Simulation of repeated measures data from a clinical trial in idiopathic pulmonary fibrosis (IPF) with the endpoint forced vital capacity (FVC)

Generation of example data for the `brms.mmrm` package

15/09/2023

Contents

About

Data from a trials in idiopathic pulmonary fibrosis (IPF) with the endpoint 'forced vital capacity' (FVC) is simulated, following an approach described by Santermans et al. (2019) and R code given in the supplement of the article.

1 Prerequisites

Load packages:

```
> packages <- c("MASS", "plyr", "data.table", "tidyr",  "dplyr", "magrittr", "stringr",
+               "simstudy" )
> invisible(lapply(packages, library, character.only = T))
```

2 Simulation function

Simulation of one trial dataset:

```
> sim.longitudinal.data <-  function(
+     seed,
+     delta,
+     missing = c("complete", "MCAR", "MAR", "MNAR"),
+     missing.p = c(0.15, 0.15),
+     switch = c("No", "Placebo.NAtoNP", "N/P.HIGHtoLOW", "N/P.HIGHtoSTOP"),
+     switch.type = c("CAR", "AR"),
+     switch.p = 0.5,
+     switch.t
+     ) {
+   missing = match.arg(missing)
+   switch = match.arg(switch)
+   switch.type = match.arg(switch.type)
+   delta52.na.low = delta[1]
+   delta52.na.high = delta[2]
+   delta52.np.low = delta[3]
+   delta52.np.high = delta[4]
+   # internal function to apply piecewise linearity
+   p.lin <- function(x, y, xnew) {
+     slopes = c(0, diff(y) / diff(x), 0)
+     tmp = sapply(xnew, function(t) {min(which(t <= x))})
+     ynew = y[tmp] + (xnew - x[tmp]) * slopes[tmp]
+     ynew
+     }
+   ## fixed parameters
+   ## ----------------
+   ## study visits
+   time = c(0, 2, 4, 8, 12, 18, 26, 34, 42, 52)
+   names(time) = c("Baseline", paste("Week", time[-1]))
+   ## timepoints source data
+   time.source = c(0, 2, 4, 6, 12, 24, 36, 52)
+   ## sample size
```

```
+   N = 400 # total
+   n = 200 # per trt group
+   n.np = round(N * 0.5) # 1/2 on nint/pirf
+   n.na = N - n.np # 1/2 on neither
+   ## baseline FVC (ml)
+   mu0 <- 2700
+   sd0 <- 800
+   ## FVC change from baseline (ml) for neither placebo at W0,2,4,8,12,18,24,30,36,44,52
+   chg.na.source = c(0, 5, 15, 40, 80, 110, 165, 205)
+   chg.na = p.lin(time.source, chg.na.source, time)
+   ## FVC change from baseline (ml) for nintendanib/pirfenidone placebo
+   chg.np.source = c(0, 2, 5, 10, 25, 50, 70, 95.1)
+   chg.np = p.lin(time.source, chg.np.source, time)
+   ## difference between active and placebo at other timepoints: linear:
+   delta.na.low = c(0, delta52.na.low / 52 * time[-1])
+   delta.na.high = c(0, delta52.na.high / 52 * time[-1])
+   delta.np.low = c(0, delta52.np.low / 52 * time[-1])
+   delta.np.high = c(0, delta52.np.high / 52 * time[-1])
+   names(delta.na.low) = names(delta.na.high) = names(delta.np.low) =
+     names(delta.np.high) = paste("delta", time[], sep ="")
+   delta.na.low[delta.na.low > chg.na] = chg.na[delta.na.low > chg.na]
+   delta.na.high[delta.na.high > chg.na] = chg.na[delta.na.high > chg.na]
+   delta.np.low[delta.np.low > chg.np] = chg.np[delta.np.low > chg.np]
+   delta.np.high[delta.np.high > chg.np] = chg.np[delta.np.high > chg.np]
+   ## average FVC for all timepoints
+   mu.na.placebo = mu0 - chg.na
+   mu.np.placebo = mu0 - chg.np
+   mu.na.low = mu.na.placebo + delta.na.low
+   mu.np.low = mu.np.placebo + delta.np.low
+   mu.na.high = mu.na.placebo + delta.na.high
+   mu.np.high = mu.np.placebo + delta.np.high
+   ## sd FVC change from baseline at week 52
+   sd.change52 = 275
+   ## correlation matrix
+   cor = function(var.i, var.j, var.change.ij) {
+     0.5 * (var.i + var.j - var.change.ij) / (sqrt(var.i) * sqrt(var.j))
+     }
+   rho <- (cor(sd0 ^ 2, sd0 ^ 2, sd.change52 ^ 2)) ^ (1 / 52)
+   cormatrix = rho ^ abs(sapply(time, function(x) time - x))
+   ## data simulation
+   ## ---------------
+   set.seed(seed)
+   ## initialization data set
+   tmp <- data.frame(
+     USUBJID = c(1:N),
+     POP = c(rep("N/P", n.np), rep("Other", n.na)),
+     TRT = rep(c("Placebo", "High"), n)
+     )
+   ## FVC data: add correlated variables for each timepoint
+   sim.na.placebo <- addCorData(
+     data.table(filter(tmp, POP == "Other", TRT == "Placebo")),
+     idname = "USUBJID",
+     mu = mu.na.placebo,
```

```
+     sigma = sd0,
+     corMatrix = cormatrix,
+     cnames = names(time)
+     )
+   sim.na.placebo$POP = "Other"
+   sim.na.placebo$TRT = "Placebo"
+   sim.np.placebo <- addCorData(
+     data.table(filter(tmp, POP == "N/P", TRT == "Placebo")),
+     idname = "USUBJID",
+     mu = mu.np.placebo,
+     sigma = sd0,
+     corMatrix = cormatrix,
+     cnames = names(time)
+     )
+   sim.np.placebo$POP = "N/P"
+   sim.np.placebo$TRT = "Placebo"
+   sim.na.high <- addCorData(
+     data.table(filter(tmp, POP == "Other", TRT == "High")),
+     idname = "USUBJID",
+     mu = mu.na.high,
+     sigma = sd0,
+     corMatrix = cormatrix,
+     cnames = names(time)
+     )
+   sim.na.high$POP = "Other"
+   sim.na.high$TRT = "High"
+   sim.np.high <- addCorData(
+     data.table(filter(tmp, POP == "N/P", TRT == "High")),
+     idname = "USUBJID",
+     mu = mu.np.high,
+     sigma = sd0,
+     corMatrix = cormatrix,
+     cnames = names(time)
+     )
+   sim.np.high$POP = "N/P"
+   sim.np.high$TRT = "High"
+   ## merging
+   sim1 = rbind(sim.na.placebo, sim.np.placebo, sim.na.high, sim.np.high)
+   sim1$USUBJID = c(1:N)
+   ## data manipulation
+   ## -----------------
+   ## long format
+   sim1 <- sim1 %>% gather(AVISIT, FVC, -USUBJID, -TRT, -POP) %>%
+     mutate(AVISIT = factor(AVISIT, levels = names(time))) %>%
+     arrange(USUBJID, AVISIT)
+   ## change from baseline
+   sim1 <- sim1 %>% group_by(USUBJID) %>%
+     mutate(CHG = FVC - FVC[AVISIT == "Baseline"])
+   ## ADY and AYR
+   sim1 <- sim1 %>%
+     mutate(AWK = ifelse(AVISIT == "Baseline", 0, as.numeric(
+       str_extract(AVISIT, "\\-*\\d+\\.*\\d*")  ))) %>%
+     mutate(AYR = AWK / 52) %>% mutate(ADY = AYR * 365.25)
```

```
+   ## Baseline FVC
+   sim1 <- sim1 %>% group_by(USUBJID) %>%
+     mutate(FVCBL = FVC[AVISIT == "Baseline"])
+   ## treatment naive placebo switching to N/P at switch.t
+   ## ----------------------------------------------------
+   if (switch == "Placebo.NAtoNP") {
+     switch.i = max(which(time <= switch.t))
+     ## assigning switch to switch.p subjects in na placebo population
+     if (switch.type == "CAR") {
+       tmp <- unique(filter(sim1, POP == "Other" & TRT == "Placebo")$USUBJID)
+       sel.id2 = sample(tmp, size = switch.p * length(tmp), replace = F)
+       sim1 <- mutate(sim1, SWITCH = USUBJID %in% sel.id2)
+       }
+     if(switch.type == "AR") {
+       tmp = filter(sim1, POP == "Other" & TRT == "Placebo" &
+                      AVISIT == levels(sim1$AVISIT)[switch.i])
+       M <- median(tmp$CHG)
+       sel.id2 = filter(tmp, CHG < M)$USUBJID
+       sim1 <- mutate(sim1, SWITCH = USUBJID %in% sel.id2)
+       }
+     ## correction in mean FVC from switch.t
+     chg.switch = c(chg.na[1:switch.i], chg.na[switch.i] +
+                      (chg.np[(switch.i + 1):length(time)] -
+                         chg.np[switch.i]))
+     correction.switch = chg.na - chg.switch
+     sim1[sim1$SWITCH, c("FVC", "CHG")] <- sim1[sim1$SWITCH, c("FVC", "CHG")] +
+       correction.switch
+     }
+   ## N/P patients switching from high to low dose
+   ## --------------------------------------------
+   if (switch == "N/P.HIGHtoLOW") {
+     switch.i = max(which(time <= switch.t))
+     ## assigning switch to switch.p subjects in N/P high dose population
+     tmp <- unique(filter(sim1, POP == "N/P" & TRT == "High")$USUBJID)
+     sel.id2 = sample(tmp, size = switch.p * length(tmp), replace = F)
+     sim1 <- mutate(sim1, SWITCH = USUBJID %in% sel.id2)
+     ## correction in mean FVC from switch.t
+     chg.switch = c((chg.np - delta.np.high)[1:switch.i],
+                    (chg.np - delta.np.high)[switch.i] +
+                      ((chg.np - delta.np.low)[(switch.i + 1):length(time)] -
+                         (chg.np - delta.np.low)[switch.i]))
+     correction.switch = (chg.np - delta.np.high) - chg.switch
+     sim1[sim1$SWITCH, c("FVC", "CHG")] <-  sim1[sim1$SWITCH, c("FVC", "CHG")] +
+       correction.switch
+     }
+   ## N/P patients switching from high dose to no IMP
+   ## -----------------------------------------------
+   if (switch == "N/P.HIGHtoSTOP") {
+     switch.i = max(which(time <= switch.t))
+     ## assigning switch to switch.p subjects in N/P high dose population
+     tmp <- unique(filter(sim1, POP == "N/P" & TRT == "High")$USUBJID)
+     sel.id2 = sample(tmp, size = switch.p * length(tmp), replace = F)
+     sim1 <- mutate(sim1, SWITCH = USUBJID %in% sel.id2)
```

```
+     ## correction in mean FVC from switch.t
+     chg.switch = c((chg.np - delta.np.high)[1:switch.i],
+                    (chg.np - delta.np.high)[switch.i] +
+                      ((chg.np)[(switch.i + 1):length(time)] - (chg.np)[switch.i]))
+     correction.switch = (chg.np - delta.np.high) - chg.switch
+     sim1[sim1$SWITCH, c("FVC", "CHG")] <- sim1[sim1$SWITCH, c("FVC", "CHG")] +
+       correction.switch
+     }
+   ## introducing missingness
+   ## -----------------------
+   n.id = length(unique(sim1$USUBJID))
+   n.visit = length(levels(sim1$AVISIT))
+   if (missing == "MCAR") {
+     tmp1 = unique(filter(sim1, TRT == "Placebo")$USUBJID)
+     tmp2 = unique(filter(sim1, TRT == "High")$USUBJID)
+     sel.id = c(sample(tmp1, missing.p[1] * length(tmp1)),
+                sample(tmp2, missing.p[2] * length(tmp2)))
+     sel.visit = runif(length(sel.id),
+                       min = min(time),
+                       max = max(time))
+     for (i in 1:length(sel.id)) {
+       sim1[sim1$USUBJID == sel.id[i] &
+              sim1$AWK >= sel.visit[i], c("FVC", "CHG")] <- NA
+       }
+     }
+   ## subjects dropout after 2 consecutive visits have more than a 10% change from
+   ## baseline
+   if (missing == "MAR") {
+     tmp <- sim1 %>% mutate(sel = -CHG / FVCBL > 0.1)
+     tmp2 <- ddply(tmp, .(USUBJID), function(x) {
+       r = rle(x$sel)
+       sel = r$values == T & r$lengths > 1
+       if (sum(sel) != 0) {
+         ind = min(which(sel == T))
+         sel.visit = sum(r$lengths[1:(ind - 1)]) + 2
+         } else{
+           sel.visit = NA
+           }
+       return(sel.visit)
+       }
+       )
+     sel.id = unique(tmp2$USUBJID[!is.na(tmp2$V1)])
+     if (length(sel.id) > missing.p * n.id) {
+       sel.id = sample(sel.id, missing.p * n.id)
+       }
+     for (i in 1:length(sel.id)) {
+       sel.visit = tmp2$V1[tmp2$USUBJID == sel.id[i]] + 1
+       if (!(sel.visit > n.visit)) {
+         sim1[sim1$USUBJID == sel.id[i] &
+           sim1$AVISIT %in%
+             levels(sim1$AVISIT)[c(sel.visit:n.visit)] , c("FVC", "CHG")] <- NA
+       }
+       }
```

```
+     }
+   ## subjects dropout before 2 consecutive visits have more than a 10% change from
+   ## baseline
+   if (missing == "MNAR") {
+     tmp <- sim1 %>% mutate(sel = -CHG / FVCBL > 0.1)
+     tmp2 <- ddply(tmp, .(USUBJID), function(x) {
+       r = rle(x$sel)
+       sel = r$values == T & r$lengths > 1
+       if (sum(sel) != 0) {
+         ind = min(which(sel == T))
+         sel.visit = sum(r$lengths[1:(ind - 1)])
+         } else{
+           sel.visit = NA}
+       return(sel.visit)}
+       )
+     sel.id = unique(tmp2$USUBJID[!is.na(tmp2$V1)])
+     if (length(sel.id) > missing.p * n.id) {
+       sel.id = sample(sel.id, missing.p * n.id)}
+     for (i in 1:length(sel.id)) {
+       sel.visit = tmp2$V1[tmp2$USUBJID == sel.id[i]] + 1
+       if (!(sel.visit > n.visit)) {
+         sim1[sim1$USUBJID == sel.id[i] &
+           sim1$AVISIT %in%
+             levels(sim1$AVISIT)[c(sel.visit:n.visit)] , c("FVC", "CHG")] <- NA
+       }
+     }
+     }
+   return(sim1)
+   }
```

3 Generate dataset

```
> ## true treatment effect
> delta52.na.low <- 60; delta52.na.high <- 120
> delta52.np.low <- 30; delta52.np.high <- 60
> delta <- c(delta52.na.low, delta52.na.high, delta52.np.low, delta52.np.high)
```

```
> ## generate dataset
> x <- sim.longitudinal.data(
+   seed = 42,
+   delta = delta,
+   missing = "MCAR",
+   missing.p = c(0.15, 0.15),
+   switch = "No"
+   )
```

```
> ## print the first 15 rows
> print.data.frame(x[1:15,], digits = 3)
   USUBJID   POP     TRT   AVISIT  FVC     CHG AWK    AYR   ADY FVCBL
1        1 Other Placebo Baseline 2222    0.00   0 0.0000   0.0  2222
```

2	1	Other Placebo	Week 2	2269	46.99	2	0.0385	14.0	2222
3	1	Other Placebo	Week 4	2177	-44.96	4	0.0769	28.1	2222
4	1	Other Placebo	Week 8	2125	-96.93	8	0.1538	56.2	2222
5	1	Other Placebo	Week 12	2065	-157.11	12	0.2308	84.3	2222
6	1	Other Placebo	Week 18	1996	-226.09	18	0.3462	126.4	2222
7	1	Other Placebo	Week 26	1872	-350.26	26	0.5000	182.6	2222
8	1	Other Placebo	Week 34	1744	-478.16	34	0.6538	238.8	2222
9	1	Other Placebo	Week 42	1687	-535.03	42	0.8077	295.0	2222
10	1	Other Placebo	Week 52	1523	-699.60	52	1.0000	365.2	2222
11	2	Other Placebo	Baseline	2626	0.00	0	0.0000	0.0	2626
12	2	Other Placebo	Week 2	2602	-23.98	2	0.0385	14.0	2626
13	2	Other Placebo	Week 4	2576	-50.11	4	0.0769	28.1	2626
14	2	Other Placebo	Week 8	2633	6.82	8	0.1538	56.2	2626
15	2	Other Placebo	Week 12	2584	-41.64	12	0.2308	84.3	2626

References

Santermans, E., Ford, P., Kreuter, M., Verbruggen, N., Meyvisch, P., Wuyts, W. A., Brown, K. K., Lederer, D. J., Byrne, A. J., Molyneaux, P. L., Sivananthan, A., Moor, C. C., Maher, T. M., and Wijsenbeek, M. (2019), "Modelling Forced Vital Capacity in Idiopathic Pulmonary Fibrosis: Optimising Trial Design," *Adv Ther*, 36, 3059–3070.