Statistics and Its Interface Volume 11 (2018) 129–139

Robust estimate of regional treatment effect in multi-regional randomized clinical trial in global drug development

Ao Yuan*, Shuxin Wang, and Ming T. Tan*

With the globalization of drug development, and thus clinical trials over multiple regions, determining and inferring the regional effects of a treatment under study are of increased interests in the global drug development, and is becoming a new research field. Existing methods mostly use subjectively specified models, and will be more or less deviated from the true one. In practice, we often have some prior knowledge of the model, but are not sure how well it will fit the data. To address this problem, we propose a semiparametric model, which is a mixture with a known parametric and an unknown nonparametric component. The parametric component represents our prior knowledge about the model, and the nonparametric part reflects our uncertainty. In this way, the prior knowledge is effectively incorporated into the robust model, due to the nonparametric component. The model parameters are estimated by maximizing the corresponding profile likelihood, and the null hypothesis of no regional effect is tested using the profile likelihood ratio statistic. We derive the asymptotic properties of the estimators. Simulation studies are then conducted to evaluate the performance of the model, and results show the clear advantages of the proposed method over existing parametric model. Then model is then used to analyze a real multi-regional clinical trial data as an illustration.

Keywords and phrases: Clinical trial, Hypothesis test, Multi-regional effects, Profile likelihood, Semiparametric model.

1. INTRODUCTION

Clinical trials are increasingly being conducted over multiple regions globally, e.g., Europe, Asia and North America. The US Food and Drug Administration (FDA) reviewed 1926 clinical trials conducted during 2001-2007, showed that 50% of them were multi-regional trials and included both US domestic and foreign sites (O'Neill, 2009). The globalization of pharmaceutical products has become the key to success for drug development. Investors in new drug development are required to do more at less cost and faster rates. However, there are also new challenges due to ethnic factors, as the pharmacodynamic or clinical data in the original population could vary with the population in the new region. Whether a disease treatment is efficacious may well depend on ethnic factors in clinical outcomes. Some regions, e.g., China and Japan, often require local trials in addition to a multi-regional clinical trial (MRCT) to support the efficacy and safety claims of the treatment in order to get the product registered in those regions. The impact of ethnic factors on the treatment is an important issue and has been intensively studied from several different perspectives. For example, the most current methods focus on the assessment of the consistency or similarity of the treatment effects among different ethnic groups. The design and conduct of MRCTs to ensure the acceptability of results for regulatory decision making presents significant challenges for industry and regulators (MHLWJ, 2007; O'Neill, 2009). Thus, to determine whether the treatment under study has regional effects, regional treatment differences, and infer such effects is of main interests in global drug development, and is becoming a new research field. O'Shea and DeMets (2001) discussed some statistical issues in this problem. Kawai et al. (2008) and Uesaka (2009) considered sample size allocation in multi-regional clinical trials. Research on this topic is growing, for example, Victers et al. (1998), Akkerhuis et al. (2000), Wedel et al. (2001), Blaire et al. (2008), Hung et al. (2010), Chen et al. (2010), Wang (2010), Quan et al. (2010), Shi (2010), Huang et al. (2012), Quan et al. (2013, 2014). Bridging clinical studies, with supplemental investigation on new regions based on foreign clinical trial data, is also on-going, as in Chow et al. (2002), Liu and Chow (2002), Hsiao et al. (2003, 2007).

Existing statistical methods on MRCT are mostly parametric. If the model is correctly specified, the model is simple to use and inference will be efficient. However, if the model is incorrect, for example in practice any subjectively specified model is more or less deviated from the true one, the inference can be biased (Huber, 1967; Pfanzagl, 1969). On the other hand, although the nonparametric model is robust to model specification, generally it is not efficient. Often we know the data model to some extent, but we are unsure of how accurate it can represent the true model. To incorporate the prior knowledge and for robustness, we propose a semiparametric model for the analysis of MRCT, in

*Corresponding authors.

which the prior knowledge about the true parametric model is implemented into the model along with a nonparametric component to reflect the uncertainty about the true model, which is a mixture of the known model and an unknown nonparametric component. The mixing proportion which is the extent of the correctness of the postulated parametric component is estimated from the data itself. It is known that parameter estimate in mixture model is difficult, and a common practice is to work on the corresponding 'complete data' non-mixture model, treating the latent memberships as missing data, then use the EM algorithm is used for estimating the parameters. The nonparametric component in our model is treated as a nuisance parameter and is profiled out. The parameter of interest, including the regional effects, is estimated via the resulting profile likelihood. The profile likelihood ratio statistic is used to test the existence of regional differences. Simulation studies are conducted to evaluate the performance of the proposed method. The results show that when the true model is deviated, as any model is more or less deviated, the proposed method can outperform the subjectively specified one. Then the model is used to analyze a real multi-regional clinical data as an illustration. Relevant technical proofs are put into the Appendix in the supplementary material.

2. THE PROPOSED METHOD

The observed data is $(y_i, \mathbf{x}_i, \boldsymbol{\delta}_i), i = 1, ..., n$ for n independent individuals from k-regions, where y_i is the quantitative response of the clinical trial under study, $\mathbf{x}_i \in R^d$ is the covariate, and $\boldsymbol{\delta}_i = (\delta_{i,1}, ..., \delta_{i,2k})'$ is the region indicator of the i-th subject. It has only one none-zero component, with $\delta_{i,j} = 1$, if this individual is from the j-th region ($j = 1, ..., k$) and with treatment 1 (control), and $\delta_{i,k+j} = 1$, if this individual is from the j-th region and with treatment 2 (new). Denote the whole observed data as $(\mathbf{Y}_n, \mathbf{X}_n, \boldsymbol{\Delta}_n)$, with $\mathbf{Y}_n = (y_1, ..., y_n)'$, $\mathbf{X}_n = (\mathbf{x}_1, ..., \mathbf{x}_n)'$ and $\boldsymbol{\Delta}_n = (\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_n)'$.

The goal is to estimate the regional effects on the treatments differences and test the consistency of the trial, or the null hypothesis of no regional effects on the difference of the two treatments. One may analyze the data for each region separately, but the joint analysis will be more informative by using all the data as there are common covariate effects among the regions, and this will help the estimation of regional effects to be more accurate.

To utilize our knowledge about the data distribution and allow for more model flexibility/robustness, we proposed the following semiparametric model.

2.1 The semiparametric model specification

We focus on the case of fixed effects, i.e., that the regional effects is a fixed vector of parameters. The case of random effects will be discussed on the basis of the former. Let $\boldsymbol{\beta} = (\beta_1, ..., \beta_d)'$ be the regression coefficients of the covariate $\mathbf{x}_i$'s, and $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{2k})'$ be the regional effects of the two treatments, i.e. $\alpha_1, ..., \alpha_k$ are the effects of the first treatment for regions $1, ..., k$; α_{k+j} are those for the second treatment. The relationship between the responses and the covariates can be written as

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \boldsymbol{\delta}_i'\boldsymbol{\alpha} + \epsilon_i, \quad E(\epsilon_i) = 0. \tag{1}$$

Often we have some prior knowledge about the model, or distribution of the error ϵ_i's, given by a known density function $g(\cdot)$, but this initial knowledge is not justified yet and it is not clear to what extent this model is true. To reflect this uncertainty, we specify the model as the additive mixture

$$f(\epsilon) = \lambda g(\epsilon) + (1 - \lambda)h(\epsilon), \qquad h \in \mathcal{H}, \tag{2}$$

where $g(\epsilon) = g(y - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha})$ and $h(\epsilon) = h(y - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha})$, $g(\cdot)$ is a known density function represent our knowledge about the model, $h(\cdot)$ is an unknown density function, $0 < \lambda \leq 1$ is a parameter (unknown) and represents the degree of our belief that g is true. When $\lambda = 1$, the specified parametric model g is perfect, and when $\lambda = 0$, the component g is totally mis-specified and model is fully nonparametric. We leave λ as a parameter to be estimated from the observed data. We define $\mathcal{H}$ to be the collection of all densities which are not a mixture of $g(\cdot)$, so that model (2) is identifiable. This requirement of $\mathcal{H}$ is natural, for if $h(\cdot)$ is a mixture with $g(\cdot)$, then we can just absorb part of $g(\cdot)$ into $\lambda g(\cdot)$.

A related proposal was in Olkin (1987), in which the mixture form is

$$\lambda g(\boldsymbol{\epsilon}|\theta) + (1 - \lambda)\hat{f}_n(\boldsymbol{\epsilon})$$

where $g(\cdot|\theta)$ is a given parametric density and $\hat{f}_n(\cdot)$ is a kernel density estimator.

2.2 Estimation of model parameters

Since $h(\cdot)$ is unknown, it is an infinite dimensional nuisance parameter. Also, it is known that the direct estimate of the parameters in the mixture model (2) is not easy. A common alternative method is to re-write (2) as a non-mixture under the 'complete data' with a missing membership indicator, then use the EM algorithm to compute the parameter estimates under the 'complete data' likelihood. To be specific, let ξ_i be 0-1 valued latent random variables, $P(\xi_i = 1) = \lambda$, $\xi_i = 1$ if ϵ_i comes from model $g(\cdot)$. Since the ξ_i's are un-observed, we treat them as "missing data". Let $\mathbf{z}_i = (y_i, \mathbf{x}_i, \boldsymbol{\delta}_i, \xi_i)$ and $\mathbf{Z}_n = (\mathbf{z}_1, ..., \mathbf{z}_n)$ be the 'complete data'. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \lambda)'$ be the vector of all the parameters to be estimated. Under the complete data, the likelihood is

$$L_n(\boldsymbol{\theta}, h|\mathbf{Z}_n) = \prod_{i=1}^{n} \left[\lambda g(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha})\right]^{\xi_i} \left[(1 - \lambda)h(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha})\right]^{1-\xi_i}$$

and the corresponding log-likelihood is

$$\ell_n(\boldsymbol{\theta}, h|\mathbf{Z}_n) = \sum_{i=1}^{n} \Big(\xi_i \big[\log g(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}) + \log \lambda \big] + (1-\xi_i) \big[\log h(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}) + \log(1-\lambda) \big] \Big) \quad (3).$$

A common method to eliminate the infinite dimensional nuisance parameter h is to find the profile log-likelihood $\tilde{\ell}_n(\boldsymbol{\theta}) = \sup_h \ell_n(\boldsymbol{\theta}, h)$. However, such supreme may not exist or may be infinity. The typical way to handle this supreme is to maximize the log-likelihood over step functions of $h(\cdot)$ with jumps at the observed points, like the empirical distribution. The Cox proportional hazards model is a typical example, in which the log-likelihood function is maximized over the nonparametric base line hazard function in step function form, to get the 'partial likelihood', which is indeed a profile likelihood. Thus we set

$$h(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}) = h_i, \qquad (i = 1, ..., n)$$

then the corresponding complete data log-likelihood is rewritten as

$$\ell_n(\boldsymbol{\theta}|\mathbf{Z}_n) = \sum_{i=1}^{n} \Big(\xi_i \big[\log g(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}) + \log \lambda \big] + (1-\xi_i) \big[\log h_i + \log(1-\lambda) \big] \Big). \quad (4)$$

To eliminate the nuisance parameters $\mathbf{h} = (h_1, ..., h_n)$, for fixed $\boldsymbol{\theta}$ and Δ, we maximize the above log-likelihood over $\mathbf{h}$ subject to $\sum_{i=1}^n h_i = 1$ and using the Lagrange multipliers. So for fixed $\boldsymbol{\theta}$ we maximize $\ell_n(\boldsymbol{\theta}) - \zeta(\sum_{i=1}^n h_i - 1)$ over $\mathbf{h}$, and get the estimates of $\mathbf{h}$ as solution $\hat{\mathbf{h}} = (\hat{h}_1, ..., \hat{h}_n)$, as (Appendix)

$$\hat{h}_i = \frac{1-\xi_i}{\sum_{j=1}^n (1-\xi_j)}, \qquad (i = 1, ..., n). \quad (5)$$

Plugging $\hat{\mathbf{h}}$ into (4), we get the profile log-likelihood

$$\tilde{\ell}_n(\boldsymbol{\theta}|\mathbf{Z}_n) = \sum_{i=1}^{n} \Big(\xi_i \big[\log g(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}) + \log \lambda \big] + (1-\xi_i) \big[\log \hat{h}_i + \log(1-\lambda) \big] \Big). \quad (6)$$

The profile MLE (Severini and Wong, 1992; Murphy and Van der Vaart, 2000) of $\boldsymbol{\theta}$ based on (6) is $\hat{\boldsymbol{\theta}}$,

$$\hat{\boldsymbol{\theta}} = \arg\sup_{\boldsymbol{\theta}} \tilde{\ell}_n(\boldsymbol{\theta}|\mathbf{Z}_n).$$

However, since $\boldsymbol{\xi} := (\xi_1, ..., \xi_n)$ is missing, the estimates $\hat{\mathbf{h}}$ and $\hat{\boldsymbol{\theta}}$ cannot be directly computed, instead we use the EM-algorithm (Dempster et al., 1977), for computation of maximum likelihood estimates under the missing data model, see also Tan, Tian and Ng (2009, chap. 2) for biomedical applications of this algorithm. For this, let $\mathbf{h}^{(0)}$ and $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)}, \boldsymbol{\alpha}^{(0)}, \lambda^{(0)})'$ be any starting values of $\mathbf{h}$ and $\boldsymbol{\theta}$ (typically $h_i^{(0)} = 1/n$ and $(\boldsymbol{\beta}^{(0)}, \boldsymbol{\alpha}^{(0)})$ can be set as the MLE of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ under model $g(\cdot)$, set $\gamma^{(0)} = 1/2$. At the r-th iteration define, in the E-step,

$$H_n(\mathbf{h}, \boldsymbol{\theta}|\mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}) = E_{\boldsymbol{\xi}}[\tilde{\ell}_n(\boldsymbol{\theta}|\mathbf{Z}_n))|\mathbf{Y}_n, \mathbf{X}_n, \mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}], \quad (7)$$

where the expectation is with respect to $\boldsymbol{\xi}$, and as if the true data is generated from parameters $\boldsymbol{\theta}^{(r)}$ and $\mathbf{h}^{(r)}$. The computation of (7) is given in the Appendix. In particular, the r-th step estimate of the ξ_i's (for $i = 1, ..., n; r = 0, 1, 2, ...$), are

$$\xi_i^{(r)} = \frac{\lambda^{(r)} g(y_i - \boldsymbol{\beta}'^{(r)}\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}^{(r)})}{\lambda^{(r)} g(y_i - \boldsymbol{\beta}'^{(r)}\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}^{(r)}) + (1-\lambda^{(r)}) h_i^{(r)}}, \qquad (r = 0, 1, 2, ...).$$

In the M-step, for fixed $\boldsymbol{\theta}$, define

$$\mathbf{h}^{(r+1)} = \mathbf{h}^{(r+1)}(\boldsymbol{\theta}) = \arg\sup_{\mathbf{h}} H_n(\mathbf{h}, \boldsymbol{\theta}|\mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}),$$

which is just (5) with the ξ_i's replaced by the $\xi_i^{(r)}$'s,

$$h_i^{(r+1)} = \frac{1-\xi_i^{(r)}}{\sum_{j=1}^n \xi_j^{(r)}}, \qquad (i = 1, ..., n).$$

Then it is known (for example, Dempster et al., 1997) that, under suitable regularity conditions, as $r \to \infty$ one has

$$\boldsymbol{\theta}^{(r)} \to \hat{\boldsymbol{\theta}}.$$

2.3 Asymptotic properties of parameter estimation

To study the asymptotic behavior of the estimators, the following notations will be used. Denote $\dot{g}(s) = dg(s)/ds$, $\dot{h}(s) = dh(s)/ds$; Similarly, $\ddot{g}(s) = d^2g(s)/ds^2$. Let $\ell(\boldsymbol{\theta}, h) = \log f(\epsilon|\boldsymbol{\theta}, h)$ with $f(\epsilon) = f(\epsilon|\boldsymbol{\theta}, h)$ given in (2), thus $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \lambda)'$ and $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0', \lambda_0)'$ be the true parameter values for generating observed data under model (2); a $b + k + 1$ dimensional parameter. The following conditions will be used.

(C1). For fixed Δ and all large n, the log-likelihood (4) has a unique profile MLE $\hat{\boldsymbol{\theta}}$.
(C2). $\int \sqrt{f(\boldsymbol{\epsilon}|\boldsymbol{\theta}_0, h_0)} d\boldsymbol{\epsilon} < \infty$.
(C3). $\{f(\cdot|\boldsymbol{\theta}, h) : \boldsymbol{\theta} \in \boldsymbol{\Theta}, h \in \mathcal{H}\}$ is bounded.
(C4). $h_0(\cdot)$ is uniformly continuous.
(C5). $\dot{g}(\cdot)$ and $\dot{h}(\cdot)$ are bounded continuous.
(C6). The Ω given in Theorem 2 is non-singular.

Theorem 1. *Assume (C1)-(C4), then for the profile MLE* $\hat{\boldsymbol{\theta}}$,

$$\hat{\boldsymbol{\theta}} \stackrel{a.s.}{\to} \boldsymbol{\theta}_0.$$

For vectors $\mathbf{a} = (a_1, ..., a_b)'$ and $\mathbf{x} = (x_1, ..., x_d)'$, denote $\mathbf{a} \otimes \mathbf{x} = (a_1 + \cdots + a_b)(\mathbf{x}', ..., \mathbf{x}')'$, a bd-dimensional column vector. When $dim(a) = 1$, $a \otimes \mathbf{x} = a\mathbf{x}$.

Let $\dot{g}(\epsilon) = \partial g(\epsilon)/\partial\epsilon$, $\dot{h}(\epsilon) = \partial h(\epsilon)/\partial\epsilon$,

$$\dot{\ell}_0(\boldsymbol{\theta}, h|\mathbf{y}, \mathbf{z})$$

$$= -\frac{\lambda\dot{g}(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha}) + (1-\lambda)\dot{h}(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha})}{f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}, h)}\mathbf{z},$$

where $\mathbf{z} = (\mathbf{x}', \boldsymbol{\delta}')'$.

Theorem 2. *Assume (C1),(C5) and (C6), then the profile MLE* $\hat{\boldsymbol{\theta}}$ *is efficient for* $\boldsymbol{\theta}$ *in the semiparametric model (2),* $(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)$ *is adaptively estimable,*

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) \stackrel{D}{\to} N(\mathbf{0}, \Omega^{-1}(\boldsymbol{\theta}_0)),$$

$$\Omega(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0, h_0}[\mathbf{i}^*\mathbf{i}^{*'}],$$

where $\mathbf{i}^*$ *is the efficient score for* $\boldsymbol{\theta}$ *in the presence of the nuisance* h*. In particular*

$$\sqrt{n}\big[(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}')' - (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0')'\big] \stackrel{D}{\to} N(\mathbf{0}, \Omega_0^{-1}(\boldsymbol{\theta}_0)),$$

$$\Omega_0(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0, h_0}[\mathbf{i}_0^*\mathbf{i}_0^{*'}],$$

where $\mathbf{i}_0^* = \dot{\ell}_0(\boldsymbol{\theta}, h|\mathbf{y}, \mathbf{z})$ *is the efficient score for* $(\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ *in the presence of the nuisance* h*.*

Note that in the model we assumed the $\boldsymbol{\delta}_i$'s are iid $\boldsymbol{\delta}$, and $\min_{1\le j\le k} P\big(\boldsymbol{\delta} = \mathbf{e}_j\big) > 0$, where $\mathbf{e}_j$ is the k-dimensional column vector with the j-th entry be 1 and others be 0's. Thus $n_j/n \to P\big(\boldsymbol{\delta} = \mathbf{e}_j\big) > 0$.

From Theorem 2, we see that $(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}')'$ is efficient for $(\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0')'$. Although the $(\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ components in $\mathbf{i}^*$ are computed explicitly, but the λ component is generally not.

2.4 Testing the null hypothesis

As mentioned before, after the parameter estimation, we need to test the null hypothesis that there is may or may not be regional effects. This null hypothesis can be written as $H_0 : \alpha_{k+j} - \alpha_j =$ Constant for $j = 1, ..., k$ (independent of j) vs the alternative H_1: not H_0. For parametric model, commonly used test statistics include the likelihood ratio statistic, score test statistic and the Wald statistic. In the parametric case the three equivalent statistics are asymptotically chi-squared distributed. Here we use the profile likelihood ratio statistic as it does not involve the computation of first and second order partial derivatives of the profile likelihood. The following theorem tells us that this test statistic parallels the Wilks (1938) result for parametric models, and can be used to test H_0 vs H_1. For this, we define the profile likelihood ratio statistic below. Recall $\hat{\boldsymbol{\theta}}$ is the profile MLE of $\boldsymbol{\theta}$ under the log profile likelihood $\tilde{\ell}_n(\boldsymbol{\theta})$, and $dim(\boldsymbol{\alpha}) = 2k$. Let $\boldsymbol{\Theta}_0 = \{(\boldsymbol{\beta}, \alpha_1, ..., \alpha_k, \alpha_1 + \gamma, ..., \alpha_k + \gamma, \lambda\}$ be the parameter space under H_0, $\tilde{\boldsymbol{\beta}}$ be the profile MLE of $\boldsymbol{\beta}_0$ under H_0,

$$\tilde{\boldsymbol{\theta}} = \arg \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \tilde{\ell}_n(\boldsymbol{\theta}).$$

Computation of $\tilde{\boldsymbol{\theta}}$ follows the same way for that of $\hat{\boldsymbol{\theta}}$. Define the profile likelihood ratio as

$$\Lambda_n = 2\big(\tilde{\ell}_n(\hat{\boldsymbol{\theta}}) - \tilde{\ell}_n((\tilde{\boldsymbol{\theta}})\big).$$

Theorem 3. *Assume conditions of Theorem 2, under* H_0*,*

$$\Lambda_n \stackrel{D}{\to} \chi^2_{k-1},$$

where χ^2_{k-1} *is the chi-squared random variable with* $k-1$ *degrees of freedom.*

The proofs of all the Theorems are in the Appendix of the supplementary material.

However, it is known that the log-likelihood ratio statistic will some times give an inflated type I error. Alternatively, the score test or Wald test can be used. Let $\dot{\tilde{\ell}}_n(\boldsymbol{\theta}|\hat{f}_n) = \partial\tilde{\ell}_n(\boldsymbol{\theta}|\hat{f}_n)/\partial\boldsymbol{\theta}$. The score test statistic is given by

$$S_n = \dot{\tilde{\ell}}_n(\hat{\boldsymbol{\theta}}_0|\hat{f}_n)\Omega(\hat{\boldsymbol{\theta}}_0)\dot{\tilde{\ell}}_n'(\hat{\boldsymbol{\theta}}_0|\hat{f}_n),$$

$$\text{under } H_0, \text{ asymptotically } \quad S_n \sim \chi^2_1.$$

For Wald test in the general case, denote $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ with $dim(\boldsymbol{\theta}) = d$ and $dim(\boldsymbol{\theta}_1) = d_1$, and $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$. Consider the null hypothesis $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$. The Wald test statistic is given by

$$W_n = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0})'Cov^{-1}(\hat{\boldsymbol{\theta}}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0}).$$

If $Cov(\hat{\boldsymbol{\theta}}_1)$ is known, $W_n \sim \chi^2_{d_1}$; If $Cov(\hat{\boldsymbol{\theta}}_1)$ is estimated, $W_n/d_1 \sim F_{d_1, n-d}$.

In our case, let $\hat{\mathbf{d}} = (\hat{d}_1, ..., \hat{d}_k)' = (\hat{\alpha}_{k+1} - \hat{\alpha}_1, ..., \hat{\alpha}_{2k} - \hat{\alpha}_k)'$, and

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & -1 \end{pmatrix},$$

then H_0 can be written as the contrast $C\mathbf{d} = 0$. So the Wald statistic in our case is

$$W_n = (C\hat{\mathbf{d}})'Cov^{-1}(C\hat{\mathbf{d}})C\hat{\mathbf{d}},$$

and we use estimated covariance, so under H_0, $W_n \sim F_{k,n-d}$, $d = dim(\boldsymbol{\theta})$.

2.5 Other issues

There are some other issues commonly encountered with MRCT, here we only give a brief account.

Random effects model. As pointed out by many authors (Senn, 1998; Fedorov et al. 2005; Quan et al., 2013;), a trial result obtained from a fixed effects model is trial specific and may not be generalizable because the expectation of $\hat{\boldsymbol{\alpha}}$ depends on the sample size configuration across the regions and the $\hat{\alpha}_i$'s. On the other hand, results from a random effects model, taking into consideration the variability among regions, may be more generalizable and applicable in a global sense. However, due to its large variability, the random effects model requires a much bigger sample size than the fixed effects model to achieve the same power to detect an overall treatment effect. We can also implement random effects into our model.

Interaction. We can also consider interactions among the sub-regions. For k sub-regions, there are $k(k+1)/2$ interactions, to reduce the number of parameters, we may consider the correlations among interactions.

Shrinkage estimation. In Quan et al. (2014), each regional effect $\Delta_j = \alpha_{2j} - \alpha_{2j-1}$ is estimated using only the data from each region. To use the full data information, they also considered shrinkage estimators. For this, let $\hat{\Delta}_j = \hat{\alpha}_{2j} - \hat{\alpha}_{2j-1}$ be the estimate of Δ_j using only the data from sub-region j, and Δ_0 be some global value. They considered shrinkage estimators of the form

$$\check{\Delta}_j = c_j\hat{\Delta}_j + (1-c_j)\Delta_0 = \Delta_0 + c_j(\hat{\Delta}_j - \Delta_0).$$

In particular, for the fixed effects model, it is assumed that $\hat{\Delta}_j \sim N(\Delta_j, \sigma^2/n_j)$ with σ^2 known. They find the optimal shrinkage estimator, in the sense of minimizing the mean squared error, to be

$$\check{\Delta}_j = c\hat{\Delta}_j + (1-c)\Delta_0, \qquad (j = 1, ..., k),$$

where $\Delta_0 = n^{-1}\sum_{j=1}^k n_j\hat{\Delta}_j$, $n = \sum_{j=1}^k n_j$, $c = \sum_{j=1}^k n_j(\hat{\Delta}_j - \Delta_0)^2/[\sum_{j=1}^k n_j(\hat{\Delta}_j - \Delta_0)^2 + k\hat{\sigma}^2]$, and $\hat{\sigma}^2$ is an estimate of σ^2.

To get variance estimate of $\check{\Delta}_j$, let $\hat{\omega}_{ij}$ be the estimated covariance of $\hat{\Delta}_i$ and $\hat{\Delta}_j$, which can be obtained from variances/covariances of $\hat{\alpha}_{2i}$, $\hat{\alpha}_{2i-1}$, $\hat{\alpha}_{2j}$, $\hat{\alpha}_{2j-1}$. Since

$$\begin{aligned}\check{\Delta}_j &= [c_j + (1-c_j)n_j/n]\hat{\Delta}_j + \sum_{i\neq j}^{k}(1-c_j)(n_j/n)\hat{\Delta}_i : \\ &= \sum_{i=1}^{k} a_{ji}\hat{\Delta}_i,\end{aligned}$$

The estimated variance of $\check{\Delta}_j$ is approximated by

$$\check{\rho}_j^2 = \sum_{i=1}^{k}\sum_{r=1}^{k} a_{ji}a_{jr}\hat{\omega}_{ir}.$$

For random effects model with $\Delta_j \sim N(\Delta, \tau^2)$, they take $c_j = \tau^2/(\tau^2 + 2\sigma^2/n_j)$, $\sigma^2 = Var(\epsilon_j)$, n_j (which is the sample size for sub-region j), $\alpha_0 = \sum_{j=1}^k w_j\hat{\Delta}_j/\sum_{j=1}^k w_j$, and $w_j = 1/(\tau^2 + 2\sigma^2/n_j)$.

3. SIMULATION STUDY AND APPLICATION

3.1 Simulation study

We simulate $n = 1000$ i.i.d. data with response y_i and with covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$. We first generate the covariates, sample the $\mathbf{x}_i$'s from the 3-dimensional normal distribution with a mean vector $\boldsymbol{\mu} = (3.1, 1.8, -0.5)'$ and covariance matrix Γ, with

$$\Gamma^{1/2} = \begin{pmatrix} 0.73 & -0.07 & 0.55 \\ 1.34 & -0.14 & 0.57 \\ 1.52 & -0.37 & 1.53 \end{pmatrix}.$$

Then we generate the response data, which, given the covariates, are from the mixture $\lambda_0 g + (1-\lambda_0)h$, with $\lambda_0 = 0.42$. The y_i's are generated as

$$y_i = \boldsymbol{\beta}_0'\mathbf{x}_i + \boldsymbol{\delta}_i'\boldsymbol{\alpha}_0 + \epsilon_i, \quad (i = 1, ..., n),$$

$\boldsymbol{\beta}_0 = (\beta_1, \beta_2, \beta_3)$ to be specified, $\boldsymbol{\alpha}_0 = (\alpha_1, ..., \alpha_{10})'$ to be specified; $\boldsymbol{\delta}_i = (\delta_{i,1}, ..., \delta_{i,10})$, $P(\delta_{ij} = 1) = 1 - P(\delta_{ij} = 0) = 0.2$ $(j = 1, ..., 10)$. We specify g be the density of $N(0, \sigma^2)$. We specify h as the density of exponential distribution with rate r.

We simulated two datasets. The first has two regions, with two contrast treatments, under four different parameter sets $\boldsymbol{\theta}_0$, representing significant regional effects and no regional effects respectively. Then we estimate the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \lambda, \sigma)$ via the profile MLE from the proposed semiparametric model, and compare the corresponding results of the MLE from the commonly used normal model, as displayed in Table 1. SHE stands for shrinkage estimator with fixed effects, which is computed for both the normal and semiparametric models. *sd* is the estimated standard deviation of the corresponding estimate.

For the normal model, there is no λ, so the corresponding column is empty. MLF (profile) is the MLE of $\boldsymbol{\theta}_0$ under the proposed profile likelihood, MLE(normal) is the MLE of $\boldsymbol{\theta}_0$ under the normal model; [*sd*] is the estimated standard deviation of the estimated parameters. For the first two data sets, we used exponential distribution with rate $r = 3$, to separate it from the normal distribution. From the above table, we see that when the true model is deviated from the commonly used normal model, for most components of $\boldsymbol{\theta}_0$, the estimates from the semiparametric mixture model are significantly better than those from the normal model. For small values of r, although the proposed model gives better parameter estimations, the estimation of λ will be far away from its true value, and tend to be close to 0.9.

Table 1. Parameter estimates under two models (simulated data)

$\boldsymbol{\theta}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	σ	λ
$\boldsymbol{\theta}_0$	(1.30, 1.20, -1.60)	(-5.60, 2.70,5.25, 0.13)	1	0.420
MLE(normal)	(0.9622, -0.3753, -1.556)	(-5.0716, 2.2804, 4.5481, 0.0021)	0.904	
[sd]	[0.2672, 0.5303, 0.4422]	[1.2294, 0.5463, 1.0413, 0.2132]	[0.310]	
SHE(normal)		(-5.07131, 2.2803, 4.5478, 0.0021)		
MLE(profile)	(1.2982, 1.2145, -1.5943)	(-5.6258, 2.6811, 4.9565, 0.1136)	$0.992 \approx 1$	0.4182
[sd]	[0.0205, 0.2250, 0.0598]	[0.4342, 0.4177, 1.2917, 0.3977]	[0.222]	[0.0167]
SHE(profile)		(-5.6256, 2.6810, 4.9563, 0.1136)		
Test H_0 vs H_1	via statistic Λ_n	1233 repetitions, 1176 rejections		Power = 0.954
$\boldsymbol{\theta}_0$	(-2.5, 2, 1.2)	(-4.5, 2.35, 6.25, -6.25)	1	0.3
MLE(normal)	(-2.4825, -1.2203, 0.4306,)	(-3.2818, 2.6916, 6.0876,-4.7985)	1.0	
[sd]	[0.2672, 0.5303, 0.4422]	[1.2294, 0.5463, 1.0413, 0.2132]	[0.370]	
SHE(normal)		(-3.2817, 2.6915, 6.0873, -4.7983)		
MLE(profile)	(-2.4999, 2.0019, 1.2001)	(-4.2894, 2.5526, 5.9702,-6.0395)	0.974	0.3058
[sd]	[0.0126, 0.1500, 0.0389]	[0.2911, 0.3127, 1.7293,0.2856]	[0.326]	[0.0254]
SHE(profile)		(-4.2893, 2.5525, 5.9700, -6.0393)		
Test H_0 vs H_1	via statistic Λ_n	1914 repetitions, 1914 rejections		Power > 0.999
$\boldsymbol{\theta}_0$	(1.30, 1.20, -1.60)	(1, 1, 1, 1)	1	0.420
MLE(normal)	(1.0331, 0.3333, -1.4754)	(-0.0091, 0.0035, -0.00001,-0.0154)	0.906	
[sd]	[0.18 , 0.3594, 0.2968]	[0.2049, 0.2065, 0.2083, 0.2124]	[0.120]	
SHE(normal)		(-0.0079, -0.0011, -0.0030, -0.0113)		
MLE(profile)	(1.2982 , 1.2009, -1.5978)	(1.5646, 1.5629, 1.5706, 1.5598)	$0.922 \approx 1$	0.4135
[sd]	[0.0108 , 0.1275, 0.0330]	[0.2480, 0.2450, 0.2469, 0.2461]	[0.110]	[0.0141]
SHE(profile)		(1.5640, 1.5636, 1.5651, 1.5630)		
Test H_0 vs H_1	via statistic Λ_n	1293 repetitions, 2 rejections		Type I error = 0.0015
$\boldsymbol{\theta}_0$	(-2.5, 2, 1.2)	(1, 1, 1, 1)	1	0.30
MLE(normal)	(-2.3309, -0.0676, 0.5171)	(-0.0059, -0.0067, -0.0082,-0.0141)	0.846	
[sd]	[0.1574, 0.3293, 0.2609]	[0.2053, 0.213, 0.2077,0.2045]	[0.134]	
SHE(normal)		(-0.0086, -0.0087, -0.0090, -0.0100)		
MLE(profile)	(-2.5001, 1.9978, 1.1997)	(1.9441, 1.9481, 1.9475,1.9455)	0.865	0.2999
[sd]	[0.0076, 0.0938, 0.0245]	[0.2052, 0.2104, 0.2072, 0.2038]	[0.127]	[0.0142]
SHE(profile)		(1.9456, 1.9459, 1.9459, 1.9457)		
Test H_0 vs H_1	via statistic Λ_n	1200 repetitions, 1 rejections		Type I error < 0.001
		$\chi^2(0.95; 4)$ =9.488		

For small r, the exponential and normal densities are not well separated. This makes the model difficult to estimate the mixing proportion λ.

Then we test the effects of multi-regions by testing the null hypothesis $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ vs $H_1 : \boldsymbol{\alpha} \neq \mathbf{0}$. Test results are also given in Table 1.

The second simulated data has three regions, with two contrast treatments. The estimation and testing results are given in Table 2.

3.2 Application to real data problem

The PLATO data has been studied and analyzed by several groups; however this data is currently unavailable to us. Instead we re-analyze the real clinical trial data from the Study of Denosumab Compared With Zoledronic Acid (Zometa) in the Treatment of Bone Metastases in Men with Hormone-Refractory Prostate Cancer, which is an international, phase 3, Randomized, Double-blind, multicenter study. The study was conducted in 39 countries and four regions (Europe, Latin America, North, America, Others) from May 2006 to December 2008. Overall, 1904 participants were enrolled. 951 of the patients received 4 mg intravenous zoledronic acid plus subcutaneous placebo and 950 patients received120 mg subcutaneous denosumab plus intravenous placebo. The patients received these treatments every 4 weeks until the primary analysis cutoff date. The main outcome (primary endpoint) is the time to the first on-study of skeletal-related event (SRE) analyzed for non-inferiority. The data set was downloaded from a public website (Project Data Sphere). This data set contains only the control group information, the total sample size is 756. Several variables (e.g. Age, Body Mass Index (BMI) and Gleason Score) were considered to be potential covariates to the First On-Study SRE. The Gleason grading system is used to help evaluate the prognosis of men with prostate cancer using samples from a prostate biopsy. Our model will be used to assess the differential regional treatment effects. It will also be used to determine the covariates influence on the treatment. This data only contains the control group information, so the effects of the regional control group

Table 2. Parameter estimates under two models (simulated data 2)

$\boldsymbol{\theta}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	λ
$\boldsymbol{\theta}_0$	(1.3, -0.52,-1.6)	(-4.5,3.5,-5.2,2.8,2.3,-5.7)	0.42
MLE(normal)	(0.5059, -1.30, -0.8088)	(-2.2922, 2.8002, -2.7499,2.353,2.0246,-3.0515)	
[sd]	[0.48844, 1.0138, 0.750]	[1.1793, 1.3882, 1.4039, 1.174,1.008,1.5542]	
MLE(profile)	(1.2861, -0.7280, -1.6237)	(-3.6726, 3.2247, -4.3664,2.9124,2.7063,-4.8692)	0.4411
[sd]	[0.1137, 1.04897, 0.2237]	[1.9321, 2.5697, 1.9553,2.251,2.0701,1.9652]	[0.0901]
Test H_0 vs H_1	via statistic Λ_n	1416 repetitions, 150 rejections	Type I error =0.106
$\boldsymbol{\theta}_0$	(1.3, 1.1,-1.6)	(-4.0,2.6,3.05,3.03,-6.23,-4.08)	0.42
MLE(normal)	(0.6735, -0.1305, -1.1005)	(-1.8094, 2.0015, 2.2611,2.2526,-3.1019,-1.8668)	
[sd]	[0.3915, 0.7262, 0.5685]	[1.0677, 1.1100, 1.2513, 1.2437,1.7803,1.1006]	
MLE(profile)	(1.2861, 1.0039, -1.6024)	(-3.5842, 2.5007, 2.7933,2.8027,-5.8118,-3.6572)	0.4376
[sd]	[0.1010, 0.9267, 0.2116]	[2.0238, 2.1236, 2.2013,2.1848,2.0776, 2.0221]	[0.0852]
Test H_0 vs H_1	via statistic Λ_n	1783 repetitions, 1607 rejections	Power =0.901

Table 3. Parameter estimates under two models (real data1)

$\boldsymbol{\theta}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	σ	λ
MLE(normal)	(-0.1237, 0.1611, -0.7777)	(-2.5361, -2.5664, -2.5084,-2.3377)	4.24	
[sd]	[0.2672, 0.5303, 0.4422]	[1.0378, 1.1601, 1.209, 1.2581]	[0.898]	
SHE(normal)		(-2.5268, -2.5513 -2.5044 -2.3662)		
MLE(profile)	(-0.0466, 0.1396, -0.0200)	(-4.2825, -4.5008, -4.5606,-4.5757)	4.45	0.7364
[sd]	[0.0370, 0.0377, 0.0463]	[0.1096, 0.1325, 0.1380,0.1533]	[0.881]	[< 0.0001]
SHE(profile)		(-4.2828, -4.5008, -4.5605, -4.5755)		

Table 4. Parameter estimates under two models (real data 2)

$\boldsymbol{\theta}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	σ	λ
MLE(normal)	(-0.1212, -0.1863, 0.1259)	(-5.3685, -5.3311, -5.407,-5.2556)	5.059	
[sd]	[0.0359, 0.0371, 0.0375]	[0.0529, 0.0751, 0.0839, 0.1407]	[0.2060]	
SHE(normal)		(-5.3685, -5.3311, -5.4070, -5.2556)		
MLE(profile)	(-0.1008, -0.1663, 0.0657)	(-5.0683, -5.9314, -5.4072,-5.6640)	5.362	0.7086
[sd]	[0.0358, 0.0371, 0.0374]	[0.0530, 0.0749, 0.0838, 0.1342]	[0.2056]	[< 0.0001]
SHE(profile)		(-5.0683, -5.9314, -5.4072, -5.6640)		

test is only hypothetical. An example of this would be, if the regional treatment group effects are approximately the same across all regions, such as (-2.5015, -2.5571, -2.5249, -2.3720). This would result in a Wald's statistic of 0.1782, which is less than the critical value of 2.39 and resulting in acceptance of the null hypothesis of no regional treatment difference. Contrarily, if the treatment effects significantly differ among the various regions, such as (-1.243, -0.9983, -2.0752, -2.4556), it would result in a larger Wald's statistic of $6.5749 > 5.99$ and the rejection of the null hypothesis.

According to the estimated coefficient for each variable, the potential covariates do not have much effects on the First On-Study SRE.

Another study is a randomized, Double-Blind, Multicenter Study of Denosumab Compared With Zoledronic Acid (Zometa) in the Treatment of Bone Metastases in Subjects with Advanced Breast Cancer. This study uses the same treatment for the case and control group for a different purpose.

For this study, the data set was also downloaded from a data share website (Project Data Sphere). Like the first real data, this data set only contains the control group information. The total sample size is 711 and several variables (i.e. pulse rate, respiratory rate, and systolic blood pressure) were considered to be potential covariates. Our model will be used to assess the differential regional treatment effects. It will also be used to check the influence the covariates influence have on the treatment. A hypothetical null hypothesis test can be made as was for the first real data.

Table 4 shows the results from our model. According to the estimated coefficients, the potential covariate give little effect on the First On-Study SRE.

Concluding remarks. We proposed a semiparametric model to analye the multi-regional clinical trial study. The model has a specified known parametric component, and an unknown nonparametric component, it accounts for covariates in the analysis, and is robust with respect to model misspecifications. The profile likelihood method is used to com-

pute the model parameters, including the regional effects. The null hypothesis of no regional effects ($\alpha_i = 0$) or no regional treatment differences ($\alpha_{k+j} - \alpha_j = \gamma, j = 1, \cdots, k$) can be tested via the semiparametric likelihood ratio statistic. Simulation studies show the method outperform the parametric model when it is not correctly specified, which is promising. The method was then used to analyze real multi-regional clinical trial data. The PLATO data is a well studied trial for this topic; however, the data is not yet available to us. An application has been submitted for access to this data. When the PLATO data is available we will apply the proposed method to analyze the data. In future studies, we will extend the model and implement random effects into the semiparametric model (2), quantify the information loss when $g(\cdot)$ is the true distribution, and the distribution of random effects with either parametric or semiparametric specifications.

ACKNOWLEDGEMENTS

The research of Tan is partially supported by R01CA164717.

APPENDIX A. PROOFS AND ASYMPTOTICS

Derivation of (5). Take derivative of $\ell_n(\boldsymbol{\theta}|\mathbf{Z}_n)$ in (4) with respect to h_i, with Lagrange multiplier $\eta(\sum_{i=1}^n h_i - 1)$ for the constraint $\sum_{i=1} h_i = 1$, and set it to zero,

$$0 = \frac{1-\xi_i}{h_i} - \eta, \qquad (i = 1, ..., n).$$

Multiply both sides of the above by h_i and sum over i, we get $\eta = \sum_{i=1}(1-\xi_i)$, and

$$\hat{h}_i = \frac{1-\xi_i}{\sum_{j=1}^n (1-\xi_j)}, \qquad (i = 1, ..., n).$$

Computation of (7). From (6),

$$\begin{aligned} &E_{\boldsymbol{\xi}}[\tilde{\ell}_n(\boldsymbol{\theta}|\mathbf{Z}_n))|\mathbf{Y}_n, \mathbf{X}_n, \mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}] \\ &= \sum_{i=1}^n \Big(E\big[\xi_i|\mathbf{Y}_n, \mathbf{X}_n, \mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}\big] \log g(y_i - \boldsymbol{\beta}'\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}) \\ &\quad + E\big[(1-\xi_i)|\mathbf{Y}_n, \mathbf{X}_n, \mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}\big] \log \hat{h}_i \Big) \end{aligned}$$

and

$$\begin{aligned} \xi_i^{(r)} &:= E\big[\xi_i|\mathbf{Y}_n, \mathbf{X}_n, \mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}\big] \\ &= P\bigg(\xi = 1 \bigg| \mathbf{Y}_n, \mathbf{X}_n, \mathbf{h}^{(r)}, \boldsymbol{\theta}^{(r)}\bigg) \\ &= P\bigg(\xi = 1 \bigg| y_i, \mathbf{x}_i, h_i^{(r)}, \boldsymbol{\theta}^{(r)}\bigg) \\ &= \frac{\lambda^{(r)} g(y_i - \boldsymbol{\beta}^{'(r)}\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}^{(r)})}{\lambda^{(r)} g(y_i - \boldsymbol{\beta}^{'(r)}\mathbf{x}_i - \boldsymbol{\delta}_i'\boldsymbol{\alpha}^{(r)}) + (1-\lambda^{(r)}) h_i^{(r)}}. \end{aligned}$$

This gives (7).

Proof of Theorem 1. Recall the density $f(\cdot) = f(\cdot|\boldsymbol{\theta}, h)$ given in (2). Let F be the distribution of f, $\hat{f}(\cdot) = f(\cdot|\hat{\boldsymbol{\theta}}, \hat{\mathbf{h}})$, and $\hat{F}$ be the distribution function of $\hat{f}$. Denote $f_0(\cdot) = f(\cdot|\boldsymbol{\theta}_0, h_0)$. Let $\mathcal{B}$ be the Borel field on R, $H(\hat{f}, f)$ be the Hellinger distance between $\hat{f}$ and f, and $||\hat{f} - f||$ be the variational distance, between $\hat{f}(\cdot)$ and $f(\cdot)$,

$$H(\hat{f}, f) = 2^{-1/2}\bigg[\int \Big(\hat{f}^{1/2}(\boldsymbol{\epsilon}) - f^{1/2}(\boldsymbol{\epsilon})\Big)^2 d\boldsymbol{\epsilon}\bigg]^{1/2},$$

$$\begin{aligned} ||\hat{f} - f|| &= 2\sup\{|\hat{F}(B) - F(B)| : B \in \mathcal{B}\} \\ &= \int |\hat{f}(\boldsymbol{\epsilon}) - f(\boldsymbol{\epsilon})| d\boldsymbol{\epsilon}. \end{aligned}$$

Recall the inequality $||\hat{f} - f|| \le 2H(\hat{f}, f)$ (see Bickel et al, 1993, p464). We will show that $H(\hat{f}, f_0) \to 0$, a.s., so that $||\hat{f} - f_0|| \to 0$, a.s., which implies $|\hat{h}_n(\cdot) - h_0(\cdot)| \to 0$ a.s. on $\mathcal{H}_n$ (defined below), and by condition (C4) $\sup_t |\hat{h}_n(t) - h_0(t)| \to 0$ a.s., and thus $\hat{f}_n(\cdot) \to f_0(\cdot)$, a.s., a.e. (L), with L being the Lebesque measure on R. Since the model is identifiable, we must have $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ (a.s.), and get the desired result.

Below we show $H(\hat{f}, f) \to 0$ a.s.. For fixed $\boldsymbol{\theta}_0$ and $h_0 \in \mathcal{H}$, let $r_{\boldsymbol{\theta},h}(\cdot) = (\sqrt{f(\cdot|\boldsymbol{\theta}, h)/f_0(\cdot)} - 1)1(f_0 > 0)$, $\mathcal{H}_n = \{h \in \mathcal{H} : h$ be of the form $h_{i,0}, h_{i,1}\}$ (as given in Section 2.2), $\mathcal{R}_n = \{r_{\boldsymbol{\theta},h} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, h \in \mathcal{H}_n\}$, and $\mathcal{R} = \{r_{\boldsymbol{\theta},h} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, h \in \mathcal{H}\}$. It is seen that $\mathcal{H}_n \subset \mathcal{H}$, $\mathcal{R}_n \subset \mathcal{R}$. let P_n and P be empirical and the true distribution of the observed data. By Lemma 1.1 of van de Geer (1993), since $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{h}})$ is the semiparametric MLE of $(\boldsymbol{\theta}_0, h_0)$ in model (2),

$$\begin{aligned} H^2(\hat{f}, f_0) &\le 2(P_n - P)\bigg(1(f_0 > 0)[\sqrt{\hat{f}/f_0} - 1]\bigg) \\ &= 2(P_n - P) r_{\hat{\boldsymbol{\theta}},\hat{h}}. \end{aligned}$$

So to show $H(\hat{f}, f) \to 0$ a.s., it suffices to show $\sup_{r\in\mathcal{R}_n} |(P_n - P)r| \to 0$ a.s., and since $\mathcal{R}_n \subset \mathcal{R}$, it suffices to show

$$\sup_{r\in\mathcal{R}} |(P_n - P)r| \to 0, \quad a.s.$$

i.e., $\mathcal{R}$ is a Glivenko-Cantelli class with respect to P.

For this, for a given probability measure P on $\mathcal{B}$, let $||g||_{L_1(P)} = \int |g(y)| P(dy)$, $N_{[\,]}(\epsilon, \mathcal{R}, L_1(P))$ be the minimum number of ϵ-brackets to cover $\mathcal{R}$ under norm $L_1(P)$, i.e. the minimum number k of pairs (l_j, u_j), $l_j, u_j \in \mathcal{R}$ such that $\forall r \in \mathcal{R}$, there is (l_j, u_j) $(1 \le j \le k)$ with $l_j \le r \le u_j$ and $||u_j - l_j||_{L_1(P)} \le \epsilon$.

Below we need to evaluate $N_{[\,]}(\epsilon, \mathcal{R}, L_1(P))$. Let $\mathcal{F} = \{f_{\boldsymbol{\theta},h} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, h \in \mathcal{H}\}$. Note that for all $f_1, f_2 \in \mathcal{F}$,

$$\begin{aligned}
&\left|\left|\left(\sqrt{\frac{f_1}{f_0}} - 1\right)1(f_0 > 0) - \left(\sqrt{\frac{f_2}{f_0}} - 1\right)1(f_0 > 0)\right|\right|_{L_1(P)} \\
&= \int \frac{|f_1^{1/2}(\boldsymbol{\epsilon}) - f_2^{1/2}(\boldsymbol{\epsilon})|}{f_0^{1/2}(\boldsymbol{\epsilon})} f_0(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\
&= \int |f_1^{1/2}(\boldsymbol{\epsilon}) - f_2^{1/2}(\boldsymbol{\epsilon})| f_0^{1/2}(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\
&= C||\sqrt{f_1} - \sqrt{f_2}||_{L_1(Q)},
\end{aligned}$$

where C is some positive finite constant, Q is the probability measure corresponding to $\sqrt{f_0}$ (after normalization), and by condition (C2) this measure is well defined. Now, let $\mathcal{F}^{1/2} = \{\sqrt{f_{\boldsymbol{\theta},h}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, h \in \mathcal{H}\}$. Since f_0 is fixed, the above equality gives $N_{[\,]}(\epsilon, \mathcal{R}, L_1(P)) \le N_{[\,]}(\epsilon/C, \mathcal{F}^{1/2}, L_1(Q))$, for some $0 < C < \infty$.

Since by (C3), $\mathcal{F}^{1/2}$ is a collection of bounded continuous functions on $(R^+)^b$, so by Corollary 2.7.4 in van der Vaart and Wellner (1996, p.158), with notations (V, d, α, r) there corresponds to $(1, 1, 1, 1)$ here,

$$\log N_{[\,]}(\epsilon, \mathcal{F}^{1/2}, L_1(Q)) = O(\frac{1}{\epsilon}).$$

Thus, for some generic positive finite constant C,

$$\begin{aligned}
N_{[\,]}(\epsilon, \mathcal{R}, L_1(P)) \le N_{[\,]}(\frac{\epsilon}{C}, \mathcal{F}^{1/2}, \\
L_1(Q)) \le \exp\{C/\epsilon\} < \infty, \qquad \forall\ \epsilon > 0,
\end{aligned}$$

and so by Theorem 2.4.1 in van der Vaart and Wellner (1996, p.122), $\mathcal{R}$ is a Glivenko-Cantelli class with respect to P, and completes the proof.

Proof of Theorem 2. Let $\mathbf{i}^*(\boldsymbol{\theta})$ be the efficient score of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \lambda)'$ in model (2). For fixed $\boldsymbol{\theta}$, let $\hat{\mathbf{h}} = \hat{\mathbf{h}}(\boldsymbol{\theta})$ be the maximizer of likelihood (3) on $\mathcal{H}_n$, with $\mathcal{H}_n$ given in the proof of Theorem 1. Note that (C5) implies $h_0(\cdot)$ is uniformly continuous, so as $n \to \infty$, $\hat{h}$ will be uniformly close to the global maximizer of h, and conditions (8)-(11) in Murphy and van der Vaart (2000, p.456) can be satisfied, and by their Theorem 1, their expressions (4) and (5) hold, and their (5) gives the desired result, see also Proposition 2 in Severini and Wong (1992), i.e.,

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \Omega^{-1}(\boldsymbol{\theta}_0)), \\
\Omega(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0}[\mathbf{i}^*(\boldsymbol{\theta}_0)\mathbf{i}^{*'}(\boldsymbol{\theta}_0)].
\end{aligned}$$

$I^*(\boldsymbol{\theta}_0)$ is the efficient Fisher information in the presence of the nuisance $h(\cdot)$. Theorem 1 in Murphy and van der Vaart (2000) also requires some other conditions, such as the score function is P-Donsker over some neighborhood of the parameters, and the Hessian matrix is P-Glivenko-Cantelli over some neighborhood of the parameters. These conditions can be easily met under mild conditions, as long as the class $\mathcal{H}$ of h's is regular. Checking conditions (8)-(11) in Murphy and van der Vaart is nontrivial (checking conditions of Proposition 2 in Severini and Wong (1992) may be just as difficult as they also require the consistency of the derivative of $\hat{h}(\cdot)$ at some rate), but it can followed by their lines for checking these conditions for the Cox model, in which the base line hazard function is maximized only at the observed data points, like our $\hat{h}$.

Although our current set up is for 1-dimensional y and ϵ, our below proofs are for multi-dimensional case for possible extension to this cases. Denote $\dot{g}(\mathbf{s}) = (\partial g(\mathbf{s})/\partial s_1, ..., \partial g(\mathbf{s})/\partial s_b)'$, $\dot{h}(\mathbf{s}) = (\partial h(\mathbf{s})/\partial s_1, ..., \partial h(\mathbf{s})/\partial s_b)'$; $\dot{g}\dot{g}'(\mathbf{s}) = \dot{g}(\mathbf{s})\dot{g}'(\mathbf{s})$; $\ddot{g}(\mathbf{s}) = (\partial^2 g(\mathbf{s})/(\partial s_i \partial s_j))_{1 \le i,j \le b}$. To compute the derivative $\partial \ell(\boldsymbol{\theta}, h)/\partial \boldsymbol{\beta}$, we treat $\boldsymbol{\beta}$ as a db-dimensional column vector $(\beta_{11}, ..., \beta_{1d}, ..., \beta_{b1}, ..., \beta_{bd})'$. Let $\mathbf{1}_b$ be the b-dimensional column vector of 1's. For vectors $\mathbf{a} = (a_1, ..., a_b)'$ and $\mathbf{x} = (x_1, ..., x_d)'$, denote $\mathbf{a} \otimes \mathbf{x} = (a_1 + \cdots + a_b)(\mathbf{x}', ..., \mathbf{x}')'$, a bd-dimensional column vector. When $dim(a) = 1$, $a \otimes \mathbf{x} = a\mathbf{x}$.

The computation of I^* is nontrivial, we only compute I_0^*. The log-likelihood for model (2) is

$$\begin{aligned}
\ell(\boldsymbol{\theta}, h) = \log f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}, h) \\
= \log\left(\lambda g(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha}) + (1 - \lambda)h(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha})\right).
\end{aligned}$$

Under the above log-likelihood, the score for $(\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ is

$$\begin{aligned}
\dot{\ell}_0(\boldsymbol{\theta}, h) = \frac{\partial \ell(\boldsymbol{\theta}, h)}{\partial(\boldsymbol{\beta}', \boldsymbol{\alpha}')'} \\
= -\frac{\lambda \dot{g}(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha}) + (1 - \lambda)\dot{h}(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\delta}'\boldsymbol{\alpha})}{f(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}, h)}\mathbf{z}.
\end{aligned}$$

Let Λ be the nuisance space of $h \in \mathcal{H}$, $\Lambda^\perp$ be its orthogonal complement, and $\Pi(\dot{\ell}_0|\Lambda)$ denote the projection of $\dot{\ell}_0$ onto Λ. The efficient score of $(\boldsymbol{\beta}', \boldsymbol{\alpha}', \lambda)'$ in the presence of the nuisance parameter h is

$$\mathbf{i}^* = \Pi(\dot{\ell}_0|\Lambda^\perp) = \dot{\ell}_0 - \Pi(\dot{\ell}_0|\Lambda).$$

Let $\mathcal{B}_0$ be the σ field generated by $\boldsymbol{\epsilon}$, since $(\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ is regression/location parameters, by (5.16) in Tsiatis (2006, p.108), or Proposition 4.3.2 in Bickel et al. (1993, p.108),

$$\begin{aligned}
\mathbf{i}_0^*(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) = \dot{\ell}_0(\boldsymbol{\theta}, h|\mathbf{y}, \mathbf{z}) - E\big(\dot{\ell}_0(\boldsymbol{\theta}, h|\mathbf{y}, \mathbf{z})\big|\mathcal{B}_0\big) \\
= \dot{\ell}_0(\boldsymbol{\theta}, h|\mathbf{y}, \mathbf{z}) - E\big(\dot{\ell}_0(\boldsymbol{\theta}, h|\mathbf{y}, \mathbf{z})\big|\mathbf{z}\big) = \dot{\ell}_0(\boldsymbol{\theta}, h|\mathbf{y}, \mathbf{z}).
\end{aligned}$$

Proof of Theorem 3. With the presence of infinite dimensional nuisance parameters, Corollary 9.11 in van der Vaart (1999, p.442-443) gives a simple proof for profile likelihood for the test $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs $\boldsymbol{\theta} \ne \boldsymbol{\theta}_0$, below we modify this proof for our result. By definition of the profile MLE $\tilde{\boldsymbol{\beta}}$, under our H_0, $\dot{\tilde{\ell}}_n(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}\mathbf{1}) = \mathbf{0}$, so

$$\tilde{\ell}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1}) - \tilde{\ell}_n(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}\mathbf{1}) = (\boldsymbol{\beta}_0' - \tilde{\boldsymbol{\beta}}', \alpha_0 - \tilde{\alpha})\dot{\tilde{\ell}}_n(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}\mathbf{1})$$

$$+\frac{1}{2}(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)\ddot{\tilde{\ell}}_n(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}\mathbf{1})(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)$$
$$+o_p\left(\sqrt{n}||(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)||\right)^2$$
$$=\frac{1}{2}(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)\ddot{\tilde{\ell}}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1})(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)'$$
$$+o_p\left(\sqrt{n}||(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)||\right)^2.$$

Let $\tilde{\mathbf{i}}_0$ be the efficient score for $(\boldsymbol{\beta}_0', \alpha_0)'$ under H_0, and $\tilde{I}_0 = E_{(\boldsymbol{\beta}_0, \alpha_0\mathbf{1}, h_0)}[\tilde{\mathbf{i}}_0\tilde{\mathbf{i}}_0']$ be the efficient Fisher information. Similarly as in the proof of Theorem 2, $(\tilde{\boldsymbol{\beta}}', \tilde{\alpha})'$ is efficient for $(\boldsymbol{\beta}_0', \alpha_0)'$ under H_0, thus one must have

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)' = \tilde{I}_0^{-1}n^{-1/2}\dot{\tilde{\ell}}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1}) + o_p(1).$$

Since $n^{-1/2}\dot{\tilde{\ell}}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1}) \xrightarrow{D} N(\mathbf{0}, \tilde{I}_0^{-1})$, we get

$$\begin{aligned}
&2\big(\tilde{\ell}_n(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}\mathbf{1}) - \tilde{\ell}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1})\big)\\
&= -(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)\ddot{\tilde{\ell}}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1})(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)'\\
&\quad + o_p\left(\sqrt{n}||(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)||\right)^2\\
&= \sqrt{n}(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)\tilde{I}_0\sqrt{n}(\tilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}_0', \tilde{\alpha}-\alpha_0)' + o_p(1)\\
&= n^{-1/2}\dot{\tilde{\ell}}_n'(\boldsymbol{\beta}_0, \alpha_0\mathbf{1})\tilde{I}_0^{-1}n^{-1/2}\dot{\tilde{\ell}}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1}) + o_p(1)\\
&\xrightarrow{D} \mathbf{Z}'\mathbf{Z} = \chi^2_{d+1},
\end{aligned}$$

where $\mathbf{Z} \sim N(\mathbf{0}, I_{d+1})$, I_{dk} is the $(d+1)$-dimensional identity matrix, χ^2_{d+1} is the chi-squared random variable with $(d+1)$ degrees of freedom.

Similarly,

$$2\big(\tilde{\ell}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\ell}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1})\big) \xrightarrow{D} \chi^2_{d+k}$$

and so

$$\begin{aligned}
2\big(\tilde{\ell}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\ell}_n(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}\mathbf{1})\big) &= 2\big(\tilde{\ell}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\ell}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1})\big)\\
-2\big(\tilde{\ell}_n(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}\mathbf{1}) - \tilde{\ell}_n(\boldsymbol{\beta}_0, \alpha_0\mathbf{1})\big) &\xrightarrow{D} \chi^2_{k-1},
\end{aligned}$$

the last step above is not straight forward as it looks, but is the same as the standard proof of Wilks' Theorem.

Received 14 July 2016

REFERENCES

Akkerhuis, K. M., Deckers, J. W., Boersma, E. et al. (2000). Geographic variability in outcomes within an international trial of glycoprotein IIb/IIIa inhibition in patients with acute coronary syndromes. *European Heart Journal*, **21**, 371–381.

Blair, J. E. A., Zannad, F., Konstam, M. A. et al. (2008). Continental differences in clinical characteristics, management, and outcomes in patients hospitalized with worsening heart failure. *Journal of the American College of Cardiology*, **52**, 1640–1648.

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

Chow, S. C., Shao, J., Hu, Y. P. (2002). Assessing sensitivity and similarity in bridging studies. *Journal of Biopharmaceutical Statistics*, **12**, 385–400. MR2190517

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Ser. B, **39**, 1–38. MR0501537

Fedorov, V., Jones, B. (2005). The design of multicentre trials. *Statistical Methods in Medical Research*, **14**, 205–248. MR2138241

Hsiao, C. F., Xu, J. Z., Liu, J. P. (2003). A group sequential approach to evaluation of bridging studies. *Journal of Biopharmaceutical Statistics*, **13**(4), 793–801.

Hsiao, C. F., Hsua, Y. Y., Tsoua, H. H., Liu, J. P. (2007). Use of prior information for Bayesian evaluation of bridging studies. *Journal of Biopharmaceutical Statistics*, **17**(1), 109–121. MR2307619

Huang, Q., Chen, G. et al. (2012). Design and sample size considerations for simultaneous globe drug development program. *Journal of Biopharmaceutical Statistics*, **22**, 1060–1073. MR2970574

Hung, H. M. J., Wang, S. J., O'Neill, R. T. (2010). Consideration of regional difference in design and analysis of multi-regional trials. *Pharmaceutical Statistics*, **9**, 173–178.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.*, **1**, 221–233. MR0216620

Kawai, N. et al. (2008). An approach to rationalize partitioning sample size into individual regions in a multiregional trial. *Drug Information Journal*, **42**, 139–147.

Liu, J. P., Chow, S. C. (2002). Bridging studies in clinical development. *Journal of Biopharmaceutical Statistics*, **12**(3), 359–367.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E. (1953), Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.

Ministry of Health, Labour and Welfare of Japan, (2007). *Basic Principles on Global Clinical Trials.*

Olkin, I. (1987). A semiparametric approach to density estimation. *Journal of the American Statistical Association*, **82**, 858–865. MR0909993

O'Neill, R. T. (2009). Multi-regional clinical trials: a regulatory perspective on issues. Presented at the *FDA/ Industry Statistics Workshop*, Washington, DC, September 24, 2009.

OShea, J. C., DeMets, D. L. (2001). Statistical issues relating to international differences in clinical trials. *American Heart Journal*, **142**, 21–28.

Pfanzagl, J. (1969). On the measurability and consistency of minimum contrast estimators. *Metrika*, **14**, 249–272.

Quan, H., Zhao, PL., Zhang, J., Rossner, M., Aizawa, K. (2010). Sample size considerations for Japanese patients in a multi-reginoal trial based on MHLW guidance. *Pharmaceutical Statistics*, **9**, 100–112.

Quan, H., Li, M., Shih. J. W., Ouyang, S. P., Chen, J., Zhang, J., Zhao, P. L. (2013). Empirical shrinkage estimator for consistency assessment of treatment effects in multi-regional clinical trials. *Statistics in Medicine*, **32**, 1691–1706. MR3060635

Quan, H., Mao, X., Chen, J., Shih, W. J., Quyang, S. P., Zhang, J., Zhao, P. L., Binkowitz, B. (2014). Multi-regional clinical trial design and consistency assessment of treatment effects. *Statistics in Medicine*, **33**(13), 2191–2205. MR3256686

Senn, S. J. (1998). Some controversies in planning and analyzing multi-center trials. *Statistics in Medicine*, **17**, 1753–1765.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics*, **20**, 1768–1802. MR1193312

Shih, W. J. (2001). Clinical trials for drug registrations in Asian-Pacific countries: Proposal for a new paradigm from a statistical perspective. *Controlled Clinical Trials*, **22**, 357–366.

Tan, M., Tian, G.-L., Ng, K. W. (2009). *Bayesian Missing Data Problems: EM, Data Augmentation and Non-iterative Computation*, London and Boca Raton, Florida: Chapman and Hall/CRC MR2562244

Tsaitis, A. A. (2006). *Semiparametric Theory and Missing Data*, Springer, New York. MR2233926

Uesaka, H. (2009). Sample size allocation to regions in a multiregional trial. *Journal of Biopharmaceutical Statistics*, **19**, 580–594. MR2751520

Wang, S. J. (2010). Multi-regional clinical trials – what are the challenges? *Pharmaceutical Statistics*, **9**, 171–172.

Vickers, A., Goyal, N., Harland, R., Rees, R. (1998). Do certain countries produce only positive results? A systematic review of controlled trials. *Controlled Clinical Trials*, **19**, 159–166.

Ao Yuan
Department of Biostatistics,
Bioinformatics and Biomathematics
Georgetown University
Washington DC 20057
USA
E-mail address: ay312@georgetown.edu

Shuxin Wang
Department of Biostatistics,
Bioinformatics and Biomathematics
Georgetown University
Washington DC 20057
USA
E-mail address: sw868@georgetown.edu

Ming T. Tan
Department of Biostatistics,
Bioinformatics and Biomathematics
Georgetown University
Washington DC 20057
USA
E-mail address: mtt34@georgetown.edu