

# ABC-XYZ Premise and Dataset Creation

Jonah Lyon (jonah@a2-ai.com)

February 4, 2024

## Table of contents

<b>1</b>	<b>Load PopPK Model Run for ABC-XYZ</b>	<b>2</b>
<b>2</b>	<b>Create population</b>	<b>2</b>
<b>3</b>	<b>Define Cohorts and Subjects</b>	<b>2</b>
<b>4</b>	<b>Random Simulation for subject covariates</b>	<b>2</b>
4.1	Demographics . . . . .	2
4.2	Labs . . . . .	3
4.3	Combine Demographics and Labs . . . . .	4
<b>5</b>	<b>Dosing History</b>	<b>4</b>
<b>6</b>	<b>28-day cycles</b>	<b>4</b>
<b>7</b>	<b>QOD for 4.8, QOD for 7.2</b>	<b>4</b>
<b>8</b>	<b>Create Base (Demographics + Dose History) Dataset</b>	<b>4</b>
<b>9</b>	<b>Test Simulation Run (validate dataset for use in mrgsolve and NONMEM)</b>	<b>5</b>

The purpose of this script is to create a dataset with an imaginary patient population for a future study.

- Lockfile written to "~/repo/jonahlyon/ABC-XYZ/renv.lock".

## 1 Load PopPK Model Run for ABC-XYZ

```
mod1 <- mread("../model/R-mrgsolve/1001.mod",  
              start = 0, end = 120, delta = 1)
```

Building 1001\_mod ... done.

## 2 Create population

- Study of  $N = 120$
- pancreatic cancer ( $HV = 0$ )
- ECOG of 0 (45%) and 1 (55%)
- obese study BMI > 30
- hepatic dysfunction (mild  $n = 40$ , moderate  $n = 60$ )
- U.S Study - 80% White, 15% Black, 3% Asian, 2% Other - 52% Female, 48% Male
- QOD, 4.8mg ( $n = 80$ ) and 7.2mg ( $n = 40$ ) Dosing Groups. All doses fasted.

## 3 Define Cohorts and Subjects

```
drug <- "ABC-XYZ"  
site <- "001"  
cohort1 <- str_c("01", str_pad(1:80, pad = 0, width = 3)) # 4 mg  
cohort2 <- str_c("02", str_pad(1:40, pad = 0, width = 3)) # 8 mg  
  
cohort1_usubjid <- str_c(drug, site, cohort1, sep = "-")  
cohort2_usubjid <- str_c(drug, site, cohort2, sep = "-")
```

## 4 Random Simulation for subject covariates

### 4.1 Demographics

```
set.seed(20240107)  
subject_dm <- data.frame(  
  USUBJID = c(cohort1_usubjid, cohort2_usubjid),  
  COHORT = c(rep(1, 80), rep(2, 40)),
```

```

SEX = sample(c("F", "M"), size = 120,
             replace = TRUE, prob = c(0.52, 0.48)),
BBMI = round(rnorm(n = 120, mean = 34.2, sd = 1.4), 1),
RACE = sample(c("White", "Black", "Asian", "Other"), size = 120,
             replace = T, prob = c(0.8, 0.15, 0.03, 0.02)),
BAGE = round(rnorm(n = 120, mean = 62, sd = 4.7)),
BECOG = sample(c(0, 1), size = 120, replace = TRUE, prob = c(0.45, 0.55))
) |>
mutate(
  SEXF = ifelse(SEX == "M", 0, 1),
  RACEN = case_when(
    RACE == "White" ~ 1,
    RACE == "Black" ~ 2,
    RACE == "Asian" ~ 3,
    RACE == "Other" ~ 4
  ),
  BHT = ifelse(SEXF == 1,
               round(rnorm(n = 120, mean = 163, sd = 6.35)/100, 2),
               round(rnorm(n = 120, mean = 175, sd = 7.62)/100, 2)),
  BWT = round(BHT^2*BBMI, 1),
  PTTYPE = "Pancreatic Cancer",
  PTTYPE = 4
)

```

## 4.2 Labs

```

subject_lb <- data.frame(
  USUBJID = c(cohort1_usubjid, cohort2_usubjid),
  BHFCC = sample(c("Mild", "Moderate"), size = 120, replace = TRUE, prob = c(0.4, 0.6)),
  BALB = round(rnorm(n = 120, mean = 3.6, sd = 0.25), 2),
  BCRCL = round(rnorm(n = 120, mean = 85, sd = 12), 1)
) |>
mutate(
  BHFC = ifelse(BHFCC == "Mild", 1, 2)
)

```

### 4.3 Combine Demographics and Labs

```
subject_dm_lb <- subject_dm |>
  left_join(subject_lb, by = "USUBJID")
```

## 5 Dosing History

### 6 28-day cycles

### 7 QOD for 4.8, QOD for 7.2

```
subject_ex <- subject_dm_lb |>
  select(USUBJID, COHORT) |>
  mutate(
    AMT = ifelse(COHORT == 1, 4.8, 7.2)
  ) |>
  crossing(
    ATFD = seq(0, 2016, by = 48) # QOD for 3 28 day cycles, in hours
  ) |>
  mutate(
    NTFD = ATFD,
    ATLD = 0,
    NTLD = 0,
    EVID = 1,
    CMT = 1,
    BLQ = 0,
    MDV = 1,
    FOOD = 0
  )
```

## 8 Create Base (Demographics + Dose History) Dataset

```
doses_demogs <- subject_ex |>
  left_join(subject_dm_lb, by = c("USUBJID", "COHORT")) |>
  mutate(
    C = NA_integer_,
```

```

LINE = row_number(),
ID = match(USUBJID, unique(USUBJID)),
STUDY = 305,
LLOQ = 0.1,
DOSEA = AMT,
DOSEN = AMT,
REGN = 2,
DV = NA_integer_,
LDV = NA_integer_
) |>
select(names(spec))

write_csv(doses_demogs,
          file = file.path(derDir, "subject_demog_dose.csv"),
          na = ".")

digest::digest(doses_demogs) |> print()

```

```
[1] "99a885e785f94ba870ccc7155c76288c"
```

## 9 Test Simulation Run (validate dataset for use in mrgsolve and NONMEM)

Note - hourly saampling is meant to demonstrate variability in PK samples, not show an example study design.

```

# validate test for simulation wiht dataset
sim1 <- mrgsim_df(mod1,
                  data = doses_demogs |>
                    filter(ATFD <= 24*16) |>
                    mutate(ETA1 = 0, ETA2 = 0, ETA3 = 0,
                           ETA4 = 0, ETA5 = 0) |>
                    mutate(TIME = ATFD),
                  etasrc = "data.all",
                  carry.out = names(spec))

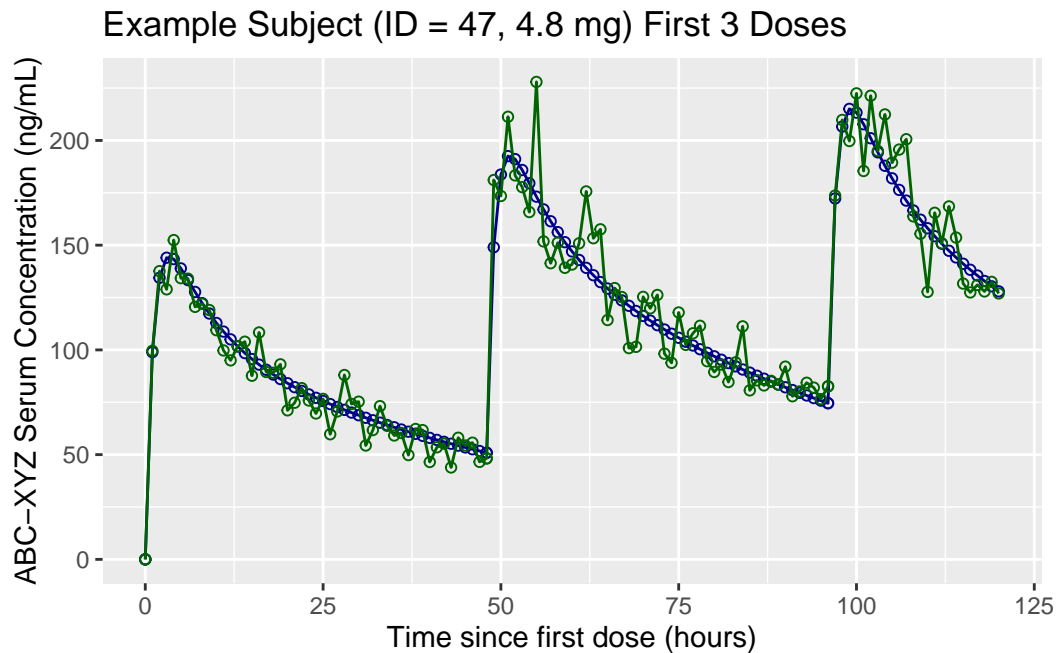
# example subject simulation
ggplot(sim1 |> filter(ID == 47), aes(TIME, IPRED)) +

```

```

geom_point(col = "darkblue", shape = 1) +
geom_line(col = "darkblue") +
geom_point(aes(y = Y), col = "darkgreen", shape = 1) +
geom_line(aes(y = Y), col = "darkgreen") +
xlim(0, 120) +
xlab("Time since first dose (hours)") +
ylab("ABC-XYZ Serum Concentration (ng/mL)") +
ggtitle("Example Subject (ID = 47, 4.8 mg) First 3 Doses")

```



BLUE = individual predicted value; GREEN = simulation of value based on known variability (residual error)