# Design of Multiregional Clinical Trials (MRCT) – Theory and Practice

Gordon Lan
Janssen R&D, Johnson & Johnson
BASS 2015, Rockville, Maryland

Collaborators: Fei Chen and Gang Li, Johnson & Johnson

Outline:

- Parameter of interest (Treatment effect difference)
- From two-sample to one-sample
- Estimation procedures under different assumptions
- Traditional clinical trials: From 1 Center to K Centers
- Models for treatment effect: Fixed effect model; Continuous Random Effects Model and Discrete Random Effects Model
- Consistency and Inconsistency
- Drop-the-min data analysis in MRCT

# Comparisons of two means $\mu_T$ versus $\mu_C$

Parameter of interest $\mu = \mu_T - \mu_C$.

In the following, we will simply this two-sample problem to a one-sample problem.

$X_T$: patient response from T

$X_C$: patient response from C

Suppose there are equal number of patients from T and C (i.e., $N_T = N_C = n$), then there are n "pairs" of measures $\{X_i = X_{Ti} - X_{Ci},\ 1, 2, \ldots, n\}$. The treatment difference $\mu = \mu_T - \mu_C$ can be estimated by $\bar{X}_T - \bar{X}_C$.

Data set = $\{X_{T1}, X_{T2},\ldots,X_{Tn}; X_{C1}, \ldots,X_{Cn}\} \rightarrow \{X_1, X_2, \ldots,X_n\}$

The mathematics behind the one-sample case is much easier to deal with (assuming $N_T = N_C$) than those for the two-samples.

What if $N_T \neq N_C$ ?

The insights learned from the one-sample case can be extended to the two-sample case with minimal modifications, but the proofs for the two-sample case involve the Brownian Motion Process.

Different ways to estimate θ --- Parameter of interest (treatment effect)

In general, we always take a weighted average approach.

Suppose that there are 3 unbiased estimates of $\theta$, say $\widehat{\theta_1}$, $\widehat{\theta_2}$, $\widehat{\theta_3}$.

What is a reasonable way to estimate the "overall" effect $\theta$?

In general, we use a "weighted" approach.

$$\widehat{\theta} = W_1\widehat{\theta_1} + W_2\widehat{\theta_2} + W_3\widehat{\theta_3}\ ,\ W_1 + W_2 + W_3 = 1.$$

Sample size weights --- $\{W_i = N_i / \sum N_j\}$.

Suppose there are 3 estimates of $\theta$, say

$\widehat{\theta_1} = 7$ ($N_1$=100) , $\widehat{\theta_2} = 8$ ( $N_2$= 200), $\widehat{\theta_3} = 9$ ($N_3$= 300).

The weights are100/600=1/6; 2/6 and 3/6,

and the estimate of the overall effect is 7*1/6 + 8*2/6+9*3/6 = 8.333.

Equal weights, mean effect= (7+8+9)/3 = 8.

Which one do you like better?

In statistics, we have a "well-established approach" to estimate $\theta$.

Unconditionally, suppose $\widehat{\theta_1}$, $\widehat{\theta_2}$, and $\widehat{\theta_3}$ are "unbised" estimates of $\theta$ with variances $\gamma_1^2$, $\gamma_2^2$ and $\gamma_3^2$. Then the "best" overall estimate of $\theta$ is the reciprocal variance estimate $\widehat{\theta} = \dfrac{\sum_1^3 \widehat{\theta}_i/\gamma_i^2}{\sum_1^3 1/\gamma_j^2}$ {that is, $W_i = \dfrac{1/\gamma_i^2}{\sum(1/\gamma_j^2)}$}

which puts more weights on "more accurate" estimates.

Next, we will introduce another factor which modifies the sample size weights.

The 3 estimates of $\theta$ are $\hat{\theta}_1 = 7$ ($N_1$=100) , $\hat{\theta}_2 = 8$ ($N_2$ = 200), $\hat{\theta}_3 = 9$ ($N_3$ = 300).

Within group variances (different) are: $\sigma_1^2 = 3$; $\sigma_2^2 = 2.6$ and $\sigma_3^2 = 2.1$.

$\hat{\theta}_1 = 7$ is based on 100 patients with variance $\hat{\gamma}_1^2 = 3/100 = 0.03$.

$\hat{\theta}_2 = 8$ is based on 200 patients with variance $\hat{\gamma}_2^2 = 2.6/200 = 0.013$;

$\hat{\theta}_3 = 9$ is based on 300 patients with variance $\hat{\gamma}_3^2 = = 2.1/300 = 0.007$.

The weightes are {0.1317, 0.3039, 0.5644} and

the reciprocal variance estimate $\hat{\theta} = \dfrac{\sum_1^3 \hat{\theta}_i / \hat{\gamma}_i^2}{\sum_1^3 1/\hat{\gamma}_j^2} = 8.4327$.

This is slightly different from the sample size weighted average 8.333.

For this specific example, the “sample size weights” and “reciprocal of variance weights” provide similar results. This is true in general unless the variance estimates $\{\widehat{\sigma}_i^2\}$ are very different.

In the second half of this talk on MRCT (Multiregional Clinical Trials), we will show you that under the Random Effects Models, the reciprocal variance weights and sample size weights could be VERY DIFFERENT.

A group sequential design requires larger sample size than a fixed (amount of information) design. We will restrict our discussion to the design and data analysis of non-sequential clinical trials in this talk.

Start with a single center study.

Consider a sequence of samples $X_1, X_2,\ldots,X_n,\ldots$ ; iid with mean $\mu$ and variance $\sigma^2$. In the large sample case, the variance can be estimated consistently by the sampled variance and mathematically, we treat $\sigma^2$ as known. ($\mu$ is fixed, unknown)

Te test $H_0$: $\mu = 0$ versus $H_A$: $\mu > 0$, we usually "choose" a simple alternative $\mu = \mu_1 > 0$ and the test statistic is

$$Z = \frac{\sum_1^N X_i}{\sigma\sqrt{N}} = \frac{\overline{X}}{\sigma\sqrt{1/N}} \text{ with rejection region } Z \geq z_\alpha (= 1.96).$$

To solve for sample size N to reach power 1-β, we choose N from the equation $EZ = z_\alpha + z_\beta$ under $\mu = \mu_1$, or,

$$N = \left(\frac{\mu_1}{\sigma}\right)^2 (z_\alpha + z_\beta)^2. \qquad \text{(Equation 1)}$$

For one-sided $\alpha = 0.025$ and 85% power,

$$z_\alpha + z_\beta = z_{0.025} + z_{0.15} = 1.96+1.04 = 3.$$

What if we have K centers?

Under a fixed effect model, the means $\mu_1=\mu_2=...=\mu_K = \mu$.

Let the sample sizes for the K centers be $\{N_1, N_2,\ldots,N_K\}$ with total sample size $N = \sum N_k$.

Then the test statistic is still

$$Z = \frac{\sum_1^N X_i}{\sigma\sqrt{N}} = \frac{\bar{X}}{\sigma\sqrt{1/N}} \text{ with rejection region } Z \geq z_\alpha .$$

And the total sample size N is computed by

$$N = \left(\frac{\mu_1}{\sigma}\right)^2 (z_\alpha + z_\beta)^2 . \qquad \text{(Equation 1)}$$

The overall mean is estimated by

$$\bar{X} = (N_1\bar{X}_1 + N_1\bar{X}_2 + \ldots + N_K\bar{X}_K) / (N_1 + N_1 + \ldots + N_K)$$

$$= \sum_{k=1}^{K} W_k \bar{X}_k \text{ where } W_k = N_k / N.$$

I had numerous discussions with colleagues on clinical trial design and data analysis of "Multi-Center" trials. It is quite common that the observed data did not support the fixed effect model assumption.

What is the alternative:  Use a different model!

There are two random effects models:

1. Continuous Random Effects Model (CREM), and
2. Discrete Random Effects Model (DREM),

CREM = DerSimonian-Laird (1986), originally introduced for meta-analysis.

CREM (originally for meta-analysis) was also considered by many statisticians and used in clinical trial data analysis. Many of my colleagues were NOT happy with the statistical characteristics of CREM. They would rather use the fixed effect model to design trials and analyze data.

In this presentation, I will elaborate on the "undesirable" (??) characteristics of CREM and suggest another random effects model (DREM) as an alternative option.

What is CREM?

Suppose we have K centers. From Center k, k = 1,2,…,K; responses $X_{kj} \sim N(\nu_k, \sigma_k^2)$. To simplify notations, we assume all the within center variances are the same; i.e., $\sigma_k^2 \equiv \sigma^2$.)

Denote the sample mean of the k-th center as $\overline{X}_k$.

Under CREM, each of the center effects $\nu_k$ is random $\sim N(\nu, \tau^2)$.

The variance component $\tau^2$ is called the between-center variance.

Conditionally, as an estimate of $\nu_k$,

$\overline{X}_k$ is normal with mean $\nu_k$ and variance $\sigma^2 / N_k$.

Unconditionally, as an estimate of the overall effect $\nu$,

$\overline{X}_k$ is normal with mean $\nu$ and variance $\sigma^2 / N_k + \tau^2$.

Denote the reciprocal variance $(\sigma^2 / N_k + \tau^2)^{-1}$ as $V_k$.

Reciprocal variance estimate:

The overall mean is estimated by

$$\hat{\nu} = \sum_{k=1}^{K} \left( \frac{V_k}{\sum_{j=1}^{K} V_j} \right) \bar{X}_k = \sum_{k=1}^{K} W_k \bar{X}_k$$

where the weights are $\{W_k = V_k / \sum_{j=1}^{K} V_j\}$.

It can be shown that $\text{Var}(\hat{\nu}) = 1 / \sum_{1}^{K} [\text{Var}(\bar{X}_k)]^{-1}$.

Note that the variance component $\tau^2$ contributes to the weights.

Fixed: $\mathrm{Var}(\overline{X}_k) = \frac{\sigma^2}{N_k}$

CREM: $\mathrm{Var}(\overline{X}_k) = \frac{\sigma^2}{N_k} + \tau^2$ = w/i variation + between variation.

When $\tau^2$ is very small, we expect the center means $\{\overline{X}_1, \overline{X}_2, ..., \overline{X}_K\}$ have small dispersion, and the statistical inferences from the "fixed effect" approach will be similar to those of the CREM approach.

For multi-regional trials, do we expect $\tau^2$ to be "very small"?

$\mathrm{Var}(\overline{X}_k) = \frac{\sigma^2}{N_k} + \tau^2$, assume $\tau^2$ is NOT negligiable.

As $N_k \to \infty$, $\mathrm{Var}(\overline{X}_k) \to \tau^2$. In the large sample case,

all the variances $\{\mathrm{Var}(\overline{X}_k)\}$ are $\approx \tau^2$ and all $\{W_k\}$ are $\approx 1/K$.

In other words, $\hat{\nu} \approx (\sum_{k=1}^{K} \overline{X}_k) / K$.

In general, CREM pushed the sample size weights

$\{N_k/N\} \to \{1/K\}$.

As $N_k \to \infty$, $\mathrm{Var}(\bar{X}_k) \to \tau^2$ which does not go to zero.

In the large sample case, $\mathrm{Var}(\bar{\bar{X}}) \approx \tau^2/K$ does not go to 0.

The power of the Z-test will not go to 1 even when all $N_k \to \infty$.

($K \to \infty$ is also required for power $\to 1$.)

Remember that CREM was introduced for meta-analysis.

Examples: $\tau^2/\sigma^2 = R$, or $\tau^2 = R\,\sigma^2$. Sample sizes =(200, 50, 300).

In the following, we consider various values of R:

Sample size weights are…… (.3636, .0909, .5454)

When R = 0.0005, CREM weights are (.3699, .0993, .5308)

When R = 0.005, CREM weights are (.3846, .1538, .4615)

When R = 0.05, CREM weights are (.3550, .2789, .3661)

When R = 1, CREM weights are (.3348, .3299, .3354)

When R becomes larger, the weight $W_2$ increases

from 0.0909 to 0.3299 ≈ 1/3.

Another example: $\tau^2/\sigma^2 = R$, or $\tau^2 = R\,\sigma^2$. Fix R = .05,

Consider different sample sizes = (200, 50, 300).

Sample size weights (.3636, .0909, .5454)

CREM weights : (.3550, .2789, .3661)

Sample sizes ↑ (1000, 250, 1500).

CREM weights: (.3389, .3200, .3411)

In general, CREM pushed the sample size weights

$\{N_k/N\} \rightarrow \{1/K\}$.

Another problem we experienced with CREM:

Consider a trial with center sizes {20, 30, 50, 400, 500}.

Pooling the first 3 centers into 1 center will seriously affect the estimate of the overall effect.

$\tau^2/\sigma^2 = R$, or $\tau^2 = R\,\sigma^2$. Fix R = .05.

Sample sizes = (20, 30, 50, 400, 500)

Sample size weights: (0.02, 0.03, 0.05, 0.4, 0.5)

CREM weights = (.134, .161, .192, .255, .258) ≈ 1/5

Pool the first three centers into one.

Sample sizes = (100, 400, 500)

Sample size weights = (0.1, 0.4, 0.5)

CREM weights = (.303, .347, .350) ≈ 1/3

Another Random Effects Approach

Discrete Random Effects Model (DREM)

The population $\Omega = \Omega_1 + \Omega_2 + \ldots + \Omega_K$

$P[\Omega_k] = W_k$ , $\sum_k W_k = 1$. Regional effects $\{\nu_k\}$ are constants.

A random patient's response $X = \mu + \varepsilon$,

where $\varepsilon \sim N(0, \sigma^2)$ and $\mu$ is a discrete distribution with $P[\mu=\nu_k]=W_k$. The mean and variance of $\mu$ are:

$E\mu = \sum W_k \nu_k = \nu$, and

$Var(\mu) = \sum W_k(\nu_k - \nu)^2 = \sum W_k \nu_k^2 - \nu^2 = \tau^2$.

In other words, conditional on a patient being taken from $\Omega_k$, the response X is normal with mean $E(X|\Omega_k) = \nu_k$ and variance $\sigma^2$.

Unconditionally, $EX = E(\mu+\varepsilon) = \sum W_k \nu_k = \nu$ and $Var(X)= \sigma^2 + \tau^2$.

Note that unconditionally, X is a mixture of normal variables and does NOT follow a normal distribution.

Also note that, under CREM, $Var(\overline{X}_k)= \frac{\sigma^2}{N_k} + \tau^2$.

Under DREM , $Var(\overline{X}_k) = \frac{\sigma^2+\tau^2}{N_k}$.

In practice, the weights $\{W_k\}$ are estimated by the sample proportions $\{w_k=N_k/N\}$. In the large sample case, $\{w_k=N_k/N\}$ are consistent estimates of $\{W_k\}$. The overall sample mean

$\overline{\overline{X}} = \frac{\sum_{k=1}^{K} N_k \overline{X}_k}{N} = \sum_{k=1}^{K} w_k \overline{X}_k$ is approximately $N(\nu, \frac{\sigma^2+\tau^2}{N})$ by CLT.

To test $H_0 : \nu = 0$ versus $H_A : \nu > 0$, the statistical test is

$Z = \frac{\overline{\overline{X}}}{\sqrt{(\sigma^2+\tau^2)/N}}$ with rejection region $Z \geq z_\alpha$.

Note that under the fixed effect model, $\overline{\overline{X}} \sim N(\nu, \frac{\sigma^2}{N})$ and the statistical test is $Z_F = \frac{\overline{\overline{X}}}{\sqrt{\sigma^2/N}}$ with rejection region $Z_F \geq z_\alpha$.

## Variance components $\tau^2$ and $\sigma^2$

Consider the pooled sample responses from the K regions together

$\{X_{11}, X_{12}, \ldots X_{k1}, X_{k2}, \ldots, X_{K1}, X_{K2}\ldots\}$ .

We may estimate $\sigma^2$ from the K regions with

$\hat{\sigma}_k^2 = \dfrac{\sum (X_{ki} - \bar{X}_k)^2}{N_k - 1}$, then use the weighted average to get

$\hat{\sigma}^2 = [\sum (N_k - 1)\sigma_k^2]/(N - K)$.

In data analysis, the fixed effect approach "ignores" the component $\tau^2$ since, under the fixed effect model, $\tau^2 = 0$. (Is this appropriate?)

As an alternative, we may use the following formulas to estimate the variance components of variance for $\overline{\overline{X}}$:

$$\widehat{\sigma^2+\tau^2} = \sum_k \sum_j (X_{kj} - \overline{\overline{X}})^2 / (N-1).$$

$$\widehat{\sigma}^2 = [\sum (N_k - 1)\sigma_k^2]/(N - K).$$

$$\widehat{\tau^2} = \widehat{\sigma^2+\tau^2} - \widehat{\sigma^2}.$$

(Stratified analysis)

Consider the following parametrization:

Express the treatment effect for Region k as

$\nu_k = \nu + \delta_k$, k = 1,…, K, with $\sum\delta_k = 0$.

Then $(\sum\nu_k)/K = [\sum(\nu + \delta_k)]/K = \nu$. ➔

The overall treatment effect $\nu$ = the unweighted average of $\{\nu_k\}$.

What if the within group variances are not the same?

It can be shown that if the within group variances are $\{\sigma_k^2\}$, then to test $H_0 : \nu = 0$ versus $H_A : \nu > 0$, the statistical test is

$$Z = \frac{\overline{\overline{X}}}{\sqrt{(\sigma^2 + \tau^2)/N}}$$ , where $\sigma^2 = \sum_{k=1}^{K} W_k \sigma_k^2$, and the rejection region is $Z \geq z_\alpha$.

Design of a trial under DREM depends upon the value of $\sigma^2+\tau^2$. If $\{\nu_k\}$ are between a and b, then $\tau^2 \leq (b-a)^2/4$. Use this upper bound to replace $\tau^2$ will lead to a conservative estimate of sample size. (Lan-Pinheiro-Chen JBS, 2015)

From one-sample to two-sample

$\sigma^2$ becomes $2\sigma^2$.

Under fixed effect model, $\mathrm{Var}[\overline{X}_{Tk} - \overline{X}_{Ck}] = 2\sigma^2(1/N_{Tk} + 1/N_{Ck})$.

Under DREM, $\mathrm{Var}[\overline{X}_{Tk} - \overline{X}_{Ck}] = (2\sigma^2 + \tau^2)(1/N_{Tk} + 1/N_{Ck})$.

DREM can be applied to comparisons of two proportions:

Difference, Relative Risk, Odds Ratio.

For survival data analysis using the logrank test, the evaluation of power under large sample theory will be slightly anti-conservative.

# Definition of Consistency: Examples

1. $P[\widehat{\nu}_k / \widehat{\nu} \geq 0.5] \geq 0.8,$ for all k or for a specific k.
2. $\widehat{\nu}_k > 0$ for all k.

Reference:

*Ministry of Health, Labour and Welfare of Japan (MHLW). Basic principles on global clinical trials, 2007.

*Quan et al 2010 Drug Information Journal.

*Liu, Tsou et al (2015) submitted for publication.

There is NO simple rule to describe all potential cases for consistency/inconsistency.

Treatment effects and "Belief" --- A review

K regional treatment effects $\nu_1, \nu_2, \ldots, \nu_K$.

How do we MODEL OUR BELIEF of treatment effects in K regions?

Assume all treatment effects are the same:

1. Fix effect model, all regional treatment effects are the same.

   $\nu_1 = \nu_2 = \ldots = \nu_K = \nu$

2. Continuous Random Effects Model (CREM).

Unconditionally, $\nu_1, \nu_2, \ldots, \nu_K$ are iid $N(\nu, \tau^2)$.

Conditionally, they could be different, but deviations come from random noise.

Example: Regions 1, 2, 3; observed treatment effects 0.4, 0.7, 0.3.
Sample sizes are equal in these 3 regions

| | Region 1 | Region 2 | Region 3 |
|---|---|---|---|
| $\nu_k$ | 0.4 | 0.7 | 0.3 |

Overall treatment effect estimate: (0.4 + 0.7 + 0.3)/3 = 0.45.

What are the treatment effects if we repeat this MRCT?

Reference for CREMs:

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–88.

## 3. DREM (Discrete Random Effects Model)

The treatment effects $\nu_1, \nu_2, \ldots, \nu_K$ are constants (could be different),

and a patient's response $X = \mu + \varepsilon$ where $\mu$ and $\varepsilon$ are independent,

$\varepsilon \sim N(0, \sigma^2)$, $\mu$ is multinomial with $P(\mu = \nu_k) = w_k = N_k / \sum N_j$; $k = 1, \ldots, K$.

Example:

| Region | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Sample size | 400 | 300 | 100 | 200 |
| Weight = $w_k$ | 0.4 | 0.3 | 0.1 | 0.2 |
| Treatment effect | 0.3 | 0.1 | 0.5 | 0.2 |

Overall treatment effect $\nu = 0.4x0.3+0.3x0.1+0.1x0.5+0.2x0.2=0.24$.

$X = \mu + \varepsilon$ where $\mu$ and $\varepsilon$ are independent,

$\varepsilon \sim N(0, \sigma^2)$, $\mu$ is multinomial with $P(\mu = \nu_k) = w_k = N_k / \sum N_j$; $k = 1,\ldots,K$.

Let $\nu = \sum w_k \nu_k = E(\mu)$ and $\tau^2 = \sum w_k(\nu_k - \nu)^2 = Var(\mu)$.

$EX = \nu$, $Var(X) = \sigma^2+\tau^2$

= within region variation + between region variation.

How to estimate the overall treament effect $\nu$?

$N = N_1 + \ ...+N_K$.

Fixed effect model: $\widehat{\nu} = \sum_1^K \frac{N_k}{N} \widehat{\nu}_k$; (weigh by sample size)

CREM: Reciprocal of variance, pushes weights to (1/K, 1/K,...,1/K).

DREM: $\widehat{\nu} = \sum_1^K \frac{N_k}{N} \widehat{\nu}_k$.

Select-the-winner and Drop-the-min

Select-the-winner

$X_1, X_2, \ldots, X_K$ iid $N(\nu,1)$. $\{X_{(1)} < X_{(2)} < \ldots < X_{(K)}\}$

$\text{Max}\{X_k\} = X_{(K)}$, what is the distribution of $X_{(K)}$?

Drop-the-min $\{X_{(1)} < X_{(2)} < \ldots < X_{(K)}\}$

How do we estimate treatment effect $\nu$ from $X_{(2)}, \ldots, X_{(K)}$?

Mathematically, these two approaches are the same when K=2.

When K = 2. Let $X_1$, $X_2$ be iid N(0,1).

Define $X_{(1)} = \min \{X_1, X_2\}$ and $X_{(2)} = \text{Max } \{X_1, X_2\}$

$E[X_{(2)}] = 0.5642$ (bias); Variance $= 1 - 0.5942^2 = 0.6817$.

In other words, $X_{(2)}$ is a biased estimate of the original mean 0.

$X_{(2)}$ - 0.5642 becomes unbiased.

Variance of $X_{(2)} = 0.6817 < \text{Var}(X_1) = \text{Var}(X_2) = 1$.

$$Z^* = \frac{X_{(2)} - \text{bias}}{\sqrt{\text{Variance of } X_{(2)}}}$$ has a skew normal distribution.

$Z^* \approx N(0,1)$. (Shun et al 2008 SIM)

Change of location parameter (Under fixed effect model)

Let $X_1$, $X_2$ be iid $N(\nu,1)$.

Define $X_{(1)} = \min\{X_1, X_2\}$ and $X_{(2)} = \text{Max}\{X_1, X_2\}$

$E[X_{(2)}] = \nu + 0.5642$ (bias); Variance $= 1 - 0.5942^2 = 0.6817$.

$$Z^* = \frac{X_{(2)} - \text{bias}}{\sqrt{\text{Variance of } X_{(2)}}}$$ has a skew normal distribution.

$Z^* \approx N(\nu,1)$.

Change of scale parameter

Let $X_1$, $X_2$ be iid $N(\nu,\eta^2)$.

Define $X_{(1)} = \min\{X_1, X_2\}$ and $X_{(2)} = \text{Max}\{X_1, X_2\}$

$E[X_{(2)}] = \nu + 0.5642\eta$ (bias); Variance = $0.6817\,\eta^2$.

That is, Bias = $0.5642\eta$, Variance = $0.6817\eta^2$.

For example, if the two regionals have the same means $\nu$ and the sample means $X_1$, $X_2$ are iid $N(\nu, \eta^2 = .0001)$. ($\eta$=1/100)

Then $X_{(2)} - 0.5642\text{x}0.01 = X_{(2)} - 0.005642$ is an unbiased estimate of $\nu$ with variance 0.6817 x .0001 = 0.00006817.

$$Z^* = \frac{X_{(2)} - \text{bias}}{\sqrt{\text{Variance of } X_{(2)}}}$$ has a skew normal distribution.

Had we started with $X_1, X_2$ iid $N(\nu, 1)$, then the numerator $(X_{(2)} - \text{bias})$ provides a unbiased estimate of $\nu$ and $Z^* \sim N(\nu, 1)$.

95% confidence interval for $\nu$ :

$$X_{(2)} - (\text{bias}) \pm 1.96 \sqrt{\text{Variance of } X_{(2)}} \ .$$

DREM: (K=2) Use observed sample means as treatment effects.

Let $X_1 \sim N(\nu_1,\eta^2)$ and $X_2 \sim N(\nu_2,\eta^2)$ be independent.

Assume $\nu_1 < \nu_2$.

Define $X_{(1)} = \min \{X_1, X_2\}$ and $X_{(2)} = \text{Max } \{X_1, X_2\}$.

$E[X_2 \mid X_{(2)} = X_2] = \nu_2 +$ (bias);

Variance = $\text{Var}[X_2 \mid X_{(2)} = X_2]$.

We can use (numerical integration or) simulation to find the bias and variance.

The 95% confidence interval for the mean $\nu_2$:

$$X_2 - (\text{bias}) \pm 1.96 \sqrt{\text{Var}\left[ X_2 \mid X_{(2)} = X_2 \right]} .$$

$R=3,\ N_1 = N_2 = N_3 = 1300;$

$\hat{\nu}_1 = 0.095;\ \hat{\nu}_2 = 0.09;\ \hat{\nu}_3 = 0.01.\ \mathrm{Var}(\hat{\nu}_1) = \mathrm{Var}(\hat{\nu}_2) = \mathrm{Var}(\hat{\nu}_3).$

Is there any inconsistency in this specific example?

Note that all observed treatment effects are positive, and overall treatment effect is “significantly larger than 0”.

What if we drop region 3?

Inconsistency example: R=3, $N_1=N_2=N_3=1300$;

$\hat{\nu}_1=0.095$; $\hat{\nu}_2=0.09$; $\hat{\nu}_3=0.01$. $\mathrm{Var}(\hat{\nu}_1) = \mathrm{Var}(\hat{\nu}_2)=\mathrm{Var}(\hat{\nu}_3)$.

1. Drop the third region from the MCRT, but

   estimate treatment effect ~ (0.095+0.09+0.01)/3 = 0.065.

2. Drop the third region from the MCRT, and

   estimate treatment effect ~ (0.095+0.09)/2 = 0.0925.

3. Modify approaches 1 and 2. HOW to modify data analysis for dropping the minimum? ←

Treatment = (0.095, 0.09, 0.01), N = (1300, 1300, 1300)

| | Fixed | DREM |
|---|---|---|
| v | 0.0625 | 0.0925 |
| √Var | 0.0194 | 0.0192 |

In general, regional sample sizes may be different; and K could be any number $\geq 2$.

Numerical integration is too complex, use simulations to evaluate Bias and Variance.

References:

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–88.

Ministry of Health, Labour and Welfare of Japan (MHLW). Basic principles on global clinical trials, 2007.

Lan KKG, Pinheiro J. Combined estimation of treatment effects under a discrete random effects model. *Statistics in Biosciences* 2012; **4**:235–244.

Lan KKG, Pinheiro J and Chen F . Designing multiregional trials under the discrete random effects model. To appear in *Journal of Biopharmaceutical Statistics*, 2015.

Hung HMJ, Wang SJ, O'Neill RT. Consideration of regional difference in design and analysis of multi-regional trials. *Pharmaceutical Statistics* 2010; **24**:173–178. DOI: 10.1002/pst.440.

Wang SJ, Hung HMJ. Ethnic sensitive or molecular sensitive beyond all regions being equal in multiregional clinical trials. *Journal of Biopharmaceutical Statistics* 2012; **22**:879-893

Chen CT, Hung HMJ, Hsiao CF. Design and evaluation of multiregional trials with heterogeneous treatment effect across regions. *Journal of Biopharmaceutical Statistics* 2012; **22**:1037–1050.

Tanaka Y, Li G, Wang Y, Chen J. Qualitative consistency of treatment effects in multiregional clinical trials. *Journal of Biopharmaceutical Statistics* 2012; **22**:988–1000.

Chen J, Quan H, Binkowitz B, Ouyang SP, Tanaka Y, Li G, Menjoge S, Ibia E, for the Consistency Workstream of the PhRMA MRCT Key Issue Team. Assessing consistent treatment effect in a multi-regional clinical trial: a systematic review. *Pharmaceutical Statistics* 2010; **9**:242–253.
DOI: 10.1002/pst.438

Quan H, Li M, Chen J, Gallo P, Binkowitz B, Lbia E, Tanaka Y, Ouyang SP, Luo X, Li G, Menjoge S, Talerico S, Ikeda K. Assessment of consistency of treatment effects in multiregional clinical trials. *Drug Information Journal* 2010; **44**:617–632

Ko FS, Tsou HH, Liu JP, Hsiao CF. Sample size determination for a specific region in a multi-regional trial. *Journal of Biopharmaceutical Statistics* 2010; **24**: 870–885.

Hui Quan and Joshua Chen (Editors). "Multi-Regional Clinical Trials for Simultaneous Global New Drug Development", to be published by Chapman & Hall in 2016.

Liu JT, Tsou HH et al (2015). Assessing the Consistency of the Treatment Effect under the Discrete Random Effects Model in Multiregional Clinical Trials. Submitted for publication.

# Backup Slides

Comments on “random sampling”

Suppose we conduct a clinical trial on treatment T (versus C) with 2 centers, one in California and another in Wisconsin.

Patients are NOT random samples from the US population. How could we extrapolate the observed treatment effect to the US population?

From California-Wisconsin to US population, treatment effect extrapolation is a medical/biological judgment.

Same for MRCT?

What if, after data have been collected, we use bootstrap (re-sample) to estimate the overall treatment effect?

Approximately, we take sample from Region k with probability $W_k = N_k/N$. DREM is just a smoothed version of the bootstrap.

$$\left( \quad N_1 \qquad\qquad N_2 \qquad\qquad N_3 \quad \right) \leftarrow \text{Sample Space}$$

$$R_1, \nu_1, N_1 \qquad R_2, \nu_2, N_2 \qquad R_3, \nu_3, N_3$$

# Thank you!