Appendix A

Proofs of Theorems

A.1 Proofs of Theorems 1-3 in Section 2

To simplify the notations, let $T_i = \sqrt{n}(\hat{\beta}_i - \beta_i)$, $T_i^* = \sqrt{n}(\beta_i^* - \hat{\beta}_i)$, $d_{\max} = \beta_{\max} - \max_{i \notin H} \beta_i$ and $V = (V_1, \ldots, V_k)$ where $V_i = \hat{\beta}_{\max} - \beta_{\max} - \hat{\beta}_i + \beta_i$. We start with two lemmas that are essential for the proof of Theorem 1.

Lemma 1.1: *Under Assumption 1.1, we have*

$$\sup_{x \in R} |P(\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max}) \leq x) - P(\max_{i \in H} T_i \leq x)| \to 0$$

as $n \to \infty$.

Proof: For any $x \in R$, we have

$$\begin{aligned}
&|P(\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max}) \leq x) - P(\max_{i \in H} T_i \leq x)| \\
=&|P(\max_{i \in H} T_i \leq x, T_j \leq x + \sqrt{n}(\beta_{\max} - \beta_j), j \notin H) - P(\max_{i \in H} T_i \leq x)| \\
\leq& 1 - P(\max_{i \notin H} T_i \leq x + \sqrt{n} d_{\max}).
\end{aligned} \tag{A.1}$$

For any $x \in R$, by Assumption 1.1, we have $1 - P(\max_{i \notin H} T_i \leq x + \sqrt{n} d_{\max}) \to 0$, so

$$|P(\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max}) \leq x) - P(\max_{i \in H} T_i \leq x)| \to 0$$

as $n \to \infty$. The result follows naturally from the property of the cdf. □

Lemma 1.2: *Under Assumptions 1.1 and 1.2, for* $0 < r < 0.5$*, we have*

$$\sup_{x \in R} |P^*(\sqrt{n}(\beta^*_{\max,\text{modified}} - \hat{\beta}_{\max}) \leq x) - P(\max_{i \in H} T_i \leq x)| \to 0$$

as $n \to \infty$*, in probability w.r.t.* P.

Proof: Similar to the proof in Lemma 1.1, we have

$$\begin{aligned}
&|P^*(\sqrt{n}(\beta^*_{\max,\text{modified}} - \hat{\beta}_{\max}) \leq x) - P^*(T_i^* \leq x + n^r V_i, i \in H)| \\
=&|P^*(T_i^* \leq x + n^r V_i, i \in H, T_j^* \leq x + n^r V_j + n^r(\beta_{\max} - \beta_j), j \notin H) - P^*(T_i^* \leq x + n^r V_i, i \in H)| \\
\leq& 1 - P^*(T_j^* \leq x + n^r d_{\max} - n^r ||V||_\infty, j \notin H).
\end{aligned} \tag{A.2}$$

When $0 < r < 0.5$, by Assumptions 1.1 and 1.2, we have $n^r||V||_\infty \to 0$ in probability and

$$1 - P^*(T_j^* \le x + n^r d_{\max} - n^r||V||_\infty, j \notin H) \to 0$$

in probability. By Assumptions 1.1 and 1.2, we have $|P^*(T_i^* \le x + n^r V_i) - P(\max_{i \in H} T_i \le x)| \to 0$ in probability so

$$|P^*(\sqrt{n}(\beta^*_{\max,\text{modified}} - \hat{\beta}_{\max}) \le x) - P(\max_{i \in H} T_i \le x)| \to 0$$

in probability w.r.t P. The result is naturally followed by the property of cdf. □

Proof of Theorem 1: It follows from Lemmas 1.1 and 1.2. □

Proof of Corollary 1.1: The result is true, because $\sqrt{n}(\beta_s - \beta_{\max}) \to 0$ in probability. Otherwise, by the definition of β_s, the probability of the event, $\max_{i \notin H} \hat{\beta}_i \ge \max_{i \in H} \hat{\beta}_i$, will not go to 0 asymptotically, which violates the consistency implied in Assumption 1.1. □

Lemma 2.1: *Under Assumptions 1.1 and 2.1, we have*

$$|E[\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max})] - E[\max_{i \in H} T_i]| \to 0$$

as $n \to \infty$.

Proof: By Assumption 2.1, we have $E[\max_{i \in H} T_i]^2 \le E[\max_{i \in H} T_i^2] \le \sum_{i \in H} ET_i^2 < \infty$ and similarly $E[\max_{i \notin H} T_i]^2 < \infty$ uniformly in n, so

$$\begin{aligned} E[\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max})]^2 =& E[\max_{i \in H} T_i + \max(0, \max_{j \notin H} \sqrt{n}(\hat{\beta}_j - \max_{i \in H} \hat{\beta}_i))]^2 \\ \le& 2\{E[\max_{i \in H} T_i]^2 + E[\max(0, \max_{j \notin H}(T_j - \max_{i \in H} T_i))]^2\} < \infty \end{aligned} \tag{A.3}$$

uniformly in n. By Assumption 1.1 and Lemma 1.1, the lemma follows from the uniform integrability for $\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max})$; i.e.

$$\begin{aligned} & E|\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max})| I_{|\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max})| > c} \\ \le& [E[\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max})]^2 P(\sqrt{n}(\hat{\beta}_{\max} - \beta_{\max}) > c)]^{1/2} \to 0 \end{aligned} \tag{A.4}$$

uniformly in n as $c \to \infty$. □

Lemma 2.2: *Under Assumptions 1.1, 1.2, and 2.2, and for any* $0 < r < 0.5$, *we have*

$$E^*[\sqrt{n}(\beta^*_{\max,\text{modified}} - \hat{\beta}_{\max})] - E \max_{i \in H} T_i \to 0$$

as $n \to \infty$*, in probability w.r.t* P.

Proof: Similar to the proof in Lemma 2.1, by Assumption 2.2, we have $E^*[\max_{i \in H} T_i^* - n^r V_i]^2 < \infty$ and $E^*[\max_{i \notin H} T_i^* - n^r V_i]^2 < \infty$ uniformly in n in probability, and then

$$\begin{aligned} & E^*[\sqrt{n}(\beta^*_{\max,\text{modified}} - \hat{\beta}_{\max})]^2 \\ =& E^*[\max_{i \in H}(T_i^* - n^r V_i) + \max(0, \max_{j \notin H}(T_j^* - n^r(\beta_{\max} - \beta_j) - n^r V_j - \max_{i \in H}(T_i^* - n^r V_i)))]^2 \\ \le& 2\{E^*|\max_{i \in H}(T_i^* - n^r V_i)|^2 + E^*|\max(0, \max_{j \notin H}(T_j^* - n^r V_j - \max_{i \in H}(T_i^* - n^r V_i)))|^2\} < \infty \end{aligned} \tag{A.5}$$

uniformly in n. The lemma then follows from Assumptions 1.1, 1.2 and Lemma 1.2 with the similar argument of uniform integrability in Lemma 2.1 but for $\sqrt{n}(\beta^*_{\text{max,modified}} - \hat{\beta}_{\text{max}})$.

Proof of Theorem 2: By Lemmas 2.1 and 2.2, the result follows. $\square$

Proof of Corollary 2.1: It follows from Theorem 2 and similar arguments of the uniform integrability in Lemma 2.2 but for $E^*[\sqrt{n}(\beta^*_{\text{max,modified}} - \hat{\beta}_{\text{max}})]$. $\square$

Proof of Theorem 3: By definition, we have

$$\begin{aligned} & E[\hat{\beta}_{\text{max,reduced,1}}(r) - \beta_i]^2 \\ =& E[\hat{\beta}_{\text{max,reduced,1}}(r) - \beta_{\text{max}}]^2 + (\beta_{\text{max}} - \beta_i)^2 + 2E[\hat{\beta}_{\text{max,reduced,1}}(r) - \beta_{\text{max}}](\beta_{\text{max}} - \beta_i) \end{aligned} \tag{A.6}$$

By Assumptions in Corollary 2.1, we have $E[\hat{\beta}_{\text{max,reduced,1}}(r) - \beta_{\text{max}}] = o(1)$, which implies the desired result. $\square$

A.2 Proofs of Theorems 4-5 in Section 3

To simplify the notation, we let $K + d = \{c : \text{distance}(c, K) \leq d\} \bigcap D$, $\overline{K + d}$ denote the complement of the set $K + d$, $(K + d) - K = (K + d) \bigcap \overline{K}$, $G_n(c) = \sqrt{n}(\hat{\beta}(c) - \beta(c))$ and $G_n^*(c) = \sqrt{n}(\beta^*(c) - \hat{\beta}(c))$. By Assumption 4.3, we know γ_{max} is achievable and K is a compact set, so all the above notations are well defined.

Lemma 4.1: *Under Assumptions 4.1 and 4.3, for any $x \in R$,*

$$\lim_{n\to\infty} P(\sqrt{n}(\hat{\gamma}_{\text{max}} - \gamma_{\text{max}}) \leq x) = P(\sup_{c\in K} G(c) \leq x).$$

Proof: By definition, we have

$$\begin{aligned} & P(\sqrt{n}(\hat{\gamma}_{\text{max}} - \gamma_{\text{max}}) \leq x) \\ =& P(\sqrt{n}\sup_{c\in D}(\hat{\beta}(c) - \beta(c) + \beta(c) - \gamma_{\text{max}}) \leq x) \\ =& P(\max(\sup_{c\in K+d}(G_n(c) + \sqrt{n}(\beta(c) - \gamma_{\text{max}})), \sup_{c\in \overline{K+d}}(G_n(c) + \sqrt{n}(\beta(c) - \gamma_{\text{max}}))) \leq x). \end{aligned} \tag{A.7}$$

From Assumption 4.3, $\sqrt{n}(\beta(c) - \gamma_{\text{max}})$ converges to negative infinity uniformly in $\overline{K + d}$, so by Assumption 4.1, we have

$$\sup_{c\in\overline{K+d}}(G_n(c) + \sqrt{n}(\beta(c) - \gamma_{\text{max}})) \to -\infty$$

in probability. Therefore, the right hand side of (A.7) is asymptotically equivalent to

$$P(\sup_{c\in K+d}(G_n(c) + \sqrt{n}(\beta(c) - \gamma_{\text{max}})) \leq x) \tag{A.8}$$

for any given d. Since for $c \in K$, $\beta(c) = \gamma_{\max}$, and for $c \in (K+d) - K$, $\beta(c) < \gamma_{\max}$, we have the following inequality.

$$P(\sup_{c \in K+d} G_n(c) \leq x) \leq (\text{A.8}) \leq P(\sup_{c \in K} G_n(c) \leq x). \tag{A.9}$$

Let $n \to \infty$, under Assumption 4.2, we have

$$P(\sup_{c \in K+d} G(c) \leq x) \leq \liminf\,(\text{A.8}) \leq \limsup\,(\text{A.8}) \leq P(\sup_{c \in K} G(c) \leq x). \tag{A.10}$$

Let $L_n(x) = P(\sqrt{n}(\hat{\gamma}_{\max} - \gamma_{\max}) \leq x)$ and recall that $L_n(x)$ is asymptotically equivalent to (A.8). Therefore, for any $d > 0$, we have

$$P(\sup_{c \in K+d} G(c) \leq x) \leq \liminf L_n(x) \leq \limsup L_n(x) \leq P(\sup_{c \in K} G(c) \leq x). \tag{A.11}$$

Under Assumptions 4.1 and 4.3, $\lim_{d \to 0} \sup_{c \in K+d} G(c) = \sup_{c \in K} G(c)$, in probability. Let $d \to 0$ in (A.11), we prove the desired result. $\square$

Lemma 4.2: *Under Assumptions 4.1, 4.2 and 4.3 and* $0 < r < 0.5$, *for any* $x \in R$,

$$\lim_{n \to \infty} P^*(\sqrt{n}(\gamma^*_{\text{max,modified}} - \hat{\gamma}_{\max}) \leq x) = P(\sup_{c \in K} G(c) \leq x)$$

in probability.

Proof: With similar arguments as those in the proof of Lemma 4.1, we have

$$\begin{aligned}
&P^*(\sqrt{n}(\gamma^*_{\text{max,modified}} - \hat{\gamma}_{\max}) \leq x) \\
=&P^*(\sup_{c \in D}(G^*_n(c) - n^r \sup_{b \in D}[\hat{\beta}(b) - \hat{\beta}(c)]) \leq x) \\
=&P^*(\sqrt{n} \max(\sup_{c \in K+d}(G^*_n(c) - n^r \sup_{b \in D}(\hat{\beta}(b) - \hat{\beta}(c))), \sup_{c \in \overline{K+d}}(G^*_n(c) - n^r \sup_{b \in D}(\hat{\beta}(b) - \hat{\beta}(c)))) \leq x).
\end{aligned} \tag{A.12}$$

Let $L^*_n(x) = P^*(\sqrt{n}(\gamma^*_{\text{max,modified}} - \hat{\gamma}_{\max}) \leq x)$. We notice that

$$n^r \sup_{c \in \overline{K+d}} \sup_{b \in D}(\hat{\beta}(b) - \hat{\beta}(c)) \leq n^r \sup_{c \in \overline{K+d}} \sup_{b \in D}(\hat{\beta}(b) - \beta(b) - (\hat{\beta}(c) - \beta(c)) - n^r \inf_{c \in \overline{K+d}}(\gamma_{\max} - \beta(c)). \tag{A.13}$$

From Assumptions 4.1, 4.2 and 4.3, the 1st term of the right hand side of (A.13) converges to 0 in probability and the second term converges to negative infinite. Therefore, $n^r \sup_{b \in D}[\hat{\beta}(b) - \hat{\beta}(c)] \to -\infty$ uniformly for $c \in \overline{K+d}$ in probability and $L^*_n(x)$ is asymptotically equivalent to

$$P^*\left(\sup_{c \in K+d}(G^*_n(c) - n^r \sup_{b \in D}[\hat{\beta}(b) - \hat{\beta}(c)]) \leq x\right)$$

in probability. Similar to (A.13), we show that

$$\sup_{c \in K+d} |n^r \sup_{b \in D}[\hat{\beta}(b) - \hat{\beta}(c)] - n^r(\gamma_{\max} - \beta(c))| \to 0$$

in probability. Therefore, we have $L_n^*(x)$ is asymptotically equivalent to

$$P^* \left(\sup_{c\in K+d} (G_n^*(c) - n^r(\gamma_{\max} - \beta(c))) \leq x \right)$$

in probability. Now, we have the following inequality in probability.

$$P^*(\sup_{c\in K+d} G_n^*(c) \leq x) \leq \liminf L_n^*(x) \leq \limsup L_n^*(x) \leq P^*(\sup_{c\in K} G_n^*(c) \leq x).$$

With similar arguments used in the proof of Lemma 4.1 and Assumption 4.3, we prove the desired result. $\square$

Proof of Theorem 4: It naturally follows from Lemmas 4.1 and 4.2 with the property of the cdf. $\square$

With $\gamma_{\max} = \sup_c \beta(c)$, we let $S_n = \{c : \beta(c) \geq \gamma_{\max} - \log(n)/n\}$ and $\gamma_{ss} = \beta(\tilde{c})$, where $\tilde{c}$ is a random variable as the value of c that achieves the minimum of $\beta(c)$ among all possible values of $argmax_{c\in D}\hat{\beta}(c)$. We further denote the smallest value of c that achieves the maximum of $\beta(c)$, $\gamma_{\max}$, by c_0, which is a well-defined fixed value from our continuous and compactness assumptions. Then, we have the following lemma to characterize the distribution of γ_{ss}.

Lemma 5.1: *Under Assumption 4.1, we have*

$$P(\gamma_{\max} - \gamma_{ss} \leq \log(n)/n) \to 1.$$

In other words, $P(\tilde{c} \in S_n) \to 1$, as $n \to \infty$.

Proof: If $\tilde{c} \notin S_n$, then, $\hat{\beta}(c_0) < \sup_{c\in\bar{S}_n} \hat{\beta}(c)$. Therefore, we have,

$$\begin{aligned}
&P(\gamma_{\max} - \gamma_{ss} > log(n)/n) \\
\leq& P\Big(\hat{\beta}(c_0) < \sup_{c\in\bar{S}_n} \hat{\beta}(c)\Big) \\
=& P\Big(\sqrt{n}(\hat{\beta}(c_0) - \gamma_{\max}) < \sqrt{n} \sup_{c\in\bar{S}_n} (\hat{\beta}(c) - \beta(c) + \beta(c) - \gamma_{\max})\Big) \\
\leq& P\Big(\sqrt{n}(\hat{\beta}(c_0) - \gamma_{\max}) < \sqrt{n} \sup_{c\in\bar{S}_n} (\hat{\beta}(c) - \beta(c)) - \sqrt{n} \inf_{c\in\bar{S}_n} (\gamma_{\max} - \beta(c))\Big).
\end{aligned} \tag{A.14}$$

Since $\sqrt{n}\sup_{c\in\bar{S}_n}(\hat{\beta}(c)-\beta(c)) \leq \sqrt{n}\sup_{c\in D}(\hat{\beta}(c)-\beta(c))$ and $\sqrt{n}\inf_{c\in\bar{S}_n}(\gamma_{\max}-\beta(c)) \geq \log(n)$, the right hand side of (A.14) converges to 0 and we finish the proof. $\square$

Lemma 5.2: *Under Assumption 4.1 and for any fixed x, we have*

$$P\Big(\sqrt{n} \sup_{c\in S_n} (\hat{\beta}(c) - \beta(c)) \leq x\Big) \to P(\sup_{c\in K} G(c) \leq x).$$

Proof: Since $S \subseteq S_n \subseteq S_2$, we can use similar techniques as those in the proof of Lemma 4.1. $\square$

Lemma 5.3: *Let q_α be the $1-\alpha$ quantile of $\sup_{c\in K} G(c)$, then, under Assumption 4.1, we have*

$$\lim_{n\to\infty} P\Big(\sqrt{n}\sup_{c\in D}(\hat\beta(c)-\gamma_{ss})\le q_\alpha\Big)=1-\alpha.$$

Proof: From the definition, we know $\hat\beta(\tilde c)=\sup_{c\in D}\hat\beta(c)$ and $\gamma_{ss}=\beta(\tilde c)$ so we have

$$\begin{aligned}
&P\Big(\sqrt{n}\sup_{c\in D}(\hat\beta(c)-\gamma_{ss})\le q_\alpha\Big)\\
=&P\Big(\sqrt{n}[\hat\beta(\tilde c)-\beta(\tilde c)](I_{\tilde c\in S_n}+I_{\tilde c\notin S_n})\le q_\alpha\Big)\\
=&P\Big(\sqrt{n}(\hat\beta(\tilde c)-\beta(\tilde c))\le q_\alpha,\tilde c\in S_n\Big)+P\Big(\sqrt{n}(\hat\beta(\tilde c)-\beta(\tilde c))\le q_\alpha,\tilde c\notin S_n\Big).
\end{aligned}\tag{A.15}$$

From Lemma 5.1, we know the second part of the right hand side of (A.15) converges to 0. Notice that

$$\{\sqrt{n}(\hat\beta(\tilde c)-\beta(\tilde c))\le q_\alpha,\tilde c\in S_n\}\supseteq\{\sqrt{n}\sup_{c\in S_n}((\hat\beta(c)-\beta(c))\le q_\alpha,\tilde c\in S_n\},$$

from Lemmas 5.1 and 5.2, we have

$$\begin{aligned}
&\liminf_{n\to\infty} P\Big(\sqrt{n}(\hat\beta(\tilde c)-\beta(\tilde c))\le q_\alpha,\tilde c\in S_n\Big)\\
\ge&\liminf_{n\to\infty} P\Big(\sqrt{n}\sup_{c\in S_n}((\hat\beta(c)-\beta(c))\le q_\alpha,\tilde c\in S_n\Big)\\
=&\liminf_{n\to\infty} P\Big(\sqrt{n}\sup_{c\in S_n}((\hat\beta(c)-\beta(c))\le q_\alpha\Big)\to 1-\alpha.
\end{aligned}\tag{A.16}$$

Since $\gamma_{ss}\le\gamma_{\max}=\sup_{c\in D}\beta(c)$, from Lemma 4.1, we have

$$\limsup_{n\to\infty} P\Big(\sqrt{n}\sup_{c\in D}(\hat\beta(c)-\gamma_{ss})\le q_\alpha\Big)\le\limsup_{n\to\infty} P\Big(\sqrt{n}\sup_{c\in D}(\hat\beta(c)-\max_{c\in D}\beta(c))\le q_\alpha\Big)=1-\alpha$$

Therefore, we complete the proof. □

Proof of Theorem 5: It naturally follows from Lemma 5.3 and Theorem 4. □

Appendix B

Synthetic Data of MONET1 Trial

To generate the synthetic data of the MONET1 trial, we consider a simple setting of n observations, $(Y_i, D_i, \delta_i, Z_i)$, $i = 1, \cdots, n$, where Y_i is the (possibly censored) survival time of the i-th subject, D_i is the treatment indicator, δ_i is the censoring indicator, and $Z_i = (Z_{i,1}, \ldots, Z_{i,K})$ is the subgroup indicator indicating whether the subject belongs to any of the $2K$ subgroups we consider.

Following the MONET1 trial in Kubota et al. (2014), we let $n = 1090$ and $K = 8$ with the following subgroups: East Asian patient or not ($Z_{i,1} = 1$ or 0), received radiotherapy or not ($Z_{i,2} = 1$ or 0), stage IIIB or not ($Z_{i,3} = 1$ or 0), Age greater than 65 or not ($Z_{i,4} = 1$ or 0), ECOG PS equal to 0 or not ($Z_{i,5} = 1$ or 0), Adenocarcinoma histology or not ($Z_{i,6} = 1$ or 0), male or female ($Z_{i,7} = 1$ or 0), and never smoked or not ($Z_{i,8} = 1$ or 0). We independently let $Z_{i,k} \sim \text{Bernoulli}(1, p_k)$, and $D_i \sim \text{Bernoulli}(1, p)$, where the parameters are estimated by the sample proportion in Table 1 and Figure 1.A in Kubota et al. (2014) and given in Table B.1. We independently generate the event time T_i by the distribution F and the censoring time C_i by the distribution G as defined in Table B.2, both of which are estimated based on Figure 1.A in Kubota et al. (2014) under the assumptions of no treatment effect. Finally, we obtain $Y_i = \min(T_i, C_i)$ and $\delta_i = I_{T_i \leq C_i}$.

Table B.1: Proportion of the subgroups p_i and the proportion p of subjects with treatment

p	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
0.5	0.208	0.143	0.139	0.339	0.362	0.817	0.615	0.281

Table B.2: Distribution for the event time and the censoring time: $F(x) = P(T = x)$ and $G(x) = P(C = x)$

x	8	16	24	32	40	48	56	64	72	80
F	0.05	0.05	0.08	0.06	0.06	0.06	0.04	0.05	0.03	0.04
G	0	0	0	0	0	1/15	1/15	1/15	1/15	1/15
x	88	96	104	112	120	128	136	144	152	160
F	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.3
G	0	0	0	0	2/15	2/15	2/15	2/15	2/15	0

Bibliography

Kubota, K et al. (2014). “Phase III study (MONET1) of motesanib plus carboplatin/paclitaxel in patients with advanced nonsquamous nonsmall-cell lung cancer (NSCLC): Asian subgroup analysis”. In: *Annals of oncology* 25(2), pp. 529–536.