Survival Analysis

Tomas Bencomo and Kyle W. Singleton, PhD

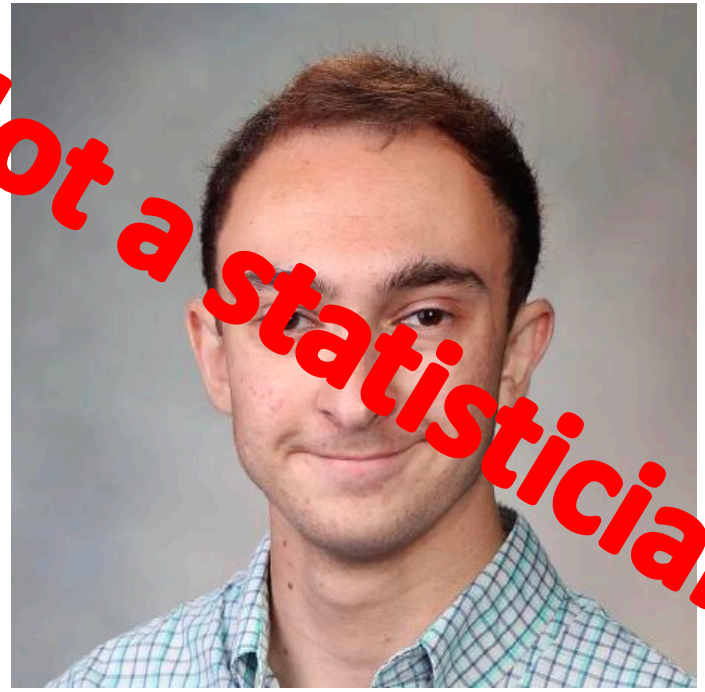
About Us

PhD Biomedical Engineering - Informatics

Graduated High School
B.S. Candidate Computer Science

A Quick Disclaimer

PhD Biomedical Engineering - Informatics



Graduated High School
B.S. Candidate Computer Science

Although we have both received formal statistics training, when in doubt consult a statistician.

Roadmap

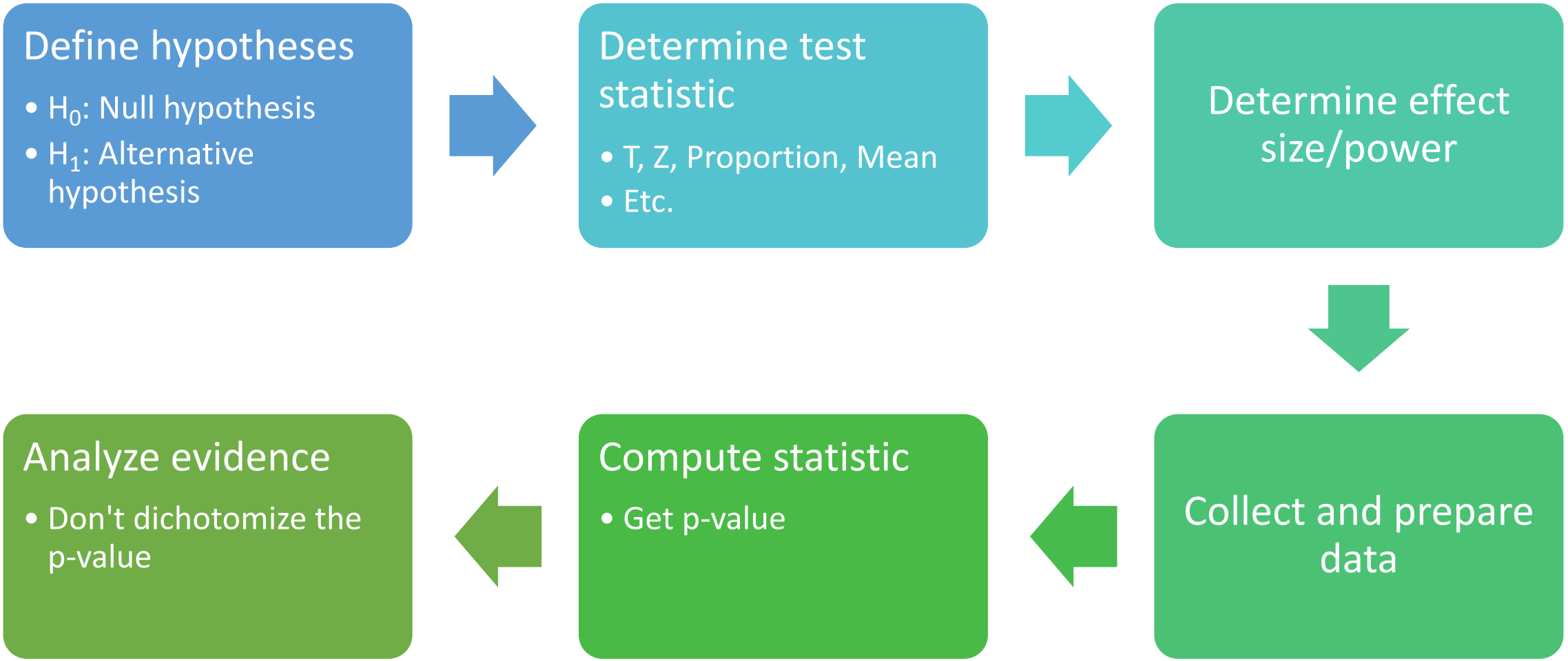
1. Statistics Review
2. Designing the Analysis
3. Analysis Methods
4. Interpreting Statistics

Why Give This Talk?

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."

- Ronald Fisher

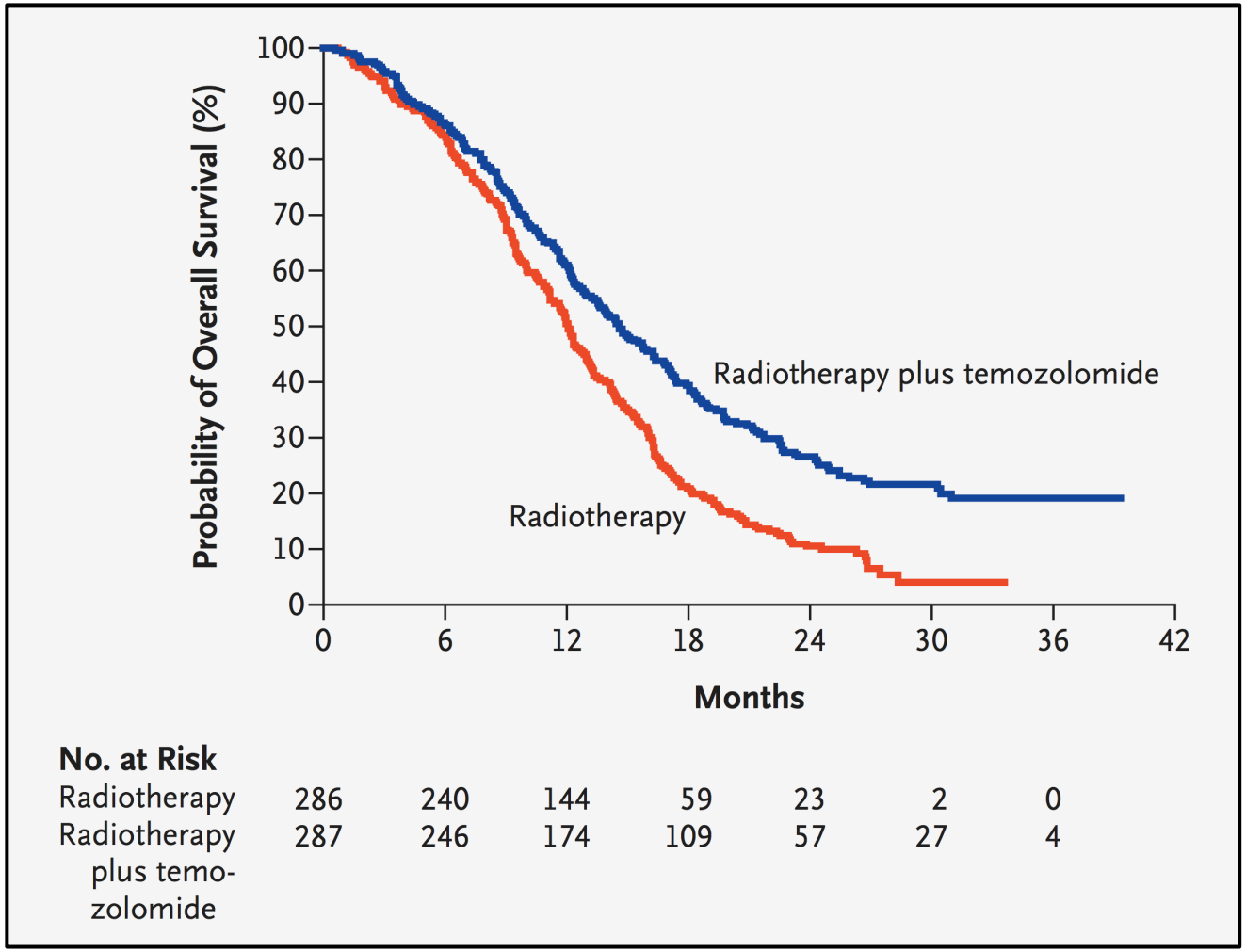
Hypothesis Testing Evaluates Evidence



Hypothesis in Survival

H_0: Adding TMZ does not improve survival

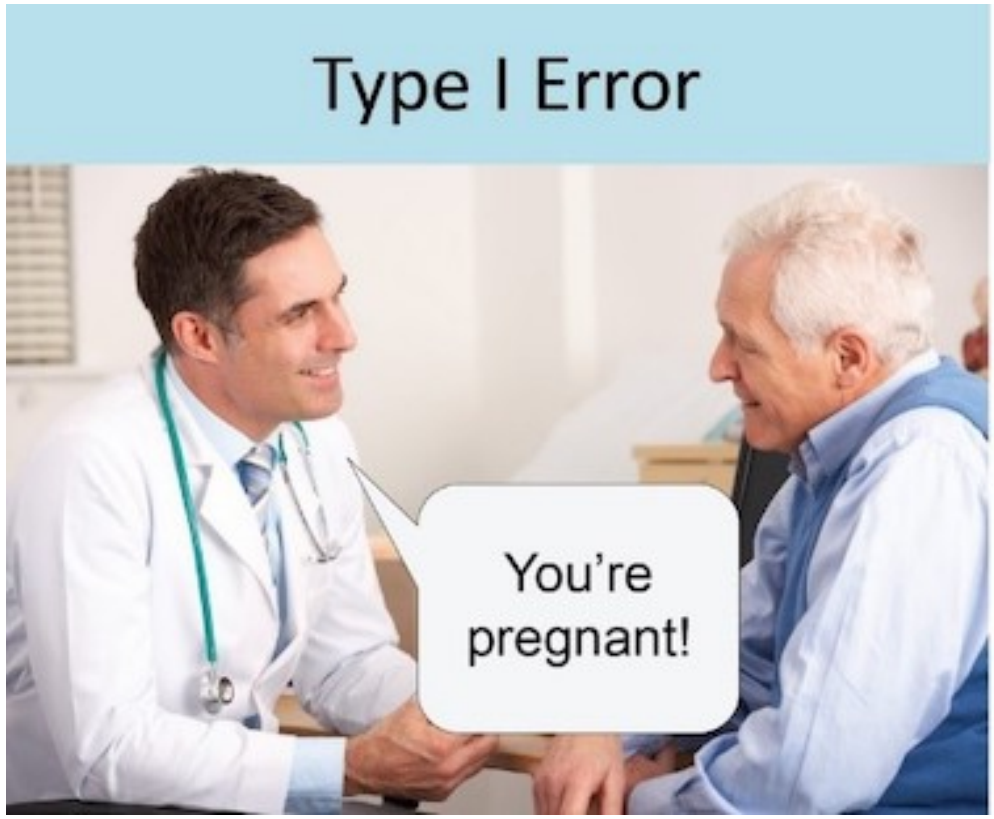
H_1: Adding TMZ does improve survival



Stupp R et al. N Engl J Med 2005;352:987-996.

Hypothesis Tests Can Make Mistakes

H_0: The patient is not pregnant
H_1: The patient is pregnant



Decision Errors

We decide to:	H_0 True	H_0 False
Reject H_0	False Positive Type I Error	True Negative
Accept H_0	True Positive	False Negative Type II Error

Experimental Design Is Critical

Observational Study

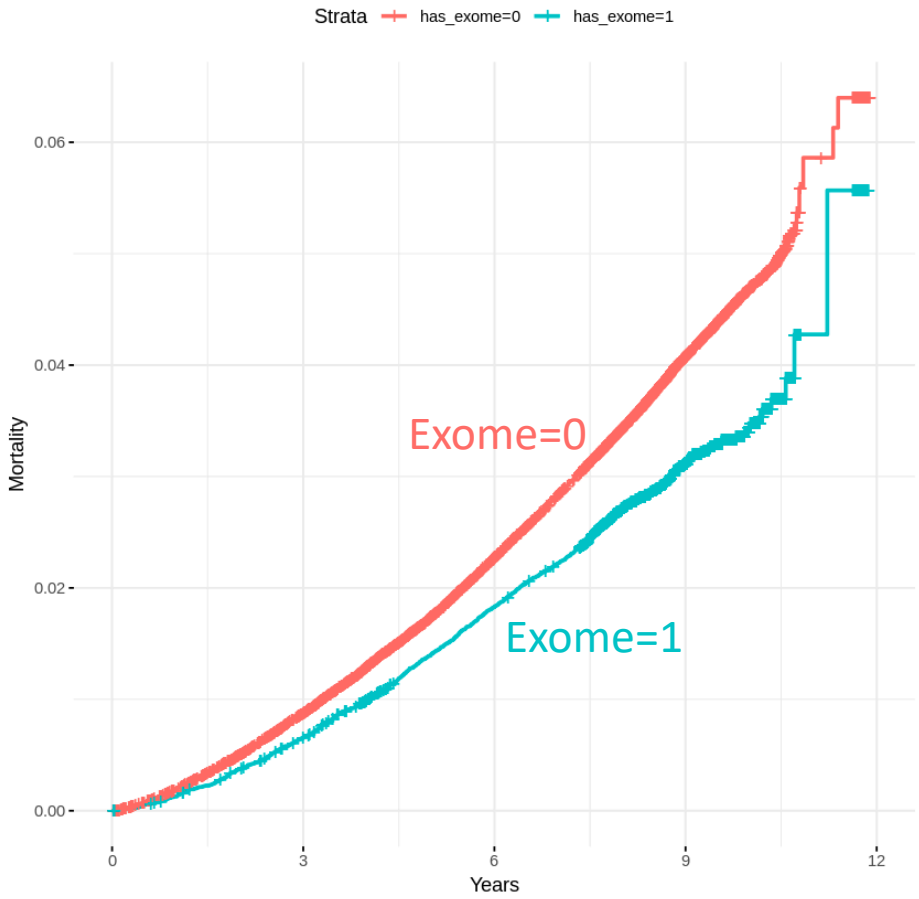
- Collect data without intervening
- Confounding and covariate imbalance major concern
- Notorious for mixed results

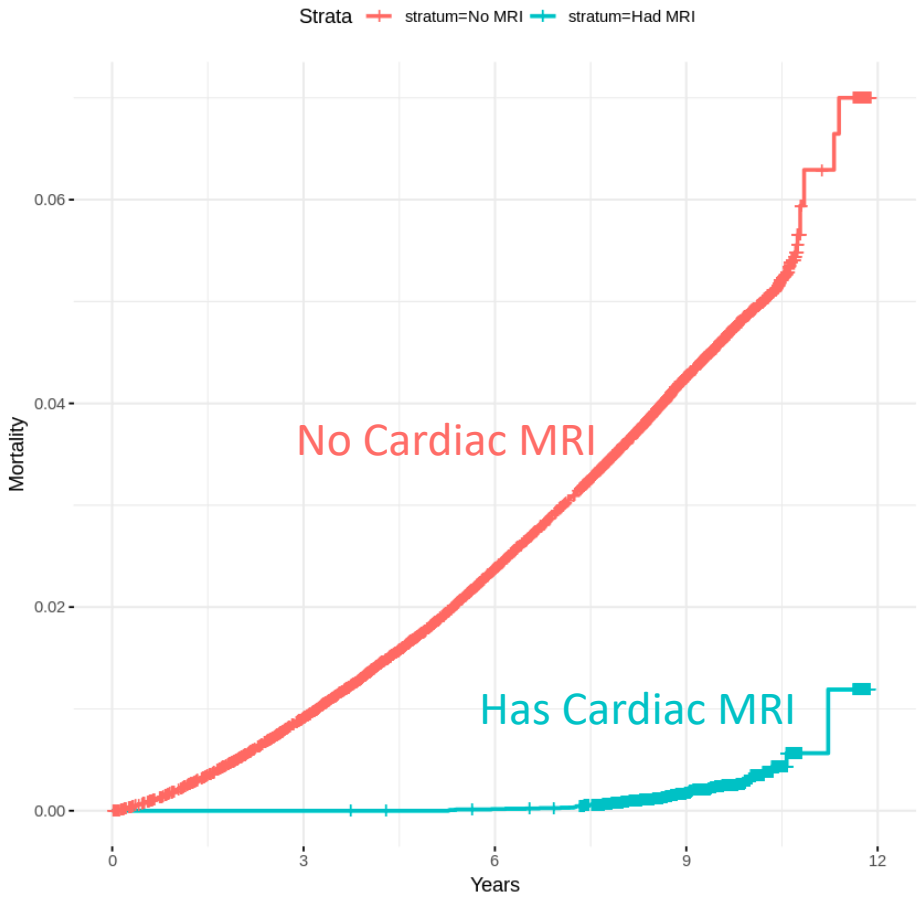
Randomized Trial

- Randomly assign patients to different groups
- Randomization deals with covariate imbalances
- Gold standard of evidence in medicine

Observational Data Is Messy

UK Biobank Data: Exome data and Cardiac MRI saves lives?



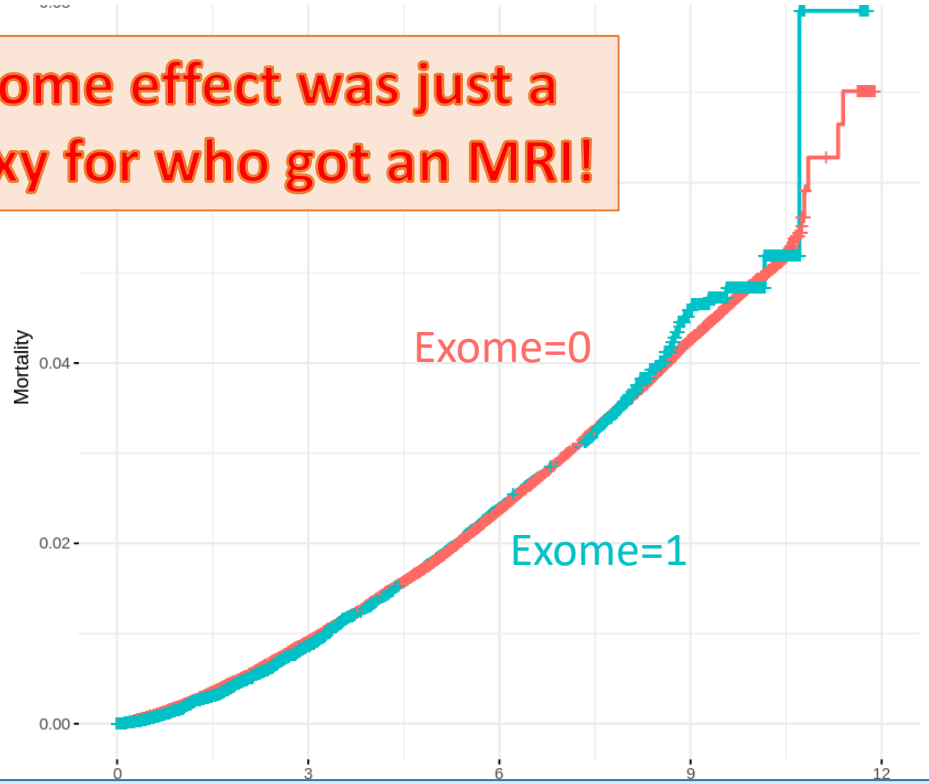


James Pirruccello Tweetorial on UK Biobank Immortal Time Bias 2019 - https://twitter.com/jpirruccello/status/1152252597723701249

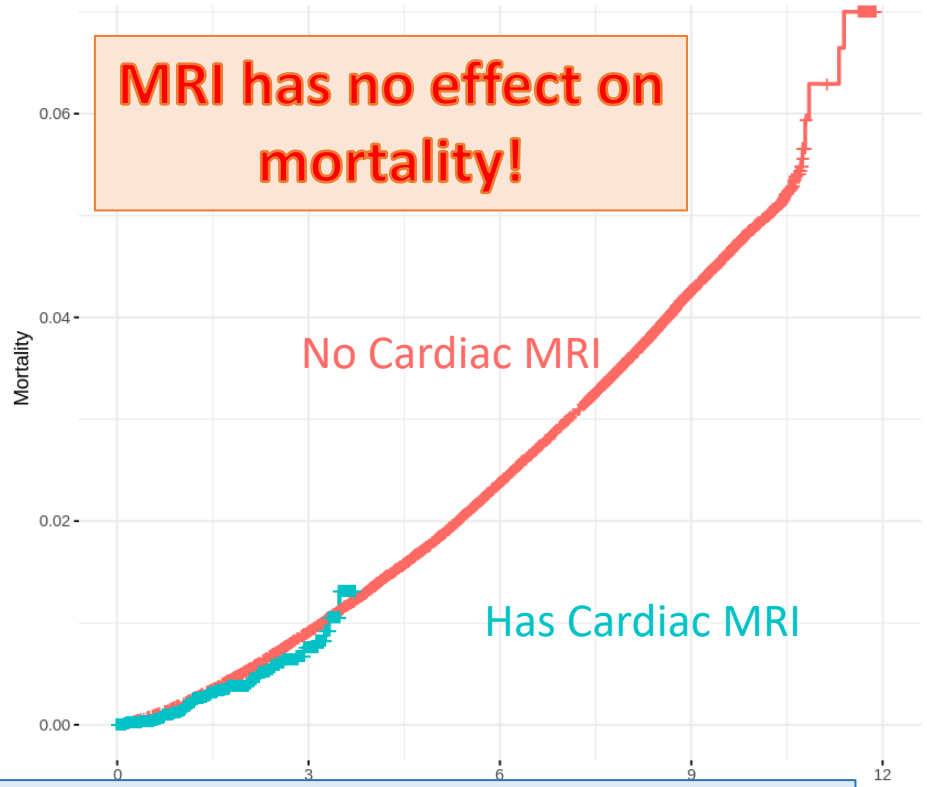
Observational Data Is Messy

Not so fast!!! We should examine from date of first MRI.

Exome effect was just a proxy for who got an MRI!



MRI has no effect on mortality!

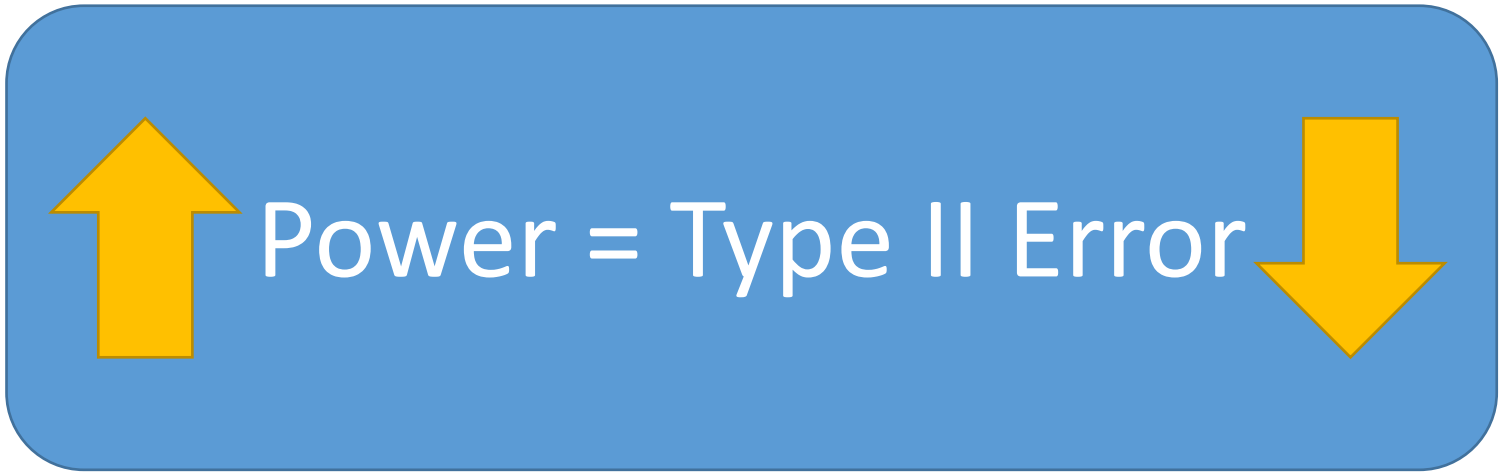


Observational studies have lots of issues that statistical tests won't catch! Be careful and don't blindly analyze data.

James Pirruccello Tweetorial on UK Biobank Immortal Time Bias 2019 - https://twitter.com/jpirruccello/status/1152252597723701249

Statistical Power

- Power = probability we reject H_0 when H_1 is true
  - i.e., how often do we find effects that actually exist
- Inversely related to rate of Type II errors



Let's Get Real About Power

- Assume treatment improves survival by 25% (HR = .75)
- GBM almost always lethal - assume 90% of patients die
- 50-50 split between treatment and control groups

Need lots of patients to have any chance of finding effect!

Sample	Power
20	9%
100	27%
400	78%
800	97%

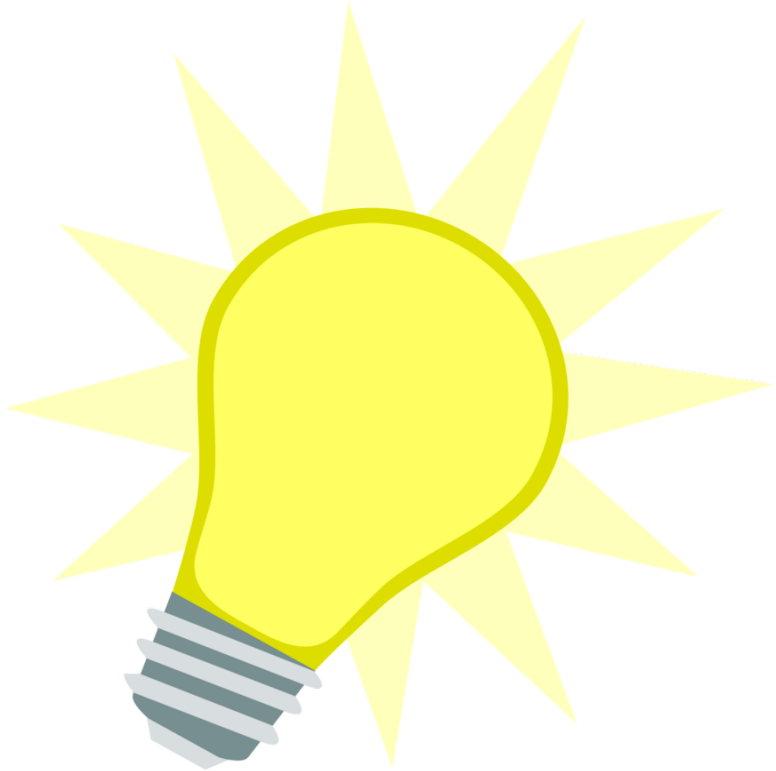
http://powerandsamplesize.com/Calculators/Test-Time-To-Event-Data/Cox-PH-2-Sided-Equality

Designing the Analysis

The most important part!

What Question Are We Trying to Answer

- Asking the right research questions is crucial!
- Question needs to be well-defined
- Define relevant effect size and power accordingly
- Confirmatory analysis – prespecified questions
- Exploratory analysis – hypothesis generating

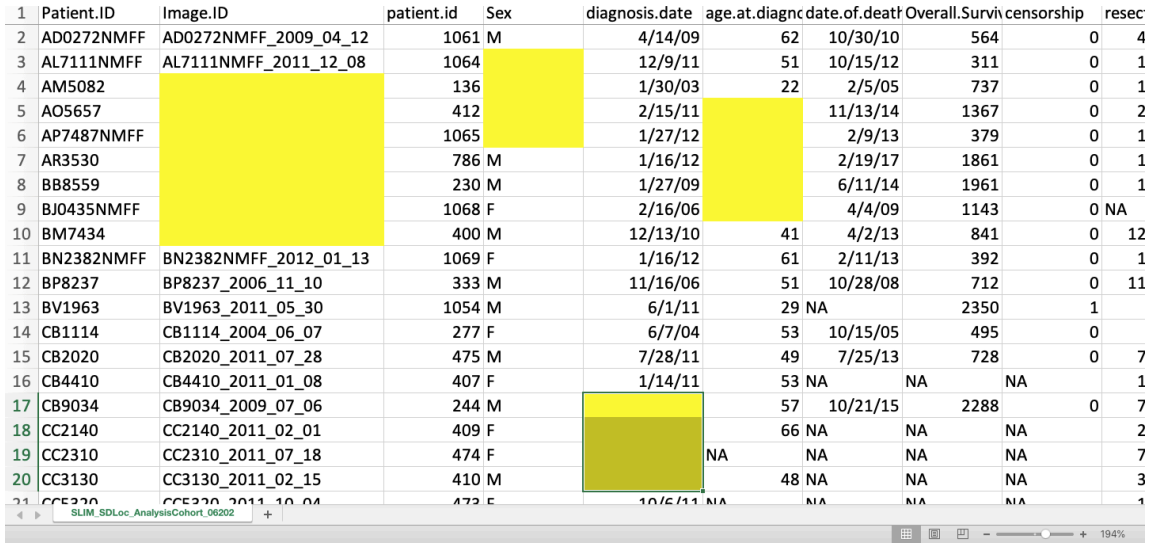


Sample Size Is Important

- Sample size affects power
  - More samples → more precision to detect smaller effect
- N = number of events NOT patients
- Adding covariates requires more events
- (Bad) rule of thumb: 15 events per variable (EPV)
  - In reality EPV much more complicated
- Subgroup heterogeneity harder to detect due to smaller N
  - Often need 4N patients

Don’t Discard Patients Missing Data

- Missing data decreases sample size
  - Smaller sample → decreased power
- Why is data missing?
  - Missing at random (MAR)
    - Relatively safe to exclude (better to impute!)
  - Informative missing (MI)
    - e.g., patients missing labs b/c died before labs could be collected
    - Simply excluding these patients could bias results
- Before fixing data, explore patterns in missing variables



	Patient.ID	Image.ID	patient.id	Sex	diagnosis.date	age.at.diagnc	date.of.death	Overall.Survi	censorship	resec
2	AD0272NMFF	AD0272NMFF_2009_04_12	1061	M	4/14/09	62	10/30/10	564	0	4
3	AL7111NMFF	AL7111NMFF_2011_12_08	1064		12/9/11	51	10/15/12	311	0	1
4	AM5082		136		1/30/03	22	2/5/05	737	0	1
5	AO5657		412		2/15/11		11/13/14	1367	0	2
6	AP7487NMFF		1065		1/27/12		2/9/13	379	0	1
7	AR3530		786	M	1/16/12		2/19/17	1861	0	1
8	BB8559		230	M	1/27/09		6/11/14	1961	0	1
9	BJ0435NMFF		1068	F	2/16/06		4/4/09	1143	0	NA
10	BM7434		400	M	12/13/10	41	4/2/13	841	0	12
11	BN2382NMFF	BN2382NMFF_2012_01_13	1069	F	1/16/12	61	2/11/13	392	0	1
12	BP8237	BP8237_2006_11_10	333	M	11/16/06	51	10/28/08	712	0	11
13	BV1963	BV1963_2011_05_30	1054	M	6/1/11	29	NA	2350	1	
14	CB1114	CB1114_2004_06_07	277	F	6/7/04	53	10/15/05	495	0	
15	CB2020	CB2020_2011_07_28	475	M	7/28/11	49	7/25/13	728	0	7
16	CB4410	CB4410_2011_01_08	407	F	1/14/11	53	NA	NA	NA	1
17	CB9034	CB9034_2009_07_06	244	M		57	10/21/15	2288	0	7
18	CC2140	CC2140_2011_02_01	409	F		66	NA	NA	NA	2
19	CC2310	CC2310_2011_07_18	474	F		NA	NA	NA	NA	7
20	CC3130	CC3130_2011_02_15	410	M		48	NA	NA	NA	3
21	CC5320	CC5320_2011_10_04	473	F	10/6/11	NA	NA	NA	NA	1

SLIM_SDLoc_AnalysisCohort_06202

194%

Imputation Gives Us The Power

Idea: Can we guess these missing values?

- Imputation lets us guess data that is MAR
- Better to guess missing values than reduce sample size
- MICE imputation guesses data from non-missing data
- Imputing multiple times maintains inference validity

```
impute_transform <- mice(patients, method = 'pmm')
S <- Surv(patients$os.time, patients$status == 0)
fit <- fit.mult.impute(S ~ rcs(age, 3) + resection + rcs(d.rho, 3) +
                       rcs(kps, 3), cph, impute_transform, data = patients)
```

Imputation Guidelines

How much is missing?	**Recommendations**
Less than 3%	Median imputation or exclude patients with missing data
Greater than 3%	MICE with max(5, 100x) imputations
Multiple predictors frequently missing	Sensitivity analysis with more imputations

Univariate Feature Selection is Bad

- Common to screen variables with univariate model
- Then only include significant variables in final model
- This is a form of stepwise variable selection
- 9/10 of statisticians agree: don't use stepwise variable selection!

Table 2. Results of Univariate Analysis of Event-free and Overall Survival Rates.*

Cohort and Marker	No. of Patients	3-Yr Event-free Survival	P Value	3-Yr Overall Survival	P Value
		%		%	
All patients					
1p36			<0.001		<0.001
No loss	689	77±2		85±2	
LOH	209	47±4		64±4	
Unb11q LOH status					
Not unbalanced	758	74±2	<0.001	83±2	<0.001
Unbalanced	151	50±5		66±5	
***MYCN* not amplified**					
1p36			<0.001		0.05
No loss	644	79±2		87±2	
LOH	100	62±6		83±5	
Unb11q LOH status			<0.001		<0.001
Not unbalanced	617	82±2		91±2	
Unbalanced	137	52±5		68±5	

* Plus–minus values are rates ±SE. Two-sided P values were calculated with the use of the log-rank test. LOH denotes loss of heterozygosity.

Attiyeh et al. 2004

What Variables Do We Include in Our Model?

- Sample size can limit number of variables to include
- Variable selection procedures can impact inference
- Select features incorrectly → exaggerate results

https://xkcd.com/1838/

How To Choose Features

Proper selection strategies:

1. Domain expert specifies relevant variables
2. Use all variables and apply shrinkage
3. Data reduction blinded to response

- Leave insignificant in model – still add to predictions
- Think about relevant confounders

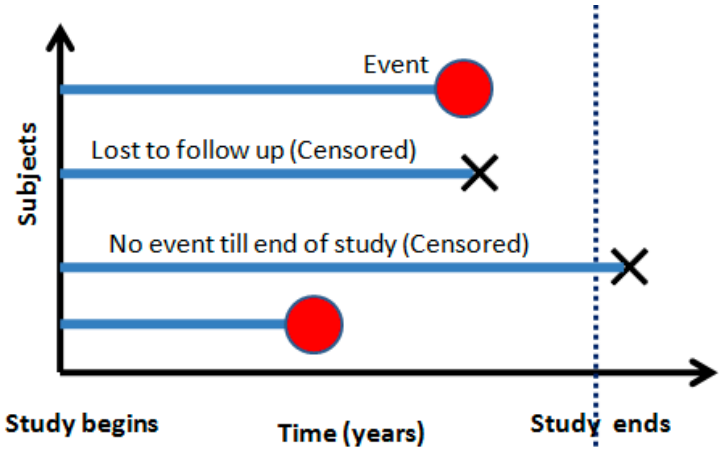
Akash Desarda – TowardsDataScience

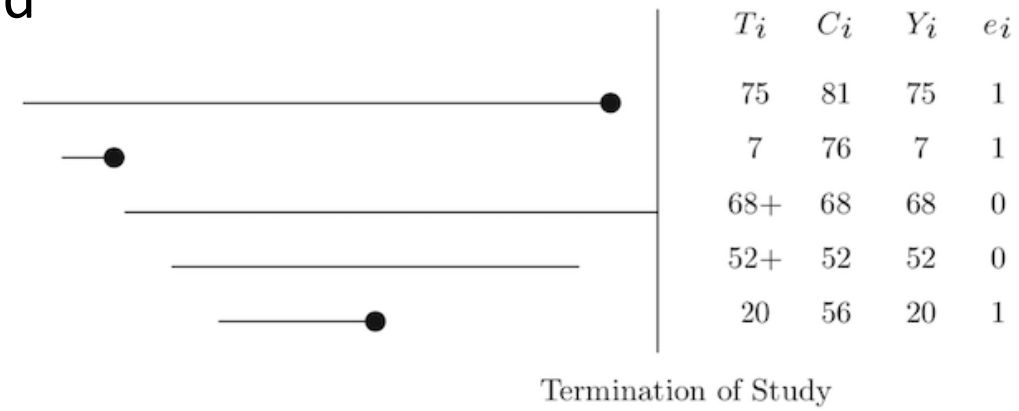
Analyzing Our Data

The fun part!

Survival Analysis

- Measuring time to an event
  - Death, Myocardial infarction, Tumor recurrence, Rubber band failure
- Patients either:
  - Experience the event
  - Are censored before the event is observed (lost to follow-up)
- Coded as 1 = event and 0 = censored
- Analyzing time survived vs binary dead/alive increases power

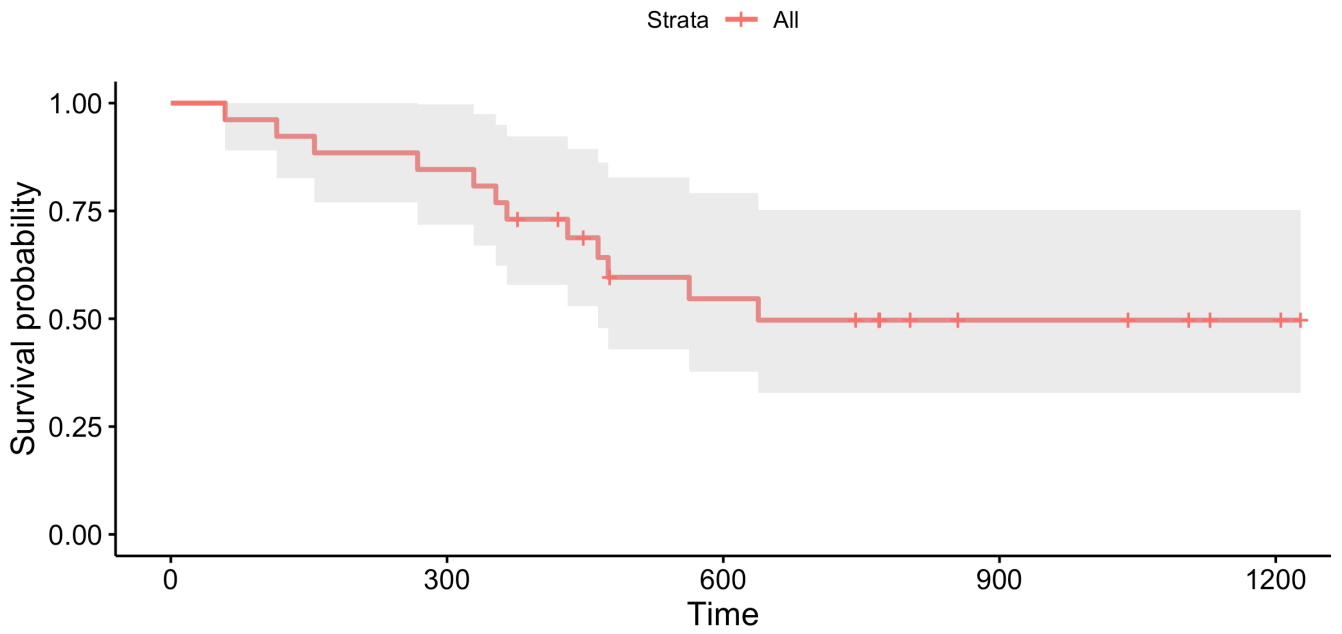




http://www.alphabetastats.com/?page_id=265

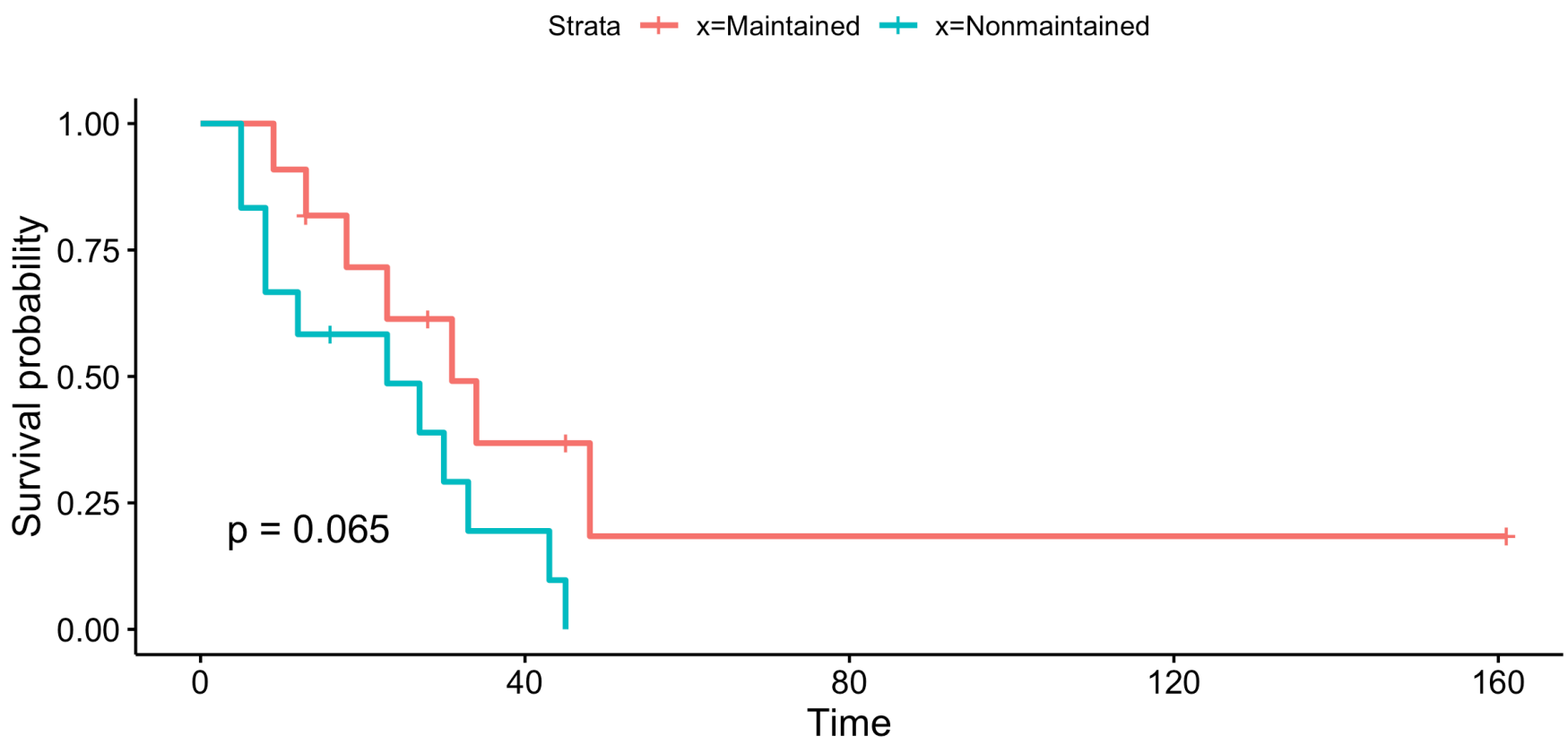
Kaplan-Meier Estimate

- Computes probability of event at time t
- Needed to handle censoring
- Great for summary statistics describing patient prognosis



Kaplan-Meier compares survival difference

- Log rank test - Test to compare survival between 2 groups



Should I use a Kaplan-Meier?

Yes!

- I have a discrete variable
- Patients were randomized
- I want to show summary statistics

- Examples
  - Male vs. Female
  - No treatment vs. treatment
  - Treatment A vs. B vs. C

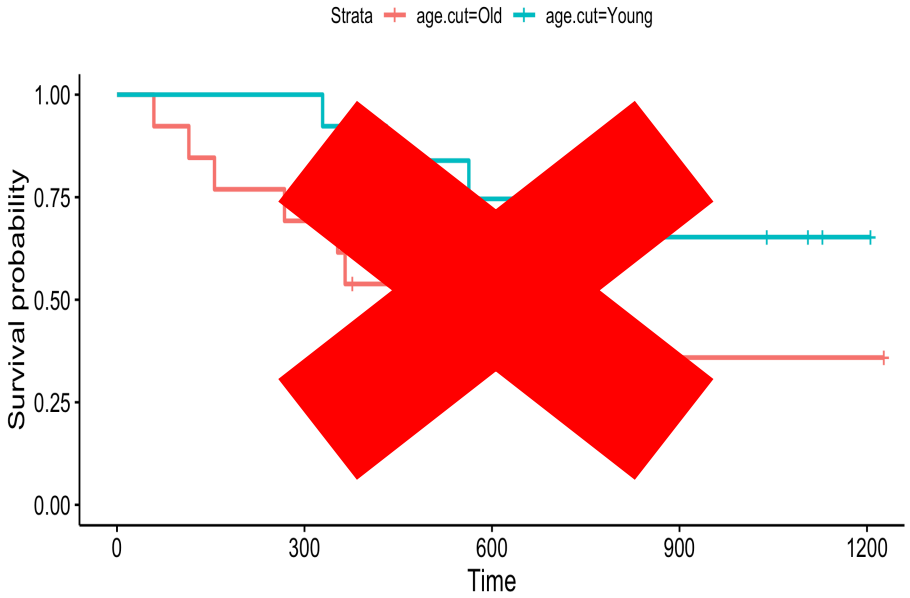
No!

- I have a continuous variable
- My variable is observational
- I need to adjust for other variables

- Examples
  - Age
  - Velocity
  - D/rho
  - Cholesterol

Dichotomization is the root of all evil

- Discards lots of info decreasing power
- Cutpoints are often arbitrary
  - Median? Mean?
  - What is high? What is low?
  - What is old? What is young?
  - What is fast? What is slow?
  - How many groups?
  - Should we trichotomize?
- Cutpoints often don’t reproduce

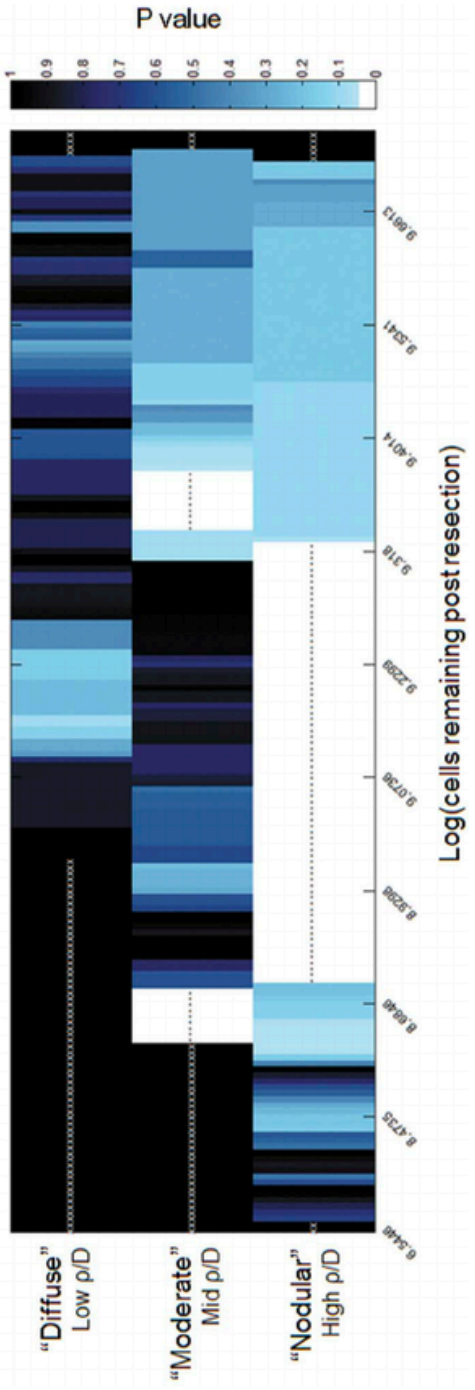


Let's Talk Iterative Kaplan-Meier

- Repeated testing inflates Type I error
- Dichotomization requires more patients to find effect
- No statistical literature on procedure
- MCMC simulation needed to correct p-values

Iterative Kaplan Meier should be avoided

Use Cox regression instead



Baldock 2014

Cox Proportional Hazards Regression

- Multiple regression model
- Model continuous variables
- Account for nonlinearity
- Adjust for covariates
- Make risk predictions

<u>Addresses many limitation of Kaplan-Meier estimates</u>

Example: cph function from *rms* package

```{r}
model <- cph(S ~ age + rx, data = prostate)
model
```

```
Frequencies of Missing Values Due to Each Variable
  S age  rx
  0   1   0

Cox Proportional Hazards Model

 cph(formula = S ~ age + rx, data = prostate)


                      Model Tests       Discrimination
                                           Indexes
Obs       501    LR chi2     21.56    R2       0.042
Events    354    d.f.            4    Dxy      0.142
Center 1.9731    Pr(> chi2) 0.0002    g        0.286
                 Score chi2  19.99    gr       1.332
                 Pr(> chi2) 0.0005

                    Coef    S.E.   Wald Z Pr(>|Z|)
age                 0.0286 0.0084  3.40  0.0007
rx=0.2 mg estrogen  0.0372 0.1451  0.26  0.7977
rx=1.0 mg estrogen -0.3513 0.1570 -2.24  0.0253
rx=5.0 mg estrogen  0.0313 0.1461  0.21  0.8304
```

Interpreting Cox Model

- Coef = ln(Hazard Ratio)

```{r}
model <- cph(S ~ age + rx, data = prostate)
model
```

```
Frequencies of Missing Values Due to Each Variable
  S age  rx
  0   1   0

Cox Proportional Hazards Model

 cph(formula = S ~ age + rx, data = prostate)


                      Model Tests       Discrimination
                                           Indexes
 Obs       501    LR chi2     21.56    R2       0.042
 Events    354    d.f.            4    Dxy      0.142
 Center 1.9731    Pr(> chi2) 0.0002    g        0.286
                  Score chi2  19.99    gr       1.332
                  Pr(> chi2) 0.0005

                     Coef    S.E.   Wald Z Pr(>|Z|)
 age                  0.0286 0.0084  3.40  0.0007
 rx=0.2 mg estrogen   0.0372 0.1451  0.26  0.7977
 rx=1.0 mg estrogen  -0.3513 0.1570 -2.24  0.0253
 rx=5.0 mg estrogen   0.0313 0.1461  0.21  0.8304
```

Interpreting Cox Model

- Coef = ln(Hazard Ratio)
- S.E. = Standard Error

```r
model <- cph(S ~ age + rx, data = prostate)
model
```

```
Frequencies of Missing Values Due to Each Variable
  S age  rx
  0   1   0

Cox Proportional Hazards Model

 cph(formula = S ~ age + rx, data = prostate)


                      Model Tests       Discrimination
                                           Indexes
Obs       501    LR chi2     21.56    R2       0.042
Events    354    d.f.            4    Dxy      0.142
Center 1.9731    Pr(> chi2) 0.0002    g        0.286
                 Score chi2  19.99    gr       1.332
                 Pr(> chi2) 0.0005

                    Coef    S.E.   Wald Z Pr(>|Z|)
age                 0.0286 0.0084  3.40  0.0007
rx=0.2 mg estrogen  0.0372 0.1451  0.26  0.7977
rx=1.0 mg estrogen -0.3513 0.1570 -2.24  0.0253
rx=5.0 mg estrogen  0.0313 0.1461  0.21  0.8304
```

Interpreting Cox Model

- Coef = ln(Hazard Ratio)
- S.E. = Standard Error
- Pr(>|Z|) = p-value

```r
model <- cph(S ~ age + rx, data = prostate)
model
```

```
Frequencies of Missing Values Due to Each Variable
  S age  rx
  0   1   0

Cox Proportional Hazards Model

 cph(formula = S ~ age + rx, data = prostate)


                      Model Tests       Discrimination
                                           Indexes
Obs       501    LR chi2     21.56    R2       0.042
Events    354    d.f.            4    Dxy      0.142
Center 1.9731    Pr(> chi2) 0.0002    g        0.286
                 Score chi2  19.99    gr       1.332
                 Pr(> chi2) 0.0005

                    Coef    S.E.   Wald Z Pr(>|Z|)
age                 0.0286 0.0084  3.40  0.0007
rx=0.2 mg estrogen  0.0372 0.1451  0.26  0.7977
rx=1.0 mg estrogen -0.3513 0.1570 -2.24  0.0253
rx=5.0 mg estrogen  0.0313 0.1461  0.21  0.8304
```

Interpreting Cox Model

- HR = 1: No difference
- HR < 1: Better survival
- HR > 1: Worse survival

Beware!

- Difference in HR vs ln(HR) scale
- Direction of effect can be flipped by coding

```{r}
model <- cph(S ~ age + rx, data = prostate)
model
```

```
Frequencies of Missing Values Due to Each Variable
  S age  rx
  0   1   0

Cox Proportional Hazards Model

 cph(formula = S ~ age + rx, data = prostate)


                      Model Tests       Discrimination
                                           Indexes
Obs       501    LR chi2     21.56    R2       0.042
Events    354    d.f.            4    Dxy      0.142
Center 1.9731    Pr(> chi2) 0.0002    g        0.286
                 Score chi2  19.99    gr       1.332
                 Pr(> chi2) 0.0005

                    Coef    S.E.   Wald Z Pr(>|Z|)
age                 0.0286 0.0084  3.40  0.0007
rx=0.2 mg estrogen  0.0372 0.1451  0.26  0.7977
rx=1.0 mg estrogen -0.3513 0.1570 -2.24  0.0253
rx=5.0 mg estrogen  0.0313 0.1461  0.21  0.8304
```

Patient Heterogeneity: Subgroup Analysis

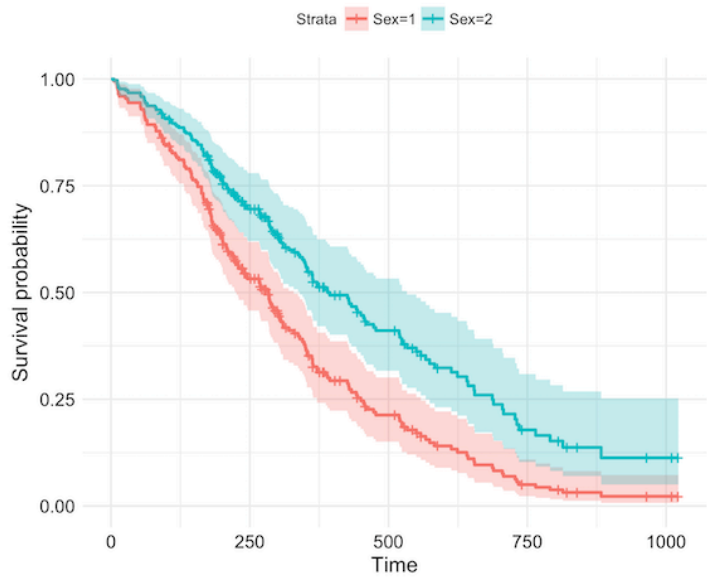
- Different treatment effect in different groups
- Use **interaction term** to test for heterogeneity
- Check for significance with anova()
- If significant must interpret effects differently!

```
cph(S ~ rcs(age, 3) + bev*sex, data = gbm)
```

```
Cox Proportional Hazards Model

 cph(formula = S ~ rcs(age, 3) + bev * sex, data = gbm)

                      Model Tests       Discrimination
                                           Indexes
 Obs       256    LR chi2     46.65    R2       0.167
 Events    219    d.f.            5    Dxy      0.300
 Center 1.2566    Pr(> chi2) 0.0000    g        0.585
                  Score chi2  50.06    gr       1.795
                  Pr(> chi2) 0.0000

                 Coef    S.E.   Wald Z Pr(>|Z|)
 age              0.0187 0.0108  1.74  0.0821
 age'             0.0199 0.0139  1.43  0.1529
 bev=Y           -0.4710 0.2854 -1.65  0.0988
 sex=M            0.2287 0.1608  1.42  0.1548
 bev=Y * sex=M -0.0724 0.3432 -0.21  0.8328
```

Check your model's assumptions

- Cox assumes hazard ratio constant over time (proportional hazards)
- If PH not met:
  - Try non-linear fits for continuous variables
  - Stratify by variable
  - Time dependent interaction
  - Switch to AFT models

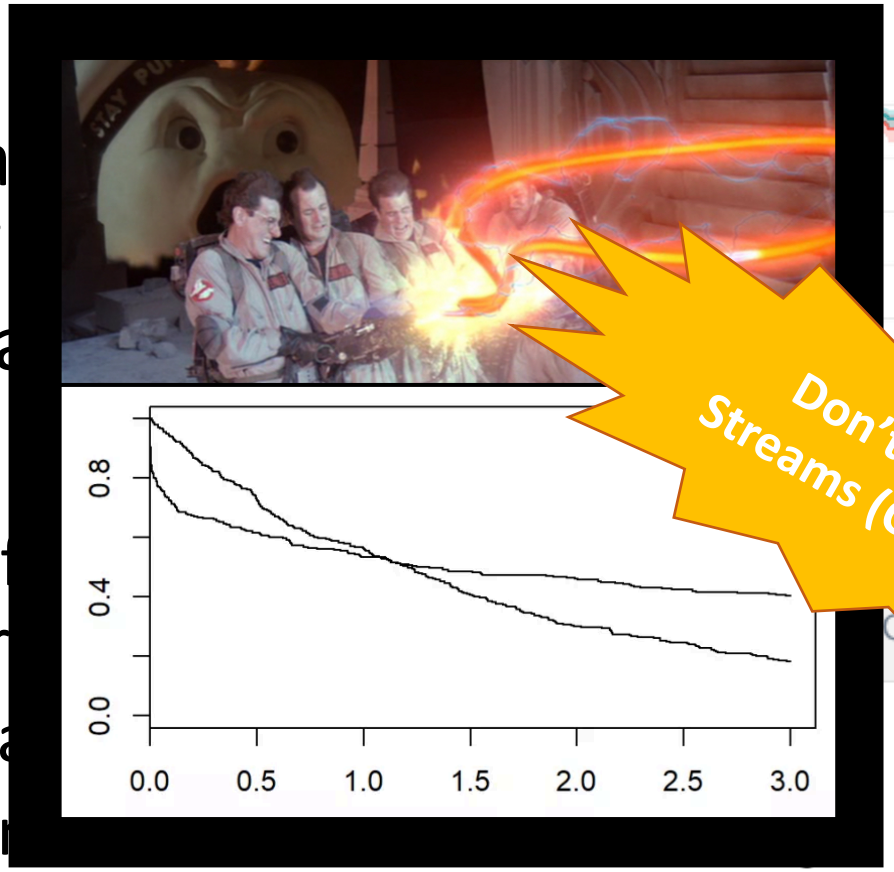


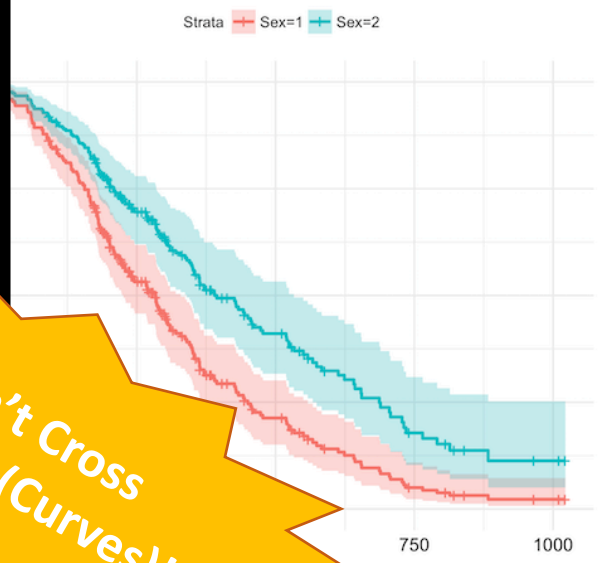

```
cox.zph(ph.fit)
```

	rho	chisq	p
age	0.00632	0.01350	0.907
age'	-0.04217	0.61655	0.432
age''	0.03370	0.40159	0.526
age'''	-0.02552	0.22934	0.632
rx=0.2 mg estrogen	-0.00200	0.00143	0.970
rx=1.0 mg estrogen	-0.08799	2.80504	0.094
rx=5.0 mg estrogen	-0.03107	0.34863	0.555
GLOBAL	NA	7.87136	0.344

Check your model's assumptions

- Cox assumes ha
  constant over ti
  (proportional ha
- If PH not met:
  - Try non-linear t
    continuous var
  - Stratify by varia
  - Time depender
  - Switch to AFT models



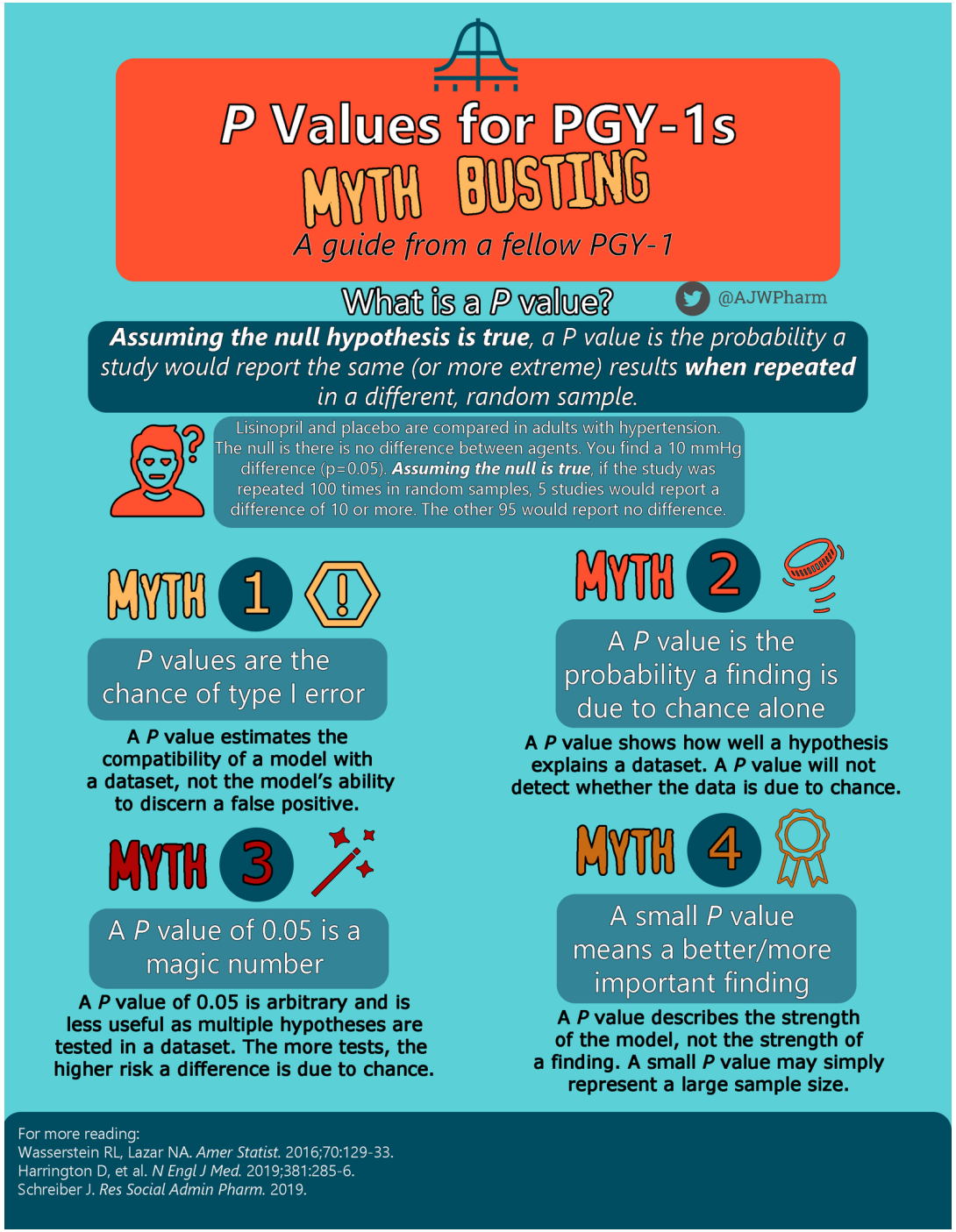


Don't Cross Streams (Curves)!

	rho	chisq	p
	0.00632	0.01350	0.907
	-0.04217	0.61655	0.432
	0.03370	0.40159	0.526
age'''	-0.02552	0.22934	0.632
rx=0.2 mg estrogen	-0.00200	0.00143	0.970
rx=1.0 mg estrogen	-0.08799	2.80504	0.094
rx=5.0 mg estrogen	-0.03107	0.34863	0.555
GLOBAL	NA	7.87136	0.344

Interpreting Results

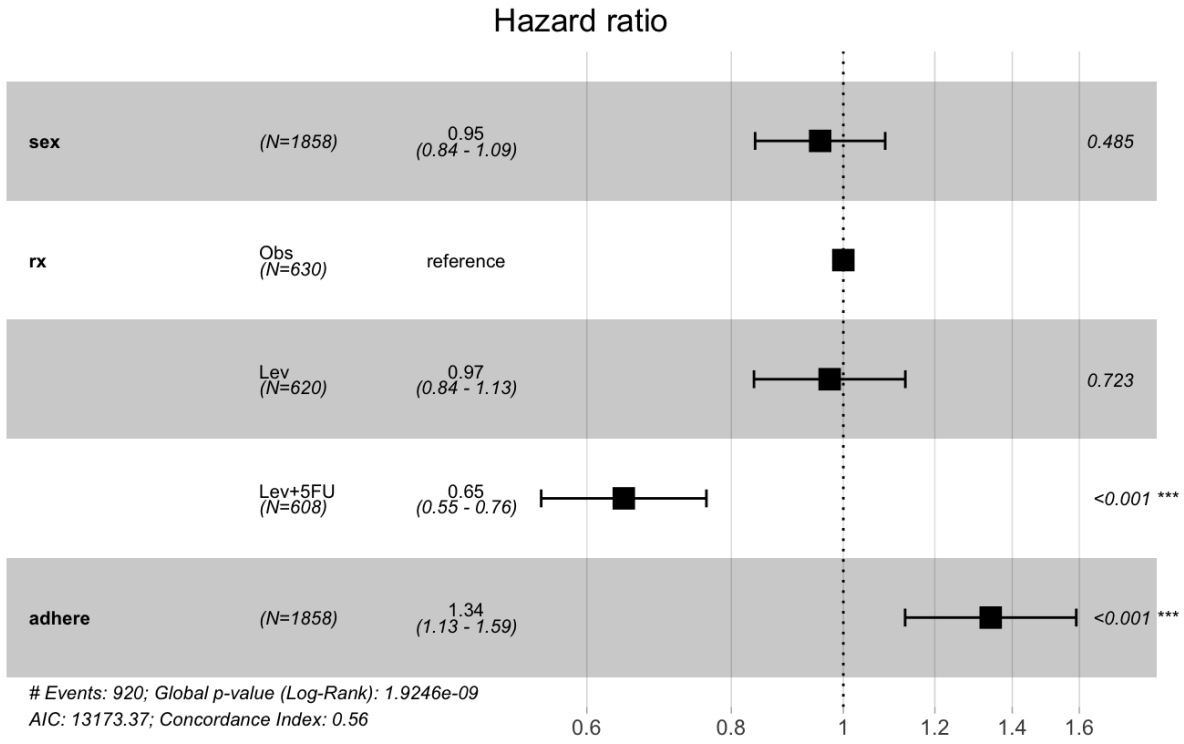
- P-value ≥ .05 doesn't mean no effect!
- Don't dichotomize findings based on p-values
- Confidence intervals give an idea of uncertainty – how wide are they?
- Think about power limitations of the study



Andy Webb PharmD on Twitter 2019

Reporting Results

- Report confidence intervals
- Don’t dichotomize findings – use CIs
- Report prespecified vs exploratory analyses



https://rpkgs.datanovia.com/survminer/reference/ggforest.html

There's Always More Out There…

- Non-linear effects
- Evaluating biomarker efficacy
- Validating prediction models
- Data reduction strategies
- Modeling in high dimensional spaces
- Simulation experiments and power calculations
- Randomized controlled trial design
- Multiple comparisons correction
- Data visualization

Ask us if you want more info on any of these topics!

Questions?

Modeling Nonlinear Effects

- Restricted cubic splines relax linearity assumptions
- Use rcs() function in R
- Number of **knots** determine how well spline fits data
- Use 3 to 5 knots depending on how much data you have

```{r}
nonlinear.fit <- cph(S ~ rcs(age, 5) + rx, data = prostate)
nonlinear.fit
```

```
Frequencies of Missing Values Due to Each Variable
  S age  rx
  0   1   0

Cox Proportional Hazards Model

 cph(formula = S ~ rcs(age, 5) + rx, data = prostate)


                      Model Tests       Discrimination
                                           Indexes
 Obs       501    LR chi2     30.12    R2       0.058
 Events    354    d.f.            7    Dxy      0.160
 Center 0.4458    Pr(> chi2) 0.0001    g        0.318
                  Score chi2  32.97    gr       1.374
                  Pr(> chi2) 0.0000

                    Coef    S.E.   Wald Z Pr(>|Z|)
 age                 0.0053 0.0278  0.19  0.8493
 age'               -0.0034 0.0601 -0.06  0.9554
 age''               1.1672 1.4950  0.78  0.4350
 age'''             -2.6326 3.4607 -0.76  0.4468
 rx=0.2 mg estrogen  0.0532 0.1455  0.37  0.7146
 rx=1.0 mg estrogen -0.3695 0.1576 -2.34  0.0190
 rx=5.0 mg estrogen  0.0110 0.1493  0.07  0.9411
```

Interpreting (Non-linear) Cox Model

- Use anova() to test if variable is associated with prognosis
- Only gives p-value not effect size
- Need to plot predictor vs HR for nonlinear variables to see effect

```{r}
anova(nonlinear.fit)
```

Wald Statistics Response: S

Factor	Chi-Square	d.f.	P
age	23.56	4	0.0001
Nonlinear	9.01	3	0.0292
rx	8.78	3	0.0324
TOTAL	32.01	7	<.0001

Plotting Predictor vs HR

- HR = 1: No difference
- HR < 1: Better survival
- HR > 1: Worse survival
- Beware HR vs ln(HR) scale