Homework: Probabilistic Bias Analysis

Due 21 April 2020

Katherine Wolf

Questions

1. Complete the following table with the hazard ratio and corresponding 95% quantile-based interval estimates in each cell (place the results for the standard analysis in the right most column of row 1). ***(10 points)***:

Table 1: Hazard ratios and 95% quantile intervals (QIs) for standard analysis and bias analysis of unmeasured confounding for the relationship between less-than-definitive therapy and breast cancer mortality.

Analysis	Systematic Error Only	Systematic and Random Error
Standard analysis	N/A	2.030 (1.182, 3.487)
Bias analysis (Uniform)	1.882 (1.597, 2.057)	1.865 (1.067, 3.273)
Bias analysis (Triangle)	1.881 (1.620, 2.046)	1.862 (1.074, 3.262)

2. From the model that assumed uniform distributions on the bias parameters:

    a. Report the median and 95% quantile interval of: 1) the parameter estimates (intercept and coefficient on `defnther`) from the logistic model for U and 2) the HR for the mortality-U relationship. ***(5 points)***

Table 2: Parameter estimates and 95% quantile intervals (QIs) (assuming Uniform distributions on the bias parameters) for 1) the logistic model modeling unmeasured confounding by indication (U), and 2) the bias (U) coefficient in the Cox proportional hazards model of breast cancer mortality.

Parameter	Estimate (95% QI)
Logistic intercept e^{β_0} (odds of $U = 1$)	0.573 (0.419, 0.775)
Logistic no definitive therapy e^{β_1} (OR)	2.052 (1.232, 3.433)
Cox mortality-U hazard ratio	1.750 (1.014, 3.051)

    b. Based on these, does the specification of the distribution of bias parameters seem to be consistent with a *confounding by indication scenario?* Why or why not? ***(10 points)***

Given the distributions of the unmeasured confounder (`U`) we specified according to receipt of definitive therapy status (`defnther`) and breast cancer mortality status (`bccause`), the logistic regression of `U` on `defnther` shows that a participant receiving the definitive therapy (`defnther = 0`) had odds 0.573 (95% QI: 0.419, 0.775) of having the unmeasured confounder (`U = 1`), and that those receving less than definitive therapy (`defnther = 1`) had 2.052 (95% QI: 1.232, 3.433) times the odds of having the unmeasured confounder (`U = 1`) compared to those who received definitive therapy (`defnther = 0`). Thus receiving less than definitive therapy appears associated with approximately a doubling of the odds of having the unmeasured confounder.

Also given the distributions of the unmeasured confounder we specified by receipt of definitive therapy and breast cancer mortality, according to the Cox proportional hazards regression of time to death from breast cancer `mortime2` on receipt of less than definitive therapy, the unmeasured confounder, and measured confounders region (`excat1`) and age (`agecat1` and `agecat2`), participants modeled to have the unmeasured confounder (for whom `U = 1`) appear to have 1.750 (95% QI: 1.014, 3.051) times the hazard of breast cancer mortality of those without the unmeasured confounder. Thus `U`, if present, does seem to be associated with an increase in the hazard of breast cancer mortality.

Since the unmeasured confounder appears positively associated with both receipt of less than definitive therapy and increased hazard of breast cancer mortality, this specification of the bias parameters does seem to be consistent with a confounding by indication scenario wherein an unmeasured factor affects both therapy assignment and breast cancer mortality.

3. Between the analyses with the Uniform distribution and the Triangular distribution on the bias parameters, which one has the more narrow bias-corrected interval estimates for the HR between `defntther` and mortality (it won't be much, but should be evident)? Why do you think that is the case? *(Hint: consider the shape of the priors on the bias parameter distributions.)* ***(10 points)***

The analyses that drew their bias parameter estimates from the Triangle distribution have narrower bias-corrected interval estimates for the HR between `defntther` and breast cancer mortality than the analyses that drew their bias parameter estimates from the Uniform distribution.

Drawing bias parameters from a Triangle instead of a Uniform distribution with the same endpoints concentrates the probability density of the bias parameter distribution in the center (pointy part!), thus reducing the variance in the bias parameter estimate going into the model. Reducing the bias parameter estimate variance then reduces the corresponding variance in its association with the outcome of interest (here, hazard of breast cancer mortality), thus decreasing the overall variance in the model and allowing more statistically efficient estimation of the exposure-outcome relationship of interest (here, the association of receiving less than definitive therapy with breast cancer mortality hazard).[1]

[1] I maybe shouldn't admit this, but I honestly don't quite understand the mathematics behind how reducing the variance of the bias parameter estimate increases the statistical efficiency of the estimator of the expoure-outcome relationship. If you have a source, I'd love to read it!

4. In no more than 4 sentences (points will be deduced for overage) provide an interpretation of the results of your bias analysis that you might include in a manuscript. Make references to 1) the original (uncorrected) estimate, 2) the treatment of the bias parameters, 3) how the bias corrected estimates compare to the original estimates, including the direction of the potential bias (not necessarily corresponding to individual sentences). You may focus on the results from the analysis assuming Uniform distributions on the bias parameters. *(Hint: see slide set for introduction to bias analysis.)* ***(15 points)***

Although Cox proportional hazards model estimates based on the observed data alone showed that exposure to receipt of less than definitive therapy was associated with 2.030 (95% confidence interval (CI): 1.182, 3.487) times the hazard of the outcome of breast cancer mortality after controlling for region and age, because therapy was assigned based on clinical indication instead of at random, we assessed the potential for confounding by indication via two probabilistic bias sensitivity analyses that accounted for both systematic and random error in the bias parameters. Based on the literature on clinical contradindications for definitive therapy, we drew random estimates of a binary unmeasured confounder variable from both Uniform (for sensitivity analysis (SA) 1) and Triangle (for SA2) distributions with assumed probabilities of occurrence ranging from 30 to 40% for unexposed participants who did not have the outcome of interest (death from breast cancer during the study period), 45 to 55% for both the unexposed who had the outcome and the unexposed who did not have the outcome, and 60 to 70% for the exposed who had the outcome. Both sensitivity analyses showed that controlling for potential confounding by indication under the bias parameter distributions we specified attenuated but did not eliminate the association between receipt of less than definitive therapy and breast cancer mortality hazard (SA1 HR: 1.865 (95% CI: 1.067, 3.273); SA2 HR: 1.862 (95% CI: 1.074, 3.262)). Thus confounding by clinical indication under the distributional assumptions we specified might have biased our results away from the null but did not fully account for the observed association between receipt of less than definitive therapy and breast cancer mortality.

R code

```
knitr::opts_chunk$set(echo = FALSE,
                      eval = TRUE,
                      results = 'hide',
                      warning = FALSE,
                      message = FALSE,
                      global.par=TRUE)


library(survival)
load("senssamp.rdata")

# rename columns?
colnames(senssamp) <- tolower(colnames(senssamp))

# fit cox proportional hazards model
obs.model <- coxph(Surv(mortime2, bccause) ~
                     defnther +
                     excat1 +
                     agecat1 +
                     agecat2,
                   ties="efron",
                   data = senssamp)
# summary(obs.model)

# prepare to sample over the distribution of bias parameters 5000 times
N.obs <- nrow(senssamp) # Determine number of observations in data

N.samp <- 5000 # Specify the number of samples

# On your own: create variables for the above parameters (named p.U00, p.U10, p.U01, and p.U11,

# The authors specify these with Uniform distributions, with p00 ranging from 30%-40%, p10 rangi

set.seed(123)

p.U00 <- runif(n = N.samp, min = 0.3, max = 0.4)
p.U10 <- runif(n = N.samp, min = 0.45, max = 0.55)
p.U01 <- runif(n = N.samp, min = 0.45, max = 0.55)
p.U11 <- runif(n = N.samp, min = 0.6, max = 0.7)

# Initialize storage vectors for the parameter estimates
HR.systematic <- # for HR corrected for systematic error only
  HR.total <- # for HR corrected for systematic + accounting for sampling error
  HR.U <- matrix(NA,
```

```
                ncol = 1,
                nrow = N.samp) # For the Y-X (defnther) and Y-U relationships

a.X <- matrix(NA,
              ncol = 2,
              nrow = N.samp) # For the U-X (defnther) relationship

attach(senssamp)

for (i in 1:N.samp){

  # Calculate probability of U for current value of bias parameters
  # given therapy (defnther) and mortality (bccause):
  p.U <- p.U00[i]*(1-defnther)*(1-bccause) +  # Definitive therapy, no death
         p.U10[i]*defnther*(1-bccause) + # Less than definitive therapy, no death
         p.U01[i]*(1-defnther)*bccause + # Definitive therapy, death
         p.U11[i]*defnther*bccause # Less than definitive therapy, death

  U <- rbinom(N.obs, 1, p.U) # Sample the unmeasured confounder

  # ***** COMPLETE THIS *****
  ##### Fit a logistic regression model here,
  ##### for the outcome U and a single predictor defnther
  U.model <- glm(U ~ defnther,
                 family = 'binomial',
                 data = senssamp)

  ##### Fit a Cox proportional hazards model here, parameterized as above
  ##### but including the unmeasured confounder just sampled.
  bias.model <- coxph(Surv(mortime2,
                           bccause) ~
                      defnther +
                      excat1 +
                      agecat1 +
                      agecat2 +
                      U,
                    ties = "efron",
                    data = senssamp)

  # ************************
  # Accumulate coefficients
  a.X[i,] <- coef(U.model) # Save coefficients from U model

  b.systematic <- coef(bias.model)[1] # beta coefficient on therapy variable
  se.systematic <- sqrt(vcov(bias.model)[1,1]) # standard error of therapy estimate

  HR.U[i] <- exp(coef(bias.model)[5]) # HR for U
```

```
  HR.systematic[i] <- exp(b.systematic) # HR accounting for systematic error only

  b.total <- rnorm(1,b.systematic, se.systematic) # Adding in random variability

  HR.total[i] <- exp(b.total) # HR accounting for systematic + random error
}

detach(senssamp)

round(quantile(HR.systematic, c(.5, .025, .975)), 3)

round(quantile(HR.total, c(.5, .025, .975)), 3)

round(quantile(HR.U, c(.5, .025, .975)), 3)

colnames(a.X) <- c("Intercept",
                   "defnther")

t(apply(a.X, 2, quantile, c(.5, .025, .975)))

### TO BE COMPLETED: REDO ABOVE WITH TRIANGULAR DISTRIBUTIONS ON BIAS PARAMETERS:

library(survival)
load("senssamp.rdata")

# rename columns?
colnames(senssamp) <- tolower(colnames(senssamp))

# fit cox proportional hazards model
t.obs.model <- coxph(Surv(mortime2,
                          bccause) ~
                     defnther +
                     excat1 +
                     agecat1 +
                     agecat2,
                   ties="efron",
                   data = senssamp)
summary(t.obs.model)

# prepare to sample over the distribution of bias parameters 5000 times
t.N.obs <- nrow(senssamp) # Determine number of observations in data

t.N.samp <- 5000 # Specify the number of samples

library(triangle)

set.seed(123)
```

```
t.p.U00 <- rtriangle(n = t.N.samp, a = 0.3, b = 0.4)
t.p.U10 <- rtriangle(n = t.N.samp, a = 0.45, b = 0.55)
t.p.U01 <- rtriangle(n = t.N.samp, a = 0.45, b = 0.55)
t.p.U11 <- rtriangle(n = t.N.samp, a = 0.6, b = 0.7)

# Initialize storage vectors for the parameter estimates
t.HR.systematic <- # for HR corrected for systematic error only
  t.HR.total <- # for HR corrected for systematic + accounting for sampling error
  t.HR.U <- matrix(NA, ncol = 1, nrow = t.N.samp) # For the Y-X (defnther) and Y-U relationships
t.a.X <- matrix(NA, ncol = 2, nrow = t.N.samp) # For the U-X (defnther) relationship

attach(senssamp)

for (i in 1:t.N.samp){

  # Calculate probability of U for current value of bias parameters
  # given therapy (defnther) and mortality (bccause):
  t.p.U <- t.p.U00[i]*(1-defnther)*(1-bccause) +  # Definitive therapy, no death
           t.p.U10[i]*defnther*(1-bccause) + # Less than definitive therapy, no death
           t.p.U01[i]*(1-defnther)*bccause + # Definitive therapy, death
           t.p.U11[i]*defnther*bccause # Less than definitive therapy, death

  t.U <- rbinom(t.N.obs, 1, t.p.U) # Sample the unmeasured confounder

  # ***** COMPLETE THIS *****
  ##### Fit a logistic regression model here,
  ##### for the outcome U and a single predictor defnther
  t.U.model <- glm(t.U ~ defnther,
                   family = 'binomial',
                   data = senssamp)

  ##### Fit a Cox proportional hazards model here, parameterized as above
  ##### but including the unmeasured confounder just sampled.
  t.bias.model <- coxph(Surv(mortime2,
                             bccause) ~
                        defnther +
                        excat1 +
                        agecat1 +
                        agecat2 +
                        t.U,
                      ties = "efron",
                      data = senssamp)

  # ************************
  # Accumulate coefficients
  t.a.X[i,] <- coef(t.U.model) # Save coefficients from U model
```

```
  t.b.systematic <- coef(t.bias.model)[1] # beta coefficient on therapy variable
  t.se.systematic <- sqrt(vcov(t.bias.model)[1,1]) # standard error of therapy estimate

  t.HR.U[i] <- exp(coef(t.bias.model)[5]) # HR for U
  t.HR.systematic[i] <- exp(t.b.systematic) # HR accounting for systematic error only

  t.b.total <- rnorm(1,t.b.systematic, t.se.systematic) # Adding in random variability

  t.HR.total[i] <- exp(t.b.total) # HR accounting for systematic + random error
}

detach(senssamp)

round(quantile(t.HR.systematic, c(.5, .025, .975)), 2)

round(quantile(t.HR.total, c(.5, .025, .975)), 2)

round(quantile(t.HR.U, c(.5, .025, .975)), 2)

colnames(t.a.X) <- c("Intercept","defntther")

t(apply(t.a.X, 2, quantile, c(.5, .025, .975)))


obs.hr <- sprintf("%.3f",
                  round(summary(obs.model)$coef['defntther',
                                                 'exp(coef)'],
                3))

obs.lo <- round(summary(obs.model)$conf.int['defntther',
                                             'lower .95'],
                 3)

obs.hi <- round(summary(obs.model)$conf.int['defntther',
                                            'upper .95'],
                3)


unif.sys.hr <- round(quantile(HR.systematic, 0.5), 3)

unif.sys.lo <- round(quantile(HR.systematic, 0.025), 3)

unif.sys.hi <- round(quantile(HR.systematic, 0.975), 3)


unif.rand.hr <- round(quantile(HR.total, 0.5), 3)
```

```
unif.rand.lo <- round(quantile(HR.total, 0.025), 3)

unif.rand.hi <- round(quantile(HR.total, 0.975), 3)


tri.sys.hr <- round(quantile(t.HR.systematic, 0.5), 3)

tri.sys.lo <- sprintf("%.3f",
                      round(quantile(t.HR.systematic, 0.025), 3))

tri.sys.hi <- round(quantile(t.HR.systematic, 0.975), 3)


tri.rand.hr <- round(quantile(t.HR.total, 0.5), 3)

tri.rand.lo <- round(quantile(t.HR.total, 0.025), 3)

tri.rand.hi <- sprintf("%.3f",
                       round(quantile(t.HR.total, 0.975), 3))


U.logist.coefs <- t(apply(a.X, 2, quantile, c(.5, .025, .975)))


U.int <- round(exp(U.logist.coefs['Intercept', '50%']), 3)

U.int.lo <- round(exp(U.logist.coefs['Intercept', '2.5%']), 3)

U.int.hi <- round(exp(U.logist.coefs['Intercept', '97.5%']), 3)


U.dft <- round(exp(U.logist.coefs['defntther', '50%']), 3)

U.dft.lo <- round(exp(U.logist.coefs['defntther', '2.5%']), 3)

U.dft.hi <- round(exp(U.logist.coefs['defntther', '97.5%']), 3)


cox.ay <- sprintf("%.3f",
              round(quantile(HR.U, c(0.5)), 3))

cox.ay.lo <- round(quantile(HR.U, c(0.025)), 3)

cox.ay.hi <- round(quantile(HR.U, c(0.975)), 3)
```