Homework: Bayesian Bias Analysis

Due 5 April 2020 at noon

Katherine Wolf

Questions

1. Complete the following table with the hazard ratio (HR) 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). ***(20 points)***:

Table 1: Posterior medians and 95% quantile-based intervals for parameters of standard analysis and bias analysis of unmeasured confounding for relationship between less-than-definitive therapy and breast cancer mortality.

Parameter	Standard Analysis	Bias Analysis 1	Bias Analysis 2
HR for < definitive therapy	2.027 (1.177, 3.491)	1.833 (1.023, 3.246)	2.038 (1.179, 3.604)
HR for U	N/A	1.792 (1.024, 3.186)	0.993 (0.481, 2.043)
Prevalence U in unexposed	N/A	0.376 (0.299, 0.456)	0.376 (0.300, 0.457)
Prevalence U in exposed	N/A	0.549 (0.451, 0.646)	0.55 (0.451, 0.645)

2. For the exposure HR (on less than definitive therapy), answer the following in a few sentences each:

a. Assuming the distribution of the parameters in **Bias Analysis 1**, what was the direction of the bias? Offer an intuitive explanation based on the imposed relationship between the confounder on the outcome (death), and its relationship with the exposure. *(Hint: Consider in the context of what the the average value of the bias parameters implies.)* **(10 points)**

In Bias Analysis 1, under the assumptions of 95% certainty that the prevalence of u among the exposed was between 0.45 and 0.65, that the prevalence of u among the unexposed was between 0.3 and 0.45, and that the hazard ratio for u was between 1 and 3, i.e., that the presence of the confounder increased the hazard of death somewhere between one and three times, the direction of the bias of the exposure HR assciated with the unmeasured confounder U was away from the null. Why? If the unmeasured confounder is more prevalent among the exposed and deadly, it will make the exposure appear more deadly than it actually is unless we control for it, since some of the extra people in the exposed group with the unmeasured confounder will die due to the effects of the unmeasured confounder and not due to the effects of the exposure. Indeed, the posterior bias analysis parameter estimates showed the unmeasured confounder to be associated with an increase in the hazard of the outcome ($e^{\beta_{yu}}$ 1.792, 95% credible interval 1.024, 3.186), and exposure to less than definitive therapy was associated with a greater hazard of death in the standard analysis (e^{β_x} 2.027, 95% credible interval 1.177, 3.491) than in Bias Analysis 1 (e^{β_x} 1.833, 95% credible interval 1.023, 3.246). In this particular case, an unmeasured confounder associated with both assignment to less than definitive therapy and death made receipt of less than definitive therapy appear to have a greater impact on probability of death than it actually does.

b. Assuming the distribution of the parameters in **Bias Analysis 2**, what was the direction of the bias? Offer an intuitive explanation based on the imposed relationship between the confounder on the outcome (death), and its relationship with the exposure. *(Hint: Consider in the context of what the the average value of the bias parameters implies.)* **(10 points)**

In Bias Analysis 2, under assumptions of 95% certainty that the prevalence of u among the exposed was still between 0.45 and 0.65, that the prevalence of u among the unexposed was still between 0.3 and 0.45, but that the hazard ratio for u was between 0.5 and 2 and centered on 1, i.e., no longer associated with the outcome of death, the unmeasured confounder did not bias the exposure HR in either direction. Why? If the unmeasured confounder does not have an effect on death, it cannot bias the relationship of the exposure to death, regardless of differences in prevalence among the exposed and unexposed. Even though we assumed the unmeasured confounder to be more prevalent among the exposed, the posterior estimates showed that it was not associated with an increase in the hazard of the outcome ($e^{\beta_{yu}}$ 0.993, 95% credible interval 0.481, 2.043), and its lack of association in either direction with the outcome meant that its effect on the exposure HR was negligible, shown by similar exposure HRs in both the unadjusted model (e^{β_x} 2.027, 95% credible interval 1.177, 3.491) and the model adjusted for bias (e^{β_x} 2.038, 95% credible interval 1.179, 3.604).

3. For the exposure HR (on less than definitive therapy), for each of the analyses (Standard, Bias Analysis 1, Bias Analysis 2): divide the upper limit of the credible interval by the lower limit, and report the 3 credible limit ratios (CrLR). Compare these CrLRs for each of the bias analysis results to the standard analysis. Describe what you see (i.e. for each, are they more/less precise than the standard analysis). Focusing on the comparison of Bias Analysis 2 vs. Standard Analysis: why do you think the pattern is as you observe (given the average bias implied by the distribution of bias parameters)? **(10 points)**

The CrLRs for the exposure HR from Bias Analyses 1, 3.173, and 2, 3.057, are both larger than the CrLR for the standard analysis, 2.966, indicating less precision in the estimate of the increase in the hazard of death associated with exposure to less than definitive therapy under the assumptions in Bias Analyses 1 and 2 than under the standard analysis assuming no unmeasured confounding. Even Bias Analysis 2 showed less precision in the exposure HR (e^{β_x}) estimate than the standard analysis, even though the point estimate itself didn't change appreciably. Why? The bias analysis adds an coefficient (bias parameter) β_{yu} to estimate the Y-U relationship, assumed to be drawn from a probability distribution with some associated variance, to the model of the hazard of the outcome Y associated with the exposure X. The bias analysis thus incorporates uncertainty in the degree to which U is biasing the estimate of the exposure-outcome relationship e^{β_x} into that estimate, increasing its uncertainty (decreasing its precision), even when U is assumed probably to not have an effect on the outcome (i.e, the distribution of $e^{\beta_{yu}}$ is assumed to be centered on 1) and thus the point estimate of the exposure-outcome relationship e^{β_x} itself doesn't change.

Table 2: Posterior credible limit ratios for the exposure hazard ratio on less than definitive therapy of standard analysis and bias analysis of unmeasured confounding for relationship between less-than-definitive therapy and breast cancer mortality.

Standard Analysis CrLR	Bias Analysis 1 CrLR	Bias Analysis 2 CrLR
2.966	3.173	3.057

R code

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


##### PB HLTH 250C: Advanced Epidemiologic Methods
##### R script for Bayesian Bias Analysis Lab Homework
##### 23 April 2020

library(R2jags)
library(epiR)
library(survival)
library(SurvRegCensCov)
library(epiR)
library(tictoc) # To time things

load("senssamp.rdata")
colnames(senssamp) <- tolower(colnames(senssamp))

# Correction to the dataset from last week's assignment:
senssamp$mortime2[senssamp$mortime2==0] <- .01 # Survival times need to be >1

########## Standard analysis of less than definitive therapy on breast cancer mortality:
##### Frequentist Analysis (Optional, for comparison):
obs.model <- survreg(Surv(mortime2, bccause) ~ defnther + excat1 + agecat1 + agecat2,
                     data=senssamp, dist="weibull")
ConvertWeibull(obs.model)$HR

##### Standard Bayesian analysis (Weibull model):
jags.weibull <- function(){
     # SAMPLING DISTRIBUTION
     for (i in 1:N) {
          log(lambda[i]) <- b[1] + b[2]*defnther[i] + b[3]*excat1[i] + b[4]*agecat1[i] +
               b[5]*agecat2[i];

          censored[i] ~ dinterval(t[i], c[i]);
          t[i] ~ dweib(shape, lambda[i]);
     }

     # Prior on betas (log-HR):
     b[1:N.x] ~ dmnorm(mu.b[1:N.x], tau.b[1:N.x, 1:N.x]) # multivariate normal prior
```

```
     # Prior on shape parameter for Weibull:
     shape ~ dlnorm(0,8); # Log-normal

     # Calculate HRs:
     for (l in 1:N.x) {
          HR[l] <- exp(b[l]);
     }
}

# Parameters and priors
N.x <- 5                        # Number of predictors in outcome model
N <- nrow(senssamp)                    # Total number of observations
mu.b <- rep(0, N.x)             # Prior mean for regression parameters
tau.b <- diag(10^-2, N.x)       # Prior precision on regression parameters

# Outcome data:
# Define death time and censoring time:
senssamp$t <- senssamp$c <- senssamp$mortime2
senssamp$t[senssamp$bccause == 0] <- NA # If censored (death indicator = 0) then death time is m
senssamp$c[senssamp$bccause == 1] <- 6 # If death (death = 1) then censoring time > death time
senssamp$censored <- 1 - senssamp$bccause # Indicator of censoring (=1)

# Data and initial values for JAGS
weibull.model.data <- list(N=N, N.x=N.x,
                           mu.b=mu.b, tau.b=tau.b,
                           t=senssamp$t, c=senssamp$c, censored=senssamp$censored,
                           defnther=senssamp$defnther, excat1=senssamp$excat1,
                           agecat1=senssamp$agecat1, agecat2=senssamp$agecat2)

set.seed(123)
tic()
standard.weibull <- jags.parallel(model=jags.weibull, data=weibull.model.data,
                        parameters.to.save = c("b","HR","shape"),
                        n.iter=100000, n.thin=10, n.chains=3,
                        jags.seed=123)
toc()
# print(standard.weibull)
# plot(as.mcmc(standard.weibull), ask=T) # Uncomment for plots
# autocorr.plot(as.mcmc(standard.weibull))

############################################################
########## Bayesian bias analysis of unmeasured confounding
############################################################

##### Bias Analysis 1
jags.weibull.conf <- function(){
        for (i in 1:N) {
```

```
                # SAMPLING DISTRIBUTION

                # ***** MODIFY THE FOLLOWING EXPRESSION FOR THE HAZARD TO
                # ***** INCLUDE THE EFFECT OF THE UNMEASURED CONFOUNDER
                # ***** ASSUMING THE LOG-HAZARD RATIO IS CALLED b.u
                # ***** AND THE COVARIATE IS CALLED U:
                log(lambda[i]) <- b[1] + b[2]*defnther[i] + b[3]*excat1[i] +
                        b[4]*agecat1[i] + b[5]*agecat2[i] + b.u*U[i];

                censored[i] ~ dinterval(t[i], c[i]);
                t[i] ~ dweib(shape, lambda[i]);

                # Distributon for the unmeasured confounder:
                pi.u[i] <- p.0*(1-defnther[i]) +   # Definitive therapy
                        p.1*defnther[i]          # Less than definitive therapy

                U[i] ~ dbin(pi.u[i],1) # Sample the unmeasured confounder
        }

        # Priors on betas:
        b[1:N.x] ~ dmnorm(mu.b[1:N.x], tau.b[1:N.x, 1:N.x]) # multivariate normal prior
        b.u ~ dnorm(mu.b.u, tau.b.u) # the prior for the bias parameter b.u

        # Shape parameter for Weibull:
        shape ~ dlnorm(0,8);

        # Priors for bias parameters:
        # ***** SPECIFY PRIORS FOR THE BIAS PARAMETERS
        # ***** THAT DEFINE THE DISTRIBUTION OF U WITH RESPECT
        # ***** TO THE EXPOSURE GROUPS.
        # ***** DEFINE THESE IN TERMS OF HYPERPARAMETERS:
        # ***** a.0, b.0 (for p.U0)
        # ***** a.1, b.1 (for p.U1)
        p.0 ~  dbeta(a.0, b.0)  # Definitive therapy, did not die
        p.1 ~  dbeta(a.1, b.1)  # Less than definitive therapy, did not die

        # Calculate HRs:
                for (l in 1:N.x) {
                        HR[l] <- exp(b[l]);
                }
        HR.u <- exp(b.u);
}

# ***** ACCORDING TO THE ABOVE DESCRIPTION OF THE BIAS PARAMETERS:
mu.b.u <- (log(3) + log(1))/2 # ***** SPECIFY THE VALUE FOR THE MEAN OF THE PRIOR FOR b.u
tau.b.u <- ((log(3) - log(1))/(2*1.96))^-2 # ***** SPECIFY THE VALUE FOR THE PRECISION FOR THE P
```

```r
# Uses epi.betabuster to obtain parameters for Beta distribution
# that satisfies the stated inputs (mode and lower quantile):
# ***** COMPLETE THE FOLLOWING TWO FUNCTION CALLS:
beta.p0 <- epi.betabuster(mode= (0.3 + 0.45)/2, conf= 0.975,
                          greaterthan = TRUE, x= 0.3)
beta.p1 <- epi.betabuster(mode= (0.45 + 0.65)/2, conf= 0.975,
                          greaterthan = TRUE, x= 0.45)

# Pulls shape parameters out of above:
a.0 <- beta.p0$shape1; b.0 <- beta.p0$shape2
a.1 <- beta.p1$shape1; b.1 <- beta.p1$shape2

# Confirm that above parameterizations yield distribution on bias
# parameters that are consistent with beliefs:
round(qlnorm(c(0.025, 0.975), mu.b.u, 1/sqrt(tau.b.u))) # Quantiles for prior HR.u
round(qbeta(c(.025, .975), a.0, b.0), 2) # Quantiles from prior for p.0
round(qbeta(c(.025, .975), a.1, b.1), 2) # Quantiles from prior for p.1

# Add these hyperparameters to the previous data list:
weibull.model.data2 <- weibull.model.data
weibull.model.data2[["mu.b.u"]] <- mu.b.u
weibull.model.data2[["tau.b.u"]] <- tau.b.u
weibull.model.data2[["a.0"]] <- a.0
weibull.model.data2[["b.0"]] <- b.0
weibull.model.data2[["a.1"]] <- a.1
weibull.model.data2[["b.1"]] <- b.1

set.seed(123)
tic()
conf.weibull <- jags.parallel(model=jags.weibull.conf, data=weibull.model.data2,
                              parameters.to.save = c("b","b.u","HR","shape","HR.u",
                                                     "p.0","p.1"),
                              n.iter=100000, n.thin=5, n.chains=3,
                              jags.seed=123)
toc()
# print(conf.weibull)
# plot(as.mcmc(conf.weibull), ask=T)




##### Bias Analysis 2
# ***** ACCORDING TO THE ABOVE DESCRIPTION OF THE BIAS PARAMETERS:
mu.b.u.2 <- (log(2) + log(0.5))/2# ***** SPECIFY THE VALUE FOR THE MEAN OF THE PRIOR FOR b.u # t
tau.b.u.2 <- ((log(2) - log(0.5))/(2*1.96))^-2# ***** SPECIFY THE VALUE FOR THE PRECISION FOR TH

# Checking
```

```
round(qlnorm(c(0.025, 0.975), mu.b.u.2, 1/sqrt(tau.b.u.2)))

weibull.model.data3 <- weibull.model.data2
weibull.model.data3[["mu.b.u"]] <- mu.b.u.2 # refreshing hyperparameters
weibull.model.data3[["tau.b.u"]] <- tau.b.u.2

set.seed(123)
tic()
conf.weibull.2 <- jags.parallel(model=jags.weibull.conf, data=weibull.model.data3,
                                parameters.to.save = c("b","b.u","HR","shape","HR.u",
                                                       "p.0","p.1"),
                                n.iter=100000, n.thin=10, n.chains=3,
                                jags.seed=123)
toc()
# print(conf.weibull.2)
# plot(as.mcmc(conf.weibull.2), ask=T)

### Summarize:
# Standard model summary
standard.model <- summary(as.mcmc(standard.weibull))$quantiles["HR[2]", c(3,1,5)]
round(standard.model,2)

# Bias analysis summary
to.keep <- c("HR[2]", "HR.u", "p.0","p.1")

BA1 <- summary(as.mcmc(conf.weibull))$quantiles[to.keep, c(3,1,5)]
BA2 <- summary(as.mcmc(conf.weibull.2))$quantiles[to.keep, c(3,1,5)]

BA.all <- cbind(BA1, BA2)
colnames(BA.all) <- rep(c("Median","2.5%","97.5%"),2)
round(BA.all,3)

st.HR <- round(standard.weibull$BUGSoutput$median$HR[2], 3)
st.HR.LL <- round(standard.weibull$BUGSoutput$summary["HR[2]", "2.5%"], 3)
st.HR.UL <- round(standard.weibull$BUGSoutput$summary["HR[2]", "97.5%"], 3)

BA1.HR <- round(BA1["HR[2]", "50%"], 3)
BA1.HR.LL <- round(BA1["HR[2]", "2.5%"], 3)
BA1.HR.UL <- round(BA1["HR[2]", "97.5%"], 3)

BA1.HR.U <- round(BA1["HR.u", "50%"], 3)
BA1.HR.U.LL <- round(BA1["HR.u", "2.5%"], 3)
BA1.HR.U.UL <- round(BA1["HR.u", "97.5%"], 3)

BA1.HR.p.0 <- round(BA1["p.0", "50%"], 3)
BA1.HR.p.0.LL <- round(BA1["p.0", "2.5%"], 3)
BA1.HR.p.0.UL <- round(BA1["p.0", "97.5%"], 3)
```

```
BA1.HR.p.1 <- round(BA1["p.1", "50%"], 3)
BA1.HR.p.1.LL <- round(BA1["p.1", "2.5%"], 3)
BA1.HR.p.1.UL <- round(BA1["p.1", "97.5%"], 3)


BA2.HR <- round(BA2["HR[2]", "50%"], 3)
BA2.HR.LL <- round(BA2["HR[2]", "2.5%"], 3)
BA2.HR.UL <- round(BA2["HR[2]", "97.5%"], 3)

BA2.HR.U <- round(BA2["HR.u", "50%"], 3)
BA2.HR.U.LL <- round(BA2["HR.u", "2.5%"], 3)
BA2.HR.U.UL <- round(BA2["HR.u", "97.5%"], 3)

BA2.HR.p.0 <- round(BA2["p.0", "50%"], 3)
BA2.HR.p.0.LL <- round(BA2["p.0", "2.5%"], 3)
BA2.HR.p.0.UL <- round(BA2["p.0", "97.5%"], 3)

BA2.HR.p.1 <- round(BA2["p.1", "50%"], 3)
BA2.HR.p.1.LL <- round(BA2["p.1", "2.5%"], 3)
BA2.HR.p.1.UL <- round(BA2["p.1", "97.5%"], 3)

# credible intervals
st.CrLR <- round(st.HR.UL/st.HR.LL, 3)
BA1.CrLR <- round(BA1.HR.UL/BA1.HR.LL, 3)
BA2.CrLR <- round(BA2.HR.UL/BA2.HR.LL, 3)
```