Using Multiple Imputation, Survival Analysis, And Propensity Score Analysis In Cancer Data With Missingness

Master's Thesis

Nathan Berliner

Department of Statistics
Rice University

11/30/2015

In an Ideal World

When exploring the effect of something

- We would have a large dataset that...
  - Was obtained from a Randomized Controlled Trial (RCT)
  - Would help answer a clearly defined question
  - Had all the covariates of scientific interest
  - Contained no missing data
- But this almost never happens!

Is This a Problem?

YES!

- Without an RCT, we can't be sure if differences in the outcome are due to the treatment or something else
- With missing data, we will be throwing away data and biasing our results

The Solution

This thesis aims to fix these problems by...

- Filling in missing data via multiple imputation
- Creating meaningful analytical models via survival analysis
- Getting a causal interpretation from observational data via propensity score analysis

Goal: To be able to apply these methods to cancer data

Plan For This Presentation

Data obtained from MDACC

Multiple Imputation: m=50 40 iterations

Survival Analysis: Kaplan-Meier Log-Rank test Cox model

Causal Analysis: IPTW Cox model

Data Explanation

- 1514 MD Anderson patients who had brain mets from breast cancer between October 2009 and December 2012
- 1242 usable cases
- 90 covariates - Missingness from 0 to 65%

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

Table 1: Data categories and examples

Questions of Interest

Want to explore survival outcomes under...

1. Chemotherapeutic drugs: Capecitabine vs other chemotherapeutic agents

2. HER2-directed therapies (Lapatinib, Trastuzumab) in HER2+ subjects

A Few Important Covariates

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

Table 2: Table of important covariates to be used in the analysis

Visualization of Missingness

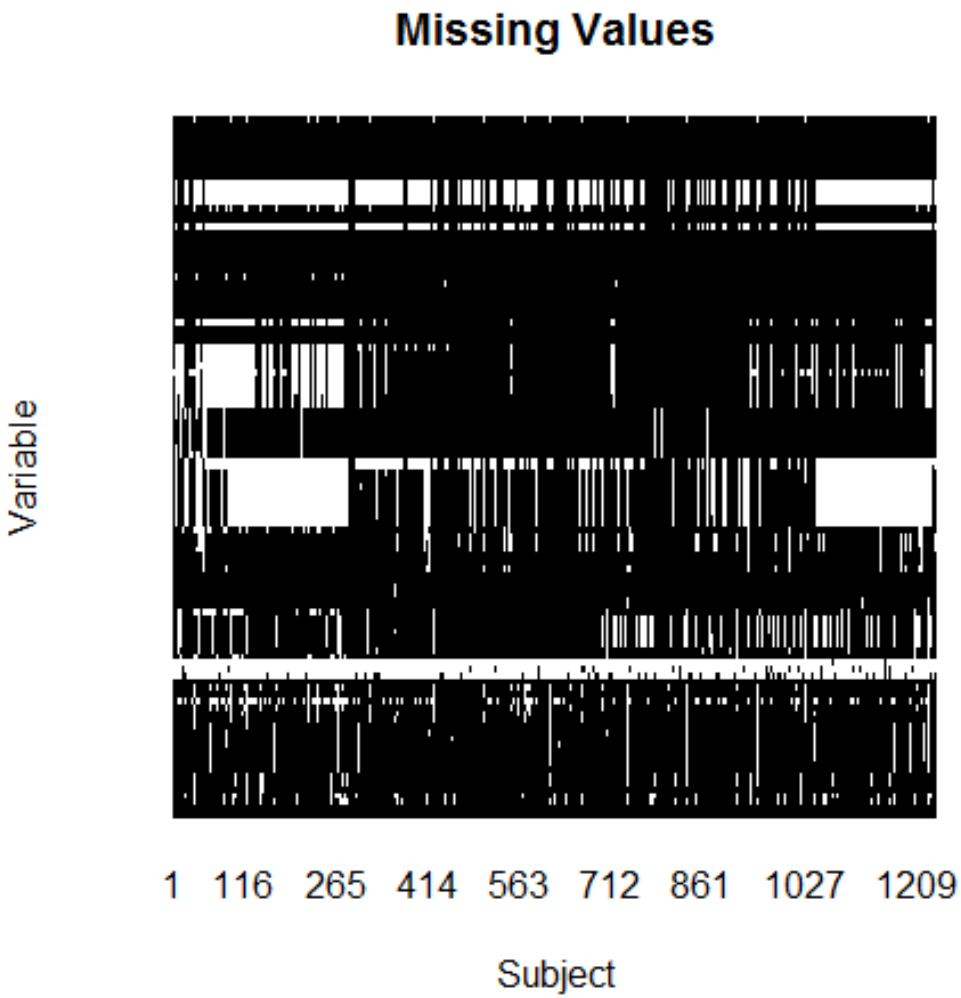
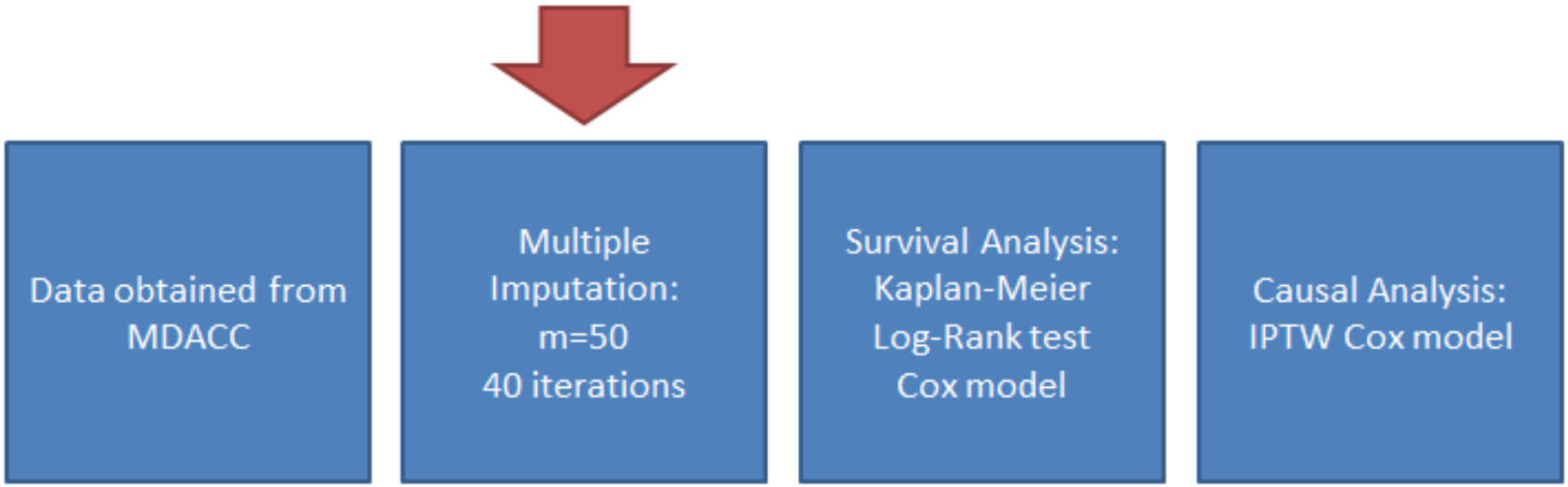


Figure 1: Visualization of missingness in the cancer dataset

Plan For This Presentation



Missing Data and Historical Approaches

- Missing data happens when we intend to collect a piece of data but don't actually get it
- Historical approaches:
  - Complete Case (CC) analysis: Throw away any record that is not complete
  - Available Case (AC) analysis: Use records so long as they are complete for the specific analysis in question
  - Single Imputation (SI): Fill in the missing value, deduct degrees of freedom to account for it

Multiple Imputation

- Theorized in the 70's and 80's by Donald Rubin to improve SI
- Instead of imputing one value, impute $m \geq 2$ values
- Draw the values from the missing data's posterior distribution given the observed data and the process that generated the missing data
- It is the gold standard method for missing data today

How Does MI work?

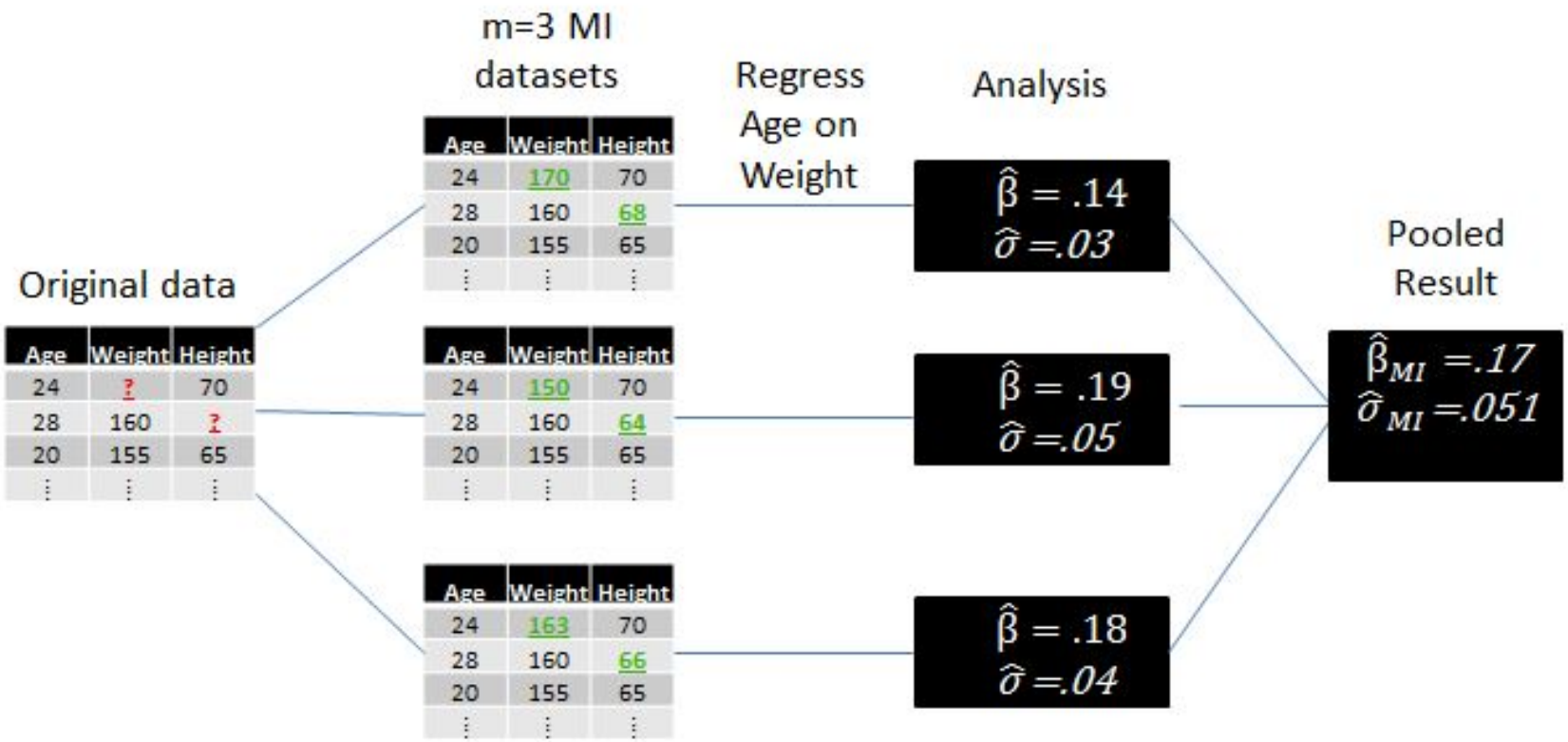


Figure 2: Visualization of MI data

Missingness is displayed by ?'s and the imputed data is shown as #'s.

MI Theory

- Missing data model: $P(R|Y_{obs}, Y_{mis}, \psi)$
  - R is the response missingness indicator
  - Y_{obs}, Y_{mis} are observed and missing parts of Y (the data)
  - ψ parameterizes the missing data model

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:

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

  - The missingness depends on both the data that we have and haven't obtained

Missing Data Mechanisms Visualization

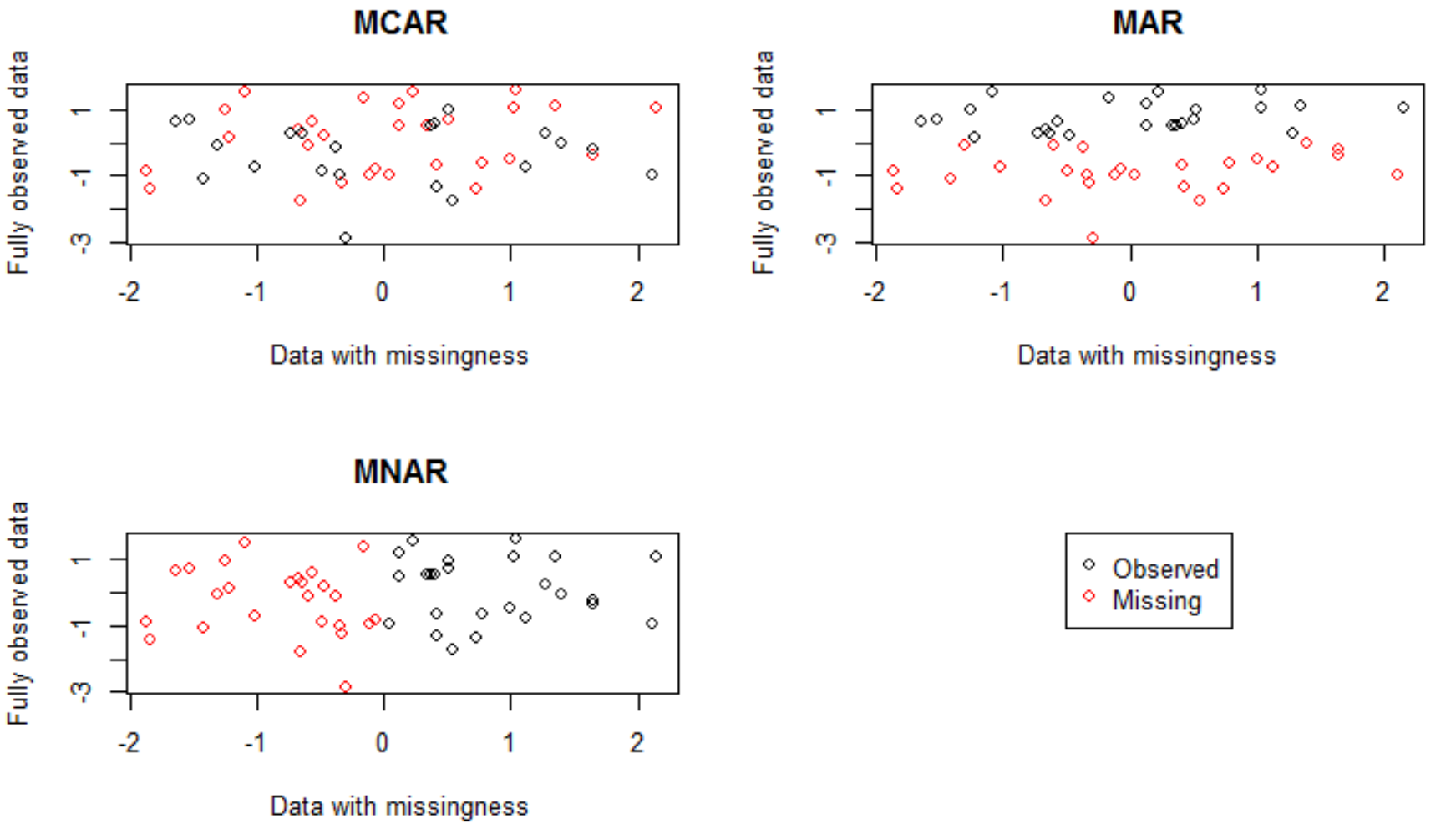


Figure 3: Visualization of Missing Data Mechanisms

Implementing MI: Full Conditional Specification (FCS)

- Assume MAR missing data mechanism
- Missing data is imputed iteratively on a variable by variable basis
- Drawing from $P(Y, R|\theta)$ through conditional densities $P(Y_j|Y_{-j}, R, \phi_j)$
  - Y_{-j}: Data without column j
  - θ: parameterizes full data model
  - ϕ_j parameterizes imputation model
- Generalization of univariate imputation
- Idea: Specify p one dimensional models to impute on the missing data columns

FCS Algorithm - MICE

1. Specify an imputation/posterior predictive model $P(Y_j^{mis}|Y_j^{obs}, Y_{-j}, R, \phi_j)$ 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 $t = 1, \ldots, T$ (number of iterations):
4. Repeat for $j = 1, \ldots, p$ (number of covariates):
   - 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 compete data except Y_j
   - Draw $\dot{\phi}_j^t \sim P(\phi_j^t|Y_{obs}, \dot{Y}_{-j}^t, R)$
   - Draw imputation $\dot{Y}_j^t \sim P(Y_j^{mis}|Y_j^{obs}, \dot{Y}_{-j}^t, R, \dot{\phi}_j^t)$
5. End repeat j
6. End repeat t

FCS imputation pseudocode, taken from [1]

FCS Pros and Cons

Pros

- Flexible
- Easy to specify models
- Handles mixed continuous categorical data
- Yields unbiased estimates with appropriate coverage

Cons

- No guarantee that full conditionals are compatible
- Takes time to set up
- Gets much harder as sample size increases to specify models

Imputation with the Cancer Data

- MAR assumption seems reasonable
- $m = 50$ datasets
- 40 iterations

Convergence

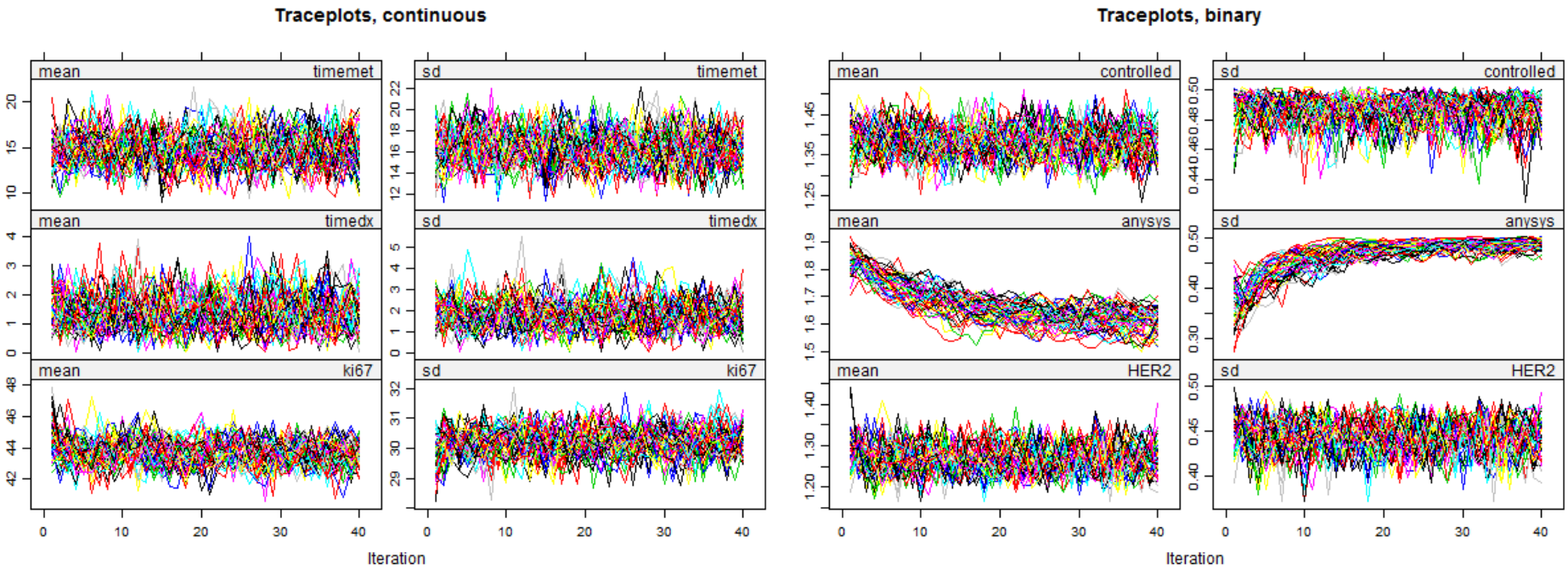


Figure 4: Selected convergence traceplots

Validity Checks

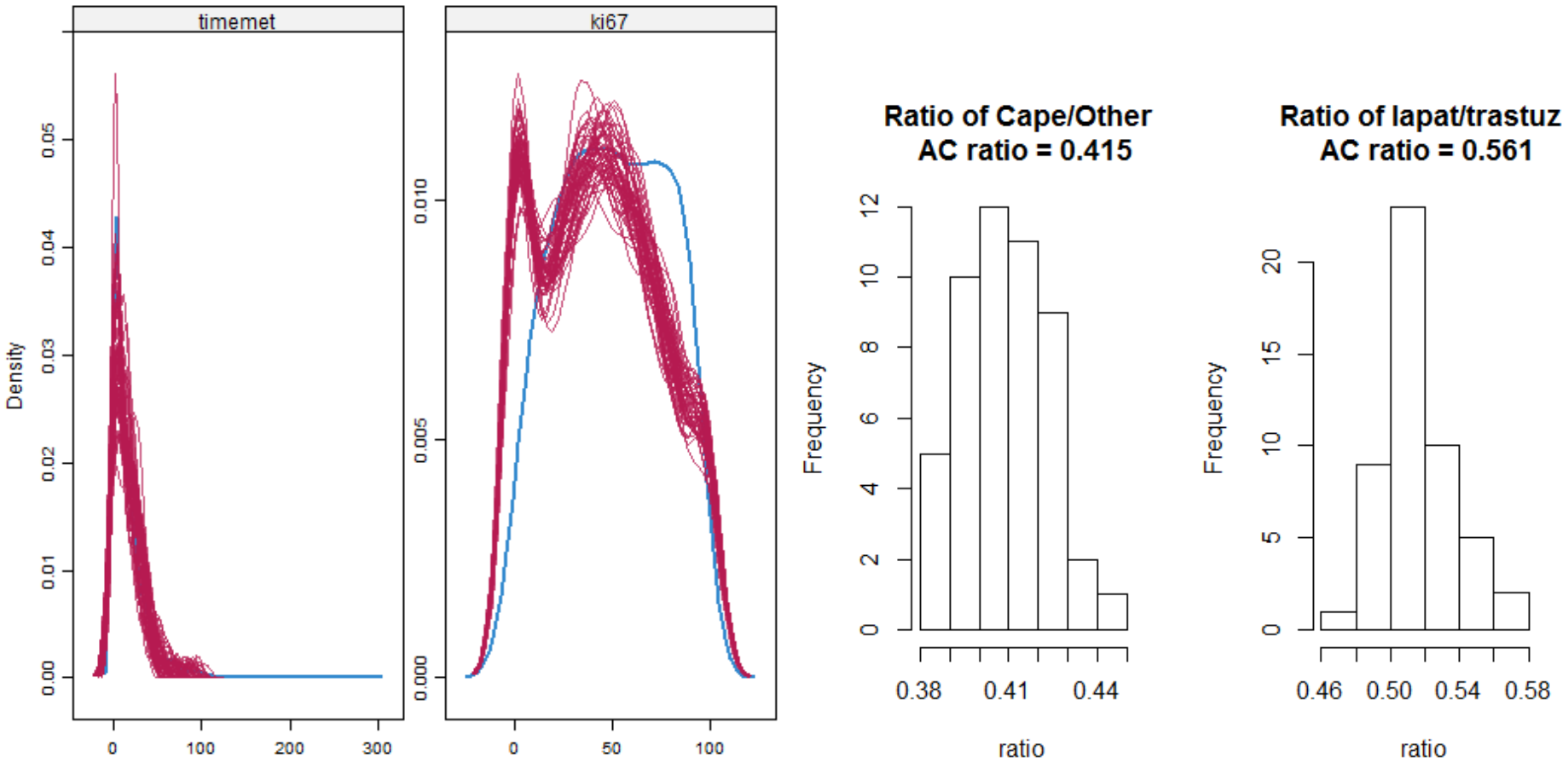


Figure 5: Selected validity checks

MI Data Breakdown

	Sys. therapy AC	Sys. therapy MI	No Sys. therapy AC	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 3: Characteristics of available case data versus MI data

Rubin's Rules

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

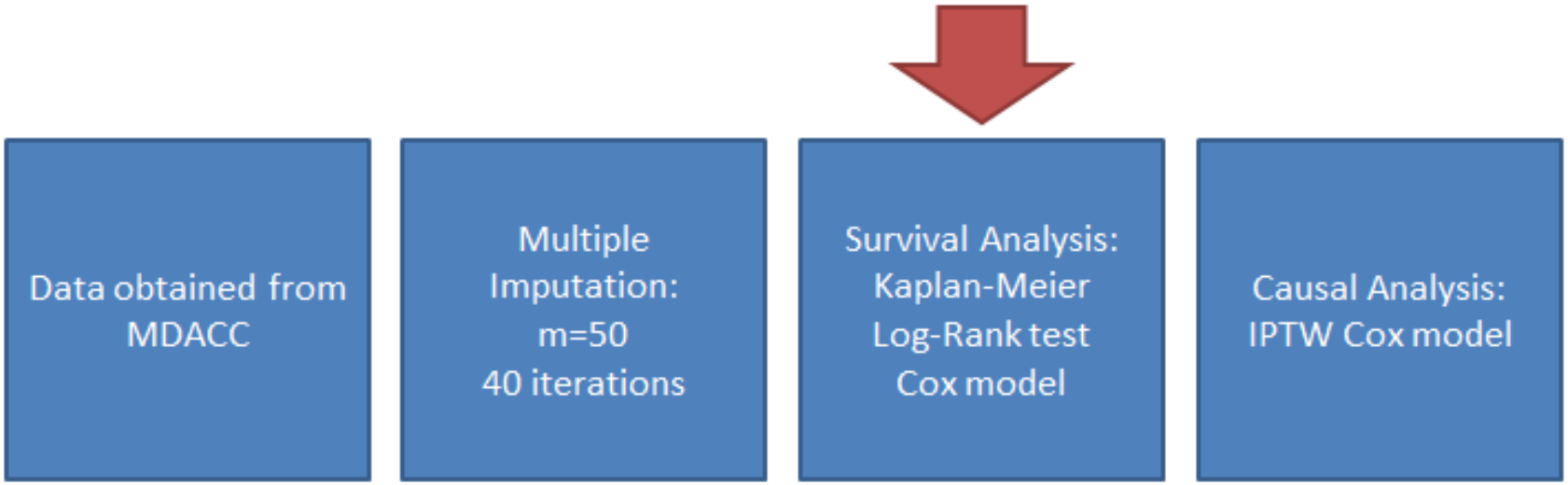
Then

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

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

- Theory developed under normality assumption [2]

Plan For This Presentation



Survival Analysis

Definition

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

Example:

- The survival of patients after brain mets from breast cancer
- Censoring/Truncation:
  - study ending and no death
  - subject dies before study starts
  - subject moves away and can't contact them
  - exact death time only known in an interval

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

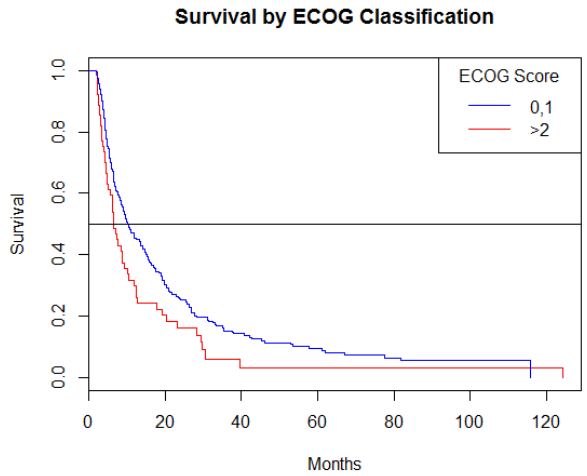


Figure 6: AC Kaplan-Meier Curve for ECOG group

Kaplan-Meier in the MI Setting

- Ensure that we have non-informative censoring
- Algorithm: Pool the complimentary log-log of the Kaplan-Meier curve via Rubin's Rules at each unique event time, get estimates, back transform [4]

Log Rank test

H_0: No difference between the survival curves of the two populations

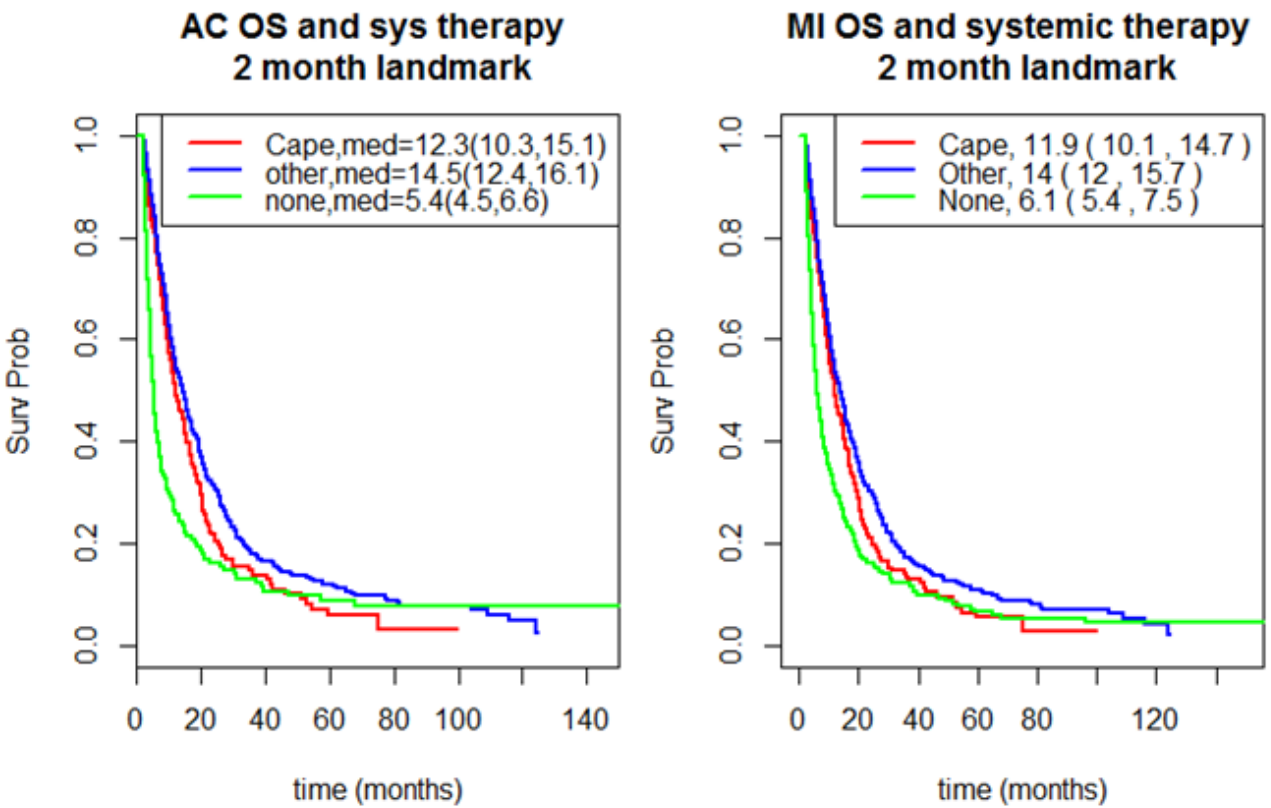
$$\frac{\sum_{j=1}^{J}(O_{1j} - E_{1j})}{\sqrt{\sum_{j=1}^{j} V_j}} \sim N(0,1)$$

- $N_j = N_{1j} + N_{2j}$ 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)
- $E_{1j} = \frac{O_j N_{1j}}{N_j}$
- $V_j = \frac{O_j(N_{1j}/N_j)(1-N_{1j}/N_j)(N_j-O_j)}{N_j-1}$

Log Rank Test in MI Setting

- Combining tests is usually a bad idea [4]
- Under no tied times, the score test on Cox regression on treatment only is equivalent to the log rank test
- Idea: Derive log rank test from Cox regression
- Algorithm: Run Wald test on Cox regression as an approximation to score test

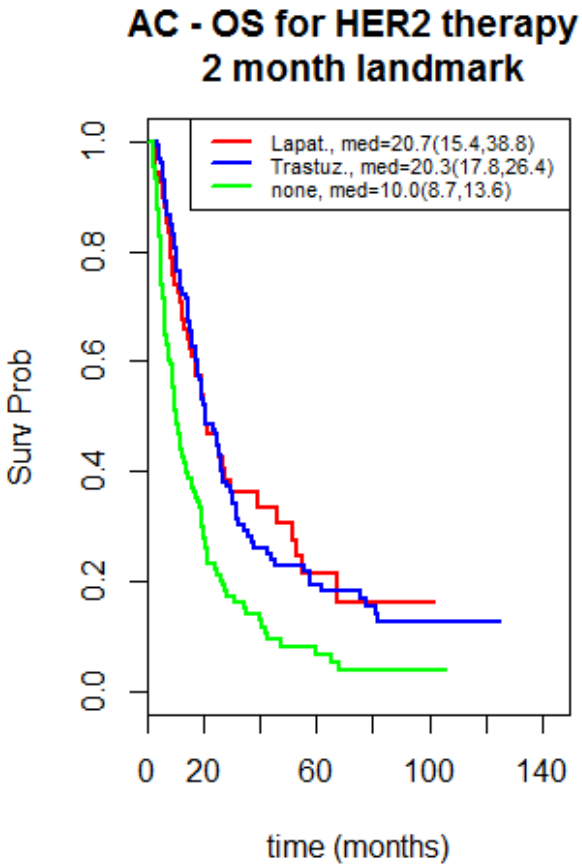
Chemo KM and Log Rank Test

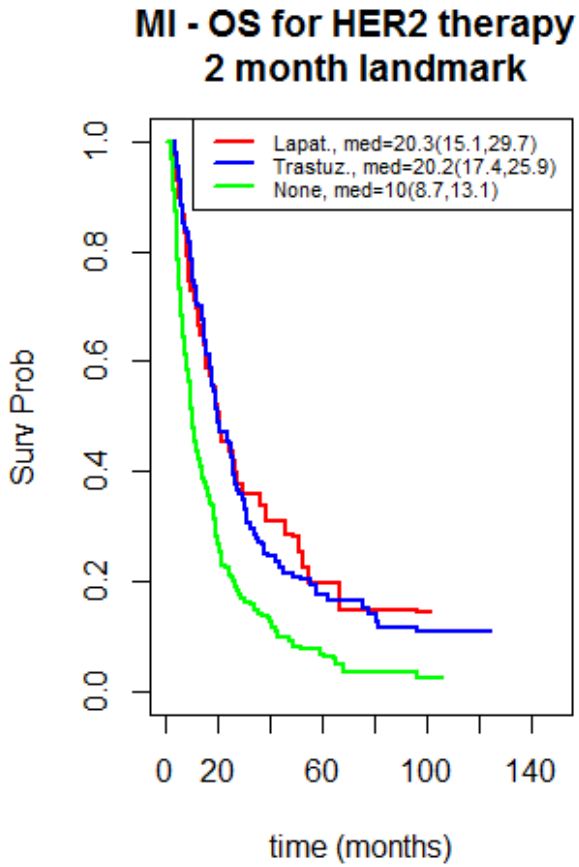


	Chemo	
	AC	MI
Cape./other/none	$<.0001$	$<.0001$
Cape./other	0.0321	0.033
Cape./none	0.00039	.0016
other/none	$<.0001$	$<.0001$

Figure 7: Chemo KM and log rank

HER2-Directed KM and Log Rank Test





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

Figure 8: HER2-Directed KM and log rank

Cox Regression: Hazard Function

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

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

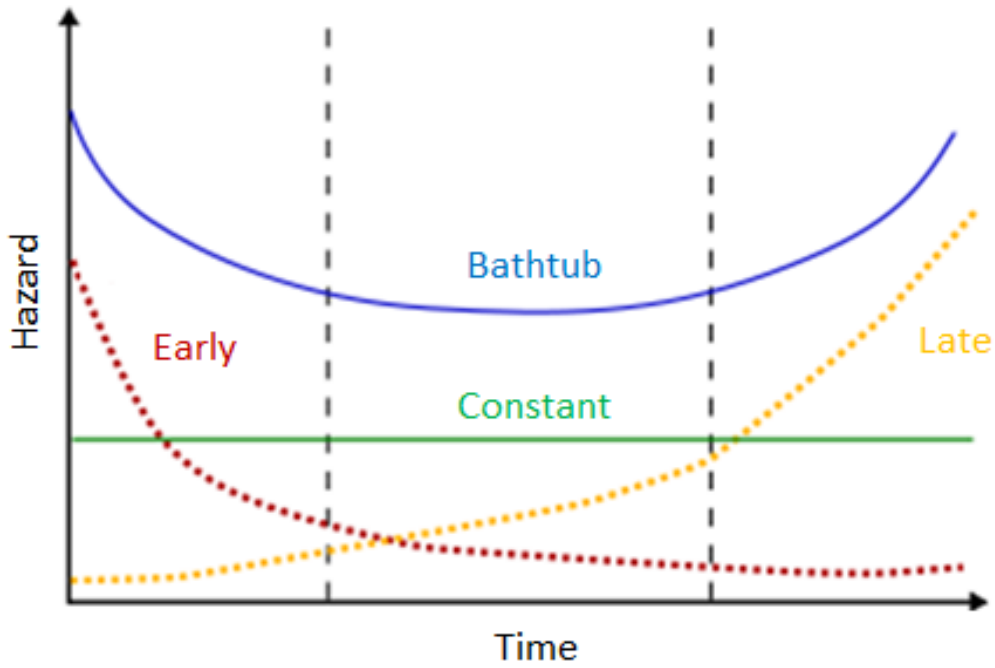


Figure 9: A few different hazard function shapes [5]

Cox Regression

- 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.

Quantity of interest: Hazard ratio $\frac{h(t|Z)}{h(t|Z^*)} = \exp(\sum_{k=1}^{p} \beta_k (Z_k - Z_k^*))$

Cox Regression in the MI Setting

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

Schoenfeld Residual Splines

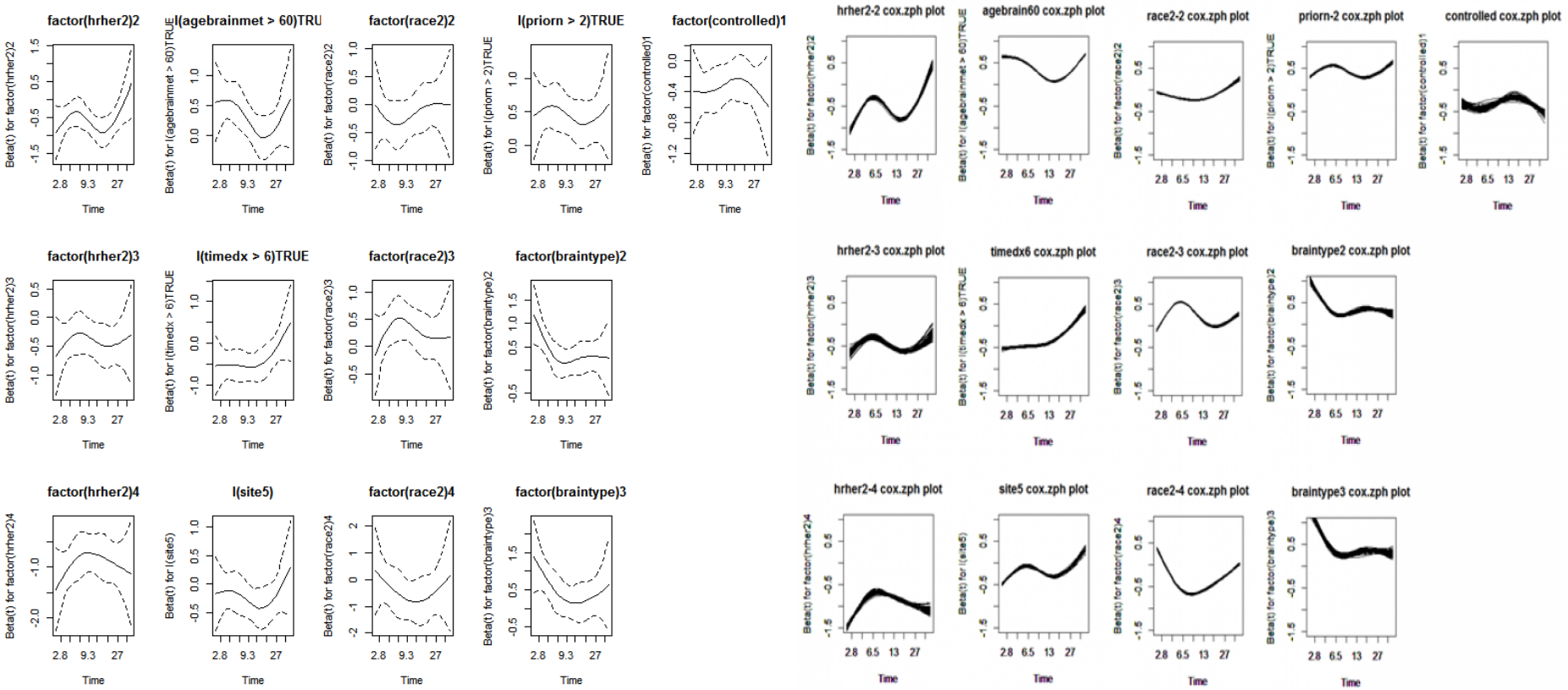


Figure 10: AC and MI Schoenfeld splines

MI Cox Regression, 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		0.70	(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 4: AC and MI Cox regression with Chemo Treatment

MI Cox Regression, HER2-Directed

			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)	0.10		1.43	(1.00,2.04)	0.047
	Other vs. White	0.70	(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	1.30	(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 5: AC and MI Cox regression with HER2 Treatment

Plan For This Presentation

Causal Analysis

The treatments were not given in an RCT

- Want to say "The treatment leads to better survival"
  - But need an RCT to say this
    - Randomization minimizes differences between groups at baseline
    - Differences in outcomes are due to treatment
- We only have observational data
  - Differences could be attributed to the drug or confounding factor, e.g.
    - Healthier patients can tolerate chemo better
    - Different cancer manifestations lead to different plans

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

Counterfactuals

- Suppose that for patient i, there are two potential outcomes
  - $Y_i(0)$ - The outcome if they had taken the control, $T_i = 0$
  - $Y_i(1)$ - The outcome if they had taken the treatment, $T_i = 1$
- The observed value for subject i: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$
- *The fundamental problem of causal inference*

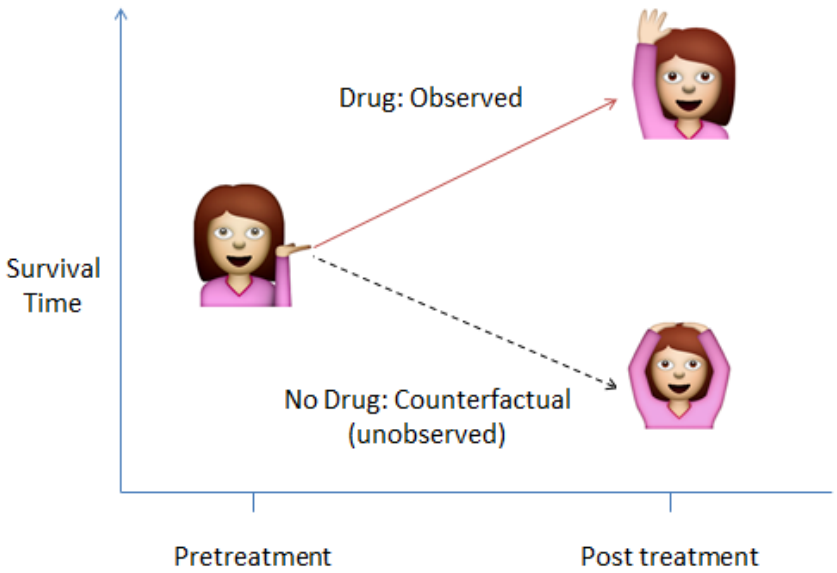


Figure 11: Example of a counterfactual

Rubin's Causal Model: Assumptions

- Stable Unit Treatment Value Assumption (SUTVA): Treatment status of another subject does not affect outcome of other units.
- Ignorability/No Unmeasured Confounders: $(Y(0), Y(1)) \perp T|X$,[6]

Estimands of Interest

- Individual Treatment Effect: $Y_i(1) - Y_i(0)$
- Average Treatment Effect (ATE): $E[Y(1) - Y(0)]$. The effect of moving entire population from untreated to treated
- Average treatment effect for the treated (ATT): $E[Y(1) - Y(0)|T = 1]$. The average treatment effect for those actually treated
- Note: $E[Y(1)|T = 1, X] \neq E[Y(1)]$, because $E[Y|T = 1, X] = E[Y_1 T + Y_0(1 - T)|T = 1, X] = E[Y_1|T = 1, X] \neq E[Y(1)]$
- If assumptions hold, ATE is unbiased estimator of true treatment effect

Propensity Score

Definition

The propensity score is the probability that the subject received the treatment given the subjects *pretreatment* covariates [6].

- Defined as $e_i(x) = P(T_i = 1|X_i)$
- Assume that the covariates X confound treatment status and outcome
- If we assume that
  $(Y(0), Y(1)) \perp T|X \implies (Y(0), Y(1)) \perp T|e(X)$, [6]
- Controlling 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
- Weighting: Weight each observation by the inverse of its propensity score, and then calculate ATE

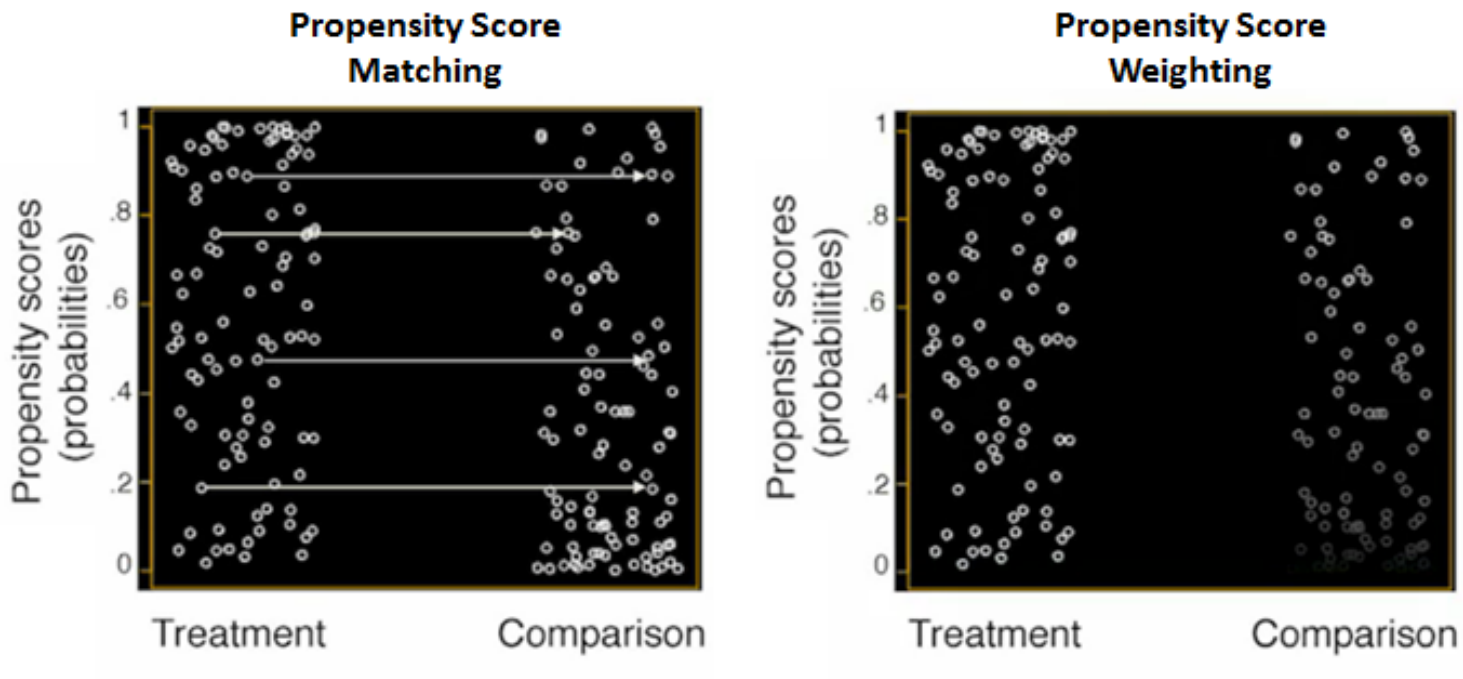


Figure 12: PS matching and weighting [7]

IPTW

- IPTW: Inverse Probability of Treatment Weights
- Idea: Weight sample by propensity score so that we get a sample where there is no confounding
- Weights: $1/e(X)$ for treatment, $1/(1-e(X))$ for control
- Can be shown that
  - $E[\frac{TY(1)}{e(X)}|T=1] = E[Y(1)]$
  - $E[\frac{(1-T)Y(0)}{1-e(X)}|T=0] = E[Y(0)]$

Propensity Score in MI Setting

Mitra and Reiter propose two methods [8]

- *Within*: Work with propensity score on each of the m MI datasets
- *Across*: Average propensity scores across the m datasets and then analyze with the averaged propensities
- Which to use: Dependent on the data

Obtaining the Propensity Score in MI

- Generating the propensity score: Generalized Boosted Model (GBM)
  - What confounders to put in propensity score model?
  - Stage, race, IDC, breast cancer surgery, HR/HER2 status, breast cancer radiation, first met site, number of prior treatments, ECOG score, localized brain mets treatment, age at brain met, type of brain mets, brain met controlled
- What estimand do we care about? ATE

Verifying Balance

For each IPTW MI dataset...

- Need the distribution of the groups to be similar
  - Standardized bias: $|\bar{X}_{k1} - \bar{X}_{k0}|/\hat{\sigma}_k$
  - Kolmogorov-Smirnov (KS) test
- Need to be sure that propensity scores are between 0 and 1

Balance Checks

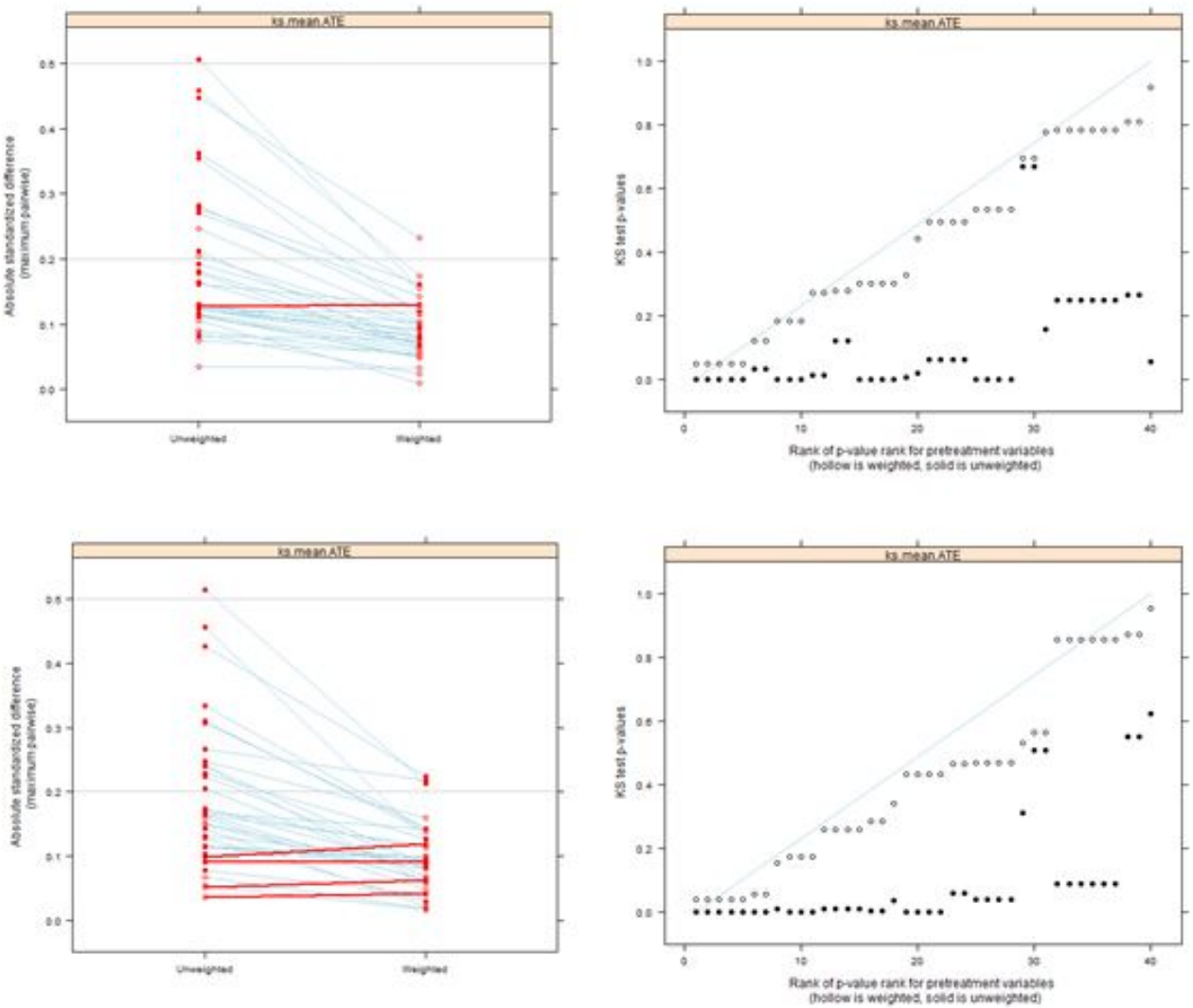


Figure 13: Selected balance diagnostics

Results of IPTW: Chemotherapeutic

	AC Unweighted			AC IPTW			MI unweighted			MI IPTW	
	HR	95% CI		HR	95% CI		HR	95% CI		HR	95% CI
Cape. vs none	0.396	(0.325,0.482)		0.655	(0.481,0.894)		0.484	(0.400,0.585)		0.702	(0.543,0.906)
Other vs none	0.336	(0.287,0.394)		0.567	(0.46,0.754)		0.413	(0.354,0.481		0.593	(0.470,0.748)
Cape. vs other	1.179	(0.983,1.416)		1.156	(0.966,1.383)		1.173	(0.981,1.402)		1.183	(0.998,1.404)

Table 6: Chemotherapeutic ATE with IPTW weights, AC and MI

	AC Unweighted			AC IPTW			MI unweighted			MI IPTW	
	HR	95% CI		HR	95% CI		HR	95% CI		HR	95% CI
Cape. vs none	0.687	(0.530,0.891)		0.603	(0.443,0.820)		0.752	(0.595,0.952)		0.701	(0.536,0.917)
Other vs none	0.521	(0.416,0.653)		0.452	(0.340,0.602)		0.579	(0.474,0.707)		0.532	(0.416,0.681)
Cape. vs other	1.318	(1.078,1.612)		1.334	(1.109,1.604)		1.300	(1.076,1.570)		1.317	(1.100,1.579)

Table 7: Chemotherapeutic ATE, Doubly Robust, AC, MI

Results of IPTW: HER2-Directed

	AC Unweighted			AC IPTW			MI Unweighted			MI IPTW	
	HR	95% CI		HR	95% CI		HR	95% CI		HR	95% CI
Lapat. vs none	0.467	(0.355,0.616)		0.571	(0.381,0.855)		0.474	(0.362,0.622)		0.485	(0.304,0.775)
Trastuz. vs none	0.488	(0.398,0.597)		0.566	(0.421,0.759)		0.506	(0.417,0.614)		0.480	(0.313,0.735)
Lapat. vs Trastuz.	0.958	(0.693,1.324)		1.009	(0.680,1.496)		0.927	(0.673,1.28)		1.011	(0.763,1.338)

Table 8: HER2 directed ATE with IPTW weights, AC and MI

	AC Unweighted			AC IPTW			MI Unweighted			MI IPTW	
	HR	95% CI		HR	95% CI		HR	95% CI		HR	95% CI
Lapat. vs none	0.468	(0.316,0.692)		0.514	(0.331,0.798)		0.524	(0.367,0.747)		0.410	(0.257,0.652)
Trastuz. vs none	0.447	(0.328,0.6089)		0.456	(0.328,0.632)		0.511	(0.381,0.685)		0.388	(0.249,0.602)
Lapat. vs Trastuz.	1.048	(0.704,1.560)		1.128	(0.726,1.754)		1.026	(0.713,1.477)		1.057	(0.788,1.417)

Table 9: HER2 directed ATE with IPTW weights, double robust

Recap

- Applied survival and causal analysis on MI cancer data
- Found overall, any treatments better than none
- Other chemotherapeutics better than Capecitabine
- Lapatinib and Trastuzumab are about the same

Critiques

- MI skeptics and method critiques
- Use of propensity scores
- Assumptions made throughout

Further Research and Extensions

- Exploring the "other chemotherapeutics"
- Competing risks
- AFT models
- Differing propensity score methods and instrumental variables
- Estimating counterfactuals as an MI problem in MI setting

Acknowledgments

- My committee - For guiding me through the process
- Dr. Hess - For advising and mentoring me
- Dr. Ibrahim and Dr. Bugano - For letting me work with their data and advising on project matters
- Margaret Poon - For ALWAYS knowing what needs to be done
- My family - For unconditionally supporting me
- My friends - For believing in me and helping critique my thesis

Supplemental Slides

- Discuss in depth topics should time allow

Propensity Score Issues

- Unmeasured confounders
- Choice of pretreatment covariates in the propensity score model
- Different models and methods may lead to different conclusions

Joint Modelling (JM)

- 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 [9]
- 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.

Figure 14: Normal JM imputation pseudocode

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 continuous missing data
- Poor with derived variables
- Can give impossible combinations

Inference with Rubin's Rules

- Assume that with complete data, inference on the estimand Q would be based on the statement $(Q - \hat{Q}) \sim N(0, U)$
  - $\hat{Q}$ is the statistic estimating Q
  - U is the variance-covariance of $(Q - \hat{Q})$
- Since true T is not known, then

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

- ν is given by [10]

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

- 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 each other to get one huge dataset
  - Will get unbiased results
  - But sample size is falsely inflated, thus cannot trust variance

KM issues in the MI setting

- Issue: Kaplan-Meier is not normally distributed
  - Solution: Complimentary log-log transformation, pool [4]
- 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

Median Survival Time

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

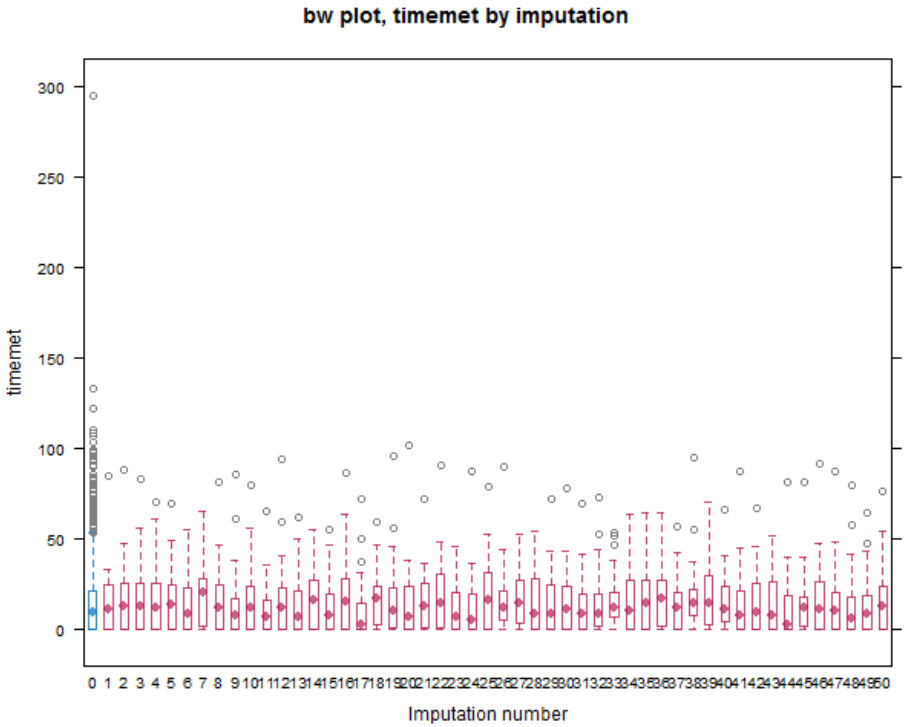
Log Rank Issues in MI setting

- Idea: Combine log rank tests from each MI dataset
  - Problem: Wastes information and is unstable [4]
  - Idea: Calculate log rank from the MI Kaplan-Meier curve
  - Problem: Risk set and deaths no longer meaningful

Setting up the model- Issues

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

Validity Checks



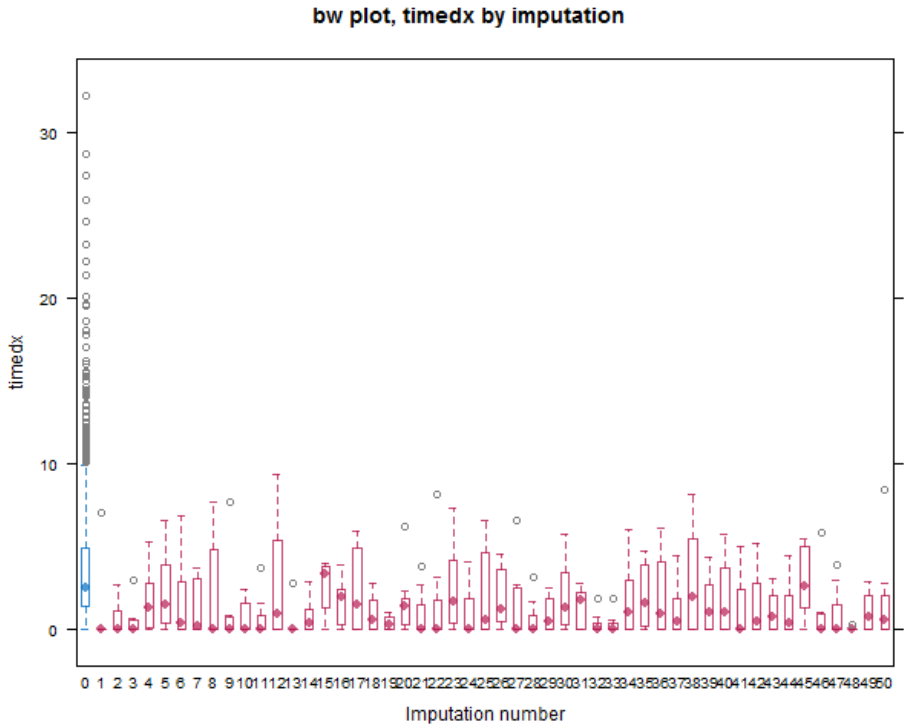


Figure 15: BW plot checks

Tabular Checks

AC	os>10	
hrher2	FALSE	TRUE
-/-	0.246	0.0967
-/+	0.0916	0.0916
+/-	0.1917	0.1154
+/+	0.0712	0.0958

AC		
controlled		
0	1	
0.6048	0.3952	

AC		
lapatrasno		
1	2	3
0.080709	0.1437	0.7756

AC		
capeothno		
1	2	3
0.206693	0.498	0.2953

[[1]]		
	os>10	
hrher2	FALSE	TRUE
-/-	0.2432	0.0974
-/+	0.091	0.0878
+/-	0.2005	0.1184
+/+	0.0692	0.0926

[[1]]		
controlled		
0	1	
0.602254	0.3977	

[[1]]		
lapatrasno		
1	2	3
0.0668	0.132	0.8011

[[1]]		
capeothno		
1	2	3
0.190821	0.4686	0.3406

[[2]]		
	os>10	
hrher2	FALSE	TRUE
-/-	0.2448	0.0966
-/+	0.0894	0.0886
+/-	0.1989	0.1192
+/+	0.0709	0.0918

[[2]]		
controlled		
0	1	
0.605475	0.3945	

[[2]]		
lapatrasno		
1	2	3
0.0684	0.1337	0.7979

[[2]]		
capeothno		
1	2	3
0.196457	0.4646	0.339

[[3]]		
	os>10	
hrher2	FALSE	TRUE
-/-	0.244	0.095
-/+	0.0894	0.0894
+/-	0.1997	0.1208
+/+	0.0709	0.091

[[3]]		
controlled		
0	1	
0.603865	0.3961	

[[3]]		
lapatrasno		
1	2	3
0.0741	0.1449	0.781

[[3]]		
capeothno		
1	2	3
0.191626	0.471	0.3374

Figure 16: Selected tabluar checks

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 10: AC and MI baseline Cox regression

Issues with Propensity Score in our Setting

- Problem: Theory was developed for binary treatments, we have ternary
  - Solution: Run each treatment as binary, then compare groups
- Propensity score model specification
  - Solution: Boosting, subject to KS statistic minimization

Propensity Score Box and Whisker Checks

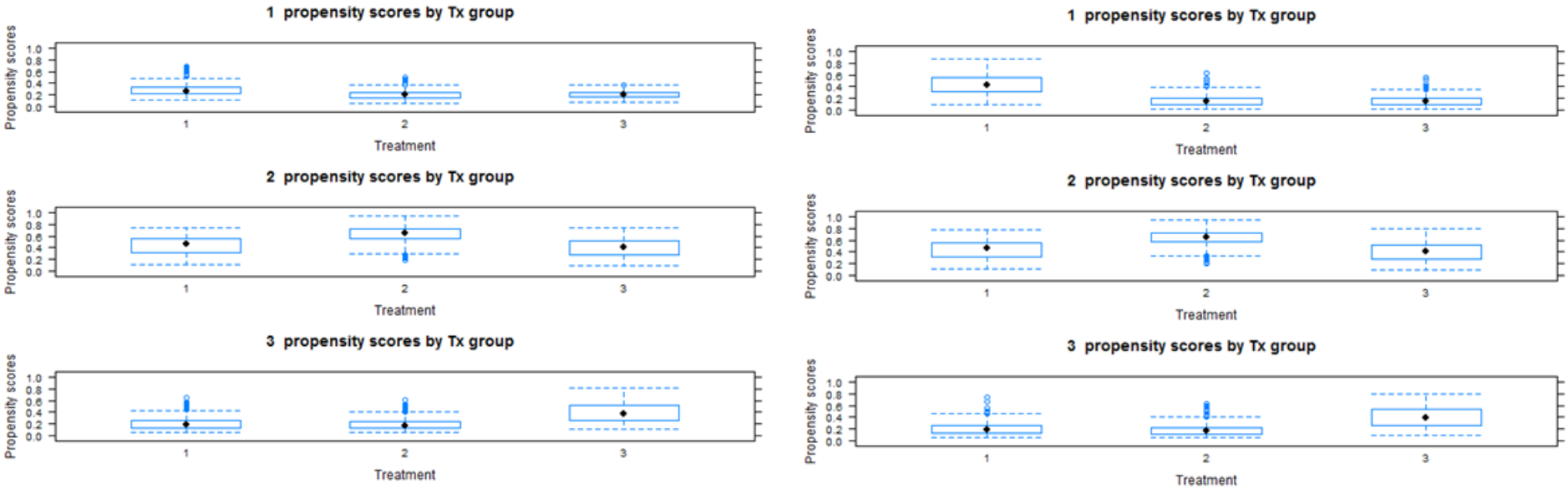
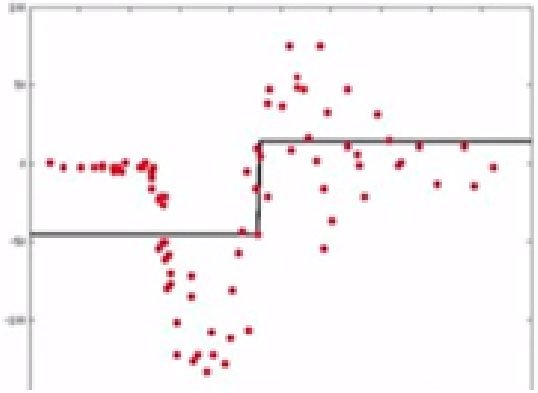


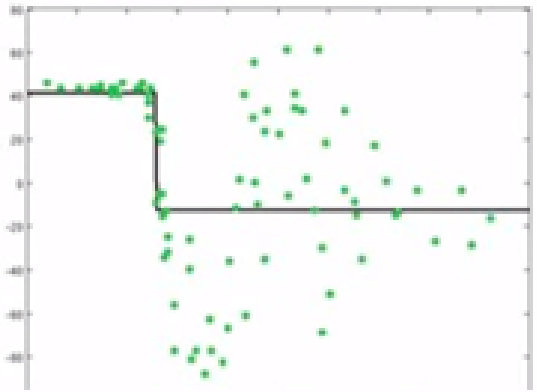
Figure 17: Selected PS BW plots

Boosting

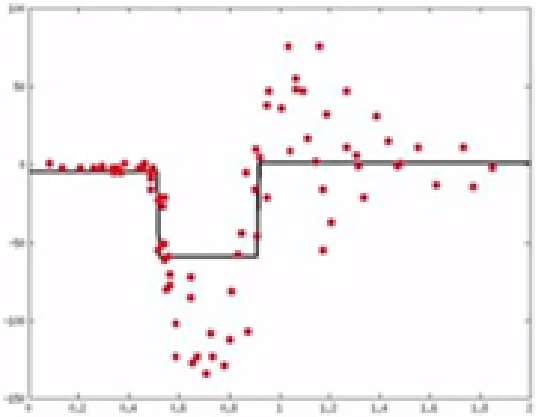
Learn a simple predictor…



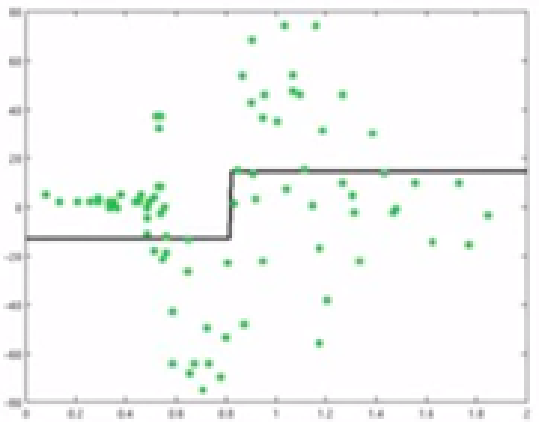
Then try to correct its errors



Combining gives a better predictor…



Can try to correct its errors also, & repeat



https://www.youtube.com/watch?v=sRktKszFmSk

References I

- S. van Buuren, *Flexible imputation of missing data*. 2012.
- D. Rubin, *Multiple Imputation for Nonresponse in Surveys*. No. JOHN WILEY & SONS, 1987.
- J. Klein and M. Moeschberger, *Techniques for Censored and Truncated Data*. 1984.
- 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.
- Wikipedia, "Bathtub curve."

References II

- 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.
- D. F. Griffin, Beth Ann McCaffrey, "TWANG Short Course."
- 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.
- 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.