Estimation of treatment effect in a multi-regional clinical trial with survival endpoint

Hsiao-Hui Tsou[1,2,*], Yu-Chieh Cheng[1], Eric Slud[3,4], Yuh-Jeng Wu[5], Chin-Fu Hsiao[1,6], Hsiao-Yu Wu[1], Ya-Ting Hsu[1], Fang-Jing Lee[7]

[1]Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County, Taiwan

[2]Graduate Institute of Biostatistics, College of Public Health, China Medical University, Taichung, Taiwan

[3]Department of Mathematics, University of Maryland, College Park, Maryland, U.S.A.

[4]Center for Statistical Research and Methodology, US Census Bureau, Washington DC, U.S.A.

[5]Department of Applied Mathematics, Chung Yuan Christian University, Taoyuan, Taiwan

[6]Division of Clinical Trial Statistics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County, Taiwan

[7]National Institute of Infectious Diseases and Vaccinology, National Health Research Institutes, Miaoli County, Taiwan

*Correspondence to: Hsiao-Hui Tsou, Institute of Population Health Sciences, National Health Research Institutes, 35 Keyan Road, Zhunan, Miaoli County 350, Taiwan.

E-mail: tsouhh@nhri.edu.tw

Abstract

Multiregional clinical trial (MRCT) has become a popular strategy in the development of new medicines. By incorporating subjects from many regions under a single protocol, an MRCT seeks regulatory approval for all participating regions. Therefore, the first goal in an MRCT is to show the overall treatment efficacy of the new therapy. In this presentation, we focus on the design and analysis of a two-arm comparative multiregional clinical trial with survival endpoint. We provide a statistical model to combine regional treatment effects for estimation of overall treatment effect. ICH E17 guidance provides some approaches to sample-size allocation in an MRCT. Thus, we evaluate those approaches by numerical examples in the presentation. The evidence of consistency in treatment effects among regions is usually required for regional approval. We further explore how to evaluate consistency in treatment effects among regions and how to make inferences of the treatment effect for a specific region.

Key Words: Multiregional clinical trials, survival endpoint, overall treatment effect,

sample-size allocation, consistency

1. Introduction

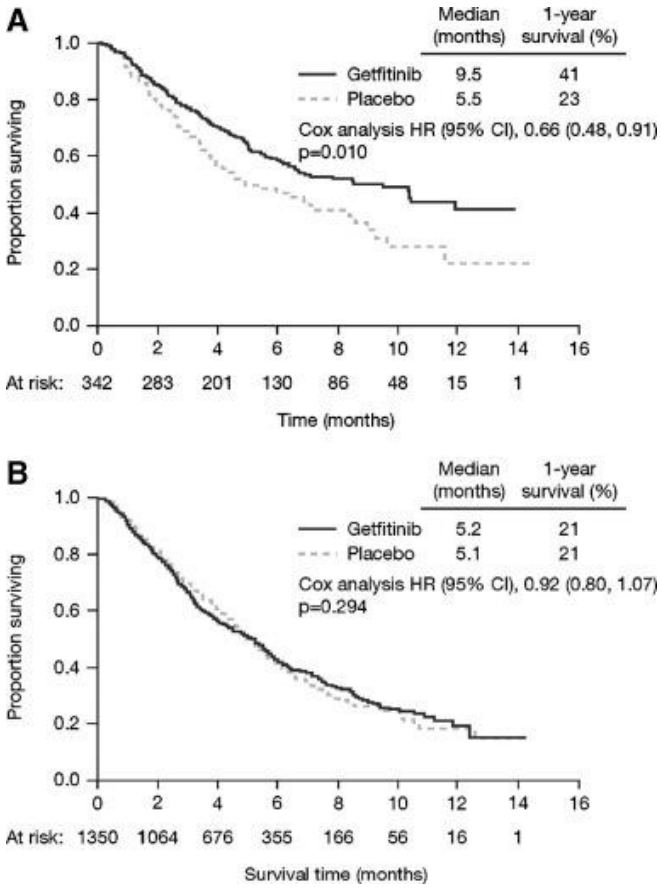
From the beginning of 21st century, multi-regional clinical trial (MRCT) has become a popular strategy for drug development. If we follow traditional clinical drug development workflow, usually a new drug (test drug) will be first submitted to a region A for seeking approval in the region A. After several years, the new drug will be submitted to another region (say region B) and try to get approval in the region B. However, it will cause a problem of drug Lag if all new drugs were developed by the traditional clinical drug development workflow. In order to resolve problem of drug Lag, International Council on Harmonisation (ICH) have published some guidance for multi-regional clinical trial. The first guidance is ICH E5: Ethnic Factors in the Acceptability of Foreign Clinical Data and the second guidance is ICH E17: General Principal for Planning & Design of Multi-Regional Clinical Trials. So what is an MRCT? According to both guidance, a multi-regional trial is a clinical trial conducted in more than one region under a common protocol. One of the key words in ICH E5 11th Q&A is hierarchy of persuasiveness. If we can get statistical significance in overall result and also get statistical significance in a region of interest, then these results are the most persuasive results. However, if the region of interest could not reach statistical significance but we can show consistent trends across regions, then these results are still persuasive results for an MRCT.

In this article, we focus on the design and analysis of a two-arm comparative multiregional clinical trial with survival endpoint which is an very important endpoint for phase III oncology trial. We were interested in what is the overall treatment effect and how to evaluate the consistency of treatment effects across regions.

2. What is the overall treatment effect?

Traditionally, an MRCT would assume a fixed effect model (FEM), that is, equal treatment effects across regions ($\theta_1 = \theta_2 = \ldots = \theta_k$). However, regional heterogeneity in MRCTs has been observed and may have an impact on the estimation of treatment effect. Some insightful articles such as Hung et al. 2010; Wang & Hung, 2012 have discussed these issues. A famous example for patients with non-small cell lung cancer is IRESSA (Chang A et al. 2006)**.** In this example, test drug IRESSA is compared to a Placebo control group in a two-arm parallel design (see figure1A). Hazard ratio (HR) for the IRESSA compared to the Placebo is 0.66 for the Asian population. However,

for the non-Asian population, the HR is not statistical significate (figure1B). Thus, in this article, we take the regional heterogeneity into consideration to design an MRCT.



Source: Chang A, Parikh P, Thongprasert S, et al. Gefitinib (IRESSA) in patients of Asian origin with refractory advanced non-small cell lung cancer: subset analysis from the ISEL study. Journal of Thoracic Oncology. 2006;1(8):847-55.

Figure 1: Ethnic difference for patients with NSCLC. (A) patients of Asian (B) patients of non-Asian.

3. Consider the survival endpoint in the MRCT

We consider a two-arm parallel MRCT consist of K disjoint clinical regions R_1, R_2, …, R_K. The primary endpoint is overall survival (OS). Assume S_0 is accrual time and S-S_0 is follow-up time. For each region, patients are allocated to a test product T or a placebo control C with a 1:1 allocation ratio within the region.

Let λ_{Ck} and λ_{Tk} be the hazard rate in the control group and treatment group in region *k*. We are interested in testing whether overall treatment effect in the treatment group is better than that in control group. Thus, the hypotheses are written as

$$\mathrm{H}_0:\sum_{k=1}^{K} w_k \log(\frac{\lambda_{Ck}}{\lambda_{Tk}})=0 \text{ vs } \mathrm{H}_\mathrm{A}:\sum_{k=1}^{K} w_k \log(\frac{\lambda_{Ck}}{\lambda_{Tk}})>0 \quad (1)$$

where

$$\log \lambda_{tk} = \mu_c + \alpha_k + \theta \mathrm{I}_{[t=\mathrm{T}]}(1+\varepsilon_k) \quad (2)$$

$t = C, T, k = 1,\ldots, K$, $\alpha_k \sim N(0,\sigma_a^2)$ and $\varepsilon_k \sim N(0,c^2)$. μ_c is called overall main effect for the control group.

The test statistic for region k is defined as

$$Z_k = \frac{\log(\hat{\lambda}_{Ck}/\hat{\lambda}_{Tk})}{\sqrt{(1/D_{Tk}+1/D_{Ck})}} \quad (3)$$

The test statistic for testing (1) is defined as

$$Z = \frac{\sum_{k=1}^{K} w_k \log(\frac{\hat{\lambda}_{Ck}}{\hat{\lambda}_{Tk}})}{\sqrt{\sum_{k=1}^{K} w_k^2\left[\frac{1}{D_{Tk}}+\frac{1}{D_{Ck}}\right]}} \quad (4)$$

where D_{Tk} and D_{Ck} be the observed number of events for treatment group (T) and control group (C).

4. Power function for benefit and sample-size determination

We consider a power function for benefit at $\theta = \frac{b}{\sqrt{N}}$ as bellow

$$PB = \mathrm{P[Benefit]} = \mathrm{P}(Z > z_{1-\alpha} \mid \theta = \frac{b}{\sqrt{N}}, \boldsymbol{\lambda_C}, \boldsymbol{\lambda_T}) \quad (5)$$

Sample size is determined such that PB = 80%.

4.1 Consistency assessment

After showing the overall treatment effect, it is interesting to evaluate the possibility of applying the overall trial results to each region. According to ICH E17, a structured exploration to examine the consistency of treatment effects across regions and subpopulations should be planned. To evaluate whether the overall treatment effect can

be applied to subjects from participating regions, we need to consider consistency assessment and sample size allocation among regions.

4.2 Sample size allocation

According to ICH E17, sample size to regions or pooled regions should be determined such that clinically meaningful differences in treatment effects among regions can be described, without substantially increasing the sample size requirements based on the primary hypothesis. Sample size allocation may take the "Consistency assessment" into account. The Japanese Ministry of Health, Labour and Welfare (MHLW) proposed a guideline "Basic Principles on Global Clinical Trials" in 2007 for the planning and implementation of global clinical studies. The guideline recommends two methods, M1 and M2, to address issues related to establishing efficacy in a specific region and consistency in efficacy among regions.

M1: $\hat{\theta}_k > \rho\hat{\theta}$ for region of interest k.

M2: $\hat{\theta}_1 > 0,\ \hat{\theta}_2 > 0, \ldots,\ \hat{\theta}_K > 0,$

In this article, we determine sample size to regions based on the consistency method M2. We define the probability of benefit and consistency in next section for determining sample size to regions in an MRCT.

4.3 Power function of benefit and consistency

The probability of benefit and consistency (PBC) is defined as

$$PBC = \mathrm{P}(Z > z_{1-\alpha},\ Z_k > 0, \text{ for all k, k=1,2,...,K} \mid \theta = \frac{b}{\sqrt{N}},\ \boldsymbol{\lambda_C},\ \boldsymbol{\lambda_T}) \quad (6)$$

Table 1: The required sample size with λ_C =0.8, ($\lambda_{T1}, \lambda_{T2}, \lambda_{T3}$) =(0.665, 0.665, 0.59), $\sigma_a = c = 0.1$

w_1	w_2	w_3	θ	b	N	D	Censoring rate	PB^*	$PBC^{\#}$
0.1	0.1	0.8	-0.28	-6.64	563	388	0.311	0.8035	0.5745
0.15	0.15	0.7	-0.27	-6.71	617	426	0.310	0.8015	0.6355
0.2	0.2	0.6	-0.26	-6.72	668	462	0.308	0.8020	0.686
0.25	0.25	0.5	-0.24	-6.62	761	529	0.305	0.8010	0.7255
0.3	0.3	0.4	-0.23	-6.7	848	593	0.301	0.8030	0.7405
0.35	0.35	0.3	-0.22	-6.55	887	622	0.299	0.8005	0.736
0.4	0.4	0.2	-0.21	-6.53	968	679	0.299	0.8015	0.725
0.45	0.45	0.1	-0.2	-6.56	1076	757	0.296	0.8015	0.666

*: empirical *PB*.

#: empirical *PBC*.

It shows the performance of sample size with given hazard rates of a control and a test groups in all regions. For simplicity, we consider K=3 (three regions participate in an MRCT), and the corresponding regional weights are (w_1, w_2, w_3) with $w_1 = w_2$. We suppose that the hazard rates of the test group are $\lambda_C = \lambda_{C1} = \lambda_{C2} = \lambda_{C3} = 0.8$ and the hazard rates of the control group are (λ_{T1}, λ_{T2}, λ_{T3}) =(0.665, 0.665, 0.59).

Table 1 shows that when w_1 increases, *N* and *D* increase. Under satisfying desired power *PB*=0.8 *PBC* has the maximized value when $w_1 \approx w_2 \approx w_3$ in Table 1. If the sample size *N* is determined for only satisfying the desired power *PB*=0.8, then the probability *PBC* (based on the *N*) may be small.

Table 2：Required sample size *N* (D) is satisfied (1) and (2) with $\lambda_C = 0.8$, $\sigma_a = c = 0.1$, $\alpha = 0.025$, and hazard ratio = 0.8

		γ_1 (PB)				
		0.7	0.75	0.8	0.85	0.9
γ_2 (PBC)	0.6	642(447)	719(499)	838(583)	960(673)	1124(784)
	0.65	666(462)	743(517)	838(581)	963(673)	1130(788)
	0.7	732(509)	752(524)	844(587)	975(681)	1136(794)
	0.75	861(600)	836(582)	856(598)	978(681)	1139(794)
	0.8	963(673)	956(667)	973(678)	984(686)	1148(802)

We consider two criteria, PB= γ_1 and PBC= γ_2, respectively, and given that γ_1 = 0.7, 0.75, 0.8, 0.85, 0.9, and γ_2 =0.6, 0.65, 0.7, 0.75, 0.8. Table 2 shows that the required *N* and corresponding *D* under given γ_1 and γ_2. As seen from Table 2, *N* (*D*) increases with increasing γ_1 or γ_2. If γ_2 increase from 0.6 to 0.8 for a larger γ_1, then *N* increase a little bit. Taking this example a bit further, if γ_2 increase from 0.6 to 0.8 for a fixed γ_1 =0.85, then *N* increases from 960 to 984.

4.4 Approaches to Sample size allocation

According to ICH E17, there is no uniformly acceptable or optimal approach to sample size allocation in an MRCT. Some approaches currently in use include:

i. Proportional Allocation: Allocation of subjects to regions in proportion to size of region/disease prevalence.
ii. Equal Allocation: Allocation of equal numbers of subjects to each region.
iii. Preservation of Effect: Allocation of subjects to one or more regions based on preserving some specified proportion of the overall treatment effect. (M1 consistency)
iv. Local Significance: Allocation of sufficient number of subjects to be able to achieve significant results within each region.

v. Fixed Minimum Number: Allocation of a fixed minimum number of subjects to a region.

In this section, we use the approaches "Preservation of Effect" and "Local Significance" introduced by ICH E17 to further derive additional version of probability of benefit and consistency, such as PBC^1, PBC^2, PBC^3, and AP_2 as follows.

$$PBC^1 = P\left(Z > Z_{1-\alpha}, Z_k > Z_{1-0.1} \text{for all } k,\ k=1,...,3\right), \tag{7}$$

$$PBC^2 = P\left(Z > Z_{1-\alpha}, Z_k > Z_{1-0.05} \text{for all } k,\ k=1,...,3\right), \tag{8}$$

$$PBC^3 = P\left(Z > Z_{1-\alpha}, Z_k > Z_{1-0.025} \text{for all } k,\ k=1,...,3\right) \tag{9}$$

PBC^2 is more stringent than PBC^1. PBC^3 is more stringent than PBC^2. PBC^3 is the most stringent.

$$AP_2 = \left(Z > Z_{1-\alpha},\ \log(\frac{\hat{\lambda}_{Cl}}{\hat{\lambda}_{Tl}}) > \ \rho\ \sum_{k=1}^{K} w_k \log(\frac{\hat{\lambda}_{Ck}}{\hat{\lambda}_{Tk}}) \right), l \text{ is a specific region} \tag{10}$$

Table 3： $\lambda_C = 0.8$ and ($\lambda_{T1}, \lambda_{T2}, \lambda_{T3}$)=(0.64, 0.64, 0.64), N=900, $\theta = -0.22$, b=−6.6, $S = 2$, and $S_0 = 1$

w_1	w_2	w_3	α^*	PB^{**}	$PBC^{\#}$	PBC^1	PBC^2	PBC^3	AP_2
0.1	0.1	0.8	0.0256	0.8156	0.5884	0.1122	0.046	0.013	0.4134
0.15	0.15	0.7	0.0286	0.8092	0.6556	0.154	0.0668	0.0300	0.4048
0.2	0.2	0.6	0.0268	0.8134	0.7052	0.1938	0.0842	0.0346	0.4186
0.25	0.25	0.5	0.0272	0.8124	0.7368	0.236	0.1098	0.0424	0.4022
0.3	0.3	0.4	0.0254	0.8130	0.7526	0.2608	0.1200	0.0498	0.4100
1/3	1/3	1/3	0.0280	0.8160	0.7630	0.2618	0.1234	0.0516	0.4174
0.35	0.35	0.3	0.0230	0.8130	0.7582	0.2552	0.1156	0.0498	0.4074
0.4	0.4	0.2	0.0260	0.8080	0.7300	0.2422	0.114	0.0476	0.4226
0.45	0.45	0.1	0.0208	0.8134	0.6776	0.1854	0.0866	0.0330	0.4134

*: empirical α.
**: empirical *PB*.
#: empirical *PBC*.
1: $P\left(Z > 1.96, Z_k > Z_{1-\alpha_2} \text{for all k, k=1,...,3}\right)$ with $\alpha_2 = 0.1$.
2: $P\left(Z > 1.96, Z_k > Z_{1-\alpha_2} \text{for all k, k=1,...,3}\right)$ with $\alpha_2 = 0.05$.
3: $P\left(Z > 1.96, Z_k > Z_{1-\alpha_2} \text{for all k, k=1,...,3}\right)$ with $\alpha_2 = 0.025$.

Table 3 shows the simulation result of *PB* and *PBC* and provides additional different regional considerations. In Table 3, we can see that PBC^3 is most stringent than others at each combination of regional weight (w_1, w_2, w_3). All of $PBC^{\#}$, PBC^1, PBC^2, and PBC^3 have maximum probability when w_k =1/3.

Table 4：Comparison of proportional allocation and equal allocation

	Proportional Allocation	**Equal Allocation**
Pros	facilitates recruitment by allocating subjects to the regions with the greatest disease burden, and absent other impediments, will generally minimize the time needed to complete enrolment.	has the advantage of optimizing the power available to detect differences in treatment effects between regions for a given overall sample size target.
Cons	The disadvantage is that some regions may end up with too few or no subjects, while other regions may dominate the outcome of the trial.	The disadvantage is that recruitment may be slowed to a possibly unacceptable level, particularly if disease prevalence or ease of recruitment varies substantially among the regions in the MRCT.

According to ICH E17, disadvantages for some approaches are

i. Preservation of Effect: Allocation to preserve a proportion of the overall effect is not practical if many regions in the trial have this requirement.
ii. Local Significance: Allocation based on achieving local significance of regional treatment effects is also not practical, as this strategy may inflate the sample size beyond feasibility and brings into question the concept of conducting an MRCT.
iii. Fixed Minimum Number: Allocating a fixed minimum sample size for regions is not recommended, if there is no scientific justification for selecting the minimum.

Conclusion

We agree with the comments presented in ICH E17 that "A balance between proportional and equal allocation is recommended, to ensure that recruitment is feasible and able to be completed in a timely fashion, but also to provide sufficient information to evaluate the drug in its regional context." Often, non-statistical reasoning — such as cost, timing of each region's joining the trial, speed of patient enrollment, or prevalence of the disease — may constrain regional sample sizes.

Acknowledgements

This work was supported by grants from the National Health Research Institutes, Zhunan, Taiwan [PH-106-PP-02, PH-107-PP-02, PH-108-PP-02] and the Ministry of Science and Technology of Republic of China [MOST 104-2118-M-400-002].

References

International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use. ICH E17: General Principles for Planning and Design of Multi-regional Clinical Trials. November 2017.

International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use. ICH E5: Ethnic Factors in the Acceptability of Foreign Clinical Data. February 1998.

Chang A, Parikh P, Thongprasert S, et al. Gefitinib (IRESSA) in patients of Asian origin with refractory advanced non-small cell lung cancer: subset analysis from the ISEL study. Journal of Thoracic Oncology. 2006;1(8):847-55.

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.

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.

Ministry of Health, Labour and Welfare of Japan (MHLW). Basic principles on global clinical trials 2007. (Available at: http://www.pmda.go.jp/files/000153265.pdf [Accessed date: September 24, 2019]