WHI Table

2023-10-21

```r
ntimes <- 100L

image_path <- '~/Documents/HW/Research/CI/causalbootstrap/images'
dat_path <- '~/Documents/HW/Research/CI/causalbootstrap/data'
whi_path <- '~/Documents/HW/Research/CI/WHI/case-study'

setwd('~/Documents/HW/Research/CI/causalbootstrap/code')

source('causal_funcs.R')

data <- foreign::read.dta(file.path(whi_path, "final_os_12 (2).dta"))
# REMOVE UNUSED LEVELS IN INCOME
data$income <- droplevels(data$income)

data <- data %>% dplyr::select(TIME_CHD, totp, f45multi,f45mvmin,ethnic,parity,
                               booph,meno,brca_f2,colon_f2,endo_f2,skin_f2,melan_f2,
                               othca10y,dvt,stroke,mi,diab,hicholrp,osteopor,cvd,cabg,
                               atrialfb,aortican,angina,hip55,smokevr,alcohol,fruits,
                               vegtabls,f60enrgy,syst_bl,dias_bl,bmix_bl,educ,income,
                               time_since_menopause, syst)

data <- data %>% dplyr::mutate(TIME_CHD = TIME_CHD/365,
                               f45multi = factor(f45multi),
                               f45mvmin = factor(f45mvmin),
                               brca_f2 = factor(brca_f2),
                               colon_f2 = factor(colon_f2),
                               endo_f2 = factor(endo_f2),
                               skin_f2 = factor(skin_f2),
                               melan_f2 = factor(melan_f2),
                               othca10y = factor(othca10y),
                               diab = factor(diab))

################################################################################
#Multiple Imputation

Xm <- data %>% dplyr::select(TIME_CHD, totp, f45multi,f45mvmin,ethnic,parity,
                             booph,meno,brca_f2,colon_f2,endo_f2,skin_f2,melan_f2,
                             othca10y,dvt,stroke,mi,diab,hicholrp,osteopor,cvd,cabg,
                             atrialfb,aortican,angina,hip55,smokevr,alcohol,fruits,
                             vegtabls,f60enrgy,syst_bl,dias_bl,bmix_bl,educ,income,
                             time_since_menopause, syst)

print(sapply(Xm, function(x) sum(is.na(x))/dim(data)[1]))
```

```
##           TIME_CHD              totp           f45multi
```

```
##         0.0000000000         0.0000000000         0.0000000000
##             f45mvmin               ethnic               parity
##         0.0000000000         0.0026915275         0.0057341239
##                booph                 meno              brca_f2
##         0.0169683258         0.0611249805         0.0008581682
##              colon_f2              endo_f2              skin_f2
##         0.0022234358         0.0010532064         0.0021064129
##              melan_f2              othca10y                  dvt
##         0.0032376346         0.0168122952         0.0007801529
##                stroke                   mi                 diab
##         0.0002340459         0.0005070994         0.0007411453
##              hicholrp              osteopor                  cvd
##         0.0191917616         0.0108831331         0.0186846622
##                  cabg              atrialfb             aortican
##         0.0137696989         0.0139647371         0.0138087065
##                angina                hip55               smokevr
##         0.0053830551         0.1802543298         0.0065922921
##               alcohol               fruits              vegtabls
##         0.0044858792         0.0000000000         0.0000000000
##               f60enrgy              syst_bl              dias_bl
##         0.0000000000         0.0008191606         0.0010141988
##                bmix_bl                 educ               income
##         0.0110781713         0.0073334374         0.0423623030
## time_since_menopause                 syst
##         0.0611249805         0.0246138243
```

```r
init  <-  mice(Xm, maxit=0)
meth  <-  init$method
predM <-  init$predictorMatrix

if(!file.exists(file.path(dat_path, 'imputed.rds'))){
  imputed <- mice(Xm, method="pmm", predictorMatrix=predM, m=5)
  imputed <- mice::complete(imputed)
  saveRDS(object = imputed, file.path(dat_path, 'imputed.Rds'))
} else{
  imputed <- readRDS(file.path(dat_path, 'imputed.rds'))
}
data_imputed <- imputed

data_imputed$totp <- recode(data_imputed$totp, 'No'=0, 'Yes'=1)

data_imputed$TIME_CHD <- data$TIME_CHD
data_imputed$failure <- data$failure
data_imputed$time_since_menopause <- data$time_since_menopause

confounds <- c('totp', 'time_since_menopause', 'TIME_CHD')
confounds <- names(data_imputed)[!names(data_imputed) %in% confounds]
for(i in confounds){
  if(class(data_imputed[[i]]) == 'factor'){
    levels(data_imputed[[i]]) <- make.names(levels(data_imputed[[i]]))
  }
}

timemen <- c('0-10', '$<$10-20', '20+')
```

```r
#Analysis
data_imputed_1 <- data_imputed %>% dplyr::filter(time_since_menopause>0 & time_since_menopause<=10)
confounders_1 <- data_imputed_1 %>% dplyr::select(-c(TIME_CHD, syst, totp, time_since_menopause))
intervention_1 <- data_imputed_1$totp

data_imputed_2 <- data_imputed %>% dplyr::filter(time_since_menopause>10 & time_since_menopause<=20)
confounders_2 <- data_imputed_2 %>% dplyr::select(-c(TIME_CHD, syst,totp, time_since_menopause))
intervention_2 <- data_imputed_2$totp

data_imputed_3 <- data_imputed %>% dplyr::filter(time_since_menopause>20)
confounders_3 <- data_imputed_3 %>% dplyr::select(-c(TIME_CHD, syst,totp,time_since_menopause))
intervention_3 <- data_imputed_3$totp

form <- paste0('totp ~ ', paste(confounds, collapse = ' + '))
methods <- c('svmLinear2', 'ps', 'cbps')
dat_list <- list(data_imputed_1, data_imputed_2, data_imputed_3)

blb_params <- expand.grid(B = c(200),
                          subsets = c(2, 4, 10),
                          gamma = c(0.8),
                          method = c('svmLinear2', 'ps', 'cbps'),
                          stringsAsFactors = FALSE)

if(!file.exists(file.path(dat_path, 'whi_cblb.rds'))){
  whi_cblb <- lapply(seq_len(nrow(blb_params)), function(didx){
    dg_row <- blb_params[didx, ]
    print(dg_row)
    cblb_dats <- lapply(dat_list, function(dt){
      time_out <- c()
      for(i in seq_len(ntimes)){
        times <- system.time({
          part <- weight_blb(data = data.table::as.data.table(dt),
                             R = dg_row$B,
                             pi_formula = as.formula(form),
                             Y = 'TIME_CHD',
                             W = 'totp',
                             subsets = dg_row$subsets,
                             b = round(nrow(dt)^dg_row$gamma),
                             method = dg_row$method,
                             type = 'obs',
                             prop_grid = data.frame(cost = 0.01))

          part <- purrr::transpose(part)
          theta_reps <- part$theta_reps

          cis <- lapply(theta_reps, function(boot_reps){
            boot:::perc.ci(t = boot_reps)[, 4:5]
          })
          cis <- Reduce(`+`, cis)/dg_row$subsets
          estim <- sapply(theta_reps, mean)
          estim <- mean(estim)
          se <- sapply(theta_reps, sd)
          se <- mean(se)
```

```
      })
      time_out[i] <- times['elapsed']
    }
    data.table(estim = estim,
               se = se,
               lower = cis[1],
               upper = cis[2],
               B = dg_row$B,
               subsets = dg_row$subsets,
               method = dg_row$method,
               time = median(time_out))

  })
  cblb_dats <- rbindlist(cblb_dats)
  cblb_dats[, `:=`(n = dg_row$n_size,
                   gamma = dg_row$gamma,
                   timemen = timemen)]
  })
  whi_cblb <- rbindlist(whi_cblb)
  saveRDS(whi_cblb, file.path(dat_path, 'whi_cblb.rds'))
} else{
  whi_cblb <- readRDS(file.path(dat_path, 'whi_cblb.rds'))
}

ipw_vals <- lapply(dat_list, function(d){
  vals <- sapply(methods, function(m){
    if(m == 'svmLinear2'){
      wts <- make_weights_ML(formula = as.formula(form),
                             data = d,
                             normed = TRUE,
                             method = m,
                             prop_grid = data.frame(cost = 0.01))
    } else{
      wts <- make_weights(formula = as.formula(form),
                          data = d,
                          method = m,
                          normed = TRUE)
    }
    Y <- d$totp*(d$TIME_CHD) - (1-d$totp)*(d$TIME_CHD)
    sum(Y*wts)
  })
  data.table(method = methods,
             full = vals)
})
ipw_vals <- rbindlist(ipw_vals)
ipw_vals[, `:=`(timemen = rep(timemen, each = length(dat_list)))]
whi_cblb <- merge(whi_cblb, ipw_vals, by = c('method', 'timemen'))
whi_cblb <- whi_cblb[gamma == 0.8 & subsets == 10]
whi_cblb[, `:=`(method = car::recode(method, "'ps'='PS';'cbps'='CBPS';'svmLinear2'='SVM'"),
                CI = paste0(' (', round(lower, 2), ', ', round(upper, 2), ')'),
                timemen = factor(timemen, levels = c('0-10', '$<$10-20', '20+')))]
whi_cblb[, `:=`(upper = NULL, lower = NULL, B = NULL, gamma = NULL, subsets = NULL)]
setcolorder(whi_cblb, c('timemen', 'method', 'full', 'estim', 'se', 'CI', 'time'))
```

```
setorder(whi_cblb, timemen)

whi_cblb %>%
  kbl(digits = 2, escape = FALSE, booktabs = TRUE,
      col.names = c('Time since menopause', 'Method', 'Full', 'Estim.', 'SE', 'CI', 'Med. Time (s)')) %>
  kable_classic(html_font = "Cambria") %>%
  save_kable(file = file.path(image_path, 'whiobs.pdf'), keep_tex = TRUE)
```