Alternative approach to evaluate treatment effects of local regions with small sample size in MRCT

Yosuke Fujii, Ami Takata, Naoki Isogawa and Shintaro Hiro
Clinical Statistics, Development Japan, Pfizer Japan Inc.

1. Introduction

In a multi-regional clinical trial (MRCT), some local regulatory authorities are interested in not only an overall treatment effect but also local treatment effects to assess if the overall result can be extrapolated to the local countries. A local treatment effect is generally estimated with subgroup analysis of the relevant region. However, low accuracy issue is sometime encountered due to small sample size. Recent MRCTs are conducted in 10-20 countries. Therefore, it is a challenging issue to ensure local sample size for keeping adequate estimation accuracy. A “borrowing information” approach has been recently proposed and it can be a good solution for this issue. Draft ICH E17 guideline also states that “If the sample size in a region is so small that the estimates of effect are unreliable, the use of other methods should be considered, including the search for options to pool regions based on commonalities, or borrowing information from other regions or pooled regions using an appropriate statistical model”. The proposed approach can be one of key approaches for the issue regarding local treatment estimates for small sample size countries. We introduce two methods in our report.

Chen, Y.H et al. (2009) and Quan, H et al. (2013) proposed to estimate local treatment effects assuming equal inter-regional variability and the local estimator is a weighted sum of the overall estimate and the crude local estimate. Guo, H et al. (2015) proposed a tree-based approach considering the similarities among countries. A hierarchal tree to determine similarities among countries is built based on certain characteristics at first. Thereafter the local treatment effect is estimated by borrowing more information from similar countries based on the tree. Guo, H et al. used Human Development Index (HDI) which assessed development results of countries in a clustering process for providing a hierarchical tree, but we attempted to use patient demographics and other baseline characteristics data in a clinical trial.

In our report, we apply the traditional method and two methods based on borrowing information to a simulated data generated by random sampling from an actual clinical trial data and compare the results, and also evaluate the usefulness of the “borrowing information” approach.

2. Method

2.1. Traditional Method

Simple subgroup analysis is the easiest and common way to evaluate the local treatment effect in each country.

Treatment effect in country \( k \) is described with raw estimates of treatment difference as follows:

\[
\theta_k = E[d_k] = E[x_k] - E[y_k].
\]

where \( x \) is data from active group, \( y \) is data from placebo group and \( k \) is country \( (k = 1, \ldots, N) \).

We often compare the location of each treatment estimate among countries.

2.2. Method 1: Borrowing Information from an overall result

The crude estimate of treatment deference at country \( k \), \( d_k \), is assumed to follow the normal distribution:

\[
d_k \sim N(\theta_k, s_k^2)
\]
$\theta_k$ is the true treatment effect at country $k$, and $s_k^2$ is a variance of $d_k$. If the true treatment effect $\theta_k$ can be assumed to be given by random sampling from following prior distribution:

$$\theta_k \sim N(\mu, \omega^2),$$

then it is regarded as a hierarchical model (Figure 1(i)). $\mu$ and $\omega^2$ are unknown parameter. Here, $\mu$ is an overall treatment effect and $\omega^2$ is inter-country variance of the true treatment effect $\theta_k$.

Given the treatment difference $d_k$, a posterior distribution of the true treatment effect $\theta_k$ will be described as following distribution based on Bayesian estimation.

$$\theta_k | d_k \sim N(B_k \mu + (1 - B_k) d_k, (1 - B_k) s_k^2)$$

$$B_k = s_k^2 / (s_k^2 + \omega^2)$$

The more homogeneous the treatment differences across the different countries, the larger the degree of utilizing the overall information ($\mu$) for local treatment evaluation.

### 2.3. Method 2: Borrowing Information from similar countries and an overall result

This method considers similarities among countries into treatment effect estimation for individual countries. Each country borrows more information from similar countries and less from dissimilar countries. The degree of the similarities among countries is provided as a hierarchical tree by a hierarchical clustering method (Figure 1(ii)). Our study attempted to use patient demographics and other baseline characteristics data in a clinical trial as a more practical way in the clustering process while Guo et al. used HDI which assessed development results of countries.

Suppose there are 4 countries where country 1 and 2 are similar and have closer relationship relative to country 3 and 4 (Figure 2).

True treatment effects of each country are $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$. $\theta_1$ and $\theta_2$ share a parent $\theta_3 \theta_4$ and $\theta_3$ have a common parent $\theta_6$. The similarities among countries are described as branch lengths between counties/clusters with $e = (e_1, e_2, e_3, e_4, e_5)$ based on clustering method with demographic/baseline data.

The raw estimates of treatment effects, $d = (d_1, d_2, d_3, d_4)$, given the true treatment effects, follow

$$\begin{pmatrix}
\theta_1 \\
\theta_2 \\
\theta_3 \\
\theta_4
\end{pmatrix} \sim N
\begin{pmatrix}
\theta_1 \\
\theta_2 \\
\theta_3 \\
\theta_4
\end{pmatrix}
\begin{pmatrix}
V_1 & V_2 & V_3 & V_4
\end{pmatrix}
$$

$$V = 
\begin{pmatrix}
V_1 & 0 & 0 & 0 \\
0 & V_2 & 0 & 0 \\
0 & 0 & V_3 & 0 \\
0 & 0 & 0 & V_4
\end{pmatrix}
$$

Further, assume that $\theta_1$, $\theta_2$ given the parent $\theta_5$, and $\theta_3$, $\theta_4$, $\theta_5$ given the parent $\theta_6$ follow

$$\begin{pmatrix}
\theta_1 \\
\theta_2
\end{pmatrix} | \theta_5 \sim N
\begin{pmatrix}
\theta_5 \\
\theta_5
\end{pmatrix}
\begin{pmatrix}
e_1 & 0 \\
0 & e_2
\end{pmatrix}
$$

$$\begin{pmatrix}
\theta_3 \\
\theta_4
\end{pmatrix} | \theta_5 \sim N
\begin{pmatrix}
\theta_5 \\
\theta_5
\end{pmatrix}
\begin{pmatrix}
e_3 & 0 \\
0 & e_4
\end{pmatrix}
$$

$$\tau \text{ is a scale parameter. True treatment effects are described as follows:}

$$\begin{pmatrix}
\theta_1 \\
\theta_2 \\
\theta_3 \\
\theta_4
\end{pmatrix} | \theta_6 \sim N
\begin{pmatrix}
\theta_6 \\
\theta_6 \\
\theta_6 \\
\theta_6
\end{pmatrix}
\begin{pmatrix}
e_1 & e_5 & e_1 + e_5 & 0 \\
e_5 & e_2 + e_5 & 0 & 0 \\
e_2 & e_5 & 0 & 0 \\
e_4 & 0 & e_3 & 0
\end{pmatrix}
$$

Unconditionally, the raw estimates follow $d \sim N(\theta_k \mathbf{1}_{4 \times 1}, \mathbf{T} + \mathbf{V})$.

Assume $\tau$, $\theta_6$ and $s_k^2$ are known, given the data of posterior estimate of $\theta$ are weighted average
of raw treatment effects in similar countries and overall treatment effects as follows:

\[
\begin{bmatrix}
\delta_1 \\
\delta_2 \\
\delta_3 \\
\delta_4
\end{bmatrix}
\begin{bmatrix}
d \\
\theta_6 \\
\tau \\
T \\
V
\end{bmatrix}
= \frac{w_{11}d_1 + w_{12}d_2 + w_{13}\theta_6}{w_{21}d_1 + w_{22}d_2 + w_{23}\theta_6}
\]

where \( \delta = \left( \frac{1}{\tau^2} + \frac{1}{\tau e_3} \right)^{-1} \left( \frac{d_3}{\tau^2} + \frac{\theta_6}{\tau e_3} \right) \) and \( \Delta = (e_1 + e_2)(e_2 + e_3) - e_3^2; w_{21} = \frac{1}{\delta} \left( \frac{1}{\tau^2} + \frac{e_3}{\tau e_4} \right); \\
w_{12} = \frac{e_1}{\tau e^2_4}; w_{13} = \frac{1}{\delta} \left( \frac{e_2}{\tau^2} + \frac{e_1 e_3}{\tau^2} + \frac{e_1 e_3}{\tau^2} \right)
\]

\( w_{11} + w_{12} + w_{13} = 1; w_{21} + w_{22} + w_{23} = 1 \)

3. Results/Conclusions

We evaluated these methods using a simulated data (incl. demographic data for Method 2) generated from random sampling of an actual clinical trial data. A total of 1000 patients were assumed to be enrolled from 22 countries in a MRCT. In the simulated study, it was assumed that the treatment effect was a difference between active group and placebo group in effective rate calculated by a binary response variable (Yes/ No). An overall result in the primary endpoint was significant. The overall and local treatment estimates by country in traditional method were provided as Figure 3(a). Number of subjects varied widely among countries. In some countries the treatment effect was not estimable due to absence of subjects in either active group or placebo group.

A result of Method 1 was shown in Figure 3(b). All estimates including both not-estimable and completely negative results in the traditional method shrank toward the overall results. 95% credible interval in each country became shorter than 95% confidence interval of the traditional method. The estimated treatment effects in countries with large population such as US and Russia were similar to those of the traditional method because these local estimates had a large influence on the overall estimate.

In the clustering analysis to identify similar countries in Method 2, Ward’s linkage with Euclidean distance was applied. Summary statistics of weight, sex, race, age, height, pre-drug treatment and pre-nondrug treatment were used to construct the cluster. Figure 4 was the clustering result. According to CCC (Cubic Clustering Criterion) which was an index to estimate the optimal number of clusters, three clusters surrounded by squares were obtained. The clusters were defined objectively, however, the result was also clinically interpretable. Treatment effects estimated in Method 2 were provided as Figure 3(c). Compared to the results in Method 1, 95% credible intervals in many countries were shorter. It was considered that variability within similar
countries defined by the clustering would be smaller than the variability across all countries. Larger treatment effects were observed in some Asian countries described as the right cluster in Figure 4 compared to the other countries.

It was suggested that the borrowing information approach introduced as Method 1 and Method 2 improved accuracy of local treatment estimates with small sample size.

Figure 3. Treatment estimates in each method (■: Point estimate, Line: 95% Confidence Interval for (a), 95% Credible Interval for (b) and (c))

Figure 4. Clustering result (Square: similar countries as a cluster, Triangles: nodes, Line: a cut point for clusters)

References