

Study Notes

*Bin Zhuo, PhD, Biostatistician
Celerion Inc.*

September 25, 2017

Contents

1	Non-inferiority test on response rate	2
1.1	Study Information	2
1.2	Background	2
1.3	Case Study	2
1.4	Statistical Analysis	3
2	Concentration, QTc and Exposure-Response(ER) analysis	4
2.1	Background	4
2.2	Summary comparison between TQT and C-QT study	4
2.3	Statistical Analysis of C-QT data	6
2.4	Result evaluation	6
2.5	Sample size evaluation using bootstrap - A real case study	6
3	Optimal Two-Stage designs	9
3.1	Background	9
3.2	Optimal two-stage designs	10
3.3	Searching optimal design parameters	10
4	Survival Analysis Notes	12
4.1	Basic concept and definition	12
4.2	Censoring and truncation	13
4.3	Nonparametric Estimation of censored and truncated data	13

1 Non-inferiority test on response rate

1.1 Study Information

- Client name: Adello biologics, LLC (Formerly known as Therapeutic Proteins, LLC)
- Celerion Study No. CA18641

NOTE FROM BIN (1): in the experiment, subjects will undergo the following steps (in order):

- **ADA screening test.** If the screening result is *NEGATIVE*, then the result will be recorded *NEGATIVE*, otherwise screening *POSITIVE* and continue
- **ADA confirmatory test.** If the result is *NEGATIVE*, then the subject will have a confirmatory *NEGATIVE* result, otherwise a *POSITIVE* result with a “titer” value recorded and continue
- **ADA neutralizing test.** If the result is *NEGATIVE*, then the subject will have a neutralizing *NEGATIVE* result, otherwise *POSITIVE* with a “titer” value recorded and there will be no more test.

NOTE FROM BIN (2): The cut-off value to determine whether a subject has *POSITIVE* or *NEGATIVE* ADA result may be different (per assay). Therefore, when (if needed) analyzing titer results, there should be a batch effect.

1.2 Background

Therapeutic proteins (sometimes also called biologics, biopharmaceuticals, biological products, or biological medicinal products) and peptides have the potential to induce immunogenicity. The consequences of product immunogenicity vary from no evidence of clinical effect to severe, life-threatening responses. Anti-drug antibodies (ADA) have been implicated in infusion reactions and anaphylaxis as well as immune complex-mediated disease. ADA have also caused secondary treatment failures (loss of efficacy) and, in rare occasions, more serious thrombocytopenia and pure red cell aplasia. Therefore, ADA are a medical concern in terms of safety and long-term efficacy of the drug and it is critical to evaluate their development in all patients during clinical studies, not just in a symptom-driven manner. With a goal of guiding medical practice, the elucidation of ADA responses and their characteristics relative to clinical consequences is vital (Shankar et al., 2014).

ADA comprises neutralizing and non-neutralizing ADA. Other terms that have been used for ADA include anti-therapeutic antibody (ATA), anti-product antibody (APA), or anti-biologic antibody (ABA).

- Neutralizing ADA (NAb): ADA that inhibits or reduces the pharmacological activity of the biologic drug molecule, as determined by an *in vitro* test or animal-based bioassay method, regardless of its *in vitro* clinical relevance (i.e., whether or not test method results relate to clinical impact in the subject).
- Non-neutralizing ADA (non-neutralizing antibody, non- NAb): ADA that binds to the biologic drug molecule but does not inhibit its pharmacological activity in an *in vitro* test or animal-based bioassay method, regardless of its *in vivo* clinical relevance (i.e., whether or not test method results relate to clinical impact in the subject).

1.3 Case Study

The goal of this study is to compare immunogenicity of Theragrastim® (new drug) and Neupogen® (reference drug) after multiple subcutaneous (SC) administrations in healthy subjects. ADA levels for Theragrastim® and Neupogen® will be estimated and compared to evaluate potential difference between the two products in the incidence of human immune responses.

This is a one center, single-blind, randomized, parallel, multiple-dose, safety and immunogenicity study. A total number of one hundred thirty four (134) healthy adult male and female subjects will be enrolled and randomized to 1 of 2 treatments (67 subjects per treatment).

The sample size is chosen based on a target of 61 subjects per arm as calculated, to which 6 subjects (~10%) were added to each arm to account for potential dropouts. With 61 subjects per arm, the trial can show, with 80% power, that the upper bound of the one-sided 95% confidence interval of the difference in ADA+ rates between the two products is below (or above) the non-inferiority margin (10%)

The power calculation for sample size is based on the following assumptions:

- The ADA+ rate of Neupogen[®] is 3.3%
- The ADA+ rate of Theragrastim[®] is 3.3%
- The mean ADA+ rate difference (δ) between the two products is zero;
- The NI margin (δ_0) is 10%.

The power calculation is based on exact method (Chan and Zhang, 1999) using δ -projected Z-statistic (i.e., the score statistic) with REML estimation procedure (Miettinen and Nurminen, 1985).

1.4 Statistical Analysis

The rate (or proportion) of subjects that have ADA+ in confirmatory test and neutralizing test (if needed) will be compared between Theragrastim[®] and Neupogen[®] treatments to determine if any differences are statistically meaningful.

1.4.1 Statistical method

The rate difference between Theragrastim[®] and Neupogen[®] will be defined as:

$$\delta = \pi_1 - \pi_2 \quad (1)$$

where π_1 is the ADA+ rate of Theragrastim[®] and π_2 is that of Neupogen[®].

In hypotheses testing, the research or alternative hypothesis represents what the study aims to show. The null hypothesis is the opposite of the research hypothesis and is what the investigator hopes to disprove (Walker and Nowacki, 2011). Therefore, the primary statistical hypothesis for the clinical trial will be tested using

$$H_0 : \pi_1 - \pi_2 \geq 0.10 \quad VS \quad H_1 : \pi_1 - \pi_2 < 0.1 \quad (2)$$

Confidence intervals (CIs) will be calculated using the Farrington-Manning method (δ -projected Z-statistic) recommended by Chan and Zhang (1999).

Note that in SAS, the null in Expression (2) is equivalently stated as

$$H_0 : \pi_2 - \pi_1 \leq -0.10 \quad VS \quad H_1 : \pi_2 - \pi_1 > -0.1 \quad (3)$$

Therefore, in the SAS output, the null is rejected (i.e., Theragrastim[®] is non-inferior to Neupogen[®] in terms of ADA+ rate) if the lower bound of the one-sided 95% CI of the difference is above NI margin (-0.10).

1.4.2 SAS implementation

The SAS code (Castelloe and Watts, 2015) is provided as follows:

```
proc FREQ data= test order=data;
  tables time*treat*Response / nopercnt nocol nocum
  riskdiff(noninf margin=0.1 method=fmscore norisks)
  alpha = 0.05;
```

```

    exact riskdiff(method = score);
    weight Count;
    ods output CrossTabFreqs = frequency
    PdiffCLs = conf
    pdiffnoninf = pvals
    ;
run;

```

Programmer note: the above code produces results as desired only if the input data set has the same structure as the data set created by the following code. If not, please sort the test data by time treat and response as this one.

```

data test;
input Time $ treat $ response $ count;
datalines;
    C2 neupogen positive 5
    C2 neupogen negative 62
    C2 theragrastim positive 3
    C2 theragrastim negative 64
    C3 neupogen positive 5
    C3 neupogen negative 62
    C3 theragrastim positive 3
    C3 theragrastim negative 64
    C4 neupogen positive 5
    C4 neupogen negative 62
    C4 theragrastim positive 3
    C4 theragrastim negative 64
    ;
Run;

```

2 Concentration, QTc and Exposure-Response(ER) analysis

2.1 Background

The International Conference on Harmonization (ICH) E14 guidance requires that all new drugs with systemic bioavailability be studied for their off-target effect on cardiac repolarization by a ‘thorough QT/QTc’ (TQT) study. With experience on exposure-response modelling on data from many of these TQT studies, results obtained from concentration-QTc analysis are gaining as much acceptance as findings from by-time analysis of the placebo-adjusted change from baseline in the QTc interval after administration of a study drug ([Panicker et al., 2016](#)).

Given the high cost of performing a TQT study, which is usually done late in phase II of clinical drug development, there has been a move towards performing intensive ECG analysis in single ascending dose (SAD) studies. While the sample size in these early-phase studies is statistically underpowered to demonstrate a placebo-adjusted change in mean QTc as described in the ICH E14 guidance, one could arguably obtain the same information by exposure-response modelling on early-phase data ([Salvi et al., 2010](#)).

2.2 Summary comparison between TQT and C-QT study

Table 1: comparison between TQT and C-QT.

	TQT	C-QT
Number of doses	At least 1 (the supratherapeutic dose) if you do not expect QT effect. However, standardly there is 2 (therapeutic and supratherapeutic doses)	At least 3 dose levels to have a adequate range of concentrations to evaluate the concentration-response of the drug
Highest Dose	Supratherapeutic dose corresponding to concentrations expected following worst case scenario in a clinical setting (i.e., renal/hepatic impairment affecting drug metabolism and/or elimination or following a DDI)	Dose level that will produce concentrations expected after 2 time your worst case scenario in a clinical setting (i.e., renal/hepatic impairment affecting drug metabolism and/or elimination or following DDI)
Positive Control (Moxifloxacin)	Always to assess assay sensitivity (Moxi plasma samples are collected to keep the study blinded and in case sponsor needs to analyze them. But samples are no required to be assay)	If you can reach 2 time your worst case scenario, a positive control is not required. If 2 time your worst case scenario is not achievable, than you need a positive control (Moxi plasma samples are collected and analyzed , in this case, since concentrations will be used in the C-QT analyses)
Study Design	4-way Xover (for drugs that show no accumulation of the parent or metabolites) N = 48 Parallel / Parallel with nested Xover (for long half-life drugs or those showing accumulation) N = 240/180	Single Ascending dose (N = 9/Cohort; 6 active - 3 placebo) Xover (N = 9; 6 active - 3 placebo) Parallel (N = 9/Group; 6 active - 3 placebo) If Moxi is added, 5 subjects will be added (1 placebo [to balance the placebo:moxi ratio] and 4 moxi per group/cohort. However need power calculation to confirm that sample size is sufficient to show assay sensitivity.)

	TQT	C-QT
Primary analysis	A mixed effect analysis of covariance (ANCOVA) is used to analyze $\Delta QTcF$ at each time point post baseline.	Explained below.

2.3 Statistical Analysis of C-QT data

The exposure-response analysis is usually performed with linear mixed models.

$$\Delta QTc_{ijk} = (\mu + TRT_j + t_k + \eta_{\mu,i}) + \theta_0(B_{QTc,i} - \bar{B}) + (\theta_1 + \eta_{c,i})C_{ijk} + \epsilon_{ijk} \quad (4)$$

Fixed term: In Equation (4), ΔQTc_{ijk} is the change from baseline QTc for subject i of treatment j at time k , μ is the overall mean, TRT_j is the effect of treatment j , t_k is the effect at time j (treated as categorical). The term $B_{QTc,i}$ and \bar{B} represent the baseline QTc value for subject i and overall mean of baseline QTc, respectively. The C_{ijk} is the drug concentration level for subject i of treatment j at time k . Note for placebo group, C_{ijk} is set to be 0 in the model.

Random term: $\eta_{\mu,i}$ is the random intercept, $\eta_{c,i}$ is the random slope for concentration, and ϵ_{ijk} is the error term. In the data analysis step, the covariance structures for both η and ϵ need to be specified.

This is based on my own understanding: the random effects has the following distribution

$$\eta = (\eta_{\mu,i}, \eta_{c,i})' \sim N(\mathbf{0}, \mathbf{G}), \quad \epsilon_{ij} \sim N(\mathbf{0}, \mathbf{\Sigma}) \quad (5)$$

where \mathbf{G} is a 2×2 covariance matrix, and $\mathbf{\Sigma}$ is an $K \times K$ covariance matrix and K is the number of time points for each subject.

2.4 Result evaluation

The study will be considered as positive (i.e., it could detect the moxifloxacin-induced QT prolongation) for an experiment if both of the following criteria are satisfied (copied from (Panicker et al., 2016) and needs to be confirmed within each study)

- the plasma concentration- $\Delta\Delta QTcF$ relationship showed a statistically significant ($P < 0.05$) positive slope, and
- the upper bound of the 90% two-sided confidence limit of the $\Delta\Delta QTcF$ at geometric mean C_{max} for the experiment exceeds 10 ms.

2.5 Sample size evaluation using bootstrap - A real case study

There is a Merck study (project number: CA15451) conducted at our site. This is an open-label, multiple-dose, randomized, 3-cohort, placebo-and active-controlled, parallel study in 54 healthy adult subjects under fasting conditions. For the purpose of detecting Moxifloxacin effect, we'll only include Cohort 2 (placebo) and Cohort 3 (Moxi) in the data analysis part.

- **Cohort 2:** Eighteen (18) subjects will receive placebo to match treatment and study procedures on Day 10 and 20 of Cohorts 1 and 3. Cardiodynamic and pharmacokinetic samples will be collected up to 24 hours following dosing on Days 10 and 20.

- **Cohort 3:** Eighteen (18) subjects will receive an oral dose of 400 mg of moxifloxacin on Day 10 and an oral dose of 800 moxifloxacin on Day 20. Subjects will also receive placebo for mirtazapine (study drug) to match the treatment and study procedures on Days 10 and 20 of Cohorts 1 and 2. Cardiodynamic and pharmacokinetic samples will be collected up to 24 hours following dosing on Days 10 and 20.

2.5.1 Data cleaning

- *Plasma concentration:* can be obtained from `pla2.sas7bdat`. For placebo group (Cohort 2), the concentration level is set to be 0 since there's no drug.
- *QTcF:* can be obtained from `ecg_core_lab.sas7bdat`. $\Delta QTcF$ is calculated as change from time-matched baseline of the average QTcF for each subject.

After data cleaning, the first few lines of the final data looks like

```
final <- read.csv("H:/Projects/CA22282/Data/final.csv", header = T)
head(final)
```

```
##   ptno   treat P_HOUR conc BaseQTcF PostQTcF changeQTcF   BQTcF   meanB
## 1    1 Placebo   0.5    0 412.3333 416.0000   3.666667 417.6667 404.787
## 2    1 Placebo   1.0    0 412.6667 421.3333   8.666667 417.6667 404.787
## 3    1 Placebo   2.0    0 411.6667 415.3333   3.666667 417.6667 404.787
## 4    1 Placebo   3.0    0 408.6667 414.0000   5.333333 417.6667 404.787
## 5    1 Placebo   4.0    0 405.0000 415.3333  10.333333 417.6667 404.787
## 6    1 Placebo   6.0    0 400.6667 401.6667   1.000000 417.6667 404.787
```

The `BQTcF` is the overall baseline, one for each subject, and `meanB` is the overall baseline for all subjects. They correspond to $B_{QTc,i}$ and \bar{B} in Equation (4), respectively.

2.5.2 Detection of Moxifloxacin effect

In this section, we first implement the model described in Equation (4), and then calculate the lower bound of the 90% two-sided CI of the $\Delta\Delta QTcF$ at geometric mean C_{max} . If the 90% lower CI is greater than 5, and the slope for the concentration is significantly positive ($P < 0.05$), then the ability to detect moxifloxacin effect can be established.

```
library(dplyr)
library(haven)
library(tidy)
library(lme4)

calculate_ci <- function(obj, data){

  result <- obj

  coef_est <- result$coefficients
  # the estimate and std for treatment difference, Moxi - placebo
  mu1 <- coef_est[rownames(coef_est)=="treatMoxi"]

  # the estimate for log concentration level
  conc_est <- coef_est[rownames(coef_est) %in% c("lgconc", "conc_um")]
  logc <- conc_est[c(1, 3)]
  # the estimated covariance matrix
  cov_est <- result$vcov
  # of the repeated measurements
```

```

rep_cov <- cov_est[1:8, 1:8]
# the estimated covariance between lgconc and treatment
cov2 <- cov_est[rownames(cov_est) %in% c("lgconc", "conc_um"), colnames(cov_est)=="treatMoxi"]

# get the cmax_mean for each data
conc <- ana_data %>% filter(treat == "Moxi") %>%
  group_by(ptno) %>% mutate(cmax = max(conc)) %>%
  ungroup() %>% select(cmax) %>% distinct() %>% t()

cmax_mean <- log(exp(mean(log(as.vector(conc))))) + 1)
#cov2 <- -12.619163
ddqtcf_mean <- mu1[1] + conc_est[1]*cmax_mean
ddqtcf_se <- sqrt(mu1[2]^2 + conc_est[2]^2*cmax_mean^2 + 2*cmax_mean*cov2)
CI <- ddqtcf_mean + c(-1, 1)*qnorm(0.95)*ddqtcf_se
report <- c(logc, ddqtcf_mean, CI)
names(report) <- c("Moxi_slope", "t_value", "est.", "lower 90%", "upper90%")
report
}

ana_data <- final %>% mutate(lgconc = log(conc + 1),
  # scaleconc = scale(conc, center = F),
  conc_um = conc/1000,
  dBQTcF = BQTcF - meanB,
  treat = factor(treat, levels = c("Placebo", "Moxi")),
  P_HOUR = as.factor(P_HOUR))

mod <- lmer(data = ana_data, changeQTcF ~ (P_HOUR-1) + treat +
  dBQTcF + lgconc + (lgconc+1|ptno) )
mod <- summary(mod)
result <- calculate_ci(mod, ana_data)
result

```

```

## Moxi_slope    t_value      est. lower 90% upper90%
##    5.402260    3.616259 12.048769  7.790722 16.306817

```

The result here shows that the estimated slope for concentration is 5.4 with t-value of 3.62. Also, at C_{max} , the 90% lower CI is 7.79, which is greater than 5. Consequently, the effect of moxifloxacin is established.

2.5.3 Bootstrap for power and sample size

In this part, we will use bootstrap to re-run the linear mixed model for power and sample size evaluation. We assume equal sample size for Moxifloxacin and Placebo group. For a given sample size, we get bootstrapped sample from data that was used in the analysis. Each subject is a sampling unit, and therefore, if a subject is chosen, then all the records belonging to this subject will be used. We will do $B = 1000$ bootstrapings to evaluate the power.

Below is an example of power calculated when 13 subjects for each treatment are to be sampled from the data.

```

### do bootstrapping for sample size

# n is the sample size
# b is the number of bootstraps

```



```

# obj is the result run from linear mixed models

bootstrap_sampsize <- function(n, b = 1000, data = ana_data){

  # do bootstrap sampling
  mox_id <- as.vector(unique(data$ptno[data$treat=="Moxi"]))
  plb_id <- as.vector(unique(data$ptno[data$treat=="Placebo"]))

  ddqtcf <- matrix(NA, ncol = 5, nrow = b)

  for(i in 1:b){

    mox_b <- sample(mox_id, n);
    plb_b <- sample(plb_id, n);
    data_b <- data[data$ptno %in% c(mox_b, plb_b), ]

    mod <- lmer(data = data_b, changeQTcF ~ (P_HOUR-1) + treat +
                dBQTcF + lgconc + (lgconc+1|ptno) )
    result <- summary(mod)

    ddqtcf[i, ] <- calculate_ci(result, data_b)
  }

  return(ddqtcf)
}

b1_s <- bootstrap_sampsize(n = 13, b = 100, data = ana_data)
mean(b1_s[, 4] > 5)

```

```
## [1] 0.89
```

The estimated power should be around 87%. Below is a table for power evaluation for different choice of sample size per treatment.

Table 2: Power evaluation for different sample size per treatment.

sample size	6	7	10	11	12	13	14	16
power	0.465	0.556	0.692	0.776	0.838	0.874	0.930	0.997

From the above table, we can see that Merck has 50% (18 used, 12 needed) more subjects than necessary to achieve a power of 80%.

3 Optimal Two-Stage designs

3.1 Background

This part is an exploration of Simon's 1989 paper (Simon, 1989). In this paper, the author discussed the optimal sample size for a two stage design, where the goal of this design is to compare the response rate of a new drug:

$$H_0 : p \leq p_0 \quad \text{VS} \quad H_1 : p \geq p_1 \quad (6)$$

In this design, the null is that the true response probability is less than some uninteresting level p_0 . If H_0 is true, then we require that the type 1 error should be less than α , whereas if H_1 is true, then the power to reject the null should be at least $1 - \beta$. In addition to these constraints, we wish to minimize the number of patients treated with a drug of low activity. The attention is restricted to two-stage designs.

3.2 Optimal two-stage designs

If the number of patients in the first and second stage are denoted by n_1 and n_2 respectively, then the expected sample size is $EN = n_1 + (1 - PET)n_2$, where PET is the probability of early termination after the first stage. With a true response probability p , we'll terminate the experiment at the end of the first stage and reject the drug (or fail to reject H_0) if r_1 or fewer responses are observed. This occurs with probability $PET = B(r_1, n_1, p)$, where B is cumulative binomial distribution. We'll reject the drug at the end of the second stage if r or fewer responses are observed. Hence the probability of rejecting a drug with success probability p is

$$B(r_1, n_1, p) + \sum_{x=r_1+1}^{\min\{n_1, r\}} b(x, n_1, p)B(r-x, n_2, p) \quad (7)$$

where b denotes the binomial probability mass function.

The design approach considered here is to specify the parameters p_0, p_1, α and β and then determine the two-stage design that satisfies the error probability constraints and minimizes the expected sample size when the response probability is p_0 . The optimization is taken over all values of n_1 and n_2 , as well as r_1 and r . Early acceptance of the drug is not permitted here.

3.3 Searching optimal design parameters

For specified values of p_0, p_1, α and β , the optimal design is found by enumeration using exact binomial probabilities. For each value of total sample size n and each value of n_1 in the range $(1, n-1)$, we determine the integer value of r_1 and r . The chosen (r_1, r) minimizes the expected sample size when $p = p_0$, which was found by searching over the range $r_1 \in (0, n_1)$. For each value of r_1 , we determine the max value of r that meets the power constraint. We then examine whether the set of parameters (n, n_1, r_1, r) satisfies the type 1 error constraint. If it does, then we compare the expected sample size to the minimum achieved by previous feasible designs and continue the search over r_1 . Keeping n fixed, we search over the range of n_1 to find the optimal two-stage design for that maximum sample size n . The search over n ranges from a lower value of about

$$\bar{p}(1 - \bar{p}) \left[\frac{z_{1-\alpha} + z_{1-\beta}}{p_1 - p_0} \right]^2 \quad (8)$$

where $\bar{p} = (p_0 + p_1)/2$.

Here's the code to implement the searching process

```
#
## Optimal two stage design
## hypothesis H0: p <= p0; VS H1: p >= p1

p0 <- 0.05; # Null
p1 <- 0.25; # alternative
alpha <- 0.1; # type 1 error
beta <- 0.1 # type 2 error
```

```

library(dplyr)
# starting sample size
start_n <- function(p0, p1, alpha, beta){
  p_bar <- (p0 + p1)/2
  z_alpha <- qnorm(1-alpha)
  z_beta <- qnorm(1-beta)
  n <- p_bar*(1-p_bar)*((z_alpha + z_beta)/(p1-p0))^2
  return(ceiling(n))
}

# calculate the probability of rejecting the null
# depending on the parameter configuration,
# if p is under the null, this calculates type 1 error;
# whereas if p following the alternative, this calculates the power.
calc_prob <- function(r1, r, n1, n2, p){
  # if rejected the drug at the first stage....
  s1 <- pbinom(r1, n1, p)
  cel <- min(n1, r)

  # passed the first stage, but rejected eventually
  part2 <- function(x){
    dbinom(x, n1, p)*pbinom(r-x, n2, p)
  }

  s2 <- 0
  # the probability of rejecting the null at the second stage....
  for (j in (r1+1:cel)){
    s2 <- s2 + part2(j)
  }

  return(1- s1 - s2)
}

expected_n <- function(r1, p, n1, n2){
  pet <- pbinom(r1, n1, p)
  EN <- n1 + (1-pet)*n2
  return(EN)
}

# for a given sample size, calculate the type 1 error and power
# this will give all possible outcomes that meets type 1 error and power constraint.
calc_alpha_beta <- function(n, p0, p1, alpha, beta){

  # all possible values of n1 given total number is n
  n1 <- seq(1, n-1, by = 1)
  # sample size in Stage 2
  n2 <- n - n1

  df_rlt <- data.frame(matrix(ncol=7))
  names(df_rlt) <- c("r1", "r", "n1", "n2", "alpha_1", "pwr1", "EN")
  # loop over n1 to find r1 and r
  for (i in 1:length(n1)){

```

```

# initiate r1
r1 <- seq(0, n1[i], by = 1)
for (k in 1:length(r1)){
  # for all possible r given the current r1
  r <- seq(r1[k], n, by = 1)
  for (l in 1:length(r)){
    pwr1 <- calc_prob(r1[k], r[l], n1 = n1[i], n2 = n2[i], p = p1)
    t1err <- calc_prob(r1[k], r[l], n1 = n1[i], n2 = n2[i], p = p0)
    en <- expected_n(r1[k], p0, n1[i], n2[i])
    rlt <- data.frame(r1[k], r[l], n1[i], n2[i], t1err, pwr1, en)
    names(rlt) <- c("r1", "r", "n1", "n2", "alpha_1", "pwr1", "EN")
    if (t1err <= alpha & pwr1 >= 1-beta){
      df_rlt <- bind_rows(df_rlt, rlt)
    }
  }
}
}
return(df_rlt[-1, ])
}

# find the optimal design
find_n <- function(df){
  df1 <- df %>% filter(r == max(r)) %>%
    filter(EN == min(EN))
  return(df1)
}

# an example
n_init <- start_n(0.05, 0.25, 0.1, 0.1)
df0 <- calc_alpha_beta(n = 24, p0 = 0.05, p1 = 0.25, alpha = 0.1, beta = 0.1)
df1 <- find_n(df0)
df1

##   r1 r  n1 n2   alpha_1   pwr1   EN
## 1  0 2   9 15 0.09312941 0.9028407 14.54626

```

4 Survival Analysis Notes

4.1 Basic concept and definition

The basic quality used to describe time-to-event phenomena is the **survival function**, the probability of an individual surviving beyond time x (experiencing an event after time x). It is defined as

$$S(x) = Pr(X > x) \quad (9)$$

Hazard function is another fundamental function in survival analysis, also known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic processes. The **hazard rate** is defined as

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x | X \geq x)}{\Delta x} \quad (10)$$

If X is a continuous random variable, then

$$h(x) = f(x)/S(x) = -d \ln[S(x)]/dx \quad (11)$$

Naturally, the cumulative hazard function is defined by the integral of the hazard function from time 0 up to x , $H(X) = \int_0^x h(t)dt = -\ln[S(x)]$.

Mean residual life (MRL) is the expected remaining lifetime, defined as $\text{mrl}(x) = E(X - x | X > x)$.

4.2 Censoring and truncation

One particular feature of time-to-event data is *censoring*, which, broadly speaking, occurs when some lifetimes are known to have occurred only within certain intervals. The types of censoring can be categorized (with subcategories) as

- right censoring
 - *Type I censoring*: The event can only be observed if it occurs prior to some pre-specified time C_r .
 - *Type II censoring*: The study continues until the failure of the first r individuals, where r is some pre-determined integer ($r < n$).
 - *Complete risks censoring*: This happens when the researcher is interested in estimation of the marginal distribution of some event but some individuals under study may experience some competing event which causes them to be removed from the study.
- left censoring: when the event time is less than a censoring time C_l . For such individuals, we know that the event occurred before time C_l , but we don't know the exact event time.
- Interval censoring: when the lifetime is only known to occur within an interval. Such censoring may happen when patients in a clinical trial or longitudinal study have periodic follow-up and the patient's event time is only known to fall in an interval $(L_i, R_i]$.

The data of this type of experiment can be represented by pairs of random variables (T, δ) , where δ indicates whether the lifetime X is observed ($\delta = 1$) or censored ($\delta = 0$), and T is equal to X if the lifetime is observed, equal to C_r if right censored, and equal to C_l if left censored.

Truncation of survival data occurs when only those individuals whose event time lies within a certain observational window (Y_L, Y_R) are observed. An individual whose event time does not fall in this interval is not observed, and therefore no information on this subject is available to the investigator. **This is contrast to censoring where there is at least partial information on each subject.** Because we are only aware of individuals with event times in the observational window, the inference for truncated data is restricted to conditional estimation.

4.3 Nonparametric Estimation of censored and truncated data

4.3.1 Right censored data

- Kaplan and Meier (1958) estimation of survival function (also called the Product-limit estimator) for *right-censored* data

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t} [1 - \frac{d_i}{Y_i}] & \text{if } t_1 \leq t \end{cases} \quad (12)$$

where Y_i is a count of the number of individuals with a time on study of t_i or more (i.e., the number of individuals who are alive at t_i or experience the event of interest at t_i), and d_i is the number of events at time t_i . The variance of Kaplan-Meier estimator is given by Greenwood's formula

$$\hat{V}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{Y_i(Y_i - d_i)}. \quad (13)$$

The product-limit method can be used to estimate the cumulative hazard function $H(t) = -\ln[S(t)]$.

- Nelson-Aalen estimator: An alternative estimator of the cumulative hazard rate

$$\tilde{H}(t) = \begin{cases} 0 & \text{if } t \leq t_1 \\ \sum_{t_i \leq t} \frac{d_i}{Y_i} & \text{if } t_1 \leq t \end{cases} \quad (14)$$

whose variance is given by

$$\sigma_H^2(t) = \sum_{t_i \leq t} \frac{d_i}{Y_i^2}. \quad (15)$$

It has better small-sample-size performance as compared to product-limit estimator.

4.3.2 Left-truncated data

The basic idea of Kaplan-Meier estimator and Nelson-Aalen estimator can be used to estimate the survival function of left-truncated data. Note that the product-limit estimator of the survival function at a time t is now an estimator of the probability of survival beyond t , conditional on survival to the smallest of the entry times L , $Pr[X > t | X \geq L] = S(t)/S(L)$. Similarly the Nelson-Aalen statistic estimates the integral of the hazard rate over the interval L to t . Note that the slope of the Nelson-Aalen estimator still provides an estimator of the unconditional hazard rate.

4.3.3 Other sampling scheme

There are other types of sampling scheme, namely, left, double, and interval censoring, right-truncation, and grouped data. Each sampling scheme provides different information about the survival function and requires a different technique for estimation. The (modified) product-limit estimator and Nelson-Aalen estimator will be applicable to those sampling schemes.

4.3.4 Univariate estimation

Kernel-smoothing technique can be used to provide a better estimator of the hazard rate (i.e., the slope of the Nelson-Aalen estimator). Sometimes, the investigator is interested in comparing the hazard rates in the treatment group to the known hazard rates in the reference group. The mortality in the treatment group may have either a multiplicative or additive effect on the reference hazard rate. The estimation of survival function could also be considered from a Bayesian perspective.

Kernel-smoothed estimators of $h(t)$ are based on the Nelson-Aalen estimator $\tilde{H}(t)$ and its variance $\hat{V}[\tilde{H}(t)]$.

Another one is the Bayes estimator. Analogous to the simple parametric case, we may use a squared-error loss function

$$L(S, \hat{S}) = \int_0^\infty [\hat{S}(t) - S(t)]^2 dw(t)$$

where $w(t)$ is a weight function. Two classes of prior distribution have been suggested for this problem. The priors are chosen because they are the conjugate prior of either the survival function or the hazard function. *A conjugate prior means the prior and the posterior distribution are in the same family.*

The first prior is for the survival function, which is assumed to be a random sample from a Dirichlet process (Teh, 2011) with a parameter function α . The Dirichlet process used here can be thought as a stick-breaking problem: starting with a stick of length 1, we break it at β_1 , assigning $1 - S(t_1)$ to be the length we just broke off; we will recursively break the other portion to obtain $S(t_2), S(t_3)$, and so forth. To assign a prior distribution to the survival function, we assume that $S(t)$ follow a Dirichlet distribution with parameter function α . It is typical to take the parameter function as $\alpha([t, \infty)) = cS_0(t)$ where $S_0(t)$ is our prior guess at the survival function and c is a measure of how much weight to put on our prior guess. With this prior, the prior mean is expressed by

$$E[S(t)] = \frac{\alpha(t, \infty)}{\alpha(0, \infty)} = \frac{cS_0(t)}{cS_0(0)} = S_0(t)$$

The data we have available to combine with our prior consists of the on study time T_j and the event indicator, δ_j . Let $0 = t_0 < t_1 < \dots < t_M < t_{M+1} = \infty$, denote the M distinct times. At time t_i , let Y_i be the number of individuals at risk, d_i the number of deaths and λ_i the number of censored observations. Let Δ_i be 1 if $d_i > 0$ and 0 if $d_i = 0$.

Combining this data with the prior, we can show that the posterior distribution of S is also Dirichlet. The parameter for the posterior distribution, α^* , is the original α plus a point mass of one at points where deaths occur. That is, for any interval (a, b) ,

$$\alpha^*([a, b]) = \alpha([a, b]) + \sum_{j=1}^n I[\delta_j > 0, a < T_j < b],$$

where $I[]$ is the indicator function. The Bayes estimator of the survival function is

$$\tilde{S}_D(t) = \frac{\alpha(t, \infty) + Y_{i+1}}{\alpha(0, \infty) + n} \prod_{k=1}^i \frac{\alpha(t_k, \infty) + Y_{k+1} + \lambda_k}{\alpha(t_k, \infty) + Y_{k+1}}, \quad \text{for } t_i \leq t < t_{i+1}, i = 0, \dots, M$$

The Bayes estimator is a continuous function between the distinct death times and has jumps at these death times. For large n this reduces to the Kaplan-Meier estimator, so that the prior information plays no role in the estimate. For small samples, the prior will dominate, and the estimator will be close to the prior guess at S .

4.3.5 Hypothesis testing

References

- Castelloe, John and Donna Watts. 2015. "Equivalence and Noninferiority Testing Using SAS/STAT Software." *SAS Institute Inc.* Paper SAS1911-2015.
- Chan, Ivan SF and Zhongxin Zhang. 1999. "Test-Based Exact Confidence Intervals for the Difference of Two Binomial Proportions." *Biometrics* 55(4):1202–1209.
- Miettinen, Olli and Markku Nurminen. 1985. "Comparative analysis of two rates." *Statistics in medicine* 4(2):213–226.
- Panicker, Gopi Krishna, Dilip R Karnad, Pramod Kadam, Fabio Badilini, Anil Damle and Snehal Kothari. 2016. "Detecting moxifloxacin-induced QTc prolongation in thorough QT and early clinical phase studies using a highly automated ECG analysis approach." *British journal of pharmacology*.

- Salvi, Vaibhav, Dilip R Karnad, Gopi Krishna Panicker and Snehal Kothari. 2010. "Update on the evaluation of a new drug for effects on cardiac repolarization in humans: issues in early drug development." *British journal of pharmacology* 159(1):34–48.
- Shankar, Gopi, S Arkin, L Cocea, V Devanarayan, S Kirshner, A Kromminga, V Quarmby, S Richards, CK Schneider, M Subramanyam et al. 2014. "Assessment and reporting of the clinical immunogenicity of therapeutic proteins and peptides-harmonized terminology and tactical recommendations." *The AAPS journal* 16(4):658–673.
- Simon, Richard. 1989. "Optimal two-stage designs for phase II clinical trials." *Controlled clinical trials* 10(1):1–10.
- Teh, Yee Whye. 2011. Dirichlet process. In *Encyclopedia of machine learning*. Springer pp. 280–287.
- Walker, Esteban and Amy S Nowacki. 2011. "Understanding equivalence and noninferiority testing." *Journal of general internal medicine* 26(2):192–196.