eAppendix

Cohort construction

Data were simulated for a hypothetical cohort of 30,000 patients with day-level object drug and precipitant drug exposures and outcome (Y).

For each patient, the following variables were generated:

- Patient ID
- Days enrolled – generated as a random variable from a gamma distribution with shape of 2 and mean of 365 days. Only the first 365 days were used (maximum days enrolled = 365)
- A binary object drug exposure was generated for 80% of the patients
    - Start date – for those with the object drug present, a start date was then generated as a random variable from a uniform distribution within each patient's enrollment period.
    - Time on object drug – for those with the object drug present, time on the drug was generated as a random variable from a gamma distribution with shape of 0.75 and mean of 130 days (median approximated 90 days). A 14-day minimum exposure was imposed.
- A binary precipitant drug exposure was generated for 90% of the patients.
    - Start date – for those with the precipitant drug present, a start date was then generated as a random variable from a uniform distribution within each patient's enrollment period.
    - The precipitant drug exposure was set to last for 14 days for all patients with the precipitant present.

One of the main assumptions for the case-crossover design is stable probability of exposure over time. In our simulated cohort, where all patients are new users of the drugs, there would be an initial increase in exposure prevalence as patients start the drugs (but few discontinue initially). To avoid bias due to a trend in exposure prevalence, we excluded the first 90 days for each patient (except in the scenario where bias due to changing precipitant drug exposure prevalence was evaluated). All patients who were exposed to a drug on day 91, were considered exposed as of that day until the end of their exposure.

Outcomes were generated as a function of a fixed baseline event probability (β_0), exposure to the object drug (β_1), exposure to the precipitant (β_2), or exposure to both (β_3). The daily probability of an outcome was calculated for each day of a patient's enrollment period using the following logistic regression model:

$$\text{Logit}\,(\Pr[Y_{ij}=1]) = -8 + \ln(2.5)\,\text{Object}_{ij} + \ln(1.0)\,\text{Precipitant}_{ij} + \ln(2.0)\,\text{Object}_{ij}*\text{Precipitant}_{ij}$$

where i is an individual patient and j is a unique day

Based on each patient's predicted probability of the outcome on a given day, we generated a binary outcome (Y_{ij} = 1, 0) for each day using a Bernoulli trial. Patients could have multiple outcomes during their enrollment period. No correlation between outcomes was modeled; thus, multiple outcomes within the same individual were independent.

Data generation and analysis R code

```
n_set<-1000 ##number of iterations
nsim<-30000 ##number of patients in each cohort
inc <--8  ##baseline rate of the outcome
a<-2.5 ##effect of the object drug
b<-1  ##effect of the precipitant
c<-2 ##effect of the interaction
results <- data.frame(simulation=seq(1:n_set))

for(i in 1:n_set){
  dat <- data.frame(ID=c(1:nsim))
  dat$enr <- round(rgamma(nsim,2,scale = 165)+121)
  dat$enr <- replace(dat$enr, dat$enr>365, 365)
  dat$obj <- rbinom(nsim, 1, .8)
  dat$obj.st <- ifelse(dat$obj==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$t.obj <- ifelse(dat$obj==1, round(rgamma(nsim,0.75, scale =170)+14), NA)
  dat$obj.end <- dat$obj.st + dat$t.obj
  dat$obj.end <- ifelse(dat$obj.end > dat$enr, dat$enr, dat$obj.end)
  dat$pre <- rbinom(nsim, 1, .9)
  dat$pre.st <- ifelse(dat$pre==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$pre.end <- dat$pre.st + 13
  dat$pre.end <- ifelse(dat$pre.end > dat$enr, dat$enr, dat$pre.end)
   ## Excluding the first 90 days from each person
  dat$obj.st <- ifelse(dat$obj.st <=90, NA, dat$obj.st-90)
  dat$obj.end <- ifelse(dat$obj.end <=90, NA, dat$obj.end-90)
  dat$pre.st <- ifelse(dat$pre.st <=90, NA, dat$pre.st-90)
  dat$pre.end <- ifelse(dat$pre.end <=90, NA, dat$pre.end-90)
  dat$obj.st <- ifelse(is.na(dat$obj.st)& dat$obj.end >0, 1, dat$obj.st)
  dat$pre.st <- ifelse(is.na(dat$pre.st)& dat$pre.end >0, 1, dat$pre.st)
  dat$enr <- dat$enr-90
  dat$obj <- ifelse(dat$obj.end >0 & !is.na(dat$obj.end), 1, 0)
  dat$pre <- ifelse(dat$pre.end >0 & !is.na(dat$pre.end), 1, 0)
  dat <- dat[dat$pre==1 & dat$obj==1,] # excluding people not exposed to both drugs
  dat <- dat[dat$pre.end >= dat$obj.st, ] # excluding people with precipitant drug exposure before object drug
  dat <- dat[dat$pre.st <= dat$obj.end, ] # excluding people with precipitant drug exposure after object drug ends
  # reshape data to long format
  dat2 <- data.frame("id"=unlist(sapply(1:nrow(dat), function(i) rep(dat$ID[i], dat$enr[i]))))
  dat2$day <- unlist(sapply(1:nrow(dat), function(i) 1:dat$enr[i]))
  dat2$obj <- 0
  dat2$pre <- 0
  for(j in 1:nrow(dat)) {
   dat2$obj[dat2$id==dat$ID[j]] <- switch(dat$obj[j]+1,
                                   rep(0, dat$enr[j]),
                                   c(rep(0, dat$obj.st[j]-1),
                                     rep(1, dat$obj.end[j]-dat$obj.st[j]+1),
                                     rep(0, dat$enr[j]-dat$obj.end[j])))
   #precipitant
```

```
  dat2$pre[dat2$id==dat$ID[j]] <- switch(dat$pre[j]+1,
                                      rep(0, dat$enr[j]),
                                      c(rep(0, dat$pre.st[j]-1),
                                        rep(1, dat$pre.end[j]-dat$pre.st[j]+1),
                                        rep(0, dat$enr[j]-dat$pre.end[j])))
 }
 #outcome generation
 dat2$logit.py <- inc + log(a)*dat2$obj + log(b)*dat2$pre + log(c)*dat2$obj*dat2$pre
 dat2$y <- rbinom(nrow(dat2), size=1, prob=plogis(dat2$logit.py))
#############################################################
 # Case-crossover take 1st outcome on the object drug
 ############################################################
 dat_obj<-dat[dat$obj==1,]
 dat3 <- dat2[dat2$obj==1,]
 dat_obj$y <- tapply(dat3$day, list(dat3$y, dat3$id), function(x) x[1])[2,]
 cases1 <-dat_obj[!is.na(dat_obj$y),]  # Restrict the dataset to events only
 cases1 <-cases1[cases1$obj.st<= cases1$y & cases1$y <= cases1$obj.end ,]
 events1<- cases1[cases1$obj.st<= cases1$y-30,] #restrict to those with outcome after at least 30 days of object exp
 events1$pre_hazard <- ifelse(events1$pre.st <= events1$y & events1$y <= events1$pre.end &
                              !is.na(events1$pre.st), 1, 0)
 events1$pre_control <- ifelse(events1$pre.st <= events1$y-30 & events1$y-30 <= events1$pre.end &
                               !is.na(events1$pre.st), 1, 0)
 events1 <- events1[,c("ID", "pre_hazard", "pre_control")]
 events1$contr <- ifelse(events1$pre_hazard== events1$pre_control,0, 1)
 results$N_contr.1st[i] <- sum(events1$contr)
 events1 <- reshape(events1,
                    varying = c("pre_control", "pre_hazard"),
                    v.names = c ("Prec"),
                    timevar = "time",
                    times = c(0, 1),
                    direction = "long")
 library(survival)
 mod1<-clogit(time ~ Prec+strata(id), data=events1)
 results$coef.CCO_1st[i] <- summary(mod1)$coefficients[1,1]
 results$se.CCO_1st[i] <- summary(mod1)$coefficients[1,3]
 results$OR.CCO_1st[i] <- summary(mod1)$coefficients[1,2]
 results$LCL.CCO_1st[i] <- exp(confint(mod1))[1,1]
 results$UCL.CCO_1st[i] <- exp(confint(mod1))[1,2]
  #############################################################
 #SCCS take 1st outcome on the object drug
 ############################################################
 cases_sccs<-cases1
 library(SCCS)
 sccs_1<-standardsccs( event~pre.st, indiv = ID, astart = obj.st, aend = obj.end, aevent = y,
                     adrug = pre.st, aedrug = pre.end, data=cases_sccs )
 results$coef.SCCS_1st[i] <- sccs_1$coefficients[1,1]
 results$se.SCCS_1st[i] <- sccs_1$coefficients[1,3]
 results$OR.SCCS_1st[i] <- sccs_1$coefficients[1,2]
 results$LCL.SCCS_1st[i] <- sccs_1$conf.int[1,3]
 results$UCL.SCCS_1st[i] <- sccs_1$conf.int[1,4]
 ############################################################
 #COUNTING ALL OUTCOMES
 ############################################################
 dat_obj$y = NULL
  count <- tapply(dat3$y, dat3$id, sum)
 for(k in 1:max(count)){
```

```
   vn <- paste0("y", k)
   dat_obj[vn] <- tapply(dat3$day, list(dat3$y, dat3$id), function(x) x[k])[2,]
  }
  dat_M <- reshape(dat_obj, varying = paste0("y", 1:max(count)),
                v.names = "y", idvar = "ID", direction = "long")
  dat_M <- dat_M[!is.na(dat_M$y),] #drop patients who don't have an outcome
  dat_M <-dat_M[dat_M$obj.st<= dat_M$y & dat_M$y <= dat_M$obj.end ,]  # Doublecheck  outcomes occur on
object drug
  dat_M$ID2 <- seq.int(nrow(dat_M))
  ##########################################################################################
  #Case-crossover ALL OUTCOMES (on object drug)
  ########################################################################################
  cases2<-dat_M
  events1<- cases2[cases2$obj.st<= cases2$y-30,]
  events1$pre_hazard <- ifelse(events1$pre.st <= events1$y & events1$y <= events1$pre.end &
                              !is.na(events1$pre.st), 1, 0)
  events1$pre_control <- ifelse(events1$pre.st <= events1$y-30 & events1$y-30 <= events1$pre.end &
                              !is.na(events1$pre.st), 1, 0)
  events1 <- events1[,c("ID","ID2", "pre_hazard", "pre_control")]
  events1$contr <- ifelse(events1$pre_hazard== events1$pre_control,0, 1)
  results$N_contr.all[i]  <- sum(events1$contr)
   events1 <- reshape(events1,
                  varying = c("pre_control", "pre_hazard"),
                  v.names = c ("Prec_all"),
                  timevar = "time",
                  times = c(0, 1),
                  direction = "long")
   modM<-clogit(time ~ Prec_all+strata(ID2), data=events1)
  results$coef.CCO_all[i] <- summary(modM)$coefficients[1,1]
  results$se.CCO_all[i] <- summary(modM)$coefficients[1,3]
  results$OR.CCO_all[i] <- summary(modM)$coefficients[1,2]
  results$LCL.CCO_all[i] <- exp(confint(modM))[1,1]
  results$UCL.CCO_all[i] <- exp(confint(modM))[1,2]
  ##############################################################
  # Case-series ALL OUTCOMES (on object drug)
  ##############################################################
  dat_M <- reshape(dat_obj, varying = paste0("y", 1:max(count)),
                v.names = "y", idvar = "ID", direction = "long")
  dat_M <- dat_M[!is.na(dat_M$y),] #drop patients who don't have an outcome
   library(SCCS)
  sccs<-standardsccs( event~pre.st, indiv = ID, astart = obj.st, aend = obj.end, aevent = y,
                   adrug = pre.st, aedrug = pre.end, data=dat_M )
    results$coef.SCCS_all[i] <- sccs$coefficients[1,1]
  results$se.SCCS_all[i] <- sccs$coefficients[1,3]
  results$OR.SCCS_all[i] <- sccs$coefficients[1,2]
  results$LCL.SCCS_all[i] <- sccs$conf.int[1,3]
  results$UCL.SCCS_all[i] <- sccs$conf.int[1,4]
}

## RESULTS
results$Bias.CCO_1st <- results$coef.CCO_1st- log(c)
results$Bias.CCO_all <- results$coef.CCO_all - log(c)
results$Bias.SCCS_1st <- results$coef.SCCS_1st - log(c)
results$Bias.SCCS_all <- results$coef.SCCS_all - log(c)
results$Bias_OR.CCO_1st <- results$OR.CCO_1st-c
results$Bias_OR.CCO_all <- results$OR.CCO_all-c
```

```
results$Bias_OR.SCCS_1st <- results$OR.SCCS_1st-c
results$Bias_OR.SCCS_all <- results$OR.SCCS_all-c
results_summary <- as.data.frame(colMeans(results))
results_summary$stdev <- apply(results, 2, sd)
results_summary<-rbind(nsim,inc,a,b,c,results_summary)
row.names(results_summary)[1]<-"Number of people"
row.names(results_summary)[2]<-"Baseline incidence"
row.names(results_summary)[3]<-"Object"
row.names(results_summary)[4]<-"Precipitant"
row.names(results_summary)[5]<-"Interaction"
results_summary<-rbind(results_summary, MSE_coef.CCO_1st = ((mean(results$Bias.CCO_1st)^2) +
(sd(results$coef.CCO_1st)^2)),
                MSE_coef.CCO_all = ((mean(results$Bias.CCO_all)^2) + (sd(results$coef.CCO_all)^2)),
                MSE_coef.SCCS_1st = ((mean(results$Bias.SCCS_1st)^2) + (sd(results$coef.SCCS_1st)^2)),
                MSE_coef.SCCS_all = ((mean(results$Bias.SCCS_all)^2) + (sd(results$coef.SCCS_all)^2))
)
```

Scenarios

Time-varying confounding

Precipitant drug confounder was simulated as a random binary variable for 40% of patients (80% in a separate scenario). For those with the covariate, the start day was generated from a uniform distribution and we set its duration (and its effect) to be 5 days. Exposure to the precipitant drug was generated as a function of exposure to the covariate so that the expected covariate prevalence among patients exposed to the precipitant was 3 times higher than prevalence among patents unexposed to the precipitant. For patients who had the covariate and the precipitant drug, precipitant drug exposure start day was the same as the start day for the covariate; for patients who did not have the covariate, the precipitant drug start day was generated as a random variable from a uniform distribution. The OR for the association between covariate and the outcome was set at 1.75.

Outcome was generated as a function of exposure to the object drug, precipitant drug, and the covariate:

$$\text{Logit}(\Pr[Y_{ij}=1]) = -8 + \ln(2.5)\,\text{Object}_{ij} + \ln(1.0)\,\text{Precipitant}_{ij} + \ln(2.0)\,\text{Object}_{ij}{*}\text{Precipitant}_{ij} + \ln(1.75)\,\text{Covariate}_{ij}$$

where *i* is an individual patient and *j* is a unique day.

Data were analyzed as described above (the confounder was treated as unmeasured; no adjustment introduced in the analysis).

R code for generating cohort for the scenario where 40% of patients were exposed to the confounder (relevant changes from base case in bold):

```
  setwd("~/Confounder/30TH")
n_set<-1000
nsim<-30000
```

```
inc <--8
a<-2.50
b<-1
c<-1
d<-1
e<-1.75
f <- 0.4

results <- data.frame(simulation=seq(1:n_set))
for(i in 1:n_set){
   dat <- data.frame(ID=c(1:nsim))
  dat$enr <- round(rgamma(nsim,2,scale = 165)+121)
  dat$enr <- replace(dat$enr, dat$enr>365, 365)
  dat$obj <- rbinom(nsim, 1, .8)
  dat$obj.st <- ifelse(dat$obj==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$t.obj <- ifelse(dat$obj==1, round(rgamma(nsim,0.75, scale =170)+14), NA)
  dat$obj.end <- dat$obj.st + dat$t.obj
  dat$obj.end <- ifelse(dat$obj.end > dat$enr, dat$enr, dat$obj.end)
   dat$conf2 <- rbinom(nsim, 1, .4)
  dat$conf2.st <- ifelse(dat$conf2==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$conf2.end <- dat$conf2.st + 4
  dat$conf2.end <- ifelse(dat$conf2.end > dat$enr, dat$enr, dat$conf2.end)
  dat$pre <- rbinom(nsim, 1, .85+0.12*dat$conf2)
  dat$pre.st <- ifelse(dat$pre==1,
                    ifelse (dat$conf2==1, dat$conf2.st, round(runif(nsim, min=1, max=dat$enr))),NA)
   dat$pre.end <- dat$pre.st + 13
  dat$pre.end <- ifelse(dat$pre.end > dat$enr, dat$enr, dat$pre.end)
    ## Excluding the first 90 days from each person
  dat$enr <- dat$enr-90
  dat$obj.st <- ifelse(dat$obj.st <=90, NA, dat$obj.st-90)
  dat$obj.end <- ifelse(dat$obj.end <=90, NA, dat$obj.end-90)
  dat$obj.st <- ifelse(is.na(dat$obj.st)& dat$obj.end >0, 1, dat$obj.st)
  dat$obj <- ifelse(dat$obj.end >0 & !is.na(dat$obj.end), 1, 0)
  dat$pre.st <- ifelse(dat$pre.st <=90, NA, dat$pre.st-90)
  dat$pre.end <- ifelse(dat$pre.end <=90, NA, dat$pre.end-90)
  dat$pre.st <- ifelse(is.na(dat$pre.st)& dat$pre.end >0, 1, dat$pre.st)
  dat$pre <- ifelse(dat$pre.end >0 & !is.na(dat$pre.end), 1, 0)
  dat$conf2.st <- ifelse(dat$conf2.st <=90, NA, dat$conf2.st-90)
  dat$conf2.end <- ifelse(dat$conf2.end <=90, NA, dat$conf2.end-90)
  dat$conf2.st <- ifelse(is.na(dat$conf2.st)& dat$conf2.end >0, 1, dat$conf2.st)
  dat$conf2 <- ifelse(dat$conf2.end >0 & !is.na(dat$conf2.end), 1, 0)
  dat <- dat[dat$pre==1 & dat$obj==1,] # get rid of people not exposed to both drug
  dat <- dat[dat$pre.end >= dat$obj.st, ] # get rid of people with precipitant drug exposure before object drug
  dat <- dat[dat$pre.st <= dat$obj.end, ] # get rid of people with precipitant drug exposure after object drug ends
  # reshape data to long format
  dat2 <- data.frame("id"=unlist(sapply(1:nrow(dat), function(i) rep(dat$ID[i], dat$enr[i]))))
  dat2$day <- unlist(sapply(1:nrow(dat), function(i) 1:dat$enr[i]))
  dat2$obj <- 0
  dat2$pre <- 0
  dat2$conf2 <- 0
  for(j in 1:nrow(dat)) {
    dat2$obj[dat2$id==dat$ID[j]] <- switch(dat$obj[j]+1,
                                   rep(0, dat$enr[j]),
                                   c(rep(0, dat$obj.st[j]-1),
                                     rep(1, dat$obj.end[j]-dat$obj.st[j]+1),
                                     rep(0, dat$enr[j]-dat$obj.end[j])))
```

```
   #precipitant
   dat2$pre[dat2$id==dat$ID[j]] <- switch(dat$pre[j]+1,
                                          rep(0, dat$enr[j]),
                                          c(rep(0, dat$pre.st[j]-1),
                                            rep(1, dat$pre.end[j]-dat$pre.st[j]+1),
                                            rep(0, dat$enr[j]-dat$pre.end[j])))
   # confounder for each day for each patient
     dat2$conf2[dat2$id==dat$ID[j]] <- switch(dat$conf2[j]+1,
                                            rep(0, dat$enr[j]),
                                            c(rep(0, dat$conf2.st[j]-1),
                                              rep(1, dat$conf2.end[j]-dat$conf2.st[j]+1),
                                              rep(0, dat$enr[j]-dat$conf2.end[j])))
  }
  #outcome generation
  dat2$logit.py <- inc + log(a)*dat2$obj + log(b)*dat2$pre + log(c)*dat2$obj*dat2$pre + log(e)*dat2$conf2
  dat2$y <- rbinom(nrow(dat2), size=1, prob=plogis(dat2$logit.py))
}
```

R code for generating cohort for the scenario where 80% of patients were exposed to the confounder:

```
 setwd("~/Confounder/30TH")
n_set<-1000
nsim<-30000
inc <--8
a<-2.50
b<-1
c<-2
d<-1
e<-1.75
f <- 0.8

results <- data.frame(simulation=seq(1:n_set))

for(i in 1:n_set){
    dat <- data.frame(ID=c(1:nsim))
  dat$enr <- round(rgamma(nsim,2,scale = 165)+121)
  dat$enr <- replace(dat$enr, dat$enr>365, 365)
  dat$obj <- rbinom(nsim, 1, .8)
  dat$obj.st <- ifelse(dat$obj==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$t.obj <- ifelse(dat$obj==1, round(rgamma(nsim,0.75, scale =170)+14), NA)
  dat$obj.end <- dat$obj.st + dat$t.obj
  dat$obj.end <- ifelse(dat$obj.end > dat$enr, dat$enr, dat$obj.end)
   dat$conf2 <- rbinom(nsim, 1, .8)
  dat$conf2.st <- ifelse(dat$conf2==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$conf2.end <- dat$conf2.st + 4
  dat$conf2.end <- ifelse(dat$conf2.end > dat$enr, dat$enr, dat$conf2.end)
  dat$pre <- rbinom(nsim, 1, .645+0.32*dat$conf2)
  dat$pre.st <- ifelse(dat$pre==1,
                    ifelse (dat$conf2==1, dat$conf2.st, round(runif(nsim, min=1, max=dat$enr))),NA)
   dat$pre.end <- dat$pre.st + 13
  dat$pre.end <- ifelse(dat$pre.end > dat$enr, dat$enr, dat$pre.end)
     ## Excluding the first 90 days from each person
  dat$enr <- dat$enr-90
  dat$obj.st <- ifelse(dat$obj.st <=90, NA, dat$obj.st-90)
```

```
 dat$obj.end <- ifelse(dat$obj.end <=90, NA, dat$obj.end-90)
 dat$obj.st <- ifelse(is.na(dat$obj.st)& dat$obj.end >0, 1, dat$obj.st)
 dat$obj <- ifelse(dat$obj.end >0 & !is.na(dat$obj.end), 1, 0)
 dat$pre.st <- ifelse(dat$pre.st <=90, NA, dat$pre.st-90)
 dat$pre.end <- ifelse(dat$pre.end <=90, NA, dat$pre.end-90)
 dat$pre.st <- ifelse(is.na(dat$pre.st)& dat$pre.end >0, 1, dat$pre.st)
 dat$pre <- ifelse(dat$pre.end >0 & !is.na(dat$pre.end), 1, 0)
 dat$conf2.st <- ifelse(dat$conf2.st <=90, NA, dat$conf2.st-90)
 dat$conf2.end <- ifelse(dat$conf2.end <=90, NA, dat$conf2.end-90)
 dat$conf2.st <- ifelse(is.na(dat$conf2.st)& dat$conf2.end >0, 1, dat$conf2.st)
 dat$conf2 <- ifelse(dat$conf2.end >0 & !is.na(dat$conf2.end), 1, 0)
  dat <- dat[dat$pre==1 & dat$obj==1,] # get rid of people not exposed to both drug
 dat <- dat[dat$pre.end >= dat$obj.st, ] # get rid of people with precipitant drug exposure before object drug
 dat <- dat[dat$pre.st <= dat$obj.end, ] # get rid of people with precipitant drug exposure after object drug ends
 # reshape data to long format
 dat2 <- data.frame("id"=unlist(sapply(1:nrow(dat), function(i) rep(dat$ID[i], dat$enr[i]))))
 dat2$day <- unlist(sapply(1:nrow(dat), function(i) 1:dat$enr[i]))
 dat2$obj <- 0
 dat2$pre <- 0
 dat2$conf2 <- 0
 for(j in 1:nrow(dat)) {
   dat2$obj[dat2$id==dat$ID[j]] <- switch(dat$obj[j]+1,
                                    rep(0, dat$enr[j]),
                                    c(rep(0, dat$obj.st[j]-1),
                                      rep(1, dat$obj.end[j]-dat$obj.st[j]+1),
                                      rep(0, dat$enr[j]-dat$obj.end[j])))
   #precipitant
   dat2$pre[dat2$id==dat$ID[j]] <- switch(dat$pre[j]+1,
                                    rep(0, dat$enr[j]),
                                    c(rep(0, dat$pre.st[j]-1),
                                      rep(1, dat$pre.end[j]-dat$pre.st[j]+1),
                                      rep(0, dat$enr[j]-dat$pre.end[j])))
   # confounder for each day for each patient
     dat2$conf2[dat2$id==dat$ID[j]] <- switch(dat$conf2[j]+1,
                                      rep(0, dat$enr[j]),
                                      c(rep(0, dat$conf2.st[j]-1),
                                        rep(1, dat$conf2.end[j]-dat$conf2.st[j]+1),
                                        rep(0, dat$enr[j]-dat$conf2.end[j])))
 }
  #outcome generation
 dat2$logit.py <- inc + log(a)*dat2$obj + log(b)*dat2$pre + log(c)*dat2$obj*dat2$pre + log(e)*dat2$conf2
 dat2$y <- rbinom(nrow(dat2), size=1, prob=plogis(dat2$logit.py))
}
```

Time trend in precipitant drug exposure

The case-crossover assumes that exposure probability is stationary over the sampled person-time in the absence of a causal relation with the outcome. We designed a scenario where the exposure to the precipitant drug increased linearly over time in the population (see figure below for exposure prevalence over the study period across 1,000 simulation iterations).

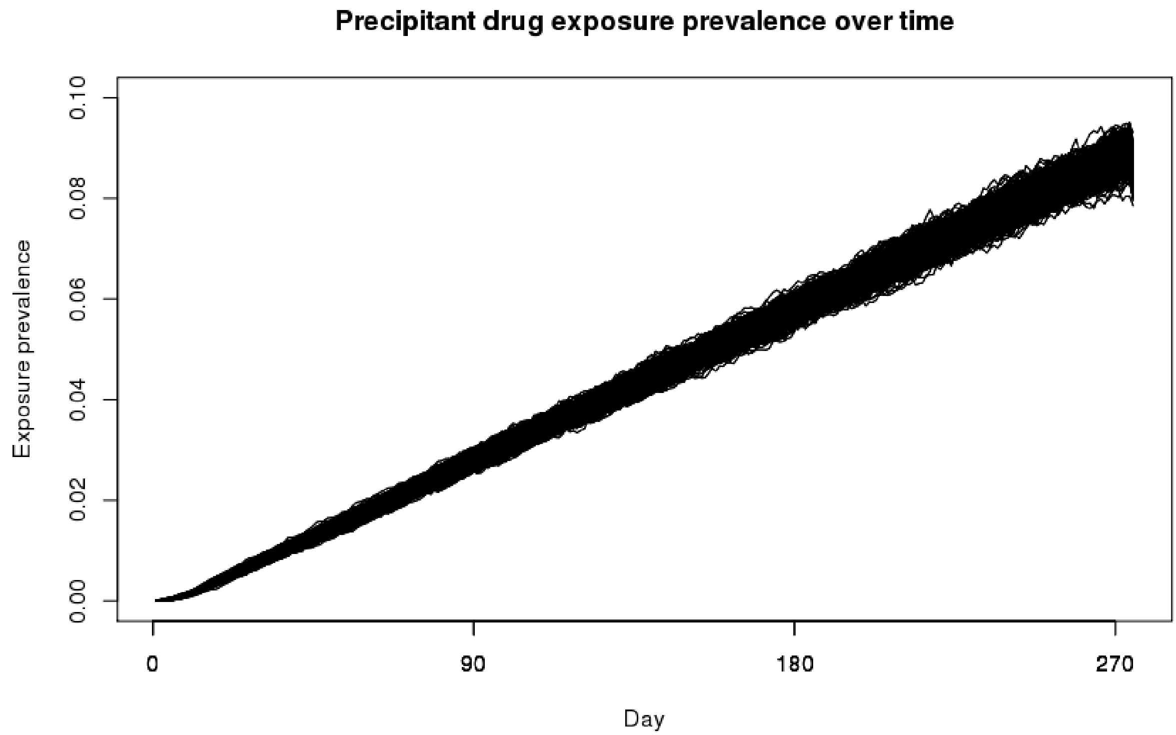
```
setwd("~/Linear/30TH")
n_set<-1000
```

```
nsim<-30000
inc <--8
a<-2.5
b<-1
c<-2
results <- data.frame(simulation=seq(1:n_set))
#ExpProb <- data.frame(cbind(1:275, matrix(rep(NA,275*n_set), nrow = 275)))
#names(ExpProb) <- c("Day", paste0("Exp_", 1:n_set))

for(i in 1:n_set){

  dat <- data.frame(ID=c(1:nsim))
  dat$enr <- round(rgamma(nsim,2,scale = 165)+121)
  dat$enr <- replace(dat$enr, dat$enr>365, 365)
  dat$obj <- rbinom(nsim, 1, .8)
  dat$obj.st <- ifelse(dat$obj==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$t.obj <- ifelse(dat$obj==1, round(rgamma(nsim,0.75, scale =170)+14), NA)
  dat$obj.end <- dat$obj.st + dat$t.obj
  dat$obj.end <- ifelse(dat$obj.end > dat$enr, dat$enr, dat$obj.end)
   ## Excluding the first 90 days from each person
  dat$obj.st <- ifelse(dat$obj.st <=90, NA, dat$obj.st-90)
  dat$obj.end <- ifelse(dat$obj.end <=90, NA, dat$obj.end-90)
  dat$obj.st <- ifelse(is.na(dat$obj.st)& dat$obj.end >0, 1, dat$obj.st)
  dat$enr <- dat$enr-90
  dat$obj <- ifelse(dat$obj.end >0 & !is.na(dat$obj.end), 1, 0)
  ## PRECIPITANT DRUG
  dat$t<-275
  u<-runif(nsim, min = 0.00001)
  dat$pre <- rbinom(nsim, 1, .9)
  dat$pre.st <- ifelse(dat$pre==1, round(sqrt(u)*dat$t), NA)
  dat$pre<- ifelse(dat$pre==1 & dat$pre.st > dat$enr, 0, dat$pre)
  dat$pre.end <- dat$pre.st + 13
  dat$pre.end <- ifelse(dat$pre.end > dat$enr, dat$enr, dat$pre.end)
   dat <- dat[dat$pre==1 & dat$obj==1,] # get rid of people not exposed to both drug
  dat <- dat[dat$pre.end >= dat$obj.st, ] # get rid of people with precipitant drug exposure before object drug
  dat <- dat[dat$pre.st <= dat$obj.end, ] # get rid of people with precipitant drug exposure after object drug ends
  # reshape data to long format
  dat2 <- data.frame("id"=unlist(sapply(1:nrow(dat), function(i) rep(dat$ID[i], dat$enr[i]))))
  dat2$day <- unlist(sapply(1:nrow(dat), function(i) 1:dat$enr[i]))
  dat2$obj <- 0
  dat2$pre <- 0
  for(j in 1:nrow(dat)) {
   dat2$obj[dat2$id==dat$ID[j]] <- switch(dat$obj[j]+1,
                                     rep(0, dat$enr[j]),
                                     c(rep(0, dat$obj.st[j]-1),
                                       rep(1, dat$obj.end[j]-dat$obj.st[j]+1),
                                       rep(0, dat$enr[j]-dat$obj.end[j])))
   #precipitant
   dat2$pre[dat2$id==dat$ID[j]] <- switch(dat$pre[j]+1,
                                     rep(0, dat$enr[j]),
                                     c(rep(0, dat$pre.st[j]-1),
                                       rep(1, dat$pre.end[j]-dat$pre.st[j]+1),
                                       rep(0, dat$enr[j]-dat$pre.end[j])))
   # repeat confounders for each day for each patient
  }
   #calculating probability of exposure on each day
```

```
  #Prob<-aggregate(dat2[, 4], list(Day=dat2$day), mean)
 #ExpProb[,i+1]<-aggregate(dat2[, 4], list(Day=dat2$day), mean)[,2]
 #outcome generation
 dat2$logit.py <- inc + log(a)*dat2$obj + log(b)*dat2$pre + log(c)*dat2$obj*dat2$pre
 dat2$y <- rbinom(nrow(dat2), size=1, prob=plogis(dat2$logit.py))
 }
```



eFigure. Daily precipitant drug exposure prevalence in 1,000 simulated cohorts (30,000 patients each) over the course of 1 year.

Non-transient precipitant drug exposure

Instead of a fixed, short-term precipitant drug exposure of 14 days, we simulated 2 scenarios, where (1) exposure to the precipitant drug lasted for 90 days for everybody; (2) 70% of patients with precipitant drug exposure were exposed for 30 days while the remaining 30% did not end their exposure (remained exposed until the end of the study period).

For the scenario with 30% of patients persistent on precipitant drug therapy, precipitant drug exposure was coded as follows:

```
  dat$pre <- rbinom(nsim, 1, .9)
 dat$pre.st <- ifelse(dat$pre==1, round(runif(nsim, min=1, max=dat$enr)), NA)
 dat$dc <- ifelse(dat$pre==1,round(rbinom(nsim, 1, .7)), 0)
 dat$pre.end <- ifelse(dat$pre==1, ifelse(dat$dc==1, dat$pre.st+29, dat$enr), NA)
 dat$pre.end <- ifelse(dat$pre.end > dat$enr, dat$enr, dat$pre.end)
```

Outcome affecting subsequent exposure to the object drug (event-based censoring)

The self-controlled case series method assumes that neither exposure nor observation periods depend on events. We simulated a scenario where exposure to the object drug is discontinued in some patients following the outcome, which is quite common with medications. In the designs nested within person-time exposed to the object drug, discontinuation of the object drug will lead to event-dependent censoring. We evaluated three such settings: (1) 10% of patients experiencing the outcome discontinue the object drug; (2) 50% of patients discontinue the object drug; and (3) 90% of patients discontinue the object drug.

```
setwd("~/Disc/30TH")
n_set<-1000
nsim<-30000
inc <--8
a<-2.5
b<-1
c<-2
d<-0.5 #percent discontinuing the object drug (0.1 and 0.9 in other two scenarios)
results <- data.frame(simulation=seq(1:n_set))

for(i in 1:n_set){
   dat <- data.frame(ID=c(1:nsim))
  dat$enr <- round(rgamma(nsim,2,scale = 165)+121)
  dat$enr <- replace(dat$enr, dat$enr>365, 365)
  dat$obj <- rbinom(nsim, 1, .8)
  dat$obj.st <- ifelse(dat$obj==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$t.obj <- ifelse(dat$obj==1, round(rgamma(nsim,0.75, scale =170)+14), NA)
  dat$obj.end <- dat$obj.st + dat$t.obj
  dat$obj.end <- ifelse(dat$obj.end > dat$enr, dat$enr, dat$obj.end)
  dat$pre <- rbinom(nsim, 1, .9)
  dat$pre.st <- ifelse(dat$pre==1, round(runif(nsim, min=1, max=dat$enr)), NA)
  dat$pre.end <- dat$pre.st + 13
  dat$pre.end <- ifelse(dat$pre.end > dat$enr, dat$enr, dat$pre.end)
   ## Excluding the first 90 days from each person
  dat$obj.st <- ifelse(dat$obj.st <=90, NA, dat$obj.st-90)
  dat$obj.end <- ifelse(dat$obj.end <=90, NA, dat$obj.end-90)
  dat$pre.st <- ifelse(dat$pre.st <=90, NA, dat$pre.st-90)
  dat$pre.end <- ifelse(dat$pre.end <=90, NA, dat$pre.end-90)
  dat$obj.st <- ifelse(is.na(dat$obj.st)& dat$obj.end >0, 1, dat$obj.st)
  dat$pre.st <- ifelse(is.na(dat$pre.st)& dat$pre.end >0, 1, dat$pre.st)
  dat$enr <- dat$enr-90
  dat$obj <- ifelse(dat$obj.end >0 & !is.na(dat$obj.end), 1, 0)
  dat$pre <- ifelse(dat$pre.end >0 & !is.na(dat$pre.end), 1, 0)
  dat <- dat[dat$pre==1 & dat$obj==1,] # get rid of people not exposed to both drug
  dat <- dat[dat$pre.end >= dat$obj.st, ] # get rid of people with precipitant drug exposure before object drug
  dat <- dat[dat$pre.st <= dat$obj.end, ] # get rid of people with precipitant drug exposure after object drug ends
  # reshape data to long format
  dat2 <- data.frame("id"=unlist(sapply(1:nrow(dat), function(i) rep(dat$ID[i], dat$enr[i]))))
  dat2$day <- unlist(sapply(1:nrow(dat), function(i) 1:dat$enr[i]))
  dat2$obj <- 0
  dat2$pre <- 0
  for(j in 1:nrow(dat)) {
   dat2$obj[dat2$id==dat$ID[j]] <- switch(dat$obj[j]+1,
```

```
                                  rep(0, dat$enr[j]),
                                  c(rep(0, dat$obj.st[j]-1),
                                    rep(1, dat$obj.end[j]-dat$obj.st[j]+1),
                                    rep(0, dat$enr[j]-dat$obj.end[j])))
   #precipitant
   dat2$pre[dat2$id==dat$ID[j]] <- switch(dat$pre[j]+1,
                                  rep(0, dat$enr[j]),
                                  c(rep(0, dat$pre.st[j]-1),
                                    rep(1, dat$pre.end[j]-dat$pre.st[j]+1),
                                    rep(0, dat$enr[j]-dat$pre.end[j])))
   }
  #outcome generation
  dat2$logit.py <- inc + log(a)*dat2$obj + log(b)*dat2$pre + log(c)*dat2$obj*dat2$pre
  dat2$y <- rbinom(nrow(dat2), size=1, prob=plogis(dat2$logit.py))
  ############################################################
  # Case-crossover take 1st outcome on the object drug
  ##########################################################
  dat_obj<-dat[dat$obj==1,]
  dat3 <- dat2[dat2$obj==1,]
  dat_obj$y <- tapply(dat3$day, list(dat3$y, dat3$id), function(x) x[1])[2,]
   ##censoring
  dat_obj$exp1 <- ifelse(!is.na(dat_obj$y), 1, 0)
  u<-nrow(dat_obj)
  dat_obj$censor1 <- rbinom(u, 1, d*dat_obj$exp1)
   dat_obj$obj.end <- ifelse(dat_obj$censor1==1, dat_obj$y, dat_obj$obj.end)
  ##analysis
  cases1 <-dat_obj[!is.na(dat_obj$y),]
  cases1 <-cases1[cases1$obj.st<= cases1$y & cases1$y <= cases1$obj.end ,]
  events1<- cases1[cases1$obj.st<= cases1$y-30,] #restrict to those with outcome after at least 30 days of object
drug
 events1$pre_hazard <- ifelse(events1$pre.st <= events1$y & events1$y <= events1$pre.end &
                              !is.na(events1$pre.st), 1, 0)
  events1$pre_control <- ifelse(events1$pre.st <= events1$y-30 & events1$y-30 <= events1$pre.end &
                               !is.na(events1$pre.st), 1, 0)
  events1 <- events1[,c("ID", "pre_hazard", "pre_control")]
  events1$contr <- ifelse(events1$pre_hazard== events1$pre_control,0, 1)
  results$N_contr.1st[i] <- sum(events1$contr)
  ##reshape the data for the matched analysis
  events1 <- reshape(events1,
                   varying = c("pre_control", "pre_hazard"),
                   v.names = c ("Prec"),
                   timevar = "time",
                   times = c(0, 1),
                   direction = "long")
  library(survival)
  mod1<-clogit(time ~ Prec+strata(id), data=events1)
  results$coef.CCO_1st[i] <- summary(mod1)$coefficients[1,1]
  results$se.CCO_1st[i] <- summary(mod1)$coefficients[1,3]
  results$OR.CCO_1st[i] <- summary(mod1)$coefficients[1,2]
  results$LCL.CCO_1st[i] <- exp(confint(mod1))[1,1]
  results$UCL.CCO_1st[i] <- exp(confint(mod1))[1,2]
  ############################################################
  #SCCS take 1st outcome on the object drug
  ##########################################################
  cases_sccs<-cases1
  library(SCCS)
```

```
 sccs_1<-standardsccs( event~pre.st, indiv = ID, astart = obj.st, aend = obj.end, aevent = y,
                    adrug = pre.st, aedrug = pre.end, data=cases_sccs )

 results$coef.SCCS_1st[i] <- sccs_1$coefficients[1,1]
 results$se.SCCS_1st[i] <- sccs_1$coefficients[1,3]
 results$OR.SCCS_1st[i] <- sccs_1$coefficients[1,2]
 results$LCL.SCCS_1st[i] <- sccs_1$conf.int[1,3]
 results$UCL.SCCS_1st[i] <- sccs_1$conf.int[1,4]
  ##########################################################
 # ALL OUTCOMES (on object drug)
 ##########################################################
 dat_obj$y = NULL
  count <- tapply(dat3$y, dat3$id, sum)
 for(k in 1:max(count)){
   vn <- paste0("y", k)
   dat_obj[vn] <- tapply(dat3$day, list(dat3$y, dat3$id), function(x) x[k])[2,]
 }
 ##censoring continuation (2nd outcomes)
  dat_obj$y2 <- ifelse(dat_obj$censor1==1, NA, dat_obj$y2)
  #2nd outcome
 dat_obj$exp2 <- ifelse(!is.na(dat_obj$y2), 1, 0)
 dat_obj$censor2 <- rbinom(u, 1, d*dat_obj$exp2)
  dat_obj$obj.end <- ifelse(dat_obj$censor2==1, dat_obj$y2, dat_obj$obj.end)
  dat_M2 <- reshape(dat_obj, varying = paste0("y", 1:max(count)),
                v.names = "y", idvar = "ID", direction = "long")
 dat_M2 <- dat_M2[!is.na(dat_M2$y),]
 dat_M2$censor <- ifelse(dat_M2$time==1 & dat_M2$censor1==1, 1,0)
 dat_M2$censor <- ifelse(dat_M2$censor==0 & dat_M2$time==2 & dat_M2$censor2==1, 1,dat_M2$censor)
 dat_M2$censor <- ifelse(dat_M2$censor==0 & dat_M2$time==3, 1,dat_M2$censor)
 dat_M2 <-dat_M2[dat_M2$obj.st<= dat_M2$y & dat_M2$y <= dat_M2$obj.end ,]
 dat_M2$ID2 <- seq.int(nrow(dat_M2))
 ###################################################################################
#Case-crossover
##################################################################################
cases2<-dat_M2
events1<- cases2[cases2$obj.st<= cases2$y-30,]  #restrict to those with outcome after at least 30 days of object drug
events1$pre_hazard <- ifelse(events1$pre.st <= events1$y & events1$y <= events1$pre.end &
                         !is.na(events1$pre.st), 1, 0)
events1$pre_control <- ifelse(events1$pre.st <= events1$y-30 & events1$y-30 <= events1$pre.end &
                          !is.na(events1$pre.st), 1, 0)
events1 <- events1[,c("ID","ID2", "pre_hazard", "pre_control")]
events1$contr <- ifelse(events1$pre_hazard== events1$pre_control,0, 1)
results$N_contr.all[i]  <- sum(events1$contr)
events1 <- reshape(events1,
                varying = c("pre_control", "pre_hazard"),
                v.names = c ("Prec_all"),
                timevar = "time",
                times = c(0, 1),
                direction = "long")
modM<-clogit(time ~ Prec_all+strata(ID2), data=events1)
results$coef.CCO_all[i] <- summary(modM)$coefficients[1,1]
results$se.CCO_all[i] <- summary(modM)$coefficients[1,3]
results$OR.CCO_all[i] <- summary(modM)$coefficients[1,2]
results$LCL.CCO_all[i] <- exp(confint(modM))[1,1]
results$UCL.CCO_all[i] <- exp(confint(modM))[1,2]
##########################################################
```

```
# Case-series
##########################################################
library(SCCS)
sccs<-standardsccs( event~pre.st, indiv = ID, astart = obj.st, aend = obj.end, aevent = y,
        adrug = pre.st, aedrug = pre.end, data=dat_M2 )
results$coef.SCCS_all[i] <- sccs$coefficients[1,1]
results$se.SCCS_all[i] <- sccs$coefficients[1,3]
results$OR.SCCS_all[i] <- sccs$coefficients[1,2]
results$LCL.SCCS_all[i] <- sccs$conf.int[1,3]
results$UCL.SCCS_all[i] <- sccs$conf.int[1,4]
}
## RESULTS
results$Bias.CCO_1st <- results$coef.CCO_1st- log(c)
results$Bias.CCO_all <- results$coef.CCO_all - log(c)
results$Bias.SCCS_1st <- results$coef.SCCS_1st - log(c)
results$Bias.SCCS_all <- results$coef.SCCS_all - log(c)
results$Bias_OR.CCO_1st <- results$OR.CCO_1st-c
results$Bias_OR.CCO_all <- results$OR.CCO_all-c
results$Bias_OR.SCCS_1st <- results$OR.SCCS_1st-c
results$Bias_OR.SCCS_all <- results$OR.SCCS_all-c
results_summary <- as.data.frame(colMeans(results))
results_summary$stdev <- apply(results, 2, sd)
results_summary<-rbind(nsim,inc,a,b,c,d,results_summary)
row.names(results_summary)[1]<-"Number of people"
row.names(results_summary)[2]<-"Baseline incidence"
row.names(results_summary)[3]<-"Object"
row.names(results_summary)[4]<-"Precipitant"
row.names(results_summary)[5]<-"Interaction"
row.names(results_summary)[6]<-"Prop. discontinued"
results_summary<-rbind(results_summary, MSE_coef.CCO_1st = ((mean(results$Bias.CCO_1st)^2) +
(sd(results$coef.CCO_1st)^2)),
          MSE_coef.CCO_all = ((mean(results$Bias.CCO_all)^2) + (sd(results$coef.CCO_all)^2)),
          MSE_coef.SCCS_1st = ((mean(results$Bias.SCCS_1st)^2) + (sd(results$coef.SCCS_1st)^2)),
          MSE_coef.SCCS_all = ((mean(results$Bias.SCCS_all)^2) + (sd(results$coef.SCCS_all)^2))
)
```

eTable 1. Comparison of parameter estimates from the case-crossover and the self-controlled case series designs across scenarios with varying duration of precipitant drug therapy and object drug discontinuation

Scenarios	Case-crossover				Self-controlled case series			
	Log estimate[1]	**Bias**	**MSE**	**OR**	**Log estimate[1]**	**Bias**	**MSE**	**IR**
Main scenarios (14-day precipitant)								
10% discontinue object drug	0.71	0.01	0.04	2.03	0.70	0.01	0.01	2.02
50% discontinue object drug	0.71	0.01	0.05	2.03	0.73	0.04	0.02	2.09
90% discontinue object drug	0.71	0.01	0.04	2.03	0.80	0.10	0.03	2.22
90-day precipitant								
10% discontinue object drug	0.70	0.01	0.02	2.02	0.71	0.02	0.01	2.03
50% discontinue object drug	0.72	0.02	0.02	2.04	0.80	0.10	0.02	2.22
90% discontinue object drug	0.74	0.04	0.02	2.09	0.97	0.28	0.09	2.64
70% exposed for 30 days and 30% do not discontinue precipitant drug (persistent use)								
10% discontinue object drug	1.06	0.36	0.16	2.87	0.74	0.05	0.01	2.11
50% discontinue object drug	1.06	0.37	0.16	2.88	1.00	0.31	0.11	2.72
90% discontinue object drug	1.07	0.37	0.16	2.90	1.40	0.70	0.51	4.04

[1]Parameter estimate on the log scale; mean value from 1,000 iterations. True value is 0.69 (OR of 2.0).

Bias calculated as the mean difference between the log estimate and the true value (averaged across 1,000 iterations); MSE – mean squared error; OR – odds ratio; IR – incidence ratio. Both OR and IR = exponential (mean log estimate).

eTable 2. Comparison of parameter estimates from the case-crossover and the self-controlled case series designs across all scenarios (analysis of 1st outcomes only)

Scenarios	Case-crossover				Self-controlled case series			
	Log estimate[1]	Bias	MSE	OR	Log estimate[1]	Bias	MSE	IR
Base case (no violation of assumptions)	0.71	0.02	0.04	2.03	0.70	<0.01	0.01	2.01
Time-varying confounder (prevalence 40%)	0.83	0.13	0.06	2.29	0.80	0.11	0.02	2.22
Time-varying confounder (prevalence 80%)	0.92	0.23	0.09	2.51	0.90	0.21	0.05	2.46
Time trend in precipitant drug exposure	0.95	0.25	0.11	2.58	0.67	-0.03	0.01	1.95
90-day precipitant	0.74	0.04	0.02	2.09	0.69	<0.01	0.01	2.00
Precipitant: 70% exposed for 30 days and 30% do not discontinue exposure	1.06	0.37	0.16	2.89	0.66	-0.03	0.01	1.94
Object drug discontinuation (10%)	0.71	0.02	0.04	2.03	0.70	0.01	0.01	2.02
Object drug discontinuation (50%)	0.71	0.02	0.05	2.04	0.74	0.04	0.02	2.09
Object drug discontinuation (90%)	0.71	0.01	0.04	2.03	0.80	0.11	0.03	2.22

[1]Parameter estimate on the log scale; mean value from 1,000 iterations. True value is 0.69 (OR of 2.0). Mean OR/IR estimated as exponential of the mean log estimate. MSE – mean squared error; OR – odds ratio; IR – incidence ratio