Supplementary materials for:

Modelling forced vital capacity in idiopathic pulmonary fibrosis: optimising trial design

Eva Santermans[1], Paul Ford[1], Michael Kreuter[2], Nadia Verbruggen[1], Paul Meyvisch[1], Wim A. Wuyts[3], Kevin K. Brown[4], David J. Lederer[5†], Adam J. Byrne[6], Philip L. Molyneaux[6], Arunon Sivananthan[6], Catharina C. Moor[7], Toby M. Maher[6,8*], Marlies Wijsenbeek[7*]

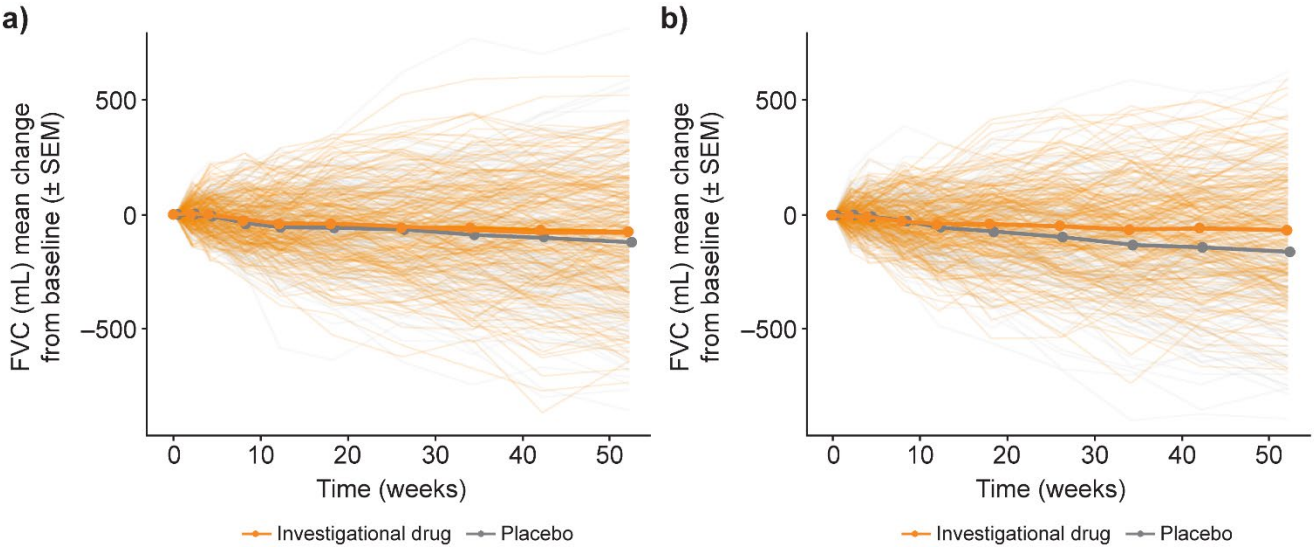
[1]Galapagos NV, Mechelen, Belgium; [2]Centre for Interstitial and Rare Lung Diseases, Pneumology, Thoraxklinik, University Hospital Heidelberg, and German Center for Lung Research, Heidelberg, Germany; [3]Unit for Interstitial Lung Diseases, Department of Pulmonary Medicine, University Hospitals Leuven, Leuven, Belgium; [4]Department of Medicine, National Jewish Health, Denver, USA; [5]Department of Medicine, Columbia University Irving Medical Center, New York, USA; [6]Fibrosis Research Group, National Heart and Lung Institute, Imperial College, London, UK; [7]Department of Respiratory Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; [8]NIHR Respiratory Clinical Research Facility, Royal Brompton Hospital, London, UK

*Joint senior author

[†]The affiliation details listed here indicate the author's affiliation at the time the study was conducted.

Corresponding author: Toby M. Maher, NIHR Respiratory Clinical Research Facility, Royal Brompton Hospital, Sydney Street, London SW3 6NP, UK
Email address: t.maher@rbht.nhs.uk
Telephone number: +44 20 7351 8121
ORCiD iD: orcid.org/0000-0001-7192-9149

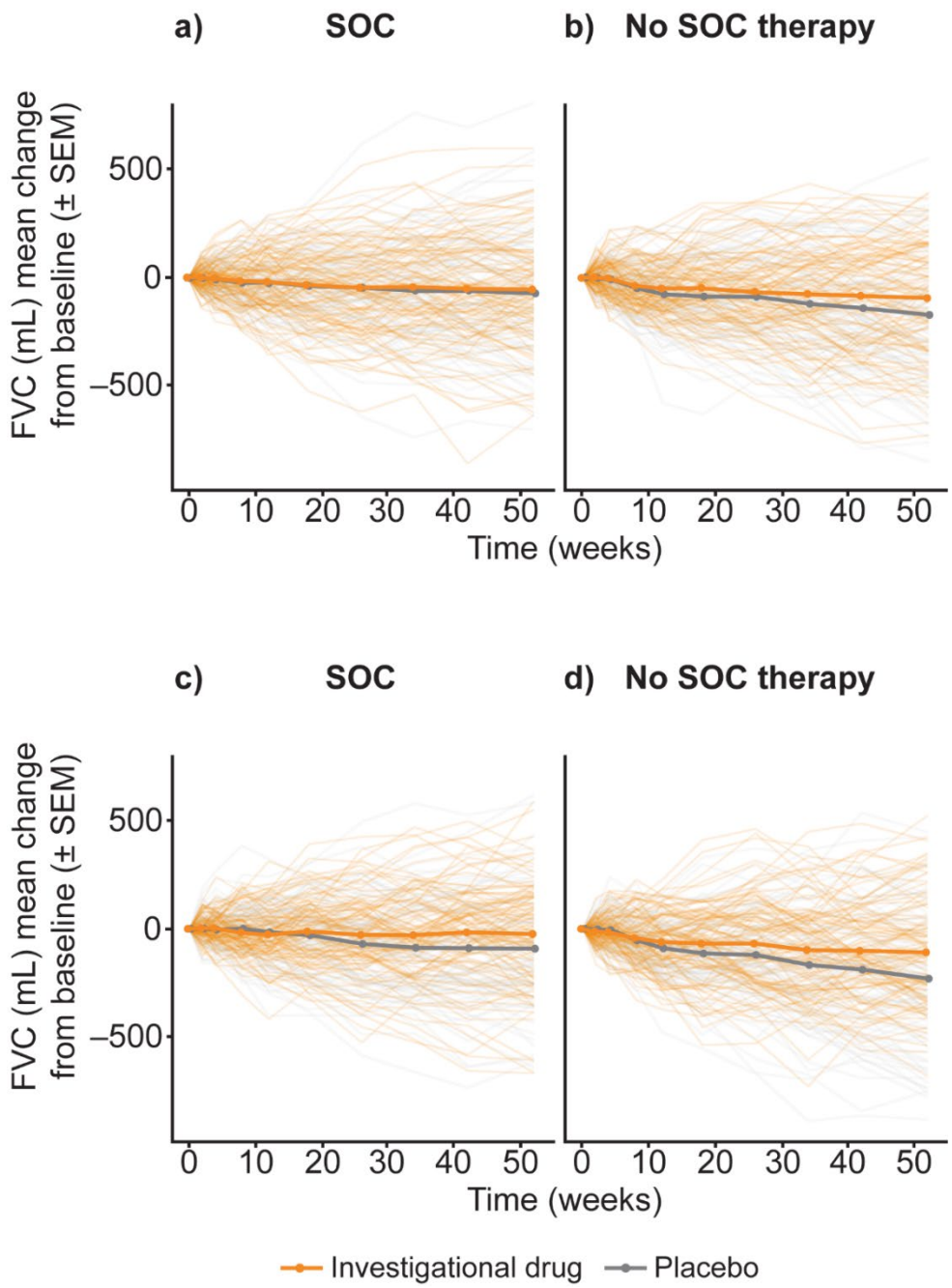
Figure S1. Examples of individual FVC data simulation outputs (with and without SOC), from two simulations (a and b). (N=400 overall; n=200 per treatment arm).



Examples of data simulation outputs using our illustrative clinical trial setting (400 subjects with idiopathic pulmonary fibrosis randomised 1:1 to investigational drug or placebo, across two strata. It was assumed that half were taking SOC (pirfenidone/nintedanib) at the start of the study and half were not).

FVC, forced vital capacity; SEM, standard error of the mean; SOC, standard of care.

Figure S2. Examples of individual FVC data simulation outputs, from two simulations, with (a and c) and without (b and d) SOC treatment. (n=100 per treatment group–strata combination).



Examples of data simulation outputs using our illustrative clinical trial setting. Results from two different simulations are shown (a and b result from simulation 1; c and d result from simulation 2). In each simulation it was assumed that half of patients were taking SOC (pirfenidone/nintedanib) at the start of the study and half were not.

FVC, forced vital capacity; SEM, standard error of the mean; SOC, standard of care.

Appendix 1. Code for statistical model (core script [pages 5–12] and functions [pages 13–17]).

```
################################################################################################
# Simulating and analyzing longitudinal FVC data
################################################################################################
# CORE SCRIPT
################################################################################################
# Last update: April 23, 2019
# Author: Eva Santermans
# R 3.5.2
################################################################################################

rm(list=ls())

## current directory
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

################################################################################################
## LIBRARIES

library("MASS")

################################################################################################
## FUNCTIONS

source("FVC-Simulations_functions.R")


################################################################################################
## DATA SIMULATION & ANALYSIS


## number of simulations
n.sim = 5000

## 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)


## base case
## ---------

missing = "complete"

## data: 5000 simulations
dat = sim.longitudinal.data.n(seed=1,n.sim,delta,missing=missing)
save(dat,file=paste("./results/analysis data/sim",n.sim,"_basecase.Rdata",sep=""))

dat = sim.longitudinal.data.n(seed=n.sim,n.sim,delta,missing=missing)
save(dat,file=paste("./results/analysis data/sim_add",n.sim,"_basecase.Rdata",sep=""))

## analysis
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis data/analysis",n.sim,"_basecase.csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis data/analysis_add",n.sim,"_basecase.csv",sep=""),na="")

## analysis mmrm
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[i]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis",n.sim,"_basecase.csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis data/analysis2_",n.sim,"_basecase.csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[i]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis_add",n.sim,"_basecase.csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis data/analysis2_add",n.sim,"_basecase.csv",sep=""),na="")

res.final = rbind(read.csv(paste("./results/analysis data/analysis2_",n.sim,"_basecase.csv",sep=""),na="")[,-1],res.tot2)
write.csv(res.final,paste("./results/analysis data/analysis_final",2*n.sim,"_basecase.csv",sep=""),na="")

## correlation
```

```
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.cor(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis data/correlation_",n.sim,"_basecase.csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.cor(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis data/correlation_add",n.sim,"_basecase.csv",sep=""),na="")
res.final = rbind(read.csv(paste("./results/analysis data/correlation_",n.sim,"_basecase.csv",sep=""),na="")[,-1],res.tot)
write.csv(res.final,paste("./results/analysis data/correlation_final",2*n.sim,"_basecase.csv",sep=""),na="")

## difference between change 0-6 and 6-12 months (paired t-test)
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.change(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis data/change",n.sim,"_basecase.csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_basecase.Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.change(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis data/change_add",n.sim,"_basecase.csv",sep=""),na="")
res.final = rbind(read.csv(paste("./results/analysis data/change",n.sim,"_basecase.csv",sep=""),na="")[,-1],res.tot)
write.csv(res.final,paste("./results/analysis data/change_final",2*n.sim,"_basecase.csv",sep=""),na="")


## scenario 1: impact of missing data
## ----------------------------------

## missing completely at random

missing = "MCAR"
missing.p = t(matrix(c(0.05,0.05,0.10,0.10,0.15,0.15,0.2,0.1,0.1,0.2),nrow=2))

## data
for (i in 1:nrow(missing.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=1*(1+i),n.sim,delta,missing=missing,missing.p=missing.p[i,])
  save(dat,file=paste("./results/analysis data/sim",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".Rdata",sep=""))
  dat=NULL
}

for (i in 1:nrow(missing.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=n.sim*(1+i),n.sim,delta,missing=missing,missing.p=missing.p[i,])
  save(dat,file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".Rdata",sep=""))
  dat=NULL
}

## analysis
for (i in 1:nrow(missing.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
    res.tot = rbind(res.tot,res[[j]])
  }
  write.csv(res.tot,paste("./results/analysis
data/analysis",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")
}

for (i in 1:nrow(missing.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
    res.tot = rbind(res.tot,res[[j]])
  }
  write.csv(res.tot,paste("./results/analysis
data/analysis_add",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")
}

## analysis mmrm
for (i in 1:nrow(missing.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
```

```
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis
data/analysis2_",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")
}

for (i in 1:nrow(missing.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis_add",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis
data/analysis2_add",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")

  res.final = rbind(read.csv(paste("./results/analysis
data/analysis2_",n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot2)
  write.csv(res.final,paste("./results/analysis
data/analysis_final",2*n.sim,"_scenario1_",paste(missing.p[i,]*100,collapse="_"),".csv",sep=""),na="")
}

## missing at random

missing = "MAR"
missing.p = c(0.15,0.15)

## data
dat = sim.longitudinal.data.n(seed=100,n.sim,delta,missing=missing,missing.p=missing.p)
save(dat,file=paste("./results/analysis data/sim",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
dat = NULL

dat = sim.longitudinal.data.n(seed=100+n.sim,n.sim,delta,missing=missing,missing.p=missing.p)
save(dat,file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
dat = NULL

## analysis
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis
data/analysis",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis
data/analysis_add",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

## analysis mmrm
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[i]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis
data/analysis2_",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[i]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis_add",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis
data/analysis2_add",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

res.final = rbind(read.csv(paste("./results/analysis
data/analysis2_",n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot2)
write.csv(res.final,paste("./results/analysis
data/analysis_final",2*n.sim,"_scenario1_MAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

## missing not at random
```

```
missing = "MNAR"
missing.p = c(0.15,0.15)

## data
dat = sim.longitudinal.data.n(seed=100,n.sim,delta,missing=missing,missing.p=missing.p)
save(dat,file=paste("./results/analysis data/sim",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
dat = NULL

dat = sim.longitudinal.data.n(seed=1000+n.sim,n.sim,delta,missing=missing,missing.p=missing.p)
save(dat,file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
dat = NULL

## analysis
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis
data/analysis",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis
data/analysis_add",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

## analysis mmrm
res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[i]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis
data/analysis2_",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (i in 1:(n.sim/1000)){
  load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".Rdata",sep=""))
  dat2 = dat[((i-1)*1000+1):(i*1000)]; rm(dat)
  res[[i]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[i]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis_add",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis
data/analysis2_add",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")

res.final = rbind(read.csv(paste("./results/analysis
data/analysis2_",n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")[,-1],res.tot2)
write.csv(res.final,paste("./results/analysis
data/analysis_final",2*n.sim,"_scenario1_MNAR_",paste(missing.p*100,collapse="_"),".csv",sep=""),na="")


## scenario 2: placebo subjects starting N/P at 26 weeks
## ------------------------------------------------------

missing = "complete"

switch = "Placebo.NAtoNP"
switch.t = 26
switch.p = c(0.25,0.5)

## data
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=20*(1+i),n.sim,delta,missing=missing,switch=switch,switch.p=switch.p[i],switch.t=switch.t)
  save(dat,file=paste("./results/analysis data/sim",n.sim,"_scenario2_",switch.p[i]*100,".Rdata",sep=""))
  dat = NULL
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=n.sim*20*(1+i),n.sim,delta,missing=missing,switch=switch,switch.p=switch.p[i],switch.t=switch.t)
  save(dat,file=paste("./results/analysis data/sim_add",n.sim,"_scenario2_",switch.p[i]*100,".Rdata",sep=""))
  dat = NULL
}

## analysis
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario2_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
```

```
    res.tot = rbind(res.tot,res[[j]])
  }
  write.csv(res.tot,paste("./results/analysis data/analysis",n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario2_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
    res.tot = rbind(res.tot,res[[j]])
  }
  write.csv(res.tot,paste("./results/analysis data/analysis_add",n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")
}

## analysis mmrm
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario2_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis",n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis data/analysis2_",n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario2_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis_add",n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis data/analysis2_add",n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")

  res.final = rbind(read.csv(paste("./results/analysis data/analysis2_",n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot2)
  write.csv(res.final,paste("./results/analysis data/analysis_final",2*n.sim,"_scenario2_",switch.p[i]*100,".csv",sep=""),na="")

}


## scenario 3: N/P subjects down-titrate to a lower dose at 3 months
## ------------------------------------------------------------------

missing = "complete"

switch = "N/P.HIGHtoLOW"
switch.t = 12
switch.p = c(0.25,0.5)

## data
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=20*(1+i),n.sim,delta,missing=missing,switch=switch,switch.p=switch.p[i],switch.t=switch.t)
  save(dat,file=paste("./results/analysis data/sim",n.sim,"_scenario3_",switch.p[i]*100,".Rdata",sep=""))
  dat = NULL
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=n.sim+30*(1+i),n.sim,delta,missing=missing,switch=switch,switch.p=switch.p[i],switch.t=switch.t)
  save(dat,file=paste("./results/analysis data/sim_add",n.sim,"_scenario3_",switch.p[i]*100,".Rdata",sep=""))
  dat = NULL
}

## analysis
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario3_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
    res.tot = rbind(res.tot,res[[j]])
  }
  write.csv(res.tot,paste("./results/analysis data/analysis",n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario3_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
    res.tot = rbind(res.tot,res[[j]])
```

```
  }
  write.csv(res.tot,paste("./results/analysis data/analysis_add",n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")
}

## analysis mmrm
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario3_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis",n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis data/analysis2_",n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario3_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis_add",n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis data/analysis2_add",n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")

  res.final = rbind(read.csv(paste("./results/analysis data/analysis2_",n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot2)
  write.csv(res.final,paste("./results/analysis data/analysis_final",2*n.sim,"_scenario3_",switch.p[i]*100,".csv",sep=""),na="")

}


## scenario 4: N/P subjects stop IMP at 3 months
## ---------------------------------------------

missing = "complete"

switch = "N/P.HIGHtoSTOP"
switch.t = 12
switch.p = c(0.25,0.5)

## data
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=20*(1+i),n.sim,delta,missing=missing,switch=switch,switch.p=switch.p[i],switch.t=switch.t)
  save(dat,file=paste("./results/analysis data/sim",n.sim,"_scenario4_",switch.p[i]*100,".Rdata",sep=""))
  dat = NULL
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  dat = sim.longitudinal.data.n(seed=n.sim+40*(1+i),n.sim,delta,missing=missing,switch=switch,switch.p=switch.p[i],switch.t=switch.t)
  save(dat,file=paste("./results/analysis data/sim_add",n.sim,"_scenario4_",switch.p[i]*100,".Rdata",sep=""))
  dat = NULL
}

## analysis
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario4_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
    res.tot = rbind(res.tot,res[[j]])
  }
  write.csv(res.tot,paste("./results/analysis data/analysis",n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario4_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.model(dat=dat2)$res
    res.tot = rbind(res.tot,res[[j]])
  }
  write.csv(res.tot,paste("./results/analysis data/analysis_add",n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")
}

## analysis mmrm
for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim",n.sim,"_scenario4_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
```

```
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis",n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis data/analysis2_",n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")
}

for (i in 1:length(switch.p)){
  print(paste("Subscenario",i,sep=" "))
  res = vector("list", n.sim/1000); res.tot=NULL
  for (j in 1:(n.sim/1000)){
    load(file=paste("./results/analysis data/sim_add",n.sim,"_scenario4_",switch.p[i]*100,".Rdata",sep=""))
    dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
    res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
    res.tot = rbind(res.tot,res[[j]])
  }
  res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis_add",n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot)
  write.csv(res.tot2,paste("./results/analysis data/analysis2_add",n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")

  res.final = rbind(read.csv(paste("./results/analysis data/analysis2_",n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")
[,-1],res.tot2)
  write.csv(res.final,paste("./results/analysis data/analysis_final",2*n.sim,"_scenario4_",switch.p[i]*100,".csv",sep=""),na="")

}


## scenario 5: MNAR + placebo subjects starting N/P
## ------------------------------------------------

missing = "MNAR"
missing.p = c(0.05,0.05)

switch = "Placebo.NAtoNP"
switch.t = 26
switch.p = 0.15

## data
dat =
sim.longitudinal.data.n(seed=50,n.sim,delta,missing=missing,missing.p=missing.p,switch=switch,switch.p=switch.p,switch.t=switch.t)
save(dat,file=paste("./results/analysis
data/sim",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".Rdata",sep=""))
dat = NULL

dat =
sim.longitudinal.data.n(seed=n.sim+50,n.sim,delta,missing=missing,missing.p=missing.p,switch=switch,switch.p=switch.p,switch.t=switch.t)

save(dat,file=paste("./results/analysis
data/sim_add",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".Rdata",sep=""))
dat = NULL

## analysis
res = vector("list", n.sim/1000); res.tot=NULL
for (j in 1:(n.sim/1000)){
  load(file=paste("./results/analysis
data/sim",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".Rdata",sep=""))
  dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
  res[[j]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[j]])
}
write.csv(res.tot,paste("./results/analysis
data/analysis",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (j in 1:(n.sim/1000)){
  load(file=paste("./results/analysis
data/sim_add",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".Rdata",sep=""))
  dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
  res[[j]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[j]])
}
write.csv(res.tot,paste("./results/analysis
data/analysis_add",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")

## analysis mmrm
res = vector("list", n.sim/1000); res.tot=NULL
for (j in 1:(n.sim/1000)){
  load(file=paste("./results/analysis
data/sim",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".Rdata",sep=""))
  dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
  res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[j]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis
data/analysis2_",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")

res = vector("list", n.sim/1000); res.tot=NULL
for (j in 1:(n.sim/1000)){
  load(file=paste("./results/analysis
data/sim_add",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".Rdata",sep=""))
  dat2 = dat[((j-1)*1000+1):(j*1000)]; rm(dat)
  res[[j]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[j]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis
data/analysis_add",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")
```

```
[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis
data/analysis2_add",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")

res.final = rbind(read.csv(paste("./results/analysis
data/analysis2_",n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")[,-1],res.tot2)
write.csv(res.final,paste("./results/analysis
data/analysis_final",2*n.sim,"_scenario5_MNAR_",paste(missing.p*100,collapse="_"),"_switch",switch.p*100,".csv",sep=""),na="")


## base case: varying sample size for W26 analysis
## ------------------------------------------------

missing = "complete"

## data: 5000 simulations
dat = sim.longitudinal.data.n(seed=1,n.sim,delta,missing=missing)
save(dat,file=paste("./results/analysis data/sim",n.sim,"_basecase220.Rdata",sep=""))
dat = NULL

## analysis
res = vector("list", n.sim/500); res.tot=NULL
for (i in 1:(n.sim/500)){
  load(file=paste("./results/analysis data/sim",n.sim,"_basecase220.Rdata",sep=""))
  dat2 = dat[((i-1)*500+1):(i*500)]; rm(dat)
  res[[i]] = sim.longitudinal.model(dat=dat2)$res
  res.tot = rbind(res.tot,res[[i]])
}
write.csv(res.tot,paste("./results/analysis data/analysis",n.sim,"_basecase250.csv",sep=""),na="")

## analysis mmrm
res = vector("list", n.sim/500); res.tot=NULL
for (i in 1:(n.sim/500)){
  load(file=paste("./results/analysis data/sim",n.sim,"_basecase250.Rdata",sep=""))
  dat2 = dat[((i-1)*500+1):(i*500)]; rm(dat)
  res[[i]] = sim.longitudinal.mmrm(dat=dat2,cor="AR1")$res
  res.tot = rbind(res.tot,res[[i]])
}
res.tot2 = rbind(read.csv(paste("./results/analysis data/analysis",n.sim,"_basecase250.csv",sep=""),na="")[,-1],res.tot)
write.csv(res.tot2,paste("./results/analysis data/analysis2_",n.sim,"_basecase250.csv",sep=""),na="")
```

```
################################################################################
# Simulating and analyzing longitudinal FVC data
################################################################################
# FUNCTIONS
################################################################################
# Last update: April 23, 2019
# Author: Eva Santermans
# R 3.5.2
################################################################################


################################################################################
## LIBRARIES

library("MASS")
library("plyr")
library("tidyverse")
library("simstudy")     # simulation of data sets
library("lme4")         # lmer function: mixed models
library("lmerTest")     # developed to mimic proc mixed output (loading this changes the summary and anova functions)
library("nlme")
library("survminer")
library("survival")
library("lsmeans")

################################################################################
## LONGITUDINAL DATA


## piecewise linear function; output ynew
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
}


## DATA SIMULATION (1 data set)
################################

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,sw
{

  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]

  ## 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)
}


## DATA SIMULATION (multiple simulations)
########################################


sim.longitudinal.data.n <-
function(seed,n.sim,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=
{

  missing = match.arg(missing)
  switch = match.arg(switch)
  switch.type = match.arg(switch.type)

  dat = vector("list", n.sim)

  for (i in 1:n.sim){
    if(i%%100==0){print(i)}
    dat[[i]] = sim.longitudinal.data(seed=seed*i,delta=delta,missing=missing,missing.p=missing.p,switch=switch,switch.type=switch.type,switch.p=switch.p,switch.t=switch.t)
  }
  return(dat)
}


## DATA ANALYSIS
################


## ANCOVA and mixed model
## ----------------------

sim.longitudinal.model <- function(dat){

  res = data.frame(simulation=integer(),duration=factor(levels=c("W26","W52")),model=factor(levels=c("Mixed","Ancova")),est=double(),p=double())

  fm1 = fm2 = fm3 = fm4 = vector("list", length(dat))
  for (i in 1:length(dat)){
    if(i%%100==0){print(i)}
    dat.tmp = dat.tmp2 = dat.tmp3 = dat.tmp4 = NULL
    tab1 = tab2 = tab3 = tab4 = NULL
    tmp1 = tmp2 = tmp3 = tmp4 = rep(NA,2)

    ## data
    dat.tmp = dat[[i]]
    dat.tmp$TRT <- factor(dat.tmp$TRT, levels=c("Placebo","High"))
    dat.tmp2 = dat.tmp %>% filter(AWK<=26)
    dat.tmp3 <- dat.tmp %>% filter(AWK==52)
    dat.tmp4 <- dat.tmp %>% filter(AWK==26)

    ## mixed model: random intercept and slope - 52 weeks
    tryCatch({
    fm1[[i]] <- lmer(FVC ~ TRT*AYR + (AYR|USUBJID), dat.tmp)
    tab1 = summary(fm1[[i]])$coefficients
    if(sum(dim(tab1)==c(4,5))==2){tmp1 = c(tab1[4,c(1,5)])}
    },error=function(e){cat("ERROR :",conditionMessage(e), "\n")})

    ## mixed model: random intercept and slope - 26 weeks
    tryCatch({
      fm2[[i]] <- lmer(FVC ~ TRT*AYR + (AYR|USUBJID), dat.tmp2)
      tab2 = summary(fm2[[i]])$coefficients
      if(sum(dim(tab2)==c(4,5))==2){tmp2 = c(tab2[4,c(1,5)])}
    },error=function(e){cat("ERROR :",conditionMessage(e), "\n")})

    ## ancova - 52 weeks
    fm3[[i]] = aov(CHG ~ FVCBL + TRT, data=dat.tmp3)
    tab3 = summary(lsmeans::lsmeans(fm3[[i]],pairwise~TRT))$contrasts
    tmp3 = c(-tab3$estimate,tab3$p.value)

    ## ancova - 26 weeks
    fm4[[i]] = aov(CHG ~ FVCBL + TRT, data=dat.tmp4)
    tab4 = summary(lsmeans::lsmeans(fm4[[i]],pairwise~TRT))$contrasts
    tmp4 = c(-tab4$estimate,tab4$p.value)

    j=i*4-3
    res[j,] <- c(i,duration="W52",model="Mixed",tmp1)
    res[j+1,] <- c(i,duration="W26",model="Mixed",tmp2)
    res[j+2,] <- c(i,duration="W52",model="Ancova",tmp3)
    res[j+3,] <- c(i,duration="W26",model="Ancova",tmp4)
    }
  return(list(fm52=fm1,fm26=fm2,fma52=fm3,fma26=fm4,res=res))
}


## MMRM
## ----

sim.longitudinal.mmrm <- function(dat,cor=c("AR1","UN")){

  cor = match.arg(cor)
  res = data.frame(simulation=integer(),duration=factor(levels=c("W26","W52")),model=factor(levels=c("MMRM_AR1","MMRM_UN")),est=double(),p=double())

  fm1 = fm2 = vector("list", length(dat))
  for (i in 1:length(dat)){
    if(i%%100==0){print(i)}
    dat.tmp = dat.tmp2 = NULL
    tab1 = tab2 = NULL
    tmp1 = tmp2 = rep(NA,2)

    ## data
    dat.tmp = dat[[i]]
    dat.tmp$TRT <- factor(dat.tmp$TRT, levels=c("Placebo","High"))
    dat.tmp <- dat.tmp[dat.tmp$AWK!=0,]
    dat.tmp$VISITNUM <- as.numeric(recode(dat.tmp$AWK,'0'='1','2'='1','4'='2','8'='3','12'='4','18'='5','26'='6','34'='7','42'='8','52'='9'))
    dat.tmp <- dat.tmp[!is.na(dat.tmp$FVC),]
    dat.tmp2 = dat.tmp %>% filter(AWK<=26)

    ## repeated measures model - 52 weeks
    tryCatch({
      if(cor=="AR1"){fm1[[i]] <- gls(CHG ~ FVCBL + TRT*AVISIT, corr=corAR1(form=~VISITNUM|USUBJID), data=dat.tmp)}else{
        fm1[[i]] <- gls(CHG ~ FVCBL + TRT*AVISIT, corr=corSymm(form=~VISITNUM|USUBJID), data=dat.tmp)
      }
      tab1 = summary(lsmeans::lsmeans(fm1[[i]],pairwise~TRT|AVISIT,data=dat.tmp))$contrasts
      if(sum(dim(tab1)==c(9,7))==2){tmp1 = c(-tab1$estimate[9],tab1$p.value[9])}
    },error=function(e){cat("ERROR :",conditionMessage(e), "\n")})

    ## repeated measures model - 26 weeks
    tryCatch({
      if(cor=="AR1"){fm2[[i]] <- gls(CHG ~ FVCBL + TRT*AVISIT, corr=corAR1(form=~VISITNUM|USUBJID), data=dat.tmp2)}else{
        fm2[[i]] <- gls(CHG ~ FVCBL + TRT*AVISIT, corr=corSymm(form=~VISITNUM|USUBJID), data=dat.tmp2)
      }
      tab2 = summary(lsmeans::lsmeans(fm2[[i]],pairwise~TRT|AVISIT,data=dat.tmp2))$contrasts
      if(sum(dim(tab2)==c(6,7))==2){tmp2 = c(-tab2$estimate[6],tab2$p.value[6])}
    },error=function(e){cat("ERROR :",conditionMessage(e), "\n")})

    j=i*2-1
```

```
    res[j,] <- c(i,duration="W52",model=paste("MMRM",cor,sep="_"),tmp1)
    res[j+1,] <- c(i,duration="W26",model=paste("MMRM",cor,sep="_"),tmp2)
  }
  return(list(fmm52=fm1,fmm26=fm2,res=res))
}


## Correlation
## -----------

sim.longitudinal.cor <- function(dat){

  res = data.frame(simulation=integer(),model=factor(levels=c("correlation")),est=double(),p=double())

  for (i in 1:length(dat)){
    if(i%%100==0){print(i)}
    dat.tmp = NULL

    ## data
    dat.tmp = dat[[i]]
    dat.tmp$TRT <- factor(dat.tmp$TRT, levels=c("Placebo","High"))
    dat.tmp <- dat.tmp[!is.na(dat.tmp$FVC),]

    dat.tmp2 <- dat.tmp %>%
      filter(POP=="Other" & TRT=="Placebo") %>%
      group_by(USUBJID) %>%
      mutate(CHG.6 = FVC-FVC[AWK==26]) %>%
      summarize(change1 = CHG[AWK==26], change2 = CHG.6[AWK==52])
    cor.res = cor.test(dat.tmp2$change1,dat.tmp2$change2,method="pearson")

    res[i,] <- c(i,model="correlation",cor.res$estimate,cor.res$p.value)
  }
  return(list(res=res))
}


## Difference between change (0-6 and 6-12 months)
## -----------------------------------------------

sim.longitudinal.change <- function(dat){

  res = data.frame(simulation=integer(),population=factor(levels=c("total","RP","SP")),est1=double(),est2=double(),p=double())

  for (i in 1:length(dat)){
    if(i%%100==0){print(i)}
    tmp1 = tmp2 = tmp3 = NULL
    dat.tmp = dat.tmp2 = dat.tmp3 = dat.tmp4 = NULL

    ## data
    dat.tmp = dat[[i]]

    dat.tmp2 <- dat.tmp %>%
      filter(POP=="Other" & TRT=="Placebo") %>%
      group_by(USUBJID) %>%
      mutate(CHG.6 = FVC-FVC[AWK==26]) %>%
      summarize(FVCBL=unique(FVCBL),change1 = CHG[AWK==26], change2 = CHG.6[AWK==52])

    j=i*3-2

    ## paired t-tests change 0-6 and 6-12 months
    ## total population
    tryCatch({tmp1 = t.test(dat.tmp2$change1,dat.tmp2$change2)},error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
    if(!is.null(tmp1)){res[j,] <- c(i,model="total",tmp1$estimate,tmp1$p.value)}else{res[j,] <- c(i,model="total",NA,NA)}
    ## RP population
    dat.tmp3 <- dat.tmp2 %>% filter(change1/FVCBL<c(-0.10))
    tryCatch({tmp2 = t.test(dat.tmp3$change1,dat.tmp3$change2)},error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
    if(!is.null(tmp2)){res[j+1,] <- c(i,model="RP",tmp2$estimate,tmp2$p.value)}else{res[j+1,] <- c(i,model="RP",NA,NA)}
    ## RP population
    dat.tmp4 <- dat.tmp2 %>% filter(change1/FVCBL>c(-0.10))
    tryCatch({tmp3 = t.test(dat.tmp4$change1,dat.tmp4$change2)},error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
    if(!is.null(tmp3)){res[j+2,] <- c(i,model="SP",tmp3$estimate,tmp3$p.value)}else{res[j+2,] <- c(i,model="SP",NA,NA)}

  }
  return(list(res=res))

}


## COMBINED: SIMULATION AND ANALYSIS
####################################

sim.longitudinal.combined <-
function(seed,n.sim,delta,missing=c("complete","MCAR","MAR","MNAR"),missing.p=0.15,switch=c("No","Placebo.NAtoNP","N/P.HIGHtoLOW","N/P.HIGHtoSTOP"),switch.type=c("CAR","AR"),switch.p=0.5,swit
{

  missing = match.arg(missing)
  switch = match.arg(switch)
  switch.type = match.arg(switch.type)

  fm1 = fm2 = fm3 = fm4 = vector("list", n.sim)
  res = data.frame(simulation=integer(),duration=factor(levels=c("W26","W52")),model=factor(levels=c("Mixed","Ancova")),est=double(),p=double())

  for (i in 1:n.sim){

    if(i%%100==0){print(i)}

    dat = dat.tmp = dat.tmp2 = dat.tmp3 = dat.tmp4 = NULL
    tab1 = tab2 = tab3 = tab4 = NULL
    tmp1 = tmp2 = tmp3 = tmp4 = rep(NA,2)

    ## data
    dat = sim.longitudinal.data(seed=seed*i,delta=delta,missing=missing,missing.p=missing.p,switch=switch,switch.type=switch.type,switch.p=switch.p,switch.t=switch.t)

    dat.tmp = dat
    dat.tmp$TRT <- factor(dat.tmp$TRT, levels=c("Placebo","High"))
    dat.tmp2 = dat.tmp %>% filter(AWK<=26)
    dat.tmp3 <- dat.tmp %>% filter(AWK==52)
    dat.tmp4 <- dat.tmp %>% filter(AWK==26)

    ## mixed model: random intercept and slope - 52 weeks
    tryCatch({
      fm1[[i]] <- lmer(FVC ~ TRT*AYR + (AYR|USUBJID), dat.tmp)
      tab1 = summary(fm1[[i]])$coefficients
      if(sum(dim(tab1)==c(4,5))==2){tmp1 = c(tab1[4,c(1,5)])}
    },error=function(e){cat("ERROR :",conditionMessage(e), "\n")})

    ## mixed model: random intercept and slope - 26 weeks
    tryCatch({
      fm2[[i]] <- lmer(FVC ~ TRT*AYR + (AYR|USUBJID), dat.tmp2)
      tab2 = summary(fm2[[i]])$coefficients
      if(sum(dim(tab2)==c(4,5))==2){tmp2 = c(tab2[4,c(1,5)])}
    },error=function(e){cat("ERROR :",conditionMessage(e), "\n")})

    ## ancova - 52 weeks
    fm3[[i]] = aov(CHG ~ FVCBL + TRT, data=dat.tmp3)
    tab3 = summary(lsmeans::lsmeans(fm3[[i]],pairwise~TRT))$contrasts
    tmp3 = c(-tab3$estimate,tab3$p.value)

    ## ancova - 26 weeks
    fm4[[i]] = aov(CHG ~ FVCBL + TRT, data=dat.tmp4)
    tab4 = summary(lsmeans::lsmeans(fm4[[i]],pairwise~TRT))$contrasts
    tmp4 = c(-tab4$estimate,tab4$p.value)

    j=i*4-3
    res[j,] <- c(i,duration="W52",model="Mixed",tmp1)
    res[j+1,] <- c(i,duration="W26",model="Mixed",tmp2)
```

```
    res[j+2,] <- c(i,duration="W52",model="Ancova",tmp3)
    res[j+3,] <- c(i,duration="W26",model="Ancova",tmp4)
    }
  return(list(res=res))
}
```