Fundamentals of Causal Inference with R: Module 4 Mediation and Time-Dependent Confounding

Babette Brumback

Professor Emerita
University of Florida, Biostatistics Department

Mediation: Introduction

- Mediation analyses address questions about the mechanism M of the causal effect of A on Y.
- Does a new treatment A act by increasing absolute lymphocyte counts M which in turn improves Covid-19 outcomes Y?
- A difficulty with mediation analyses is that although we can randomize A, we typically cannot randomize M.
- Therefore, we must consider adjusting for confounding by H_2 of the effect of M on Y.
- Sometimes, we cannot randomize A either.
- In that case, we additionally must consider confounding by H_3 of the effect of A on M and confounding by H_1 of the effect of A on Y.

Introduction

- Assumptions needed for confounding adjustment are depicted in the causal DAG of Figure 1.
- In the randomized trial, confounders H_1 and H_3 are absent, and we need only adjust for H_2.
- The next section presents the theory behind causal methods for mediation analyses assuming the causal DAG of Figure 1 holds and that the confounders H_1, H_2, and H_3 are observed.
- We reference Vanderweele (2015) for these ideas and more about causal mediation.

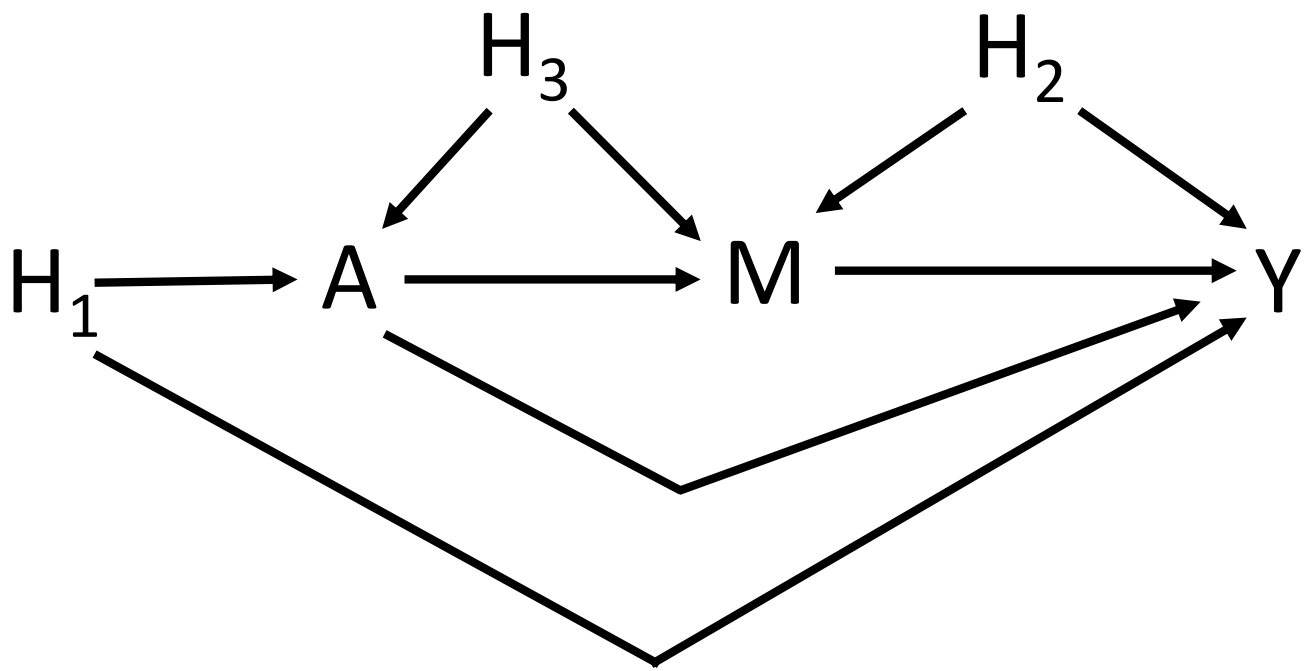

Figure 1: Causal DAG for Mediation by M of the Effect of A on Y

Theory

- We introduce potential outcomes notation for both Y and M. We let $M(a)$ be the potential outcome of the mediator M to assigning treatment $A = a$.
- We let $Y(a, m)$ be the potential outcome for the outcome Y to assigning treatment $A = a$ and mediator $M = m$.
- We also consider potential outcomes of the form $Y(a, M(a^*))$, that is, we assign treatment $A = a$ and then we assign the mediator M to its potential outcome had we assigned treatment $A = a^*$.
- We make consistency assumptions for all potential outcomes.
- Analysis of mediation partitions the total effect into direct and indirect effects.

Theory

- The *total effect* (TE) compares the potential outcome for Y to assigning treatment $A = a$ with that of assigning treatment $A = a^*$:

$$TE(a, a^*) = Y(a, M(a)) - Y(a^*, M(a^*)).$$

- The *controlled direct effect* (CDE) compares the potential outcome to assigning treatment $A = a$ with that of assigning treatment $A = a^*$ while controlling the mediator at $M = m$:

$$CDE(a, a^*, m) = Y(a, m) - Y(a^*, m).$$

- The *natural direct effect* (NDE) compares the potential outcome to assigning treatment $A = a$ with that of assigning treatment $A = a^*$ while assigning $M = M(a^*)$:

$$NDE(a, a^*; a^*) = Y(a, M(a^*)) - Y(a^*, M(a^*)).$$

- The *natural indirect effect* (NIE) compares the potential outcome to assigning $M = M(a)$ with that of assigning $M = M(a^*)$ while assigning $A = a$:

$$NIE(a, a^*; a) = Y(a, M(a)) - Y(a, M(a^*)).$$

- Note that $TE = CDE + CIE$ and also that $TE = NDE + NIE$.

Theory

- We cannot identify these effects on an individual level, but we can identify them on a population level in terms of their expected values.
- For estimation, we introduce the following four assumptions about the confounders:
  1. There are no unmeasured treatment-outcome confounders given H (where $H = (H_1, H_2, H_3)$)
  2. There are no unmeasured mediator-outcome confounders given (H, A)
  3. There are no unmeasured treatment-mediator confounders given H
  4. There is no mediator-outcome confounder that is affected by treatment (i.e. no arrow from A to H_2)
- For the CDE and CIE, only assumptions (1) and (2) are needed. For the NDE and NIE, we need all four assumptions. Assumptions (1) and (3) are guaranteed when A is randomized.

Theory

- To construct estimators of our causal quantities, we formalize these assumptions in terms of our potential outcomes:
  1. $Y(a, m) \amalg A|H$
  2. $Y(a, m) \amalg M|H, A$
  3. $M(a) \amalg A|H$
  4. $Y(a, m) \amalg M(a^*)|H$
- Again, for the CDE and CIE, only (1) and (2) are needed, and once again, assumptions (1) and (3) are guaranteed when A is randomized.

Theory

- Under assumption (1), the total effect conditional on H is given by

$$E(TE(a, a^*)|H = h) = E(Y|A = a, H = h) - E(Y|A = a^*, H = h).$$

- We have already proved the two parts of this result in Module 1.

Theory

- Under assumptions (1) and (2), the controlled direct effect conditional on H is given by:

$$E(CDE(a, a^*; m)|H = h) =$$
$$E(Y|A = a, M = m, H = h) - E(Y|A = a^*, M = m, H = h).$$

- To prove this, recall that $CDE(a, a^*; m) = Y(a, m) - Y(a^*, m)$.
- By assumption (1), $E(Y(a, m)|H) = E(Y(a, m)|A, H)$.
- By assumption (2), $E(Y(a, m)|A, H) = E(Y(a, m)|A, M, H)$.
- By consistency, $E(Y(a, m)|A = a, M = m, H) = E(Y|A = a, M = m, H)$.

Theory

- Under assumptions (1)-(4), the natural direct effect conditional on $H = h$ is

$$E(NDE(a, a^*; a^*)|H = h) =$$
$$\Sigma_m \{E(Y|A = a, m, h) - E(Y|A = a^*, m, h)\} P(M = m|A = a^*, h). \quad (1)$$

- Under assumptions (1)-(4), the natural indirect effect conditional on $H = h$ is:

$$E(NIE(a, a^*; a)|H = h) =$$
$$\Sigma_m E(Y|A = a, m, h) \{P(M = m|A = a, h) - P(M = m|A = a^*, h)\}. \quad (2)$$

Theory

- First, we prove (1).
  1. Recall, $NDE(a, a^*; a^*) = Y(a, M(a^*)) - Y(a^*, M(a^*))$.
  2. By the law of total expectation, $E(Y(a, M(a^*))|H) = \Sigma_m E(Y(a, M(a^*))|M(a^*) = m, H)P(M(a^*) = m|H)$.
  3. By consistency, the right hand side equals $\Sigma_m E(Y(a, m)|M(a^*) = m, H)P(M(a^*) = m|H)$.
  4. By assumptions (3) and (4), the right hand side equals $\Sigma_m E(Y(a, m)|H)P(M(a^*) = m|A = a^*, H)$.
  5. By consistency and assumption (1), the right hand side equals $\Sigma_m E(Y(a, m)|A = a, H)P(M = m|A = a^*, H)$.
  6. By assumption (2), the right hand side equals $\Sigma_m E(Y(a, m)|A = a, M = m, H)P(M = m|A = a^*, H)$.
  7. By consistency, the right hand side equals $\Sigma_m E(Y|A = a, M = m, H)P(M = m|A = a^*, H)$.
  8. Similarly, we can show $E(Y(a^*, M(a^*))|H) = \Sigma_m E(Y|A = a^*, M = m, H)P(M = m|A = a^*, H)$.

- Second, we prove (2).
  1. Recall, $NIE(a, a^*; a) = Y(a, M(a)) - Y(a, M(a^*))$.
  2. By the previous proof we can show $E(Y(a, M(a))|H) = \Sigma_m E(Y|A = a, M = m, H)P(M = m|A = a, H)$.
  3. By the previous proof we can also show $E(Y(a, M(a^*))|H) = \Sigma_m E(Y|A = a, M = m, H)P(M = m|A = a^*, H)$.
  4. Thus

$$E(NIE(a, a^*; a)|H) =$$
$$\Sigma_m E(Y|A = a, M = m, H)\{P(M = m|A = a, H) - P(M = m|A = a^*, H)\}.$$

Theory

- We can estimate all of these conditional effects by substituting parametric or nonparmetric estimators of the probabilities and expectations in the expressions involving the observed data.
- To estimate unconditional effects, we can average over the distribution of H in the population. For sampling variability, we can use the bootstrap.

Theory

- We may often want to know how much of the total effect is mediated by M.
- Using the CDE and CIE partition, we can estimate the proportion of the effect that would be eliminated if we fixed the mediator to a specific level $M = m$:

$$PE = (TE - CDE(m))/TE.$$

- Using the NDE and NIE partition, we can estimate the proportion mediated as

$$PM = NIE/TE.$$

- In some applications, these measures can be less than zero or greater than one.
- In those cases, we do not use them, but rather restrict our description of the mediation to the CDE and CIE or the NDE and NIE.

Theory

- In several applications, the dimension of H requires parametric methods.
- For example, we could use parametric models that accomodate treatment-mediator interaction:

$$E(Y|A=a, M=m, H=h) = \beta_0 + \beta_1 a + \beta_2 m + \beta_3 am + \beta_4^T h \tag{3}$$

-

$$E(M|A=a, H=h) = \alpha_0 + \alpha_1 a + \alpha_2^T h \tag{4}$$

- Under assumptions (1) through (4), we can combine the estimates from the two models to obtain the following formulas for unconditional direct and indirect effects, comparing exposure levels a and a^*:

$$\begin{aligned} CDE(a, a^*; m) &= (\beta_1 + \beta_3 m)(a - a^*) \\ NDE(a, a^*; a^*) &= (\beta_1 + \beta_3(\alpha_0 + \alpha_1 a + \alpha_2^T E(H)))(a - a^*) \\ NIE(a, a^*; a) &= (\beta_2 \alpha_1 + \beta_3 \alpha_1 a)(a - a^*) \end{aligned}$$

Traditional Parametric Methods

- Two traditional parametric methods assume models (3) and (4) but without an interaction in (3), that is, with $\beta_3 = 0$.
- In that case, we have In that case, we have

$$CDE(a, a^*; m) = \beta_1(a - a^*)$$

and

$$NIE(a, a^*; a) = \beta_2\alpha_1(a - a^*).$$

- Letting $a = 1$ and $a^* = 0$, the NIE is estimated by $\hat{\beta}_2\hat{\alpha}_1$. This is called the *product method* in the literature.

Traditional Parametric Methods

- Alternatively, we can estimate the NIE by comparing the coefficient of a in

$$E(Y|A = a, H = h) = \beta_0 + \beta_1 a + \beta_2(\alpha_0 + \alpha_1 a + \alpha_2^T h) + \beta_4^T h,$$

  which equals $\beta_1 + \beta_2\alpha_1$, to the coefficient of a in

$$E(Y|A = a, M = m, H = h) = \beta_0 + \beta_1 a + \beta_2 m + \beta_4^T h,$$

  which equals β_1.
- The idea is to see if the coefficient $\beta_1 + \beta_2\alpha_1$ is significantly reduced by adding M to the regression.
- We see it is reduced by the difference of those two coefficients, which is $\beta_2\alpha_1$, which equals the NIE of the product method.
- This method is known as the *difference method* in the literature.
- We estimate sampling variability using the bootstrap.

Examples

- We present two examples, letting $a = 1$ and $a^* = 0$ in the estimation of the unconditional TE, CDE, CIE, NDE, and NIE.
- First, we investigate whether the effect of naltrexone on unsuppressed viral load is mediated by reduced drinking, adjusting for confounding by baseline antiretroviral adherence, using data from the Double What-If? Study.
- We reassign variables so that A is naltrexone, H is baseline adherence, M is reduced drinking, and Y is unsuppressed viral load.
- For teaching purposes, we compute the causal estimates two different ways.
- First, we use `mediation.r`, and second, we use `nonparamediation.r`.

Examples

```
> mediation.r
function(dat=meddoublewhatifdat,ids=c(1:nrow(meddoublewhatifdat)))
{
dat<-dat[ids,]
PH0<-1-mean(dat$H)
PH1<-mean(dat$H)
EY1M0H0<-(mean(dat$Y[(dat$A==1)&(dat$M==0)&(dat$H==0)])*
(1-mean(dat$M[(dat$A==0)&(dat$H==0)]))
+ mean(dat$Y[(dat$A==1)&(dat$M==1)&(dat$H==0)])*
(mean(dat$M[(dat$A==0)&(dat$H==0)])))
EY1M0H1<-(mean(dat$Y[(dat$A==1)&(dat$M==0)&(dat$H==1)])*
(1-mean(dat$M[(dat$A==0)&(dat$H==1)]))
+ mean(dat$Y[(dat$A==1)&(dat$M==1)&(dat$H==1)])*
(mean(dat$M[(dat$A==0)&(dat$H==1)])))
EY1M0<-EY1M0H0*PH0 + EY1M0H1*PH1
EY0M0H0<-(mean(dat$Y[(dat$A==0)&(dat$M==0)&(dat$H==0)])*
(1-mean(dat$M[(dat$A==0)&(dat$H==0)]))
+ mean(dat$Y[(dat$A==0)&(dat$M==1)&(dat$H==0)])*
(mean(dat$M[(dat$A==0)&(dat$H==0)])))
EY0M0H1<-(mean(dat$Y[(dat$A==0)&(dat$M==0)&(dat$H==1)])*
(1-mean(dat$M[(dat$A==0)&(dat$H==1)]))
+ mean(dat$Y[(dat$A==0)&(dat$M==1)&(dat$H==1)])*
(mean(dat$M[(dat$A==0)&(dat$H==1)])))
EY0M0<-EY0M0H0*PH0 + EY0M0H1*PH1
```

Examples

```
EY1M1H0<-(mean(dat$Y[(dat$A==1)&(dat$M==0)&(dat$H==0)])*
(1-mean(dat$M[(dat$A==1)&(dat$H==0)]))
+ mean(dat$Y[(dat$A==1)&(dat$M==1)&(dat$H==0)])*
(mean(dat$M[(dat$A==1)&(dat$H==0)])))
EY1M1H1<-(mean(dat$Y[(dat$A==1)&(dat$M==0)&(dat$H==1)])*
(1-mean(dat$M[(dat$A==1)&(dat$H==1)]))
+ mean(dat$Y[(dat$A==1)&(dat$M==1)&(dat$H==1)])*
(mean(dat$M[(dat$A==1)&(dat$H==1)])))
EY1M1<-EY1M1H0*PH0 + EY1M1H1*PH1
CDE1H0<-(mean(dat$Y[(dat$A==1)&(dat$M==1)&(dat$H==0)])
- mean(dat$Y[(dat$A==0)&(dat$M==1)&(dat$H==0)]))
CDE1H1<-(mean(dat$Y[(dat$A==1)&(dat$M==1)&(dat$H==1)])
- mean(dat$Y[(dat$A==0)&(dat$M==1)&(dat$H==1)]))
CDE1<-CDE1H0*PH0 + CDE1H1*PH1
CDE0H0<-(mean(dat$Y[(dat$A==1)&(dat$M==0)&(dat$H==0)])
- mean(dat$Y[(dat$A==0)&(dat$M==0)&(dat$H==0)]))
CDE0H1<-(mean(dat$Y[(dat$A==1)&(dat$M==0)&(dat$H==1)])
- mean(dat$Y[(dat$A==0)&(dat$M==0)&(dat$H==1)]))
CDE0<-CDE0H0*PH0 + CDE0H1*PH1
NDE<-EY1M0-EY0M0
NIE<-EY1M1-EY1M0
TE<-NDE+NIE
PM<-(NIE/TE)
PE1<-((TE-CDE1)/TE)
PE0<-((TE-CDE0)/TE)
c(TE,CDE0,CDE1,NDE,NIE,PE0,PE1,PM)
}
```

Examples

```
> nonparamediation.r
function(dat=meddoublewhatifdat,ids=c(1:nrow(meddoublewhatifdat)))
{
dat<-dat[ids,]
out<-glm(Y~A*M*H,family=binomial,data=dat)
med<-glm(M~A*H,family=binomial,data=dat)
dat10<-dat00<-dat01<-dat11<-dat
dat10$A<-1
dat10$M<-0
dat00$A<-0
dat00$M<-0
dat01$A<-0
dat01$M<-1
dat11$A<-1
dat11$M<-1
EY1M0H<-(predict(out,newdata=dat10,type="response")*
(1-predict(med,newdata=dat00,type="response"))
+ predict(out,newdata=dat11,type="response")*
predict(med,newdata=dat01,type="response"))
EY1M0<-mean(EY1M0H)
EY0M0H<-(predict(out,newdata=dat00,type="response")*
(1-predict(med,newdata=dat00,type="response"))
+ predict(out,newdata=dat01,type="response")*
predict(med,newdata=dat01,type="response"))
EY0M0<-mean(EY0M0H)
```

Examples

```
EY1M1H<-(predict(out,newdata=dat10,type="response")*
(1-predict(med,newdata=dat10,type="response"))
+ predict(out,newdata=dat11,type="response")*
predict(med,newdata=dat11,type="response"))
EY1M1<-mean(EY1M1H)
CDE1H<-(predict(out,newdata=dat11,type="response")-
predict(out,newdata=dat01,type="response"))
CDE1<-mean(CDE1H)
CDE0H<-(predict(out,newdata=dat10,type="response")-
predict(out,newdata=dat00,type="response"))
CDE0<-mean(CDE0H)
NDE<-EY1M0-EY0M0
NIE<-EY1M1-EY1M0
TE<-NDE+NIE
PM<-(NIE/TE)
PE1<-((TE-CDE1)/TE)
PE0<-((TE-CDE0)/TE)
c(TE,CDE0,CDE1,NDE,NIE,PE0,PE1,PM)
}
```

Examples

- These two functions both implement a nonparametric mediation analysis and return exactly the same answers, presented in Table 1.
- The second function is easily modified to incorporate a higher-dimensional H and also parametric modeling assumptions. Table 2 presents results of a parametric mediation analysis, letting

```
out<-glm(Y~A+M+H,family=binomial,data=dat)
```

```
med<-glm(M~A+H,family=binomial,data=dat)
```

Examples

Table 1: Nonparametric Mediation Analysis of the Double What-If? Study

Measure	Estimate	95% CI
TE	-0.151	(-0.202, -0.101)
CDE(0)	-0.042	(-0.111, 0.026)
CDE(1)	0.047	(-0.123, 0.217)
NDE	-0.036	(-0.101, 0.030)
NIE	-0.116	(-0.154, -0.078)
PE(0)	0.720	(0.308, 1.132)
PE(1)	1.310	(0.087, 2.534)
PM	0.765	(0.365, 1.166)

Examples

Table 2: Mediation Analysis of the Double What-If? Study with Parametric Assumptions

Measure	Estimate	95% CI
TE	-0.151	(-0.202, -0.101)
CDE(0)	-0.039	(-0.099, 0.022)
CDE(1)	-0.039	(-0.100, 0.022)
NDE	-0.037	(-0.096, 0.021)
NIE	-0.114	(-0.144, -0.084)
PE(0)	0.744	(0.383, 1.106)
PE(1)	0.743	(0.379, 1.108)
PM	0.753	(0.403, 1.103)

Examples

- We note that the results of the nonparametric and parametric mediation analyses are almost the same, except for the controlled direct effects, which are nearly the same and both negative in the parametric analysis due to the absence of an interaction between treatment and mediator in the outcome model.
- We note that the PE(0), PE(1), and PM are all quite high, and correspondingly that the estimated direct effects are not statistically signifiant.
- Thus, the effect of naltrexone on unsuppressed viral load appears to be mediated almost entirely by reduced drinking.
- We know from the program that generated the data, `doublewhatifsim.r` in Chapter 1, that the effect is 100% mediated.
- Finally, we note that the confidence intervals for PE(0), PE(1), and PM include values greater than one, which in this case we should interpret as truncated at one.

Examples

- For our second example, we simulated data to show that mediation analysis can lead to direct and indirect effects with opposite signs, which is perhaps nonintuitive.
- We used `simmed.r`, on the next slide.

Examples

```
> simmed.r
function ()
{
set.seed(44444)
A<-rbinom(n=1000,size=1,prob=.5)
H<-rbinom(n=1000,size=1,prob=.5)
tmppm<-A+H
pm<-exp(tmppm)/(1+exp(tmppm))
M<-rbinom(n=1000,size=1,prob=pm)
tmppy<-H-M+A
py<-exp(tmppy)/(1+exp(tmppy))
Y<-rbinom(n=1000,size=1,prob=py)
dat<-cbind(H,A,M,Y)
dat<-data.frame(dat)
dat
```

Examples

- To motivate the simulation, we let A be randomization to a diet that does not work, in fact even causing weight gain for some participants.
- We let H indicate genetics inducing a propensity for weight gain.
- Let M be adoption of a stricter diet midway through the study.
- Let Y be weight at the end of study.
- We see that $A = 1$ and $H = 1$ are randomized with probability 0.5, and that they both influence $M = 1$.
- However, the end of study weight tends to be higher for those with a propensity for weight gain, with $H = 1$, and for those randomized to the diet, with $A = 1$, with some mitigation from adopting the stricter diet, $M = 1$.

Examples

Table 3: Nonparametric Mediation Analysis of the Data Generated by `simmed.r`

Measure	Estimate	95% CI
TE	0.160	(0.102, 0.218)
CDE(0)	0.279	(0.183, 0.376)
CDE(1)	0.168	(0.097, 0.239)
NDE	0.209	(0.153, 0.265)
NIE	-0.049	(-0.071, -0.028)
PE(0)	-0.747	(-1.478, -0.017)
PE(1)	-0.049	(-0.324, 0.225)
PM	-0.308	(-0.533, -0.083)

Examples

- We see that overall, the initial diet $A = 1$ causes weight gain, and that moreover, there is a direct effect of weight gain no matter which method we use to quantify the direct effect.
- However, the indirect effect is for weight loss via adoption of $M = 1$.
- The opposite signs of the direct and indirect effects lead to PE(0), PE(1), and PM that are not meaningful, and we see that their estimates are negative.
- Nevertheless, the mediation analysis is helpful in terms of teasing apart the direct and indirect effects of the initial diet $A = 1$.

A Real Life Example

- I'd like to acknowledge the stimulating and synergistic collaboration with my high school classmate, Dr. Erik Kulstad, MD, MS, Professor of Emergency Medicine at UT Southwestern Medical Center and CMO of Attune Medical, Chicago, IL, USA.
- Thermal injury to the esophagus can lead to a life-threatening complication, called atrial-esophageal fistula (AEF), after ablation for atrial fibrillation.
- The ensoETM device (Attune Medical, Chicago, IL, USA) is routinely used to control body temperature in at risk patients in an intensive care setting and in patients whose body temperature must be lowered to protect an injured brain.
- As it does this by warming or cooling the lumen of the esophagus and stomach, physicians hypothesized that it could also protect the esophagus during radiofrequency (RF) ablation.
- We analyze pooled data from three randomized clinical trials designed to estimate the effect of cooling on esophageal lesions graded as 0, 1, 2, and 3. Data from one trial leads to a statistically significant difference, whereas data from the other two do not.
- The primary analysis based on the pooled data failed to show a difference. However, greater RF ablation time was associated with greater thermal injury, and we found that it also differed across the trials.
- Including RF ablation time, which has the status of a potential mediator because it is a post-treatment variable, as an effect modifier presents a different picture of the pooled data, and it illustrates some of the challenges of working with an ordinal outcome that is non-zero for only 29 of 164 patients in the combined data.

Analyzing Pooled Data from 3 Clinical Trials

- Previous analysis in a de novo submission to the FDA of the pooled data from 3 clinical trials did not reach statistical significance. A Bayesian hierarchical proportional odds model with and without covariates was used to analyze the effect of cooling on lesion grades of 0, 1, 2, or 3. Grade 4 is AEF, but, due to the smaller size of the clinical trials, no AEFs were observed.
- Linear regression with robust standard error showed that in clinical trial A (120 patients) cooling reduced the expected lesion grade by -0.267 (P=0.006); in clinical trial B, (40 patients) it did not have a statistically significant effect on lesion grade (increase of 0.100, P=0.751); in clinical trial C (6 patients) it did not have a statistically significant effect on lesion grade (decrease of -0.333, P=0.643).
- Pooling the three trials also does not show a statistically significant effect on lesion grade (decrease of -0.181, P=0.11).
- The analyses in the de novo submission found that radiofrequency ablation time (RF Time) had a statistically significant effect of increasing lesion grades (the other covariates considered did not).
- These analyses did not differentiate between pre- and post- treatment covariates; RF Time was the only post- treatment covariate considered. My colleague asked me if any post-treatment covariates might be mediating the treatment effect, and RF Time was the most natural candidate for me to consider.

Analyzing Pooled Data from 3 Clinical Trials

- Mediation analyses are easier to think about in terms of linear models rather than proportional odds models, because the latter require more careful analysis due to non-collapsibility of odds ratios.
- Therefore, I started with linear models of lesion grade on treatment arm and RF Time and their interaction in the pooled data.
- I used the bootstrap to assess statistical significance, due to likely violations of homoscedasticity.
- The coefficient of interaction was almost statistically significant (it was statistically significant using the naive standard errors). It is important to note, however, that that test of the coefficient represents the contrast of a treatment effect at a given RF Time versus the treatment effect at RF Time equal to zero.
- It is still possible for there to be a statistically significant treatment effect at higher RF Times, even though when we contrast it with the effect at time zero the difference is not statistically significant. This is because the difference can be more variable than the effect itself.

Analyzing Pooled Data from 3 Clinical Trials

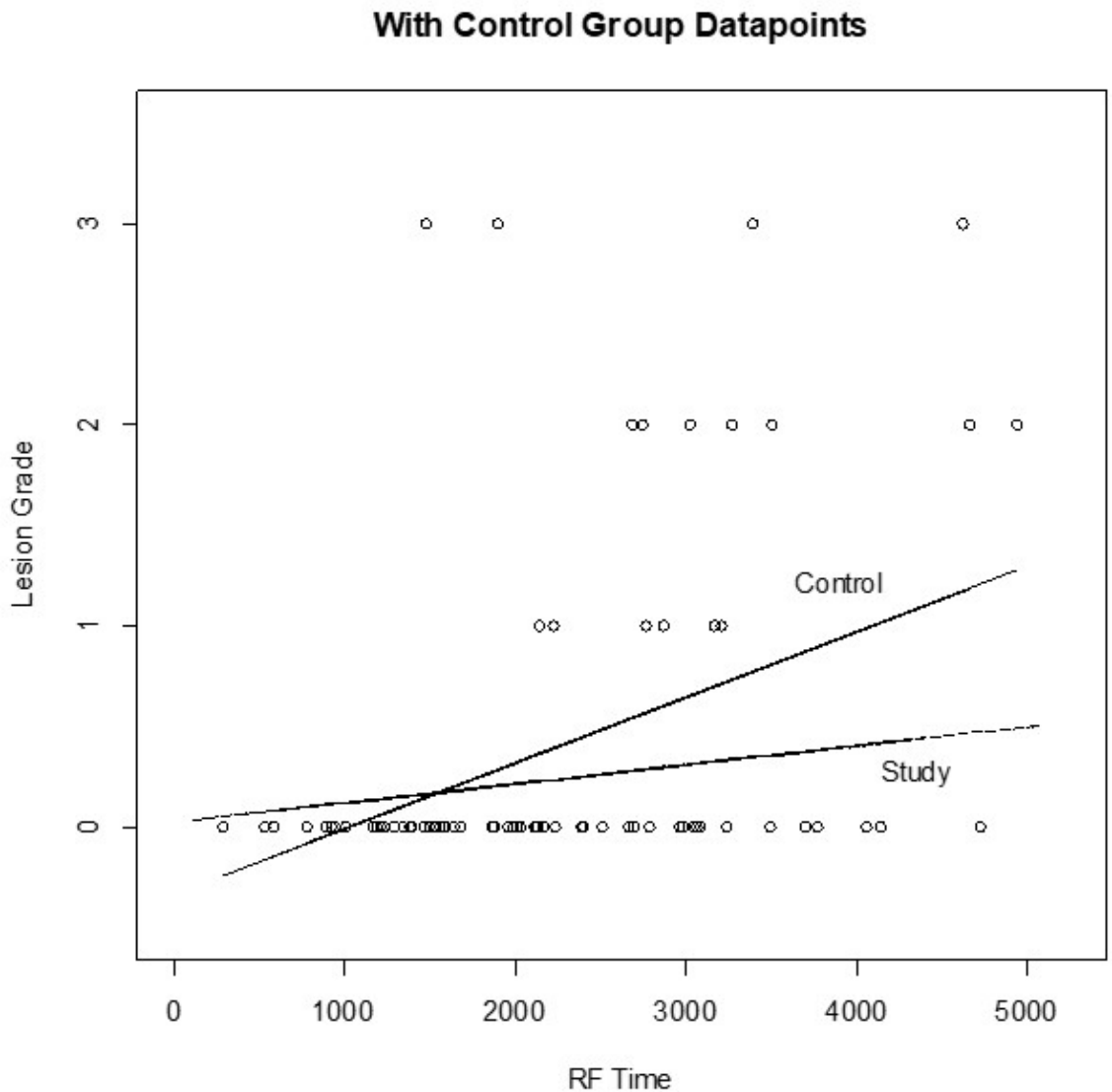

Figure 2: Linear Model with Control Group Datapoints Shown

Analyzing Pooled Data from 3 Clinical Trials

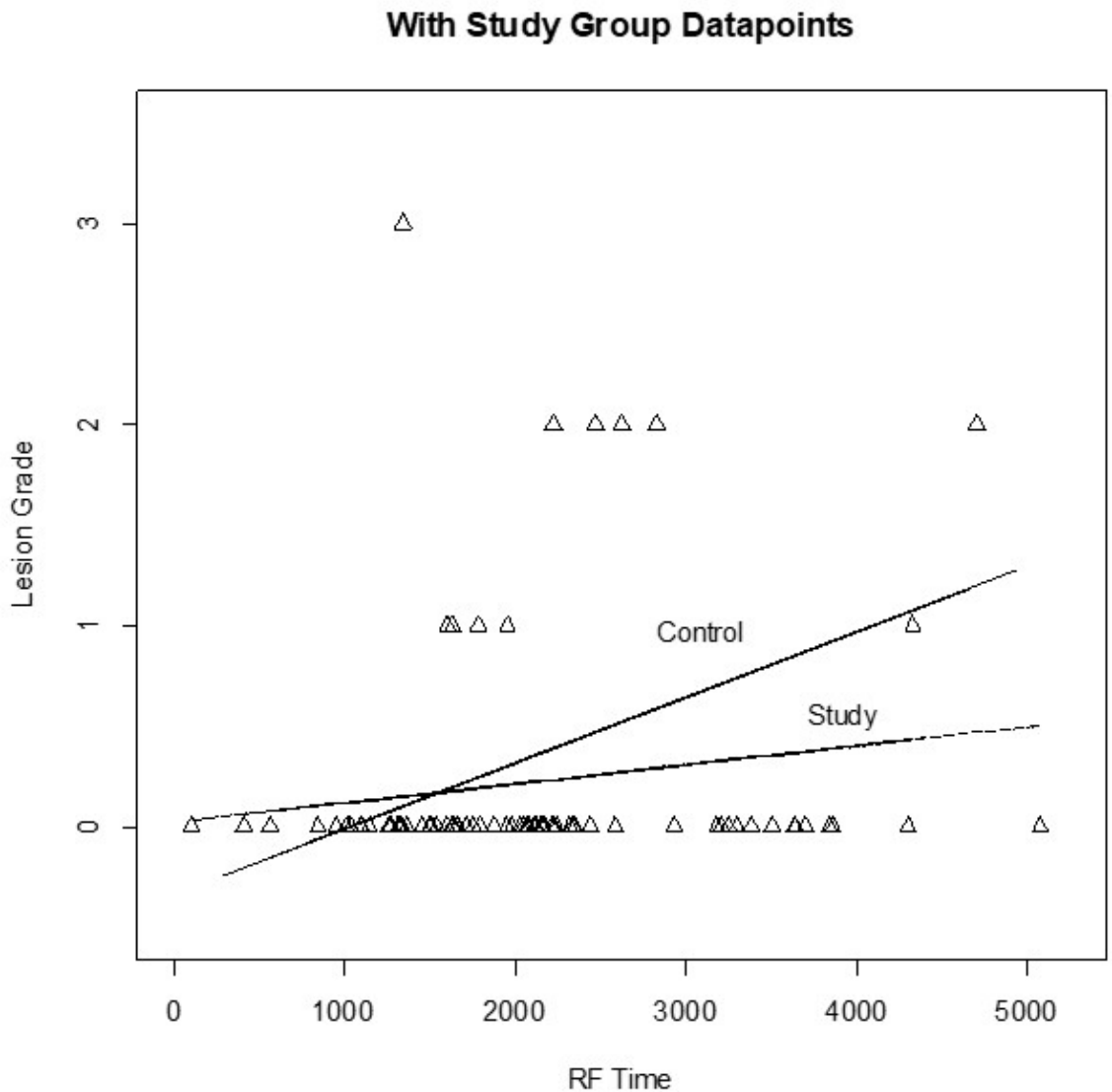

Figure 3: Linear Model with Study Group Datapoints Shown

Analyzing Pooled Data from 3 Clinical Trials

- Because the expected lesion grade was negative for small RF Times, I also fit a proportional odds model.
- Note it is not hard to compute the expected lesion grade using the fitted probabilities from the output.
- The bootstrap did not work for the proportional odds model, but the jackknife did, so I used that to assess sampling variability.

Analyzing Pooled Data from 3 Clinical Trials

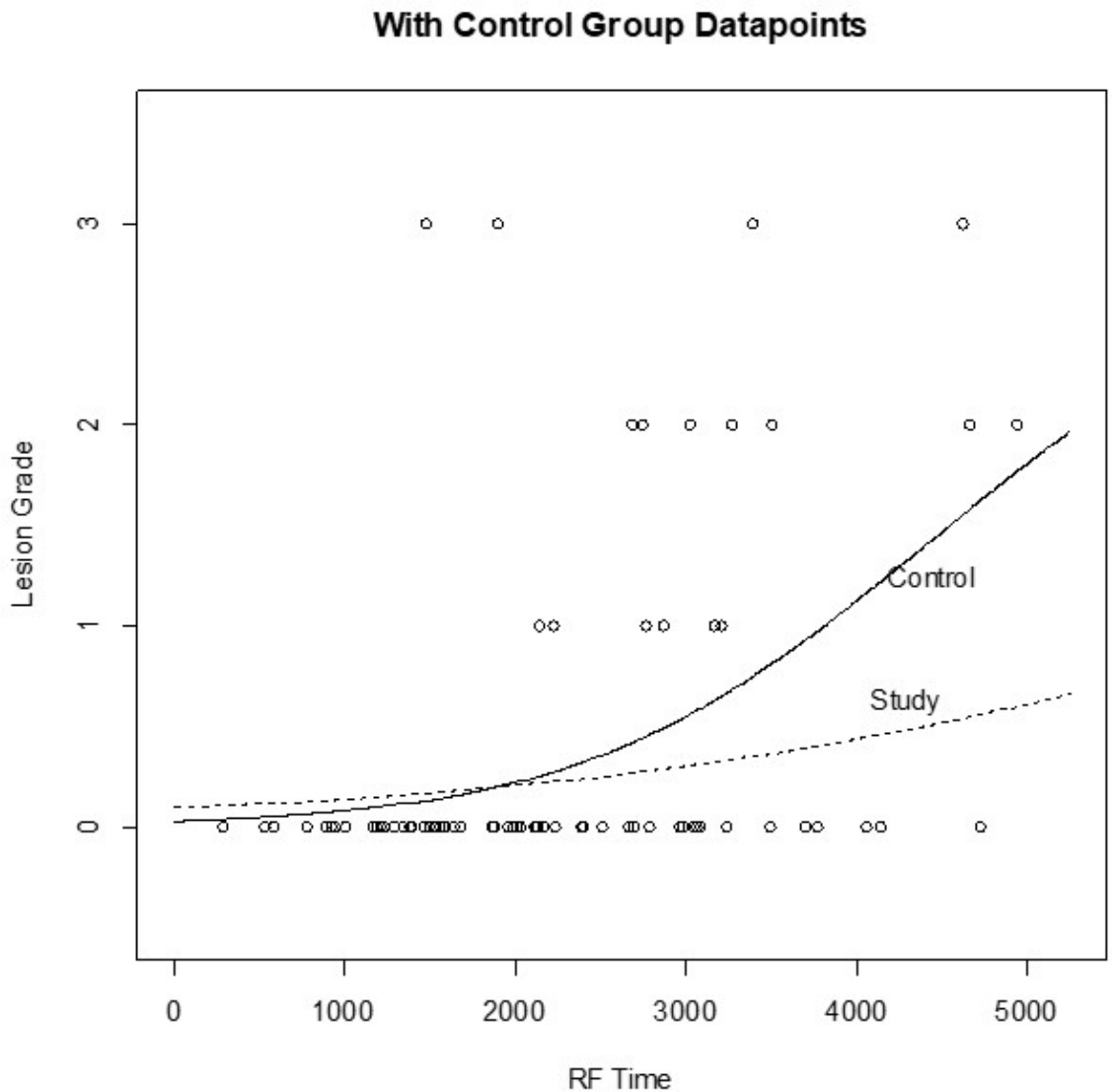

Figure 4: Proportional Odds Model with Control Group Datapoints

Analyzing Pooled Data from 3 Clinical Trials

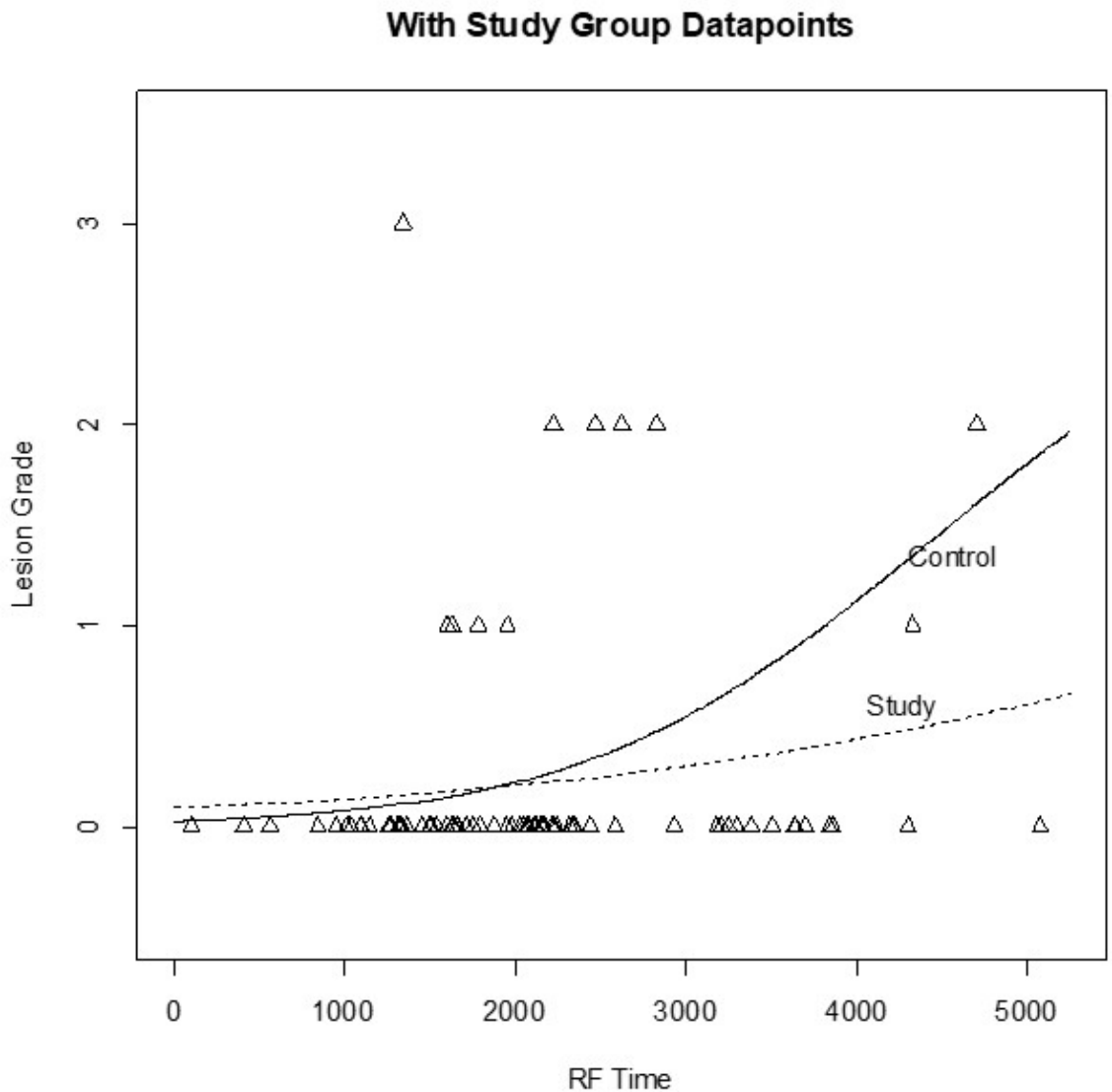

Figure 5: Proportional Odds Model with Study Group Datapoints Shown

Analyzing Pooled Data from 3 Clinical Trials

- Next, I computed the difference between the two curves and 95% pointwise confidence bands using the bootstrap for the linear model and the jackknife for the proportional odds model.
- Then, I determined the shortest RF Time with a statistically significant treatment effect.

Analyzing Pooled Data from 3 Clinical Trials

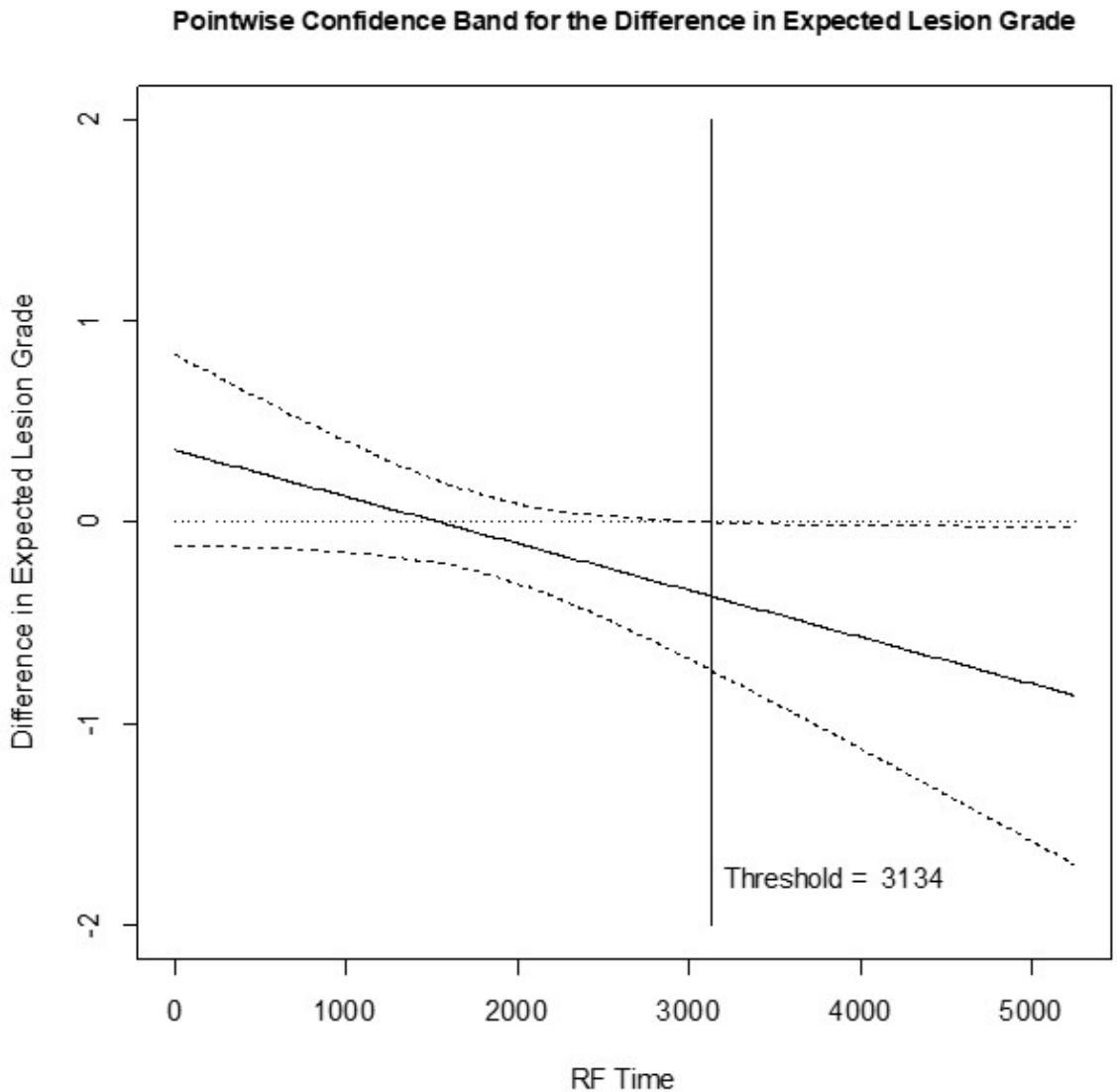

Figure 6: Treatment Effect as a Function of RF Time for the Linear Model

Analyzing Pooled Data from 3 Clinical Trials

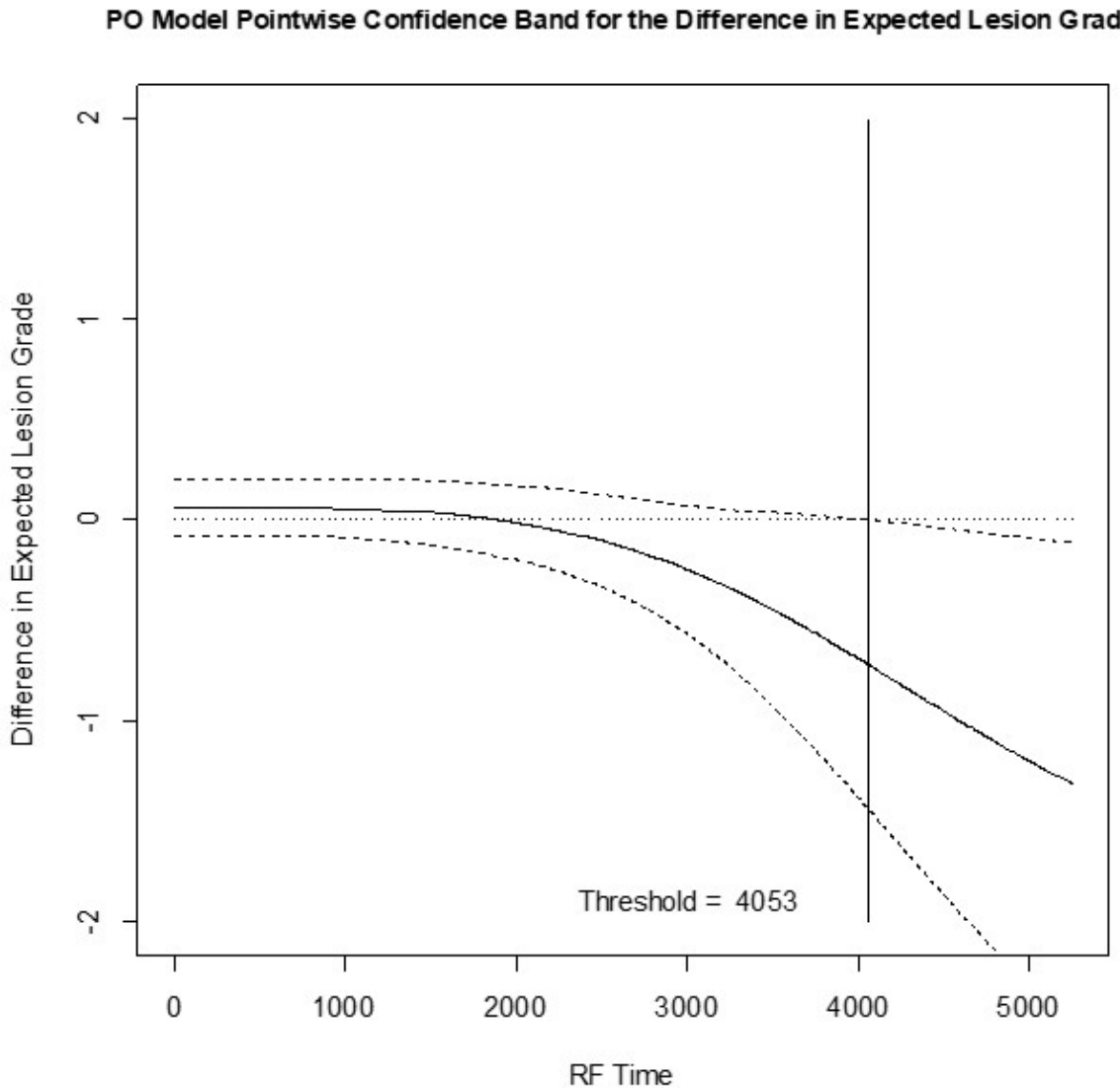

Figure 7: Treatment Effect as a Function of RF Time for the PO Model

Analyzing Pooled Data from 3 Clinical Trials

- Next, I looked at this as a mediation analysis.
- For either model, the total effect of cooling is to decrease the expected lesion grade by -0.171 (-0.400, 0.054), which is not quite statistically significant.
- For the linear model, if we adjust the expected grades in the treatment group to counterfactual values had the distribution of RF time been as it was in the control group (i.e., somewhat higher), the direct effect of cooling is to decrease the expected lesion grade by -0.158 (-0.381, 0.065), which again, is not quite statistically significant.
- The difference between the total effect and the direct effect is the indirect effect, which we see is small. Therefore most of the effect of cooling is due to differences at fixed RF times, which we capture in detail with the confidence bands above, which reach statistical significance for longer RF times.
- For comparison, with the proportional odds model, the direct effect is -0.144 (-0.357, 0.069).
- We note that the direct effect point estimates of -0.158 and -0.144 are averages with respect to the distribution of RF times observed in the control group, whereas we achieve statistically significant direct effects at all RF times higher than the threshold of 3134 seconds with the linear model and 4053 seconds with the proportional odds model.

Analyzing Pooled Data from 3 Clinical Trials

- A reviewer asked how we justified pooling the data across the three clinical trials.
- The decision to pool the data across the three clinical trials was prompted by the primary analysis presented in the de novo submission, which pooled the data to estimate a main effect of treatment accompanied by random study effects.
- That analysis effectively targets a main treatment effect estimand that weights the clinical trials equally.
- As the clinical trial samples can be viewed as convenience samples, rather than as representative of the typical AF population presenting for ablation, we see no reason to weight the clinical trials equally; thus, we opt to weight them according to their respective numbers of participants.
- Analyses that pool the data without including random effects do just that, and using the bootstrap to assess sampling variation is robust to cross-study differences as well as heteroscedasticity due to the form of outcome.

Analyzing Pooled Data from 3 Clinical Trials

- We note that differences in estimated treatment effects for Trial B and Trial A presented in the pooled de novo submission did not include an interaction between Study and RF Time.
- The average of RF Time was higher for Trial A than for Trial B (2220 for Trial A and 1777 for Trial B); this could explain the differences observed in the de novo submission.
- Furthermore, the Trial B sample is smaller, with only 40 patients, leading to substantial sampling variability in the de novo estimates.

Discussion

- We presented a re-analysis to show that one could analyze the data to find that the ensoETM device protects the esophagus from thermal injury during radiofrequency ablation.
- The pooled re-analysis of the clinical trial data suggests the treatment effect may be stronger for longer RF Times.
- We found this to be an interesting mediation analysis, in which effect-modification by the potential mediator was also present.

Time-Dependent Confounding: Introduction

- These methods are motivated by a cancer clinical trial, with a hypothetical version of the data that has been modified for expository purposes.
- To analyze those data, we will assume the causal DAG in Figure 8 holds.

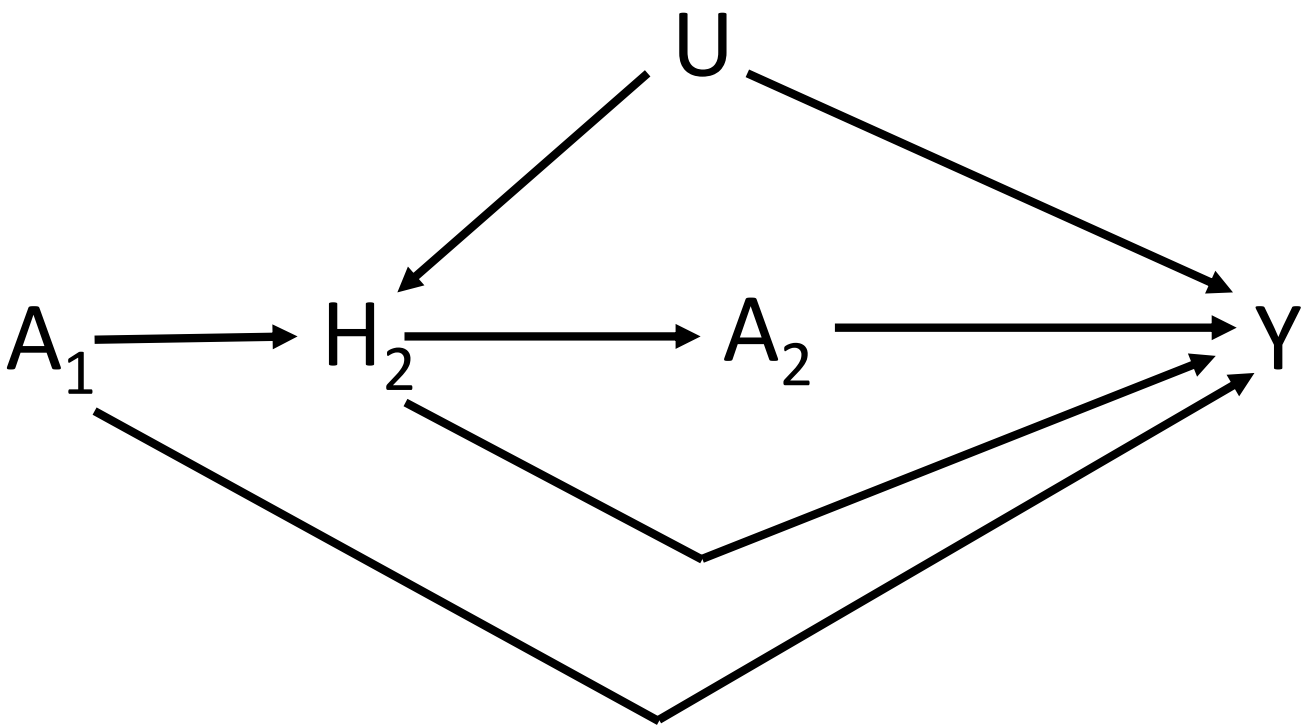

Figure 8: H_2 is a Time-Dependent Confounder

A Cancer Clinical Trial

- To illustrate *time-dependent confounding*, we introduce Children's Oncology Group (COG) (Pediatric Oncology Group at the time) study P9462.
- P9462 was a phase II study of relapsed neuroblastoma patients randomized to either Topotecan (TOPO, i.e. $A_1 = 0$), the standard therapy, or Topotecan + Cytoxan (TOPO/CTX, $A_1 = 1$), the experimental therapy.
- The primary outcome was response ($H_2 = 1$), originally assessed via a two-stage group sequential design. Although statistical significance was not achieved, there was a trend toward a higher response rate in those patients treated with the experimental therapy.
- Although response was the primary endpoint of the P9462 study, a comparison of the proportion of patients surviving at two years ($Y = 1$) is also of interest and allows us to illustrate the analysis of *sequential treatments*, which can be subject to time-dependent confounding.

A Cancer Clinical Trial...

Table 4: A Hypothetical Cancer Clinical Trial

A_1	H_2	A_2	n	n with $Y = 1$	proportion with $Y = 1$
0	0	0	410	120	0.29
0	0	1	30	0	0.00
0	1	0	160	30	0.19
0	1	1	30	20	0.67
1	0	0	280	30	0.11
1	0	1	20	10	0.50
1	1	0	190	80	0.42
1	1	1	70	20	0.29

A Cancer Clinical Trial...

- A factor which complicates the survival analysis is that some patients received a bone marrow transplant (BMT) ($A_2 = 1$) after the response assessment. Nine patients in the TOPO/CTX group had a post-treatment BMT, versus only three in the TOPO group.
- Patients who achieved a response were more likely to have the post-treatment BMT than those who did not, in either treatment group. This raises the interesting question of how to adjust for the second treatment (the post-treatment BMT) in the comparison of survival across the first two treatments.
- The variable H_2, which is thought to influence both A_2 and Y, is itself influenced by A_1, and therefore meets the criteria to be a time-dependent confounder. Because A_2 is a postrandomization event, i.e. it occurs after A_1, adjusting for it in the analysis of the effect of A_1 on Y requires us to make use of the time-dependent confounder H_2 in a complex way.
- Two classes of statistical models which enable this are called *marginal structural models* and *structural nested mean models*.
- We might also be interested in knowing how to combine A_1 and A_2 to optimize survival, making use of H_2 in the decision for A_2; that is, we might wish to discover the *optimal dynamic treatment regime*.

A Cancer Clinical Trial...

- Another wrinkle in our analyses is the problem that the trial did not produce any patients randomized to standard therapy ($A_1 = 0$) who did not have a response ($H_2 = 0$) but who had a BMT ($A_2 = 1$). Therefore, we had no data on the expected outcome Y following this sequence of events. We will suppose none of these patients were assigned $A_2 = 1$ due to an adverse expected outcome.
- For expository purposes, our analyses of these data will ignore the group sequential design, instead assuming an ordinary randomized clinical trial.
- Additionally, we increased the sample size by a factor of 10, so that our analyses will not be limited by the small sample sizes. Finally, we added data on thirty patients with $A_1 = 0$, $H_2 = 0$, $A_2 = 1$, and $Y = 0$, to reflect our supposition that the unobserved sequence would have led to an adverse outcome.
- As in this example, it is often the case that adjustment for time-dependent confounding is hindered by small sample sizes. In that event, such analyses are akin to *qualitative methods*.
- We present the analyses in this context for learning a new way to think and for planning analyses of similar studies that might be expected to produce data with the same structure.

Introduction

- In the DAG, the variable H_2 is a *time-dependent confounder*, because it is a confounder for treatment A_2 at time two, but not for treatment A_1 at time one.
- In fact, H_2 is an *intermediate variable* for the effect of treatment A_1 on the outcome Y, because it is a mediator of that effect.
- This dual role of H_2 as a confounder with respect to one treatment but as an intermediate variable with respect to another treatment makes it tricky to adjust for it in the analysis of the joint effects of A_1 and A_2 on Y.
- In the following subsections, we will present three approaches to analyzing causal effects of A_1 and A_2 on Y.
- As we will see, the three approaches require three different potential outcomes frameworks for validity.
- In what follows, we will assume A_1 and A_2 are binary, noting that modifications exist for more general versions.

Marginal Structural Models

- Marginal structural models (MSMs) require consistency of four potential outcomes $Y(a_1, a_2)$, for $a_1 = 0, 1$ and $a_2 = 0, 1$, to treatment with $A_1 = a_1$ followed by $A_2 = a_2$.
- We also assume sequential randomization,

$$A_1 \amalg Y(a_1, a_2) \tag{5}$$

and

$$A_2 \amalg Y(a_1, a_2) | A_1, H_2. \tag{6}$$

Marginal Structural Models

- A marginal structural model is a model for the marginal means of the potential outcomes, such as the saturated model

$$E(Y(a_1, a_2)) = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \beta_3 a_1 a_2. \tag{7}$$

- To estimate the parameter $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$, we need to relate $E(Y(a_1, a_2))$ to the observed data.
- One way to do that is through the relation

$$E(Y(a_1, a_2)) = E_{H_2|A_1=a_1} E(Y|A_1 = a_1, H_2, A_2 = a_2), \tag{8}$$

which we prove as follows.

- By (5), we have that

$$E(Y(a_1, a_2)) = E(Y(a_1, a_2)|A_1 = a_1).$$

Marginal Structural Models

- Then by the double expectation theorem, we have that

$$E(Y(a_1, a_2)|A_1 = a_1) = E_{H_2|A_1=a_1} E(Y(a_1, a_2)|A_1 = a_1, H_2).$$

- Finally, by (6), the right-hand side equals

$$E_{H_2|A_1=a_1} E(Y(a_1, a_2)|A_1 = a_1, H_2, A_2 = a_2),$$

which equals the right-hand side of (8) by consistency.

Marginal Structural Models

- Thus, one way to estimate β is to use (8), which involves the outcome model $E(Y|A_1, H_2, A_2)$.
- However, a more popular method arises from an exposure-modeling approach, which involves a model for

$$e(H_2, A_1) = E(A_2|H_2, A_1).$$

- We can use proofs directly analogous to those of exposure-model standardization in Module 2 to show that

$$E(Y(a_1, 1)) = E\left(\frac{A_2 Y}{e(H_2, A_1)}|A_1 = a_1\right)$$

and

$$E(Y(a_1, 0)) = E\left(\frac{(1 - A_2)Y}{1 - e(H_2, A_1)}|A_1 = a_1\right).$$

- We also use proofs directly analogous to those of Module 2 to show that weighted averages of Y with weights equal to

$$W = \frac{1}{A_2 e(H_2, A_1) + (1 - A_2)(1 - e(H_2, A_1))}$$

within the four $(A_1 = a_1, A_2 = a_2)$ subgroups are valid estimates of $E(Y(a_1, a_2))$.

Marginal Structural Models

- As in Module 2, we can compute these estimates using the `weighted.mean` function or using the `glm` or gee functions to fit a weighted linear model; the method is often called *inverse-probability of treatment weighted* (IPTW) estimation.
- We implement it with the `glm` function and the code `msm.r`, which outputs an estimate of β as well *contrasts* of β.
- For example, the contrast $\beta_1 + \beta_3$ compares administration of both $A_1 = 1$ and $A_2 = 1$ with administration of just $A_2 = 1$, whereas the contrast $\beta_2 + \beta_3$ compares $A_1 = 1$ and $A_2 = 1$ with $A_1 = 1$ and $A_2 = 0$.
- Finally, $\beta_1 + \beta_2 + \beta_3$ compares administration of both treatments with administration of neither treatment.

Marginal Structural Models

```
> msm.r
function(dat=cogdat,ids=1:nrow(cogdat))
{
dat<-dat[ids,]
out<- glm(A2 ~ A1*H2, data = dat, family = binomial)
e<- glm(A2 ~ A1*H2, data = dat, family = binomial)$fitted
W<-1/(e*dat$A2 + (1-e)*(1-dat$A2))
msm<- summary(glm(formula = Y ~ A1 + A2 + A1 * A2,
        data = dat, family = gaussian, weights = W))
beta<-msm$coef[,1]
c(beta,(beta[2]+beta[4]),(beta[3]+beta[4]),(beta[2]+beta[3]+beta[4]))
}
```

Marginal Structural Models

- We estimate the sampling distribution using the bootstrap.
- The hypothetical cancer clinical trial data are stored in `cogdat`.
- Results of applying marginal structural models with IPTW estimation to the cancer data are presented in Table 5.

Table 5: IPTW of MSM parameters for the Hypothetical Cancer Clinical Trial

Parameter	Estimate	95% CI
β_0	0.261	(0.224, 0.298)
β_1	-0.008	(-0.063, 0.047)
β_2	-0.060	(-0.129, 0.009)
β_3	0.208	(0.056, 0.359)
$\beta_1 + \beta_3$	0.199	(0.056, 0.342)
$\beta_2 + \beta_3$	0.148	(0.008,0.287)
$\beta_1 + \beta_2 + \beta_3$	0.140	(-0.001,0.280)

Marginal Structural Models

- We observe that the estimates of β_1 and β_2 are not statistically significantly different from zero.
- Therefore, either treatment on its own does not appear to influence survival at two years versus administration of neither treatment.
- However, the statistical significance of the estimates of $\beta_1 + \beta_3$ and $\beta_2 + \beta_3$ suggests that administering $A_1 = 1$ followed by $A_2 = 1$ results in increased survival relative to administering just one treatment.
- To compare joint administration to administration of neither treatment, we need to estimate the contrast $\beta_1 + \beta_2 + \beta_3$, which we see is not quite statistically significant, although it is very close.
- Our point estimates suggest that administering neither treatment results in a 26.1% chance of survival for two years, whereas administering both results in a 26.1%+14.0%=40% chance of survival for two years.
- These survival probabilities refer to the entire population that participants were sampled from; that is, they are marginal rather than conditional probabilities.

Structural Nested Mean Models

- Structural nested mean models (SNMM) require consistency of two potential outcomes: Y_1 is the potential outcome to treatment with the observed A_1 and then treatment with $A_2 = 0$, whereas Y_0 is the potential outcome to $A_1 = 0$ followed by $A_2 = 0$.
- We also assume sequential randomization,

$$A_1 \amalg Y_0 \tag{9}$$

and

$$A_2 \amalg Y_1 | A_1, H_2. \tag{10}$$

Structural Nested Mean Models

- For our example, there are two levels to the 'nest' of the structural nested mean model.
- The second level reflects the effect of changing A_2 from its observed value to $A_2 = 0$:

$$E(Y - Y_1|A_2, H_2, A_1) = f(A_1, H_2; \beta_2)A_2, \tag{11}$$

where $f(\cdot)$ is specified by the analyst, whereas the first level reflects the effect of further changing A_1 from its observed value to $A_1 = 0$:

$$E(Y_1 - Y_0|A_1) = \beta_1 A_1. \tag{12}$$

- We observe that these models target conditional causal effects, with the first one conditional on A_2, H_2, and A_1; the second one conditional on A_1.
- The parameters β_2 and β_1 quantify elements of those conditional causal effects.

Structural Nested Mean Models

- To estimate β_1 and β_2, we need to relate them to the observed data.
- We can use the relations

$$E(Y|A_2, H_2, A_1) = f(A_1, H_2; \beta_2)A_2 + h(H_2, A_1) \tag{13}$$

and

$$E(Y - f(A_1, H_2; \beta_2)A_2|A_1) = \alpha + \beta_1 A_1, \tag{14}$$

where $h(\cdot)$ must be correctly specified by the analyst.

- For estimation of β_2 using the first relation, we can fit a non-parametric model if possible, or a parametric model if necessary.

Structural Nested Mean Models

- For estimation of β_1 in the second relation, we (a) substitute into the left-hand side the estimate $\hat{\beta}_2$ from the fitting the model specified by first relation, and (b) fit the model specified by the second relation to estimate β_1.
- To prove the first relation (13), it suffices to show that

$$E(Y - f(A_1, H_2; \beta_2)A_2 | A_2, H_2, A_1) \tag{15}$$

  does not depend on A_2.
- By (11), we have that (15) equals $E(Y_1 | A_2, H_2, A_1)$, which, by (10), does not depend on $A2$.

Structural Nested Mean Models

- To prove the second relation (14), it suffices to show that

$$E(Y - f(A_1, H_2; \beta_2)A_2 - \beta_1 A_1 | A_1) \tag{16}$$

  does not depend on A_1.
- By the double expectation theorem, we have that (16) equals

$$E_{A_2, H_2|A_1} E(Y - f(A_1, H_2)A_2 | A_2, H_2, A_1) - \beta_1 A_1,$$

  which, by (11), equals

$$E_{A_2, H_2|A_1} E(Y_1 | A_2, H_2, A_1) - \beta_1 A_1 = E(Y_1 | A_1) - \beta_1 A_1.$$

- By (12), this equals $E(Y_0|A_1)$, which equals $E(Y_0)$ by (9), which does not depend on A_1.

Structural Nested Mean Models

- We estimated the parameters of the structural nested mean model for the hypothetical cancer clinical trial using `snmm.r`, with bootstrap estimates of the sampling distribution.
- We fit the models

$$E(Y - Y_1|A_2, H_2, A_1) = (\beta_{20} + A_1\beta_{21} + H_2\beta 22 + A_1 * H_2\beta_{23})A_2,$$

and

$$E(Y_1 - Y_0|A_1) = \beta_1 A_1.$$

Structural Nested Mean Models

```
> snmm.r
function(dat=cogdat,ids=1:nrow(cogdat))
{
dat<-dat[ids,]
nest2.out<-summary(lm(Y~A1+H2+A2 +
A1*H2 + A1*A2 + H2*A2 + A1*H2*A2,data=dat))
b2<-nest2.out$coef[,1]
Y1hat<-dat$Y-b2[4]*dat$A2-b2[6]*dat$A1*dat$A2 - b2[7]*dat$H2*dat$A2 -
b2[8]*dat$A1*dat$H2*dat$A2
nest1.out<-summary(lm(Y1hat~A1,data=dat))
b1<-nest1.out$coef[2,1]
c(b2[c(4,6:8)],b1)
}
```

Table 6: SNMM Estimation for the Hypothetical Cancer Clinical Trial

Contrast	Estimate	95% CI
β_{20}	-0.293	(-0.337, -0.248)
$\beta_{20} + \beta_{22}$	0.479	(0.290, 0.669)
$\beta_{20} + \beta_{21}$	0.393	(0.159, 0.626)
$\beta_{20} + \beta_{21} + \beta_{22} + \beta_{23}$	-0.135	(-0.259, -0.011)
β_1	-0.008	(-0.063, 0.047)

Structural Nested Mean Models

- From estimation of β_1, we see that the effect of A_1 is not statistically significant when A_2 will not be administered afterwards.
- From estimation of the contrasts of β_2, we see that the effect of A_2 is to decrease survival when $A_1 = 0$ and $H_2 = 0$ ($\hat{\beta}_{20} = -0.293$), to increase survival when $A_1 = 0$ and $H_2 = 1$ ($\hat{\beta}_{20} + \hat{\beta}_{22} = 0.479$), to increase survival when $A_1 = 1$ and $H_2 = 0$ ($\hat{\beta}_{20} + \hat{\beta}_{21} = 0.393$), and to decrease survival when $A_1 = 1$ and $H_2 = 1$ ($\hat{\beta}_{20} + \hat{\beta}_{21} + \hat{\beta}_{22} + \hat{\beta}_{23} = -0.135$).
- Therefore, the decision about whether to treat with $A_2 = 0$ or $A_2 = 1$ should make use of information on A_1 and H_2, where possible.

Optimal Dynamic Treatment Regimes

- Marginal structural models provided information on optimal *static treatment regimes*, that is, the optimal sequence of A_1 and A_2 when a static decision must be made at baseline.
- Structural nested mean models provided some information on optimal *dynamic treatment regimes*; we learned how to choose the optimal A_2 when a dynamic decision could be made after observing A_1 and H_2.
- However, we were left hanging about the optimal choice of A_1.
- In this subsection, we learn how to estimate the *optimal dynamic treatment regime*, which will tell us how first to select A_1 and second to select A_2, taking into account A_1 and H_2, in order to obtain the best results for the population under study.

Optimal Dynamic Treatment Regimes

- We require consistency of the eight potential outcomes $Y(a_1, d_2)$, where a_1 is a static treatment regime recording the setting of A_1 to 0 or 1, and d_2 is one of four possible dynamic treatment regimes recording the setting of A_2 to 0 or 1 depending on a_1 and H_2.
- For example, supposing $a_1 = 0$, then one of the four possible dynamic treatment regimes for d_2 is to set A_2 to 0 if $H_2 = 0$ or to 1 if $H_2 = 1$, whereas another possible one is to set A_2 to 0 regardless of H_2.
- We also assume sequential randomization,

$$A_1 \amalg Y(a_1, d_2) \tag{17}$$

and

$$A_2 \amalg Y(a_1, d_2) | A_1, H_2. \tag{18}$$

Optimal Dynamic Treatment Regimes

- Given a_1, we determine the optimal d_2, which is a function of H_2, by maximizing

$$E(Y(a_1, d_2)|A_1 = a_1, H_2).$$

- By (18), this is equivalent to maximizing

$$E(Y(a_1, d_2)|A_1 = a, H_2, A_2 = d_2),$$

- which, by consistency, is equivalent to maximizing

$$E(Y|A_1 = a, H_2, A_2 = d_2), \tag{19}$$

  which is a function of the observed data.
- In this way, we can determine the optimal d_2, which we denote by d_2^{opt}, as a function of a_1.

Optimal Dynamic Treatment Regimes

- To find the optimal a_1, or a_1^{opt}, we find the a_1 that maximizes

$$E(Y(a_1, d_2^{opt})).$$

- By (17), this is equivalent to maximizing

$$E(Y(a_1, d_2^{opt})|A_1 = a_1),$$

which, by the double expectation theorem, equals

$$E_{H_2|A_1=a_1} E(Y(a_1, d_2^{opt})|A_1 = a_1, H_2),$$

which, by (18), equals

$$E_{H_2|A_1=a_1} E(Y(a_1, d_2^{opt})|A_1 = a_1, H_2, A_2 = d_2^{opt}),$$

which, by consistency, is equivalent to maximizing

$$E_{H_2|A_1=a_1} E(Y|A_1 = a_1, H_2, A_2 = d_2^{opt}), \tag{20}$$

which is a function of the observed data.

Optimal Dynamic Treatment Regimes

- Therefore, to find the optimal dynamic treatment regime (a_1^{opt}, d_2^{opt}), we first find the d_2 that maximizes (19) for each possible value of a_1, and we second find the a_1 that maximizes (20) with d_2^{opt} set equal to that d_2.
- We compute the optimal dynamic treatment regime using the series of `R` functions which follow. We start with the data set `cogdat`.

Optimal Dynamic Treatment Regimes

```
> head(cogdat)
   A1 H2 A2 Y
1   0  0  0 1
11  0  0  0 1
12  0  0  0 1
13  0  0  0 1
14  0  0  0 1
15  0  0  0 1
```

Optimal Dynamic Treatment Regimes

- Next we use the function `mkcogtab.r` to convert the data into table form.

Optimal Dynamic Treatment Regimes

```
> mkcogtab.r
function(dat=cogdat,ids=1:nrow(cogdat))
{
dat<-dat[ids,]
tab<-data.frame(xtabs(~A2+H2+A1,data=dat))
tmp<-data.frame(xtabs(Y~A2+H2+A1,data=dat))
tab$prop<-tmp$Freq/tab$Freq
tab
}
> mkcogtab.r()
  A2 H2 A1 Freq    prop
1  0  0  0  410 0.29268
2  1  0  0   30 0.00000
3  0  1  0  160 0.18750
4  1  1  0   30 0.66667
5  0  0  1  280 0.10714
6  1  0  1   20 0.50000
7  0  1  1  190 0.42105
8  1  1  1   70 0.28571
```

Optimal Dynamic Treatment Regimes

- Note that `prop` is the proportion of participants in the A_2, H_2, A_1 stratum with $Y = 1$.
- Then, we find d_2^{opt} using the function `A2opt.r`.

Optimal Dynamic Treatment Regimes

```
> A2opt.r
function(tab)
{
A2opt<-NULL
propopt<-NULL
a<-which(tab[1:2,"prop"]==max(tab[1:2,"prop"]) ) -1
A2opt<-c(A2opt,a,a)
b<-tab$prop[(1+a)]
propopt<-c(propopt,b,b)
a<-which(tab[3:4,"prop"]==max(tab[3:4,"prop"]) ) -1
A2opt<-c(A2opt,a,a)
b<-tab$prop[(3+a)]
propopt<-c(propopt,b,b)
a<-which(tab[5:6,"prop"]==max(tab[5:6,"prop"]) ) -1
A2opt<-c(A2opt,a,a)
b<-tab$prop[(5+a)]
propopt<-c(propopt,b,b)
a<-which(tab[7:8,"prop"]==max(tab[7:8,"prop"]) ) -1
A2opt<-c(A2opt,a,a)
b<-tab$prop[(7+a)]
propopt<-c(propopt,b,b)
tab$A2opt<-A2opt
tab$propA2opt<-propopt
tab
}
```

Optimal Dynamic Treatment Regimes

```
> A2opt.r(mkcogtab.r())
  A2 H2 A1 Freq    prop A2opt propA2opt
1  0  0  0  410 0.29268     0   0.29268
2  1  0  0   30 0.00000     0   0.29268
3  0  1  0  160 0.18750     1   0.66667
4  1  1  0   30 0.66667     1   0.66667
5  0  0  1  280 0.10714     1   0.50000
6  1  0  1   20 0.50000     1   0.50000
7  0  1  1  190 0.42105     0   0.42105
8  1  1  1   70 0.28571     0   0.42105
```

Optimal Dyanmic Treatment Regimes

- Note that `A2opt` is d_2^{opt} as a function of A_1 and H_2.
- This enables us to find a_1^{opt} using `A1opt.r`.

Optimal Dynamic Treatment Regimes

```
> A1opt.r
function(tab)
{
a<-prop.table(xtabs(Freq~H2+A1,data=tab),2)
A10prop<-a[1,1]*tab$propA2opt[1] + a[2,1]*tab$propA2opt[3]
A11prop<-a[1,2]*tab$propA2opt[5] + a[2,2]*tab$propA2opt[7]
if (A10prop > A11prop){
A1opta<-0
propA1opta<-A10prop
}
if (A11prop > A10prop){
A1opta<-1
propA1opta<-A11prop
}
A1opt<-rep(A1opta,8)
propA1opt<-rep(propA1opta,8)
tab$A1opt<-A1opt
tab$propA1opt<-propA1opt
tab
}
> A1opt.r(A2opt.r(mkcogtab.r()))
  A2 H2 A1 Freq    prop A2opt propA2opt A1opt propA1opt
1  0  0  0  410 0.29268     0   0.29268     1   0.46335
2  1  0  0   30 0.00000     0   0.29268     1   0.46335
3  0  1  0  160 0.18750     1   0.66667     1   0.46335
4  1  1  0   30 0.66667     1   0.66667     1   0.46335
5  0  0  1  280 0.10714     1   0.50000     1   0.46335
6  1  0  1   20 0.50000     1   0.50000     1   0.46335
7  0  1  1  190 0.42105     0   0.42105     1   0.46335
8  1  1  1   70 0.28571     0   0.42105     1   0.46335
```

Optimal Dynamic Treatment Regimes

- Note that `A1opt` is a_1^{opt}, which is estimated as 1.
- Putting this together with the results of `A2opt`, we observe that the optimal dynamic treatment regime is to set $A_1 = 1$ and then to set $A_2 = 1$ if $H_2 = 0$ or $A_2 = 0$ if $H_2 = 1$.
- The marginal survival probability following implementation of the optimal dynamic treatment regime is estimated at 0.463.
- Conditional on H_2, this survival probability increases to 0.500 if $H_2 = 0$ and decreases to 0.421 if $H_2 = 1$.

Optimal Dynamic Treatment Regimes

- Finally, we can use the bootstrap to determine the proportion of bootstrap samples with the bootstrap estimate of a_1^{opt} equal to the observed estimate of a_1^{opt}, and likewise we can compute bootstrap confidence intervals for the survival probability corresponding to the optimal treatments.
- Because we are using the bootstrap to compute estimates not returned by the `boot` function, we need to write our own bootstrap function using the `sample` command.
- The functions `optimal.r` and `bootoptimal.r` provide the estimates for us.

Optimal Dynamic Treatment Regimes

```
> optimal.r
function(tab)
{
propA1opt<-tab$propA1opt
A1opt<-tab$A1opt[1]
propA1opt<-tab$propA1opt[1]
if (A1opt==0)
{
A2optH20<-tab$A2opt[1]
propA2optH20<-tab$propA2opt[1]
A2optH21<-tab$A2opt[3]
propA2optH21<-tab$propA2opt[3]
}
if (A1opt==1)
{
A2optH20<-tab$A2opt[5]
propA2optH20<-tab$propA2opt[5]
A2optH21<-tab$A2opt[7]
propA2optH21<-tab$propA2opt[7]
}
c(A1opt,propA1opt,A2optH20,propA2optH20,A2optH21,propA2optH21)
}
```

Optimal Dynamic Treatment Regimes

```
> bootoptimal.r
function()
{
orig<-optimal.r(A1opt.r(A2opt.r(mkcogtab.r())))
dat<-cogdat
Nboot<-1000
A1opt<-NULL
propA1opt<-NULL
propA2optH20<-NULL
propA2optH21<-NULL
for (i in 1:Nboot)
{
ids<-sample(c(1:nrow(cogdat)),replace=T)
out<-optimal.r(A1opt.r(A2opt.r(mkcogtab.r(dat,ids))))
A1opt<-c(A1opt,out[1])
propA1opt<-c(propA1opt,out[2])
propA2optH20<-c(propA2optH20,out[4])
propA2optH21<-c(propA2optH21,out[6])
}
pA1opt<-mean(A1opt)
SEpropA1opt<-sqrt(var(propA1opt))
```

Optimal Dynamic Treatment Regimes

```
lclpropA1opt<-orig[2]-1.96*SEpropA1opt
uclpropA1opt<-orig[2]+1.96*SEpropA1opt
SEpropA2optH20<-sqrt(var(propA2optH20))
lclpropA2optH20<-orig[4]-1.96*SEpropA2optH20
uclpropA2optH20<-orig[4]+1.96*SEpropA2optH20
SEpropA2optH21<-sqrt(var(propA2optH21))
lclpropA2optH21<-orig[6]-1.96*SEpropA2optH21
uclpropA2optH21<-orig[6]+1.96*SEpropA2optH21
list(pA1opt=pA1opt,lclpropA1opt=lclpropA1opt,
uclpropA1opt=uclpropA1opt,lclpropA2optH20=lclpropA2optH20,
uclpropA2optH20=uclpropA2optH20,lclpropA2optH21=lclpropA2optH21,
uclpropA2optH21=uclpropA2optH21)
}
```

Optimal Dynamic Treatment Regimes

```
> bootoptimal.r()
$pA1opt
[1] 0.805
$lclpropA1opt
[1] 0.36094
$uclpropA1opt
[1] 0.56575
$lclpropA2optH20
[1] 0.25783
$uclpropA2optH20
[1] 0.74217
$lclpropA2optH21
[1] 0.18804
$uclpropA2optH21
[1] 0.65406
```

Optimal Dynamic Treatment Regimes

- We observe that a_1^{opt} equals one for 80.5% of the bootstrap samples.
- We now can attach confidence intervals to our estimated survival probabilities.
- The marginal survival probability following implementation of the optimal dynamic treatment regime is estimated at 0.463 (0.361, 0.566).
- Conditional on H_2, this survival probability increases to 0.500 (0.258, 0.742) if $H_2 = 0$ and decreases to 0.421 (0.188, 0.654) if $H_2 = 1$.
- We observe that the survival probabilities conditional on H_2 are quite variable.