Supplements
Lessons Lessoned from Multi-regional Trials with Signals of Treatment Effect Heterogeneity

1. Probability of observing a negative treatment difference in a subgroup

Suppose the primary variable of a trial follows a normal distribution. Assuming the common standard deviation of the treatment and control is σ, to detect a between-treatment difference of δ with power, $1-\beta$, at the 2-sided significance level 0.05, the total sample size under 1:1 randomization is:

$$N_\beta = 4(1.96 + z_\beta)^2 \left(\frac{\sigma}{\delta}\right)^2,$$

where z_β is the $(1-\beta)$th-quantile of the standard normal distribution.
Under the alternative, for a subgroup with a total sample size n, the chance to observe a negative result, i.e., mean difference <0, is $\Phi\left(-\sqrt{n}\frac{\delta}{2\sigma}\right)$, where Φ is the cumulative distribution function of the standard normal distribution.

For a trial with a power of PWR= $1-\beta$, if the subgroup sample size is a proportion r of N, i.e., n=rN_{PWR}, the chance to observe a negative result on this subgroup is $\Phi\left(-\sqrt{n}\frac{\delta}{2\sigma}\right) = \Phi\left(-(1.96 + z_\beta)\sqrt{r}\right)$. Table S1 presents the probability of observing a negative treatment diifference in a subgroup of size *n* as a proportion of the total planned sample sizes.

Table S1. Chance to observe a negative treatment difference in a subgroup with a sample size *n=rN*

	Power of the Trial		
r=n/N	**80%**	**90%**	**95%**
5%	27%	23%	21%
10%	19%	15%	13%
15%	14%	10%	8%
20%	11%	7%	5%

Suppose an MRCT involves K regions with a fraction of sample size $r_k$for the kth region. The probability of observing a negative result in at least one region is

$$1 - \Phi\left((1.96 + z_\beta)\sqrt{r_1}\right) \times \ldots \times \Phi\left((1.96 + z_\beta)\sqrt{r_K}\right).$$

This probability is minimized when $r_1 = \cdots = r_K = \frac{1}{K}$, i.e., the probability of observing a negative result is $\geq 1 - \left[\Phi\left((1.96 + z_\beta)\sqrt{1/K}\right)\right]^K$.

2. Ad hoc probability of observing a negative treatment difference in a subgroup

Quan, et al (2017) used an approach to calculate the ad hoc probability of observing a negative result in a subgroup based the study result. In a MRCT with K regions, suppose the sample size per treatment arm, the treatment effect estimates and their variances of region k are $N_k, \hat{\delta}_k$, and $\hat{\sigma}_k^2$, respectively. The true overall treatment effect δ is its estimate $\hat{\delta}$. The ad hoc probability of observing a negative result in a subgroup is

$$1 - P(\hat{\delta}_1 > 0, \ldots, \hat{\delta}_K > 0|\delta) = 1 - \prod_{k=1}^K P(\hat{\delta}_k > 0|\delta)$$
$$= 1 - \prod_{k=1}^K \Phi\left(\hat{\delta}/\sqrt{N_k\hat{\sigma}_k^2}\right).$$

3. Formulae for conversion of summary statistics for survival data

For an event driven MRCT, let λ_1 and λ_0 be the true overall hazard rates of the active treatment and placebo groups. Moreover, suppose λ_{1k} and λ_{0k} are the true hazard rates, E_{1k} and E_{0k} are the expected numbers of events, $\hat{E}_{1k}$ and $\hat{E}_{0k}$ are the observed numbers of events, U_{1k} and U_{0k} are the total exposures (patient-years) for the active treatment and placebo for the kth region/subgroup, k=1,…, K, respectively. Without data of individual patients from publications, summary statistics will be used to derive the asymptotic results. First, $\hat{\lambda}_{ik} = \hat{E}_{ik}/U_{ik}$ is the estimate of the overall regional hazard rate λ_{ik} (it can be treated as the average hazard rate if it is not constant over time). Note that $Var(\hat{\lambda}_{ik}) = \lambda_{ik}/U_{ik}$. Thus, based on a delta method, the estimate of variance of $\log(\hat{\lambda}_{ik})$ is $\left(\frac{\hat{E}_{ik}}{U_{ik}}\right)\frac{1}{(\hat{\lambda}_{ik})^2} = \frac{1}{\hat{E}_{ik}}$. Asymptotically (with the expected number of events replaced by the observed number of events), the log

estimate of the hazard ratio of the active treatment versus placebo for subgroup *k*

$$\log(\hat{\theta}_k) = \log(\hat{\lambda}_{1k} / \hat{\lambda}_{0k}) \sim N(\log(\lambda_{1k} / \lambda_{0k}), \frac{1}{\hat{E}_{1k}} + \frac{1}{\hat{E}_{0k}}). \quad (1)$$

The 95% confidence interval for $\log(\lambda_{1k} / \lambda_{0k})$ is

$$(L_k, U_k) = (\log(\hat{\lambda}_{1k} / \hat{\lambda}_{0k}) - z_{0.975}\sqrt{1/\hat{E}_{1k} + 1/\hat{E}_{0k}},\ \log(\hat{\lambda}_{1k} / \hat{\lambda}_{0k}) + z_{0.975}\sqrt{1/\hat{E}_{1k} + 1/\hat{E}_{0k}})$$

and the 95% confidence interval for hazard ratio $\lambda_{1k} / \lambda_{0k}$ is

$$(\exp(L_k), \exp(U_k)). \quad (2)$$

We can also use (1) to derive the p-value for testing $H_{0k} : \delta_k = \lambda_{1k} / \lambda_{0k} \geq 1$ versus

$H_{ak} : \delta_k = \lambda_{1k} / \lambda_{0k} < 1$. Note that under $H_{0k} : \delta_k = 1$

$$t = \log(\hat{\lambda}_{1k} / \hat{\lambda}_{0k}) / \sqrt{1/\hat{E}_{1k} + 1/\hat{E}_{0k}} \sim N(0,1).$$

Thus, the p-value is

$$\Pr(Z < t)$$

where Z is a random variable with a standard normal distribution.

The overall estimate of the log hazard rate of individual treatment group combining data across the subgroups asymptotically follows

$$\log(\hat{\lambda}_i) = \frac{\sum_{k=1}^{K} \hat{E}_{ik} \log(\hat{\lambda}_{ik})}{\sum_{k=1}^{K} \hat{E}_{ik}} \sim N(\frac{\sum_{k=1}^{K} E_{ik} \log(\lambda_{ik})}{\sum_{k=1}^{K} E_{ik}}, \frac{1}{\sum_{k=1}^{K} E_{ik}}) \quad (3)$$

with the proportions of the inverses of the variances as the weights. Confidence intervals for the overall hazard rates of individual treatments can be derived based on (3). If a stratified approach with the subgroup factor as a stratification factor is used for deriving the overall hazard ratio, the overall estimate of the hazard ratio as a weighted combination of the individual hazard ratios will be

$$\hat{\theta} = \frac{\sum_{k=1}^{K} \log(\hat{\lambda}_{1k} / \hat{\lambda}_{0k}) / \hat{\sigma}_k^2}{\sum_{k=1}^{K} 1/\hat{\sigma}_k^2}$$

where $\hat{\sigma}_k^2 = \frac{1}{\hat{E}_{1k}} + \frac{1}{\hat{E}_{0k}}$ is the estimate of the variance of the hazard ratio for the kth region/subgroup (if a publication directly provides $\hat{\sigma}_k^2$ derived based on individual patients' data, this $\hat{\sigma}_k^2$ should be used. If the publication provides the 95% confidence interval for $\lambda_{1k} / \lambda_{0k}$, via (2), we can also derive $\hat{\sigma}_k^2$.) Again, the proportions of the inverses of the variances are used as the weights. The asymptotic distribution for the overall estimate of the hazard ratio

$$\hat{\theta} \sim N\left(\frac{\sum_{k=1}^{K} \log(\lambda_{1k} / \lambda_{0k}) / \hat{\sigma}_k^2}{\sum_{k=1}^{K} 1/\hat{\sigma}_k^2}, \frac{1}{\sum_{k=1}^{K} 1/\hat{\sigma}_k^2}\right). \quad (4)$$

This asymptotic distribution can be used to derive the 95% confidence interval and p-value for the overall hazard ratio.

To calculate the probability of observing at least one negative regional treatment effect, we need to evaluate

$$\Pr(\log(\hat{\theta}_1) > 0, \text{ or } \ldots, \log(\hat{\theta}_K) > 0 \mid \theta_1 = \ldots = \theta_K = \theta) \quad (5)$$

based on the asymptotic distribution of $\log(\hat{\theta}_k)$ (see (1)) given the true value $\theta_k = \theta$ which can be the estimate of the overall hazard ratio of the whole MRCT. Note that $\log(\hat{\theta}_k)$'s are independent. Probability (5) can be easily calculated based on (1). If probability (5) is large, the chance of observing negative regional treatment effects will be large as well. Other probabilities can also be calculated.

For formal analyses performed by the study sponsor, individual patients' data should be used to derive the variances and asymptotic distributions.

4. The details of the application of the drop-min approach in ISEL study

Table S2: Survival Data from the ISEL Study

		Gefitinib	Placebo	HR	95% CI	p-value
Non-Asian	N	894	456	0.93	0.81,1.08	0.364
	Death	536	269			
Asian	N	235	107	0.66	0.48, 0.91	0.011
	Death	98	75			
Sources: Carroll (2004) and Thatcher, et al 2005.						

Table S3: Simulated Bias and Adjusted Test Statistic for Asian Patient Survival Data from the ISEL Study

(For Asian origin patients, ln(HR)= −0.416, SD of ln(HR)= 0.163, p=0.011)

	Bias	SD of bias	Adjusted ln(HR)	Adjusted HR	Adjusted z-value	Adjusted p-value
FEM	−0.102	0.149	−0.314	0.73	−2.110	0.035
DREM	−0.010	0.154	−0.406	0.67	−2.637	0.008

5. The details of the application of the drop-min approach in BiDil example

Table S4: Survival Data from the V-HeFT I Study

		BiDil	Placebo	p-value	HR	ln(HR)	SD of ln(HR)
White	N	132	192	0.48	0.8855	−0.1216	0.1721
	Death	56	85				
Black	N	49	79	0.041	0.5322	−0.6306	0.3086
	Death	15	35				
Overall	N	186	183	0.093	0.7785	−0.2504	0.1491
	Death	72	120				
N, number of death, and log-rank p-value were from Carlson, et al (1999); ln(HR), HR, and SD of ln(HR) were calculated from formulas in section 3 of this Supplement.							

Table S5: Simulated Bias and Adjusted Test Statistics for Black Patient Survival Data from the V-HeFT I Study

(For black patients, ln(HR) = -0.6306, SD of ln(HR)= 0.3086)

	Bias	SD of bias	Adjusted ln(HR)	Adjusted HR	Adjusted z-value	Adjusted p-value
FEM	−0.141	0.206	−0.490	0.61	−2.373	0.018
DREM	−0.042	0.279	−0.589	0.55	−2.114	0.035

Table S6: Survival Data from the Studies V-HeFT I and II Combined

		BiDiL V-HeFT I-II combined	Placebo V-HeFT I	HR	ln(HR)	SD of ln(HR)	p-value
White	N	414	192	0.83	−0.1922	0.1331	0.1496
	Death	168	85				
	AMR	15.5%	18.8%				
Black	N	158	79	0.62	−0.4712	0.2170	0.030
	Death	54	35				
	AMR	10.8%	17.3%				

N, number of death, and AMR (i.e., Annual Mortality Rate) were from Carlson, et al (1999);
HR is estimated as the ratio of AMR;
ln(HR), SD of ln(HR), and p-value were calculated from formulas in Supplement of this manuscript.

Table S7: Simulated bias and adjusted test statistic for black patients' survival data from the V-HeFT I & II Combined

(For black patients, ln(HR) = −0.4712, SD of ln(HR) = 0.2170, p = 0.030)

	Bias	SD of bias	Adjusted ln(HR)	Adjusted HR	Adjusted z-value	Adjusted p-value
FEM	−0.102	0.149	−0.370	0.69	−2.486	0.013
DREM	−0.047	0.188	−0.424	0.65	−2.255	0.024