, R)$ for variable Y_j with $j = 1, \ldots, p$.
2. For each j, fill in starting imputations $\dot{Y}_j^0$ by random draws from Y_j^{obs}.
3. Repeat for $t = 1, \ldots, T$:
4. Repeat for $j = 1, \ldots, p$:
5. Define $\dot{Y}_{-j}^t = (\dot{Y}_1^t, \ldots, \dot{Y}_{j-1}^t, \dot{Y}_{j+1}^{t-1}, \ldots, \dot{Y}_p^{t-1})$ as the currently complete data except Y_j.
6. Draw $\dot{\phi}_j^t \sim P(\phi_j^t|Y_j^{\text{obs}}, \dot{Y}_{-j}^t, R)$.
7. Draw imputations $\dot{Y}_j^t \sim P(Y_j^{\text{mis}}|Y_j^{\text{obs}}, \dot{Y}_{-j}^t, R, \dot{\phi}_j^t)$.
8. End repeat j.
9. End repeat t.

FCS Pros and Cons

Pros

- Flexible
- Easy to specify models
- Handles mixed continous categorical

Cons

- No guarantee that full conditionals are compatible
- Slow
- Gets much harder as sample size increases to specify models

Decision

- Both are not as good as having complete data
- Cancer and survival data present challenges for JM
- FCS offers us the most ease and flexibility

Setting Up The Model

- Specify the models
- Specify the predictors for each model
- Determine number of iterations and datasets to impute
  - This is a topic of hot debate
  - Old literature suggested 5 imputations, 5 iterations, but more now

Checking The Imputations

Convergence

- Chains should be freely intermingled with no pattern
- Convergence when variance between chains is no larger than variance withing each chain
- Formal tests like Gelman/Rubin $\hat{R}$ proposed to check convergence

Validation

- "Does the data look like it could have come from real data had it not been missing"?
  - Requires intimate knowledge of the data
- Graphical checks
  - Density plots
  - Conditional scatter plots
  - Box and whisker
  - etc.

Pooling

- We now have m imputed datasets
- Run the analysis on each of the m complete datasets
- But we want one analysis, not m

Pooling Notation

Let

- $\hat{Q}_i$ be the scientific estimand from the i^{th} MI dataset
- U_i be the variance-covariance matrix of the i^{th} MI estimand

Then

- The MI estimate is given by

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i$$

- The MI "within" variance is given by

$$\bar{U} = \frac{1}{m}\sum_{i=1}^{m}U_i$$

- the MI "between" variance is given by

$$B = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_i - \bar{Q})(\hat{Q}_i - \bar{Q})'$$

Rubin's Rules

- Total variance given by [2]

$$T = \bar{U} + B + \frac{B}{m}$$

- To do inference, assume that the complete sample estimate $\hat{Q} \sim N(Q, U)$, where U is the variance-covariance of $(Q - \hat{Q})$
- Since true T is not known, then

$$\frac{Q - \bar{Q}}{\sqrt{T}} \sim t_\nu$$

- ν is given by [5]

$$\nu = \frac{\nu_{old}\nu_{obs}}{\nu_{old} + \nu_{obs}}$$

- Where $\nu_{obs} = \frac{\nu_{com}+1}{\nu_{com}+3}\nu_{com}(1 - \frac{B+B/m}{T})$
- ν_{com} is the hypothetical complete sample degrees of freedom
- $\nu_{old} = \frac{m-1}{(\frac{B+B/m}{T})^2}$

The Stack Method

- Rubin's Rules work well, but not always
  - Ex: partitioning the MI data on an imputed variable
  - Taking the average is not a good idea
- Solution: Stack the MI datasets on top of eachother to get one huge dataset
  - Will get unbiassed results
  - But sample size is falsley inflated, thus cannot trust variance

Outline

Kaplan-Meier in the MI Setting

- Clearly define the population, groups, and events of interest
- Ensure that we have noninformative censoring
- Issue: Kaplan-Meier is not normally distributed
  - Solution: Complimentary log log transformation, pool [6]
- Issue: Imputations leave one KM curve much shorter than the rest
  - Solution 1: Truncate all curves at the lowest time
  - Solution 2: Extend the curves out to the longest time
  - Solution 3: Use the stacked method
- Algorithm: Pool the complimentary log log of the Kaplan-Meier curve, get estimates, back transform

Median Survival Time

- Want a measure of central tendency
  - Survival distributions often skewed, so mean is poor choice
- Median: smallest time such that $S(t) \leq .5$
- Algorithm: Take MI Kaplan-Meier curve, observe first time it goes below 50%
- Confidence interval at median: Take the median of the upper and lower confidence bands

Log Rank Test

- Idea: Combine log rank tests from each MI dataset
  - Problem: Wastes information and is unstable [6]
  - Idea: Calculate log rank from the MI Kaplan-Meier curve
  - Problem: Risk set and deaths no longer meaningful
- Solution: Under no tied times, Cox Regression on a treatment is equivalent to the log rank test
  - And very similar under tied times
- Idea: Derive log rank test from Cox model
  - Pooling LRT and Score test is unstable [6]
  - Wald test is asymptotically equivalent
- Final Solution: Run the Wald test on Cox model as an approximation

Cox Model in the MI Setting

- Goal: To get a "baseline" Cox model, then add treatment variables
- Need to check for proportional hazards assumption
  - Problem: MI cox model doesn't have residuals
  - Solution: Check assumptions on stacked dataset or each MI dataset individually
- Cox model is normally distributed, use Rubin's Rules to pool
- Add treatment covariates, rerun models, pool

Outline

Propensity Score

- How to use the propensity score
  - Matching, weighting, covariate adjustment, stratification
- Generating the propensity score
  - Logistic regression, Probit regression, CART, etc.
- Verifying balance

Propensity Score in MI Setting

- Mitra and Reiter propose two methods [7]
- Within: Work with propensity score on each of the m MI datasets
- Across: Average propensity scores across the m datasets and then analyze
- Which to use: Dependent on your data

Outline

Data Explanation

- 1514 MD Anderson patients who had brain mets from breast cancer
- 90 covariates
  - Missingness from 0 to 65%

Type	Example
Subject data	Age range, race, date of birth
Cancer data	TNM staging, type, receptor status
Pre brain mets data	Treatment types
Post brain mets clinical observations	Seizures, headache, nausea
Post brain mets data	Treatment type, type of brain mets
Survival data	Survival time after brain mets, censoring indicator

Table: Data Categories and Examples

Important Covariates

Name	Percent Missing	Meaning
hrher2	5	Categorical variable: The hormonal receptor and HER2 receptor status of the subject
agebrainmet	0	Indicator: Age greater or less than 60 at time of brain mets
timedx	1	Indicator: Time (years) from breast cancer diagnosis to brain mets diagnosis greater or less than 6 years
site5	1	Indicator: First metastasis was to brain
race2	0	Categorical: White, Black, Hispanic, other
priorn	0	Indicator: Number of prior treatments in metastatic setting before brain mets
braintype	4	Categorical: Single, multiple, Leptomeningeal disease
controlled	12	Indicator: Extracranial progression of brain mets
capeothno	18	Indicator: Capecitabine, other, or no chemotherapeutic treatment. Treatment variable 1
lapatrasno	18	Indicator: Lapatinib, Trastuzumab, or no HER2 treatment. Treatment variable 2
os	0	Overall survival (months)
dead	0	Indicator: death indicator
her2	10	Indicator: HER2 receptor status

Table: Table of important covariates to be used in the analysis

Visualization of Missingness

Figure: Visualization of missingness in the cancer dataset

Imputation

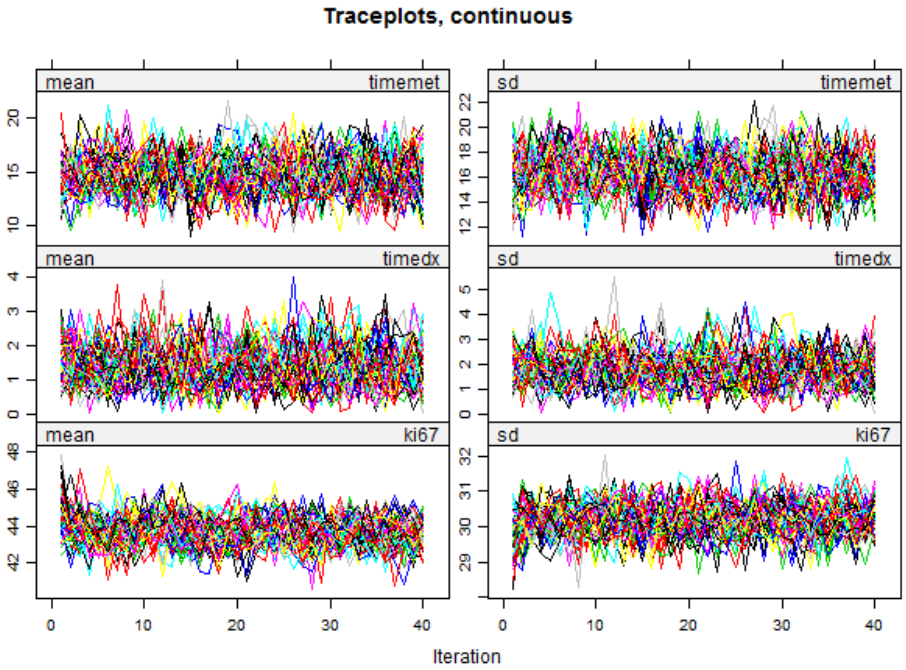
- MAR assumption seems reasonable
- FCS over JM due to nature of data
- Need to set up models and predictors
- Check for convergence and validity

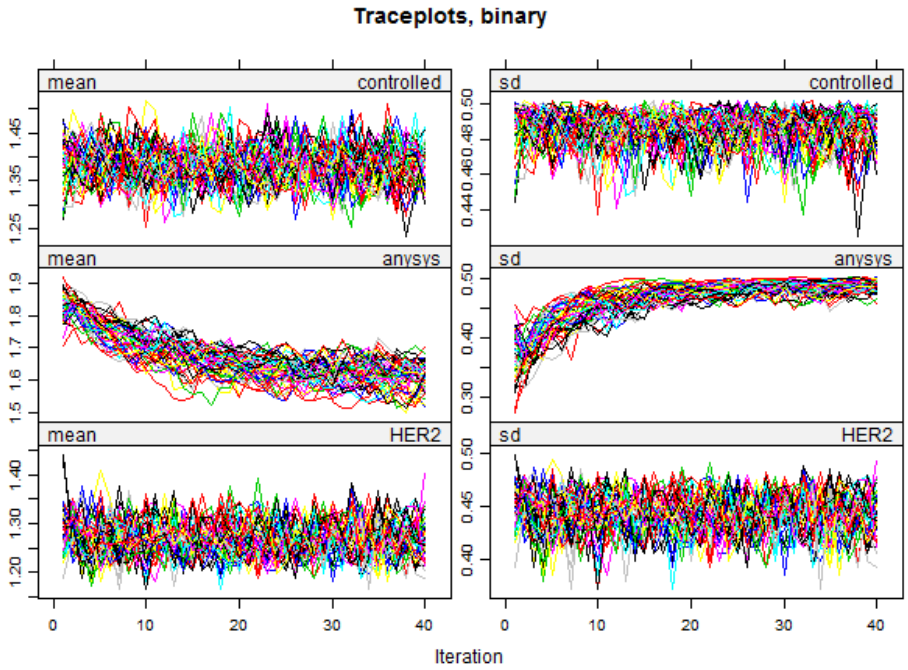
Setting up the model

Issues

- Many categorical variables
- Collinearity between predictors
- Variables with poor influx/outflux [1]
- How many iterations and imputations to draw?

Convergence

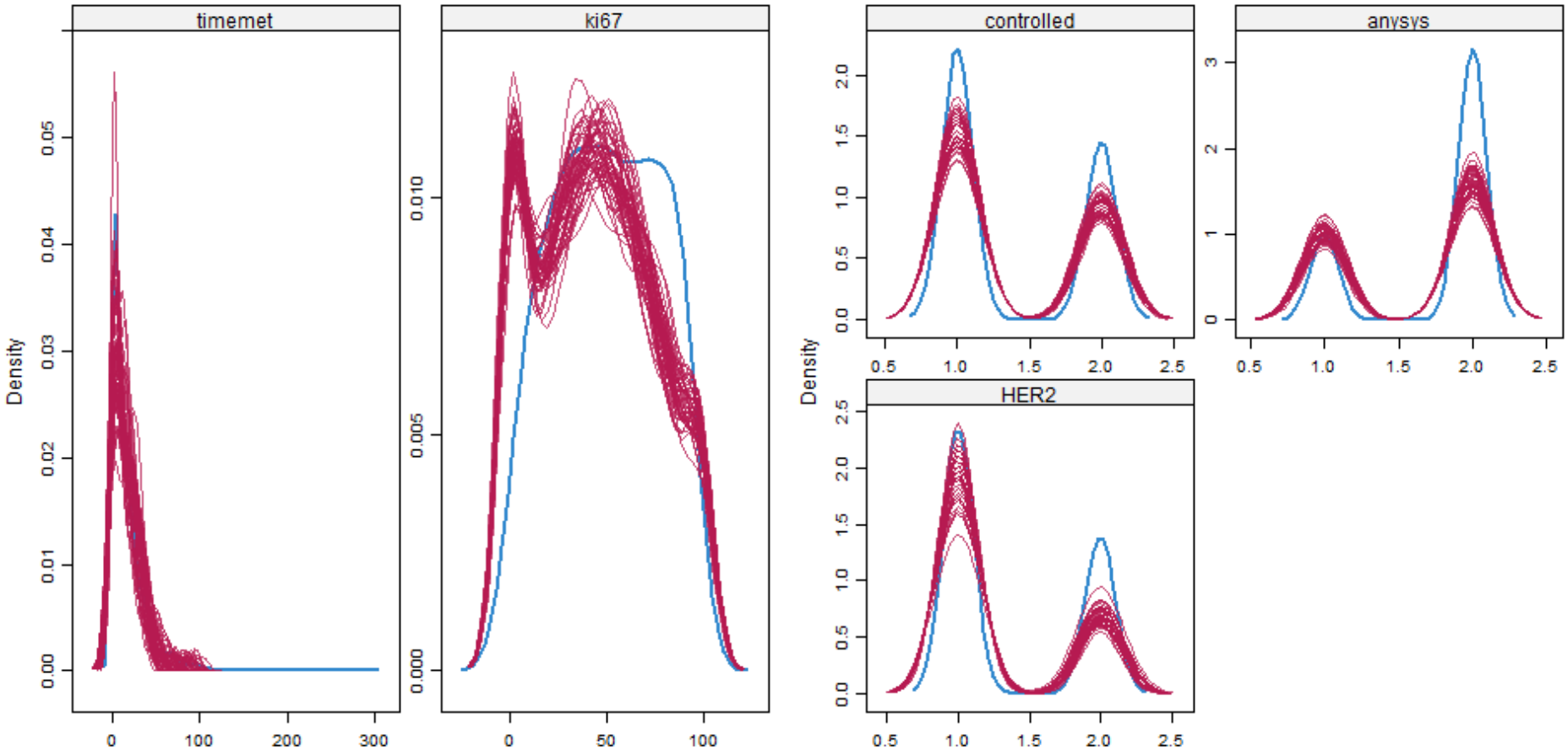




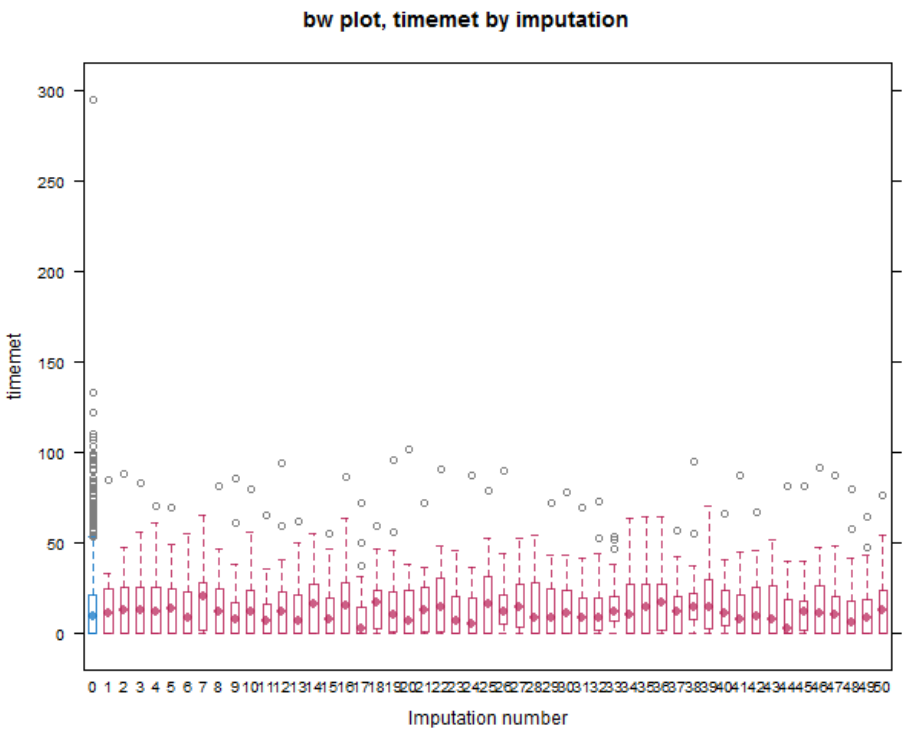
Validity

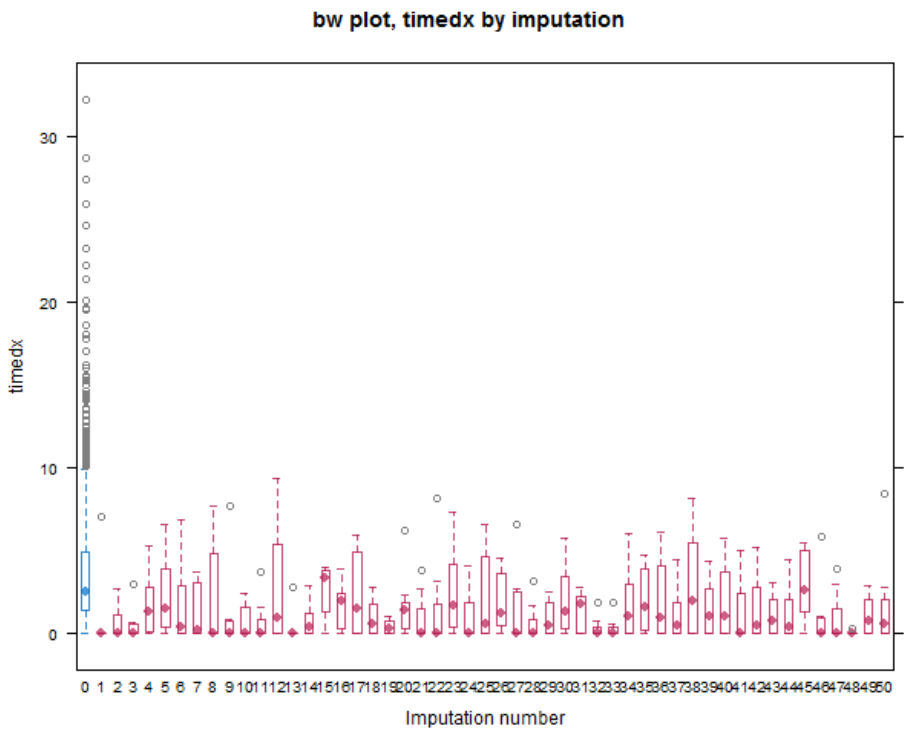
- Lots of tools for continuous imputations
- not many for categorical
  - Solution: look at tables to verify validity

Validity Checks



Validity Checks





Tabluar Checks

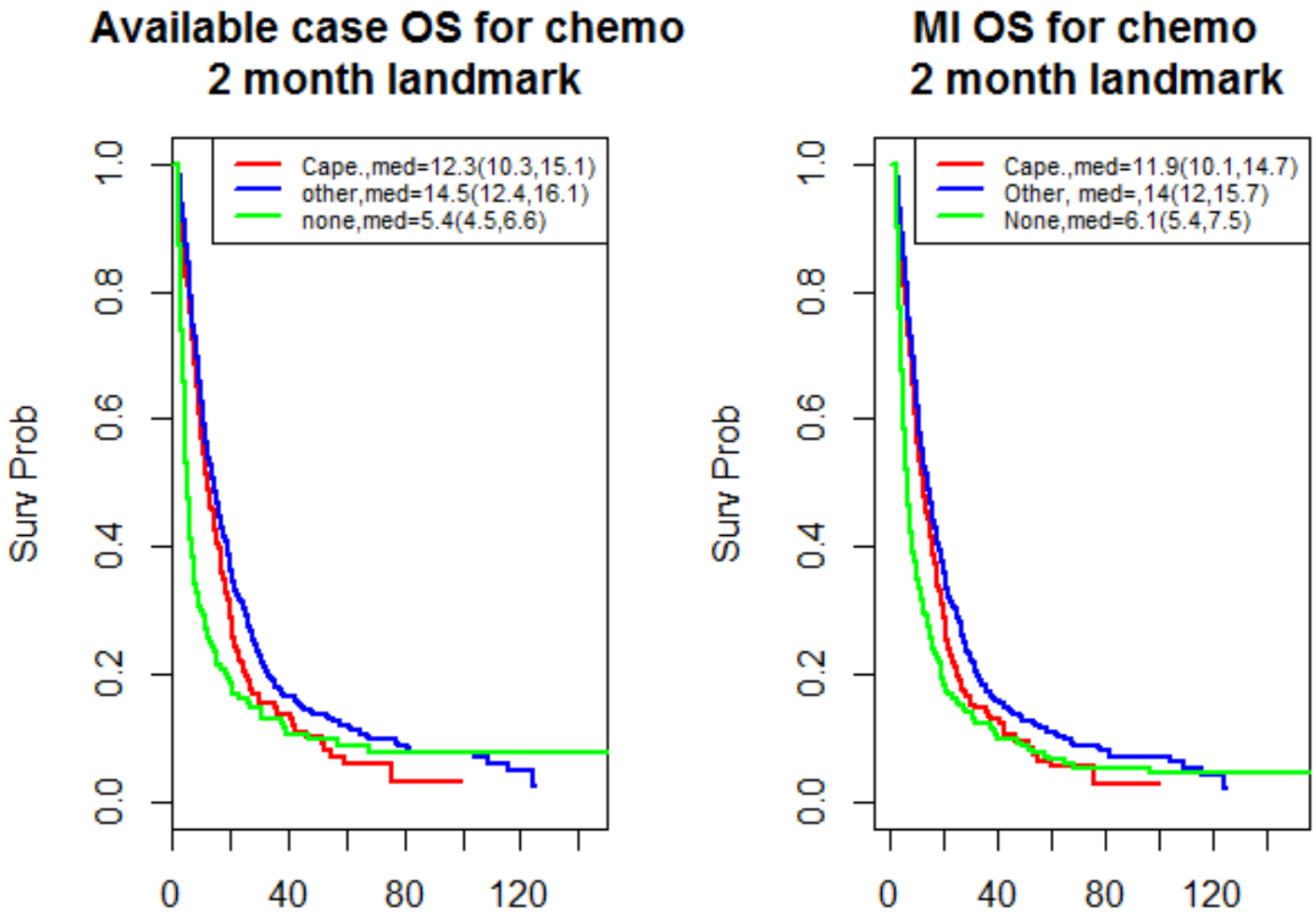
MI data Breakdown

	Sys therapy available case	Sys therapy MI	No Sys therapy available case	No Sys therapy MI
Age (mean,sd)	51.4(10.8)	51.2(10.9)	52.7(11.9)	52.9(11.4)
Breast Cancer subtype				
HR+/HER2-	27%	31%	28%	33%
HR+/HER2+	19%	18%	12%	13%
HR-/HER2+	22%	20%	15%	12%
Triple negative	32%	32%	45%	42%
Prior therapies for stage 4	1(0-3)	2(0-4)	2(0-4)	2(0-4)
Single brain lesion	25%	23%	23%	20%
Controlled extra-cranial	40%	40%	35%	36%
ECOG 0-1	84%	70%	53%	40%
Local Therapy				
Resection Alone	5%	5%	9%	7%
SBRT alone	13%	12%	9%	8%
WBRT	60%	59%	52%	53%
Resection/SBRT+WBRT	12%	14%	10%	8%
no local therapy	10%	10%	20%	23%

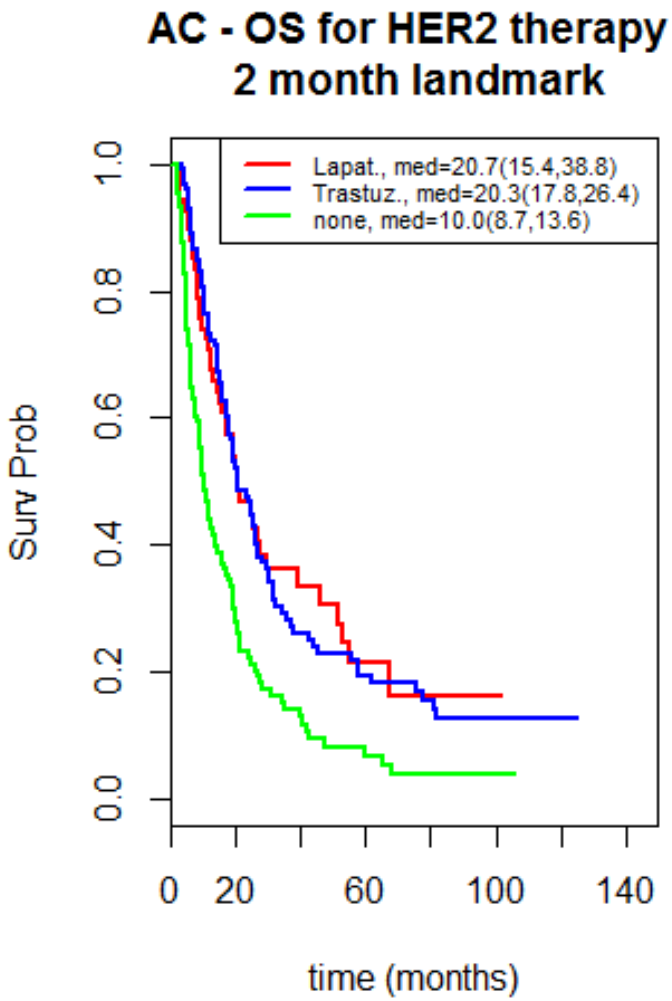
Table: Characteristics of available case data versus MI data

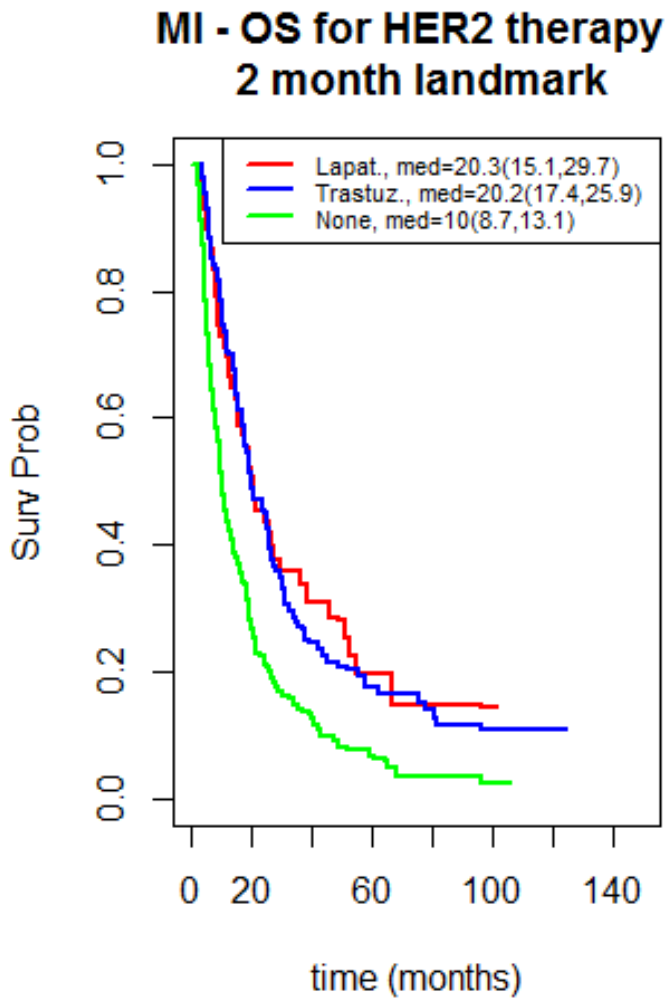
Kaplan-Meier in MI

- Non-informative censoring reasonable
- Pooled by Rubin's Rules on Complimentary log-log



Kaplan-Meier in MI





Log Rank Test

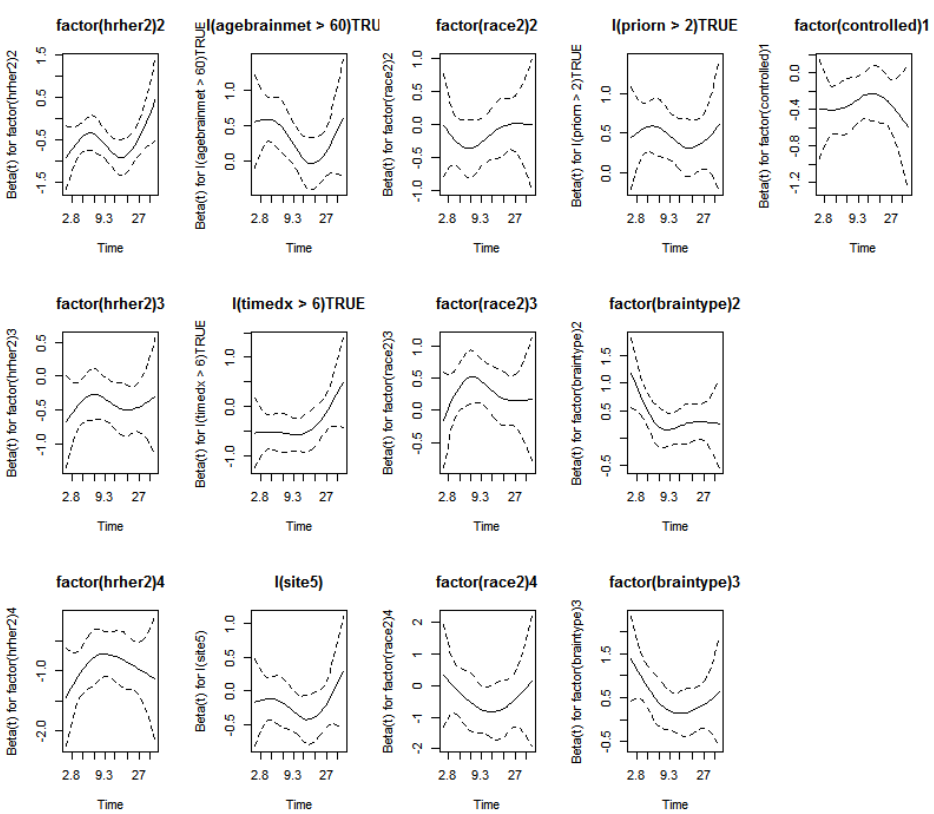
	Chemo	
	AC	MI
cape/other/none	$<.0001$	$<.0001$
cape/other	0.0321	0.033
cape/none	0.00039	.0016
other/none	$<.0001$	$<.0001$

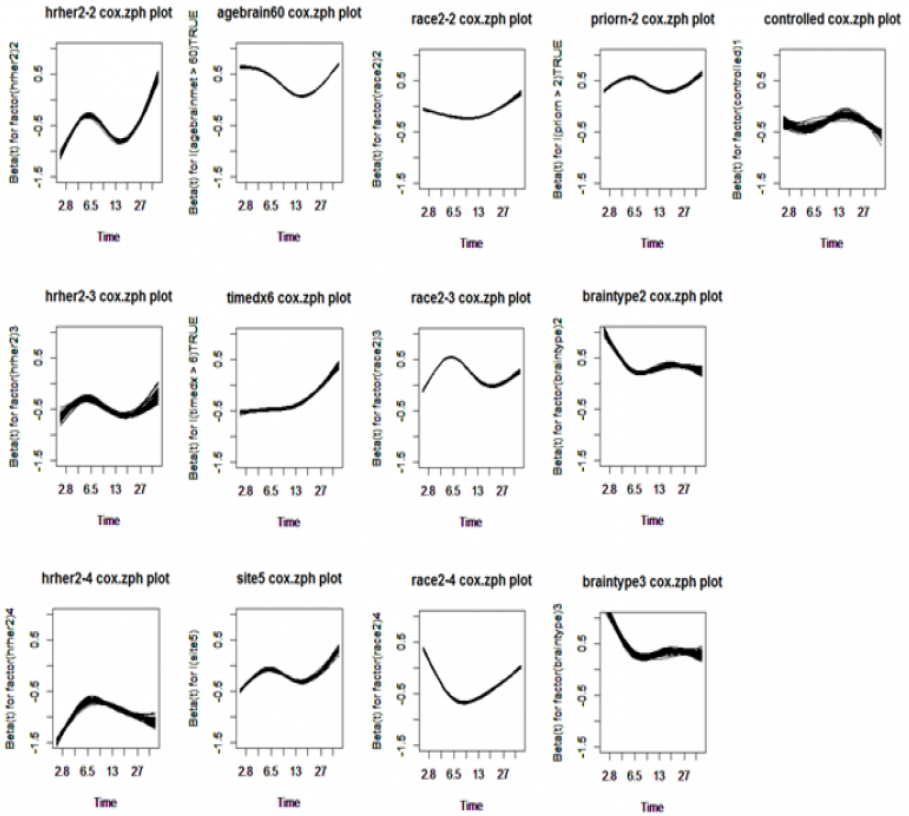
	HER2	
	AC	MI
Lapat/Traztuz/none	$<.0001$	$<.0001$
Lapat/Trastuz	.87	.81
Lapat/none	.00017	.00018
Trastuz/none	$<.0001$	$<.0001$

Cox Model in MI

- Get good baseline model
- Need to check proportional hazards
- Add treatment variables

Schoenfeld Residual Splines





Base model

			AC n= 845				MI	
Variable	Contrast	HR	95% CI	pvalue		HR	95% CI	pvalue (t test)
HR/HER2	-/+ vs. -/-	0.57	(0.46,0.71)	<0.0001		0.59	(0.48,0.72)	<0.0001
	+/- vs. -/-	0.66	(0.54,0.81)	<0.0001		0.63	(0.52,0.76)	<0.0001
	+/+ vs. -/-	0.4	(0.31,0.50)	<0.0001		0.4	(0.32,0.50)	<0.0001
Age	>60 vs. <60	1.37	(1.13,1.65)	0.0011		1.45	(1.22,1.72)	<0.0001
Dx to BM	>6 vs. <6	0.66	(0.54,0.82)	0.00013		0.71	(0.59,0.86)	0.0002
First DM	Brain vs. Oth	0.8	(0.66,0.97)	0.026		0.83	(0.70,0.99)	0.02
Race	Hisp. Vs. White	0.85	(0.68,1.07)	0.17		0.88	(0.71,1.08)	0.11
	Black vs. White	1.31	(1.06,1.63)	0.014		1.25	(1.02,1.52)	0.015
	Other vs. White	0.65	(0.40,1.04)	0.075		0.7	(0.45,1.07)	0.05
# prior Rx	>2 vs. 0-2	1.58	(1.31,1.91)	<0.0001		1.53	(1.29,1.82)	<0.0001
BM type	Mult. Vs. Single	1.45	(1.20,1.76)	<0.0001		1.48	(1.24,1.76)	<0.0001
	LMD vs. Single	1.6	(1.21,2.13)	0.001		1.58	(1.25,2.00)	<0.0001
Sys. Cont.	Yes vs. No	0.71	(0.61,0.83)	<0.0001		0.73	(0.63,0.85)	<0.0001

Table: AC and MI baseline Cox model

MI Cox Model, Chemo

		AC n= 745				MI		
Variable	Contrast	HR	95% CI	p-value		HR	95% CI	p-value (t test)
HR/HER2	-/+ vs. -/-	0.62	(0.49,0.79)	<.0001		0.63	(0.51,0.77)	<.0001
	+/- vs. -/-	0.65	(0.53,0.81)	0.00011		0.64	(0.53,0.78)	<.0001
	+/+ vs. -/-	0.41	(0.31,0.53)	<.0001		0.42	(0.34,0.53)	<.0001
Age	>60 vs. <60	1.34	(1.10,1.64)	0.0041		1.44	(1.21,1.72)	<.0001
Dx to BM	>6 vs. <6	0.72	(0.58,0.90)	0.0032		0.71	(0.58,0.86)	0.00039
First DM	Brain vs. Oth	0.77	(0.63,0.95)	0.014		0.81	(0.68,0.96)	0.016
Race	Hisp. Vs. White	0.77	(0.61,0.98)	0.034		0.86	(0.69,1.06)	0.15
	Black vs. White	1.29	(1.02,1.63)	0.032		1.23	(1.01,1.51)	0.043
	Other vs. White	0.76	(0.47,1.25)	0.28		0.7	(0.45,1.08)	0.11
# prior Rx	>2 vs. 0-2	1.61	(1.32,1.98)	<.0001		1.53	(1.28,1.82)	<.0001
BM type	Mult. Vs. Single	1.46	(1.20,1.78)	0.00017		1.51	(1.27,1.81)	<.0001
	LMD vs. Single	1.45	(1.04,2.03)	0.029		1.41	(1.11,1.80)	0.0049
Sys. Cont.	Yes vs. No	0.57	(0.48,0.68)	<.0001		0.69	(0.59,0.80)	<.0001
Chemo	Cape. vs. none	0.69	(0.53,0.89)	0.0046		0.75	(0.60,0.95)	0.018
	other vs. none	0.52	(0.42,0.65)	<.0001		0.58	(0.47,0.71)	<.0001

Table: AC and MI Cox model with Chemo Treatment

AC and MI Cox Model with HER2 Treatment

			AC n=292				MI n between 391 and 415	
Variable	Contrast	HR	95% CI	p-value		HR	95% CI	p-value (t test)
HR/HER2	+/+ vs. -/+	0.65	(0.49,0.87)	0.0036		0.66	(0.51,0.85)	0.0015
Age	>60 vs. <60	1.38	(0.95,2.01)	0.092		1.58	(1.15,2.18)	0.0054
Dx to BM	>6 vs. <6	0.64	(0.43,0.97)	0.033		0.69	(0.49,0.99)	0.041
First DM	Brain vs. Oth	0.84	(0.58,1.20)	0.34		0.86	(0.62,1.17)	0.34
Race	Hisp. Vs. White	0.69	(0.46,1.02)	0.064		0.76	(0.53,1.09)	0.14
	Black vs. White	1.41	(0.94,2.11)	0.1		1.43	(1.00,2.04)	0.047
	Other vs. White	0.7	(0.32,1.53)	0.38		0.83	(0.46,1.52)	0.55
# prior Rx	>2 vs. 0-2	1.88	(1.34,2.63)	0.00028		1.71	(1.28,2.28)	0.00028
BM type	Mult. Vs. Single	1.3	(0.92,1.86)	0.14		1.25	(0.91,1.70)	0.16
	LMD vs. Single	2.15	(1.20,3.88)	0.011		1.77	(1.10,2.83)	0.018
Sys. Cont.	Yes vs. No	0.73	(0.55,0.97)	0.029		0.78	(0.60,1.01)	0.063
HER2 therapy	Lapat vs. none	0.47	(0.32,0.69)	0.00015		0.52	(0.37,0.75)	0.00036
	Trastuz vs. none	0.45	(0.33,0.61)	<.0001		0.51	(0.38,0.68)	<.0001

Table: AC and MI Cox model with HER2 Treatment

Causal Analysis

- Need to get propensity score from *pretreatment* covariates
- Check the balance and standardized bias
- Run IPTW analysis on each dataset
- Use Mitra and Reiter's within method to get each MI dataset ATE [7]
- Pool via Rubin's Rules

Critiques

- MI doubters and method critiques
- Assumptions in the survival section
- Propensity score in general and model choices

Further research

- Differing imputation methods
- Competing risks
- AFT models
- Differing propensity score methods
- Estimating counterfactuals as an MI problem in MI setting

Any Questions?

- Any questions?

References I

- S. van Buuren, *Flexible imputation of missing data*. 2012.
- D. Rubin, *Multiple Imputation for Nonresponse in Surveys*. No. JOHN WILEY & SONS, 1987.
- P. Rosenbaum and D. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," vol. 70, no. 1, pp. 41–55, 1983.
- H. Demirtas and D. Hedeker, "Imputing continuous data under some non-Gaussian distributions," *Statistica Neerlandica*, vol. 62, no. 2, pp. 193–205, 2008.
- J. Barnard and D. Rubin, "Small-sample degrees of freedom with multiple imputation," *Biometrika*, vol. 86, no. 4, pp. 948–955, 1999.

References II


A. Marshall, D. G. Altman, R. L. Holder, and P. Royston, "Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines.," *BMC medical research methodology*, vol. 9, p. 57, 2009.


R. Mitra and J. P. Reiter, "A comparison of two methods of estimating propensity scores after multiple imputation," *Statistical Methods in Medical Research*, pp. 1–17, 2012.