Group Sequential Designs for Non-proportional Hazards Alternatives

Cyrus Mehta, Ph.D and Pranab Ghosh, Ph.D

Cytel Inc. Cambridge, MA

email: mehta@cytel.com – web: www.cytel.com – tel: 617-661-2011

Acknowledgement

Joint work with

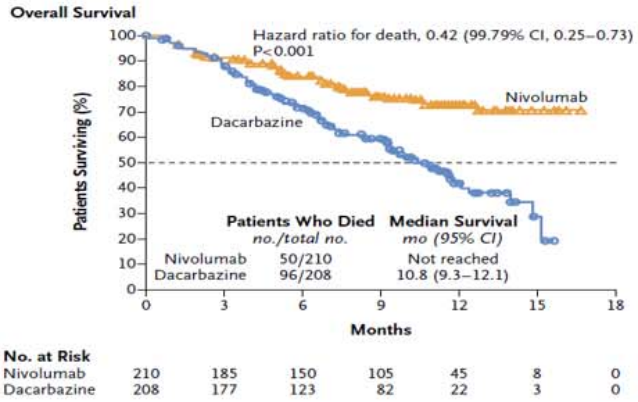
Christopher Jennison, Ph.D

University of Bath, UK

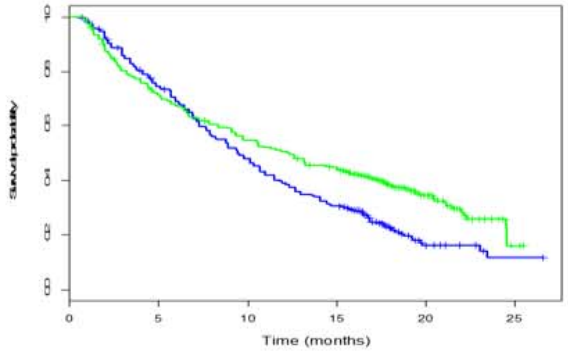
Presentation Outline

- Logrank test the industry standard for time-to-event trials
- But logrank test loses power for immuno-oncology trials
- Single-look Max-Combo test was proposed by Industry Consortium (Feb 2018)
- We have extended Max-Combo to group-sequential setting
- Power comparisons

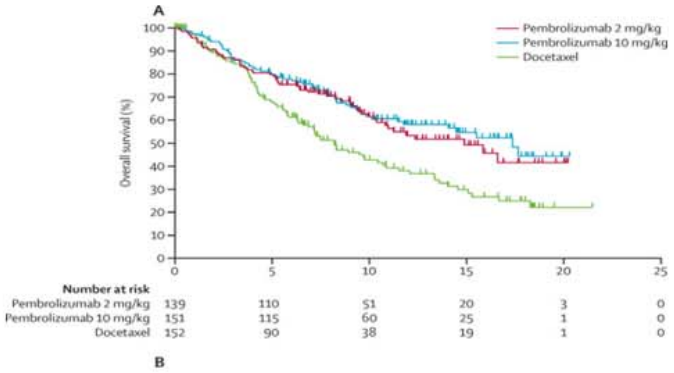
Recent IO trials have brought the concept of NPH in the forefront....



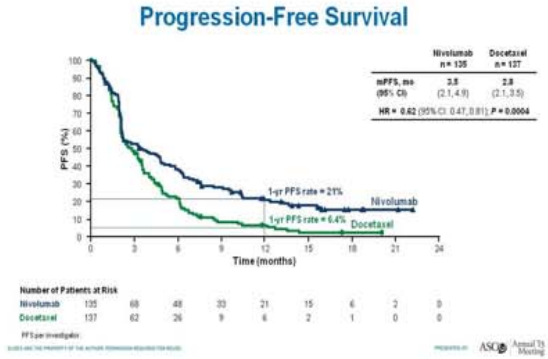
OS: Nivo in melanoma



OS: Nivo in NSCLC



OS: Pembro in NSCLC



PFS: Nivo in NSCLC

33

- Logrank test will lose power against these alternatives
- Slide presented at Duke-FDA public meeting, 5 Feb 2018

Alternative: Max Combo Test

- Weighted logrank with Harrington-Fleming $G^{\rho,\gamma}$ class of weights: $\hat{Q}(t) = \hat{S}(t)^{\rho}(1-\hat{S}(t))^{\gamma}$
- At k^{th} analysis time suppose d_k events were observed.
- H-F Score Statistics : $S_k^{\rho,\gamma} = \sum_{j=1}^{d_k} \hat{Q}(T_j)\,(d_{1j} - E(d_{1j}))$
- Standardized H-F Statistics : $Z_k^{\rho,\gamma} = \frac{S_k^{\rho,\gamma}}{\sqrt{Var(S_k^{\rho,\gamma})}}$
- Rather than rely on a single H-F statistic, take the max of three; representing equal, early, and late separation
- Max Combo Stat: $M_k = \max(Z_k^{0,0}, Z_k^{1,0}, Z_k^{0,1})$ at look k

Two-Stage Design

- Industry consortium (Duke-FDA public meeting, 2018) focused on single stage max combo test. Find c such that

$$P_0\{\max(Z^{0,0}, Z^{1,0}, Z^{0,1}) \geq c\} = \alpha$$

- But two-stage designs with early stopping may be more ethical due to possible prolonged survival in I-O trials

- Find c_1 and c_2 such that

$$P_0\{\max(Z_1^{0,0}, Z_1^{1,0}, Z_1^{0,1}) \geq c_1\} = \alpha_1$$

$$P_0\{\max(Z_1^{0,0}, Z_1^{1,0}, Z_1^{0,1}) < c_1, \max(Z_2^{0,0}, Z_2^{1,0}, Z_2^{0,1}) \geq c_2\} = \alpha - \alpha_1$$

Distribution of $\max(Z_k^{0,0}, Z_k^{1,0}, Z_k^{2,0})$

1. **Covariance between H-F stats is known (Karrison, 2016)**

$$Cov(Z_k^{\rho_1,\gamma_1}, Z_k^{\rho_2,\gamma_2}) = \frac{Var(S_k^{(\rho_1+\rho_2)/2,(\gamma_1+\gamma_2)/2})}{\sqrt{Var(S_k^{\rho_1,\gamma_1})Var(S_k^{\rho_2,\gamma_2})}}$$

where $Var(S_k^{\rho,\gamma}) = \sum_{j=1}^{d_k} \hat{Q}(t_j)^2 Var(d_{1j})$

2. **H-F statistics have independent increments (Tsiatis, 1982)**

$$Cov(S_{k+1}^{\rho_1,\gamma_1}, S_k^{\rho_2,\gamma_2}) = Cov(S_k^{\rho_1,\gamma_1}, S_k^{\rho_2,\gamma_2}) = Var(S_k^{(\rho_1+\rho_2)/2,(\gamma_1+\gamma_2)/2})$$

3. **These two results, combined with Ghosh et al (2018), provide joint distribution of $\max(Z_k^{0,0}, Z_k^{1,0}, Z_k^{2,0})$, $k = 1, 2$**

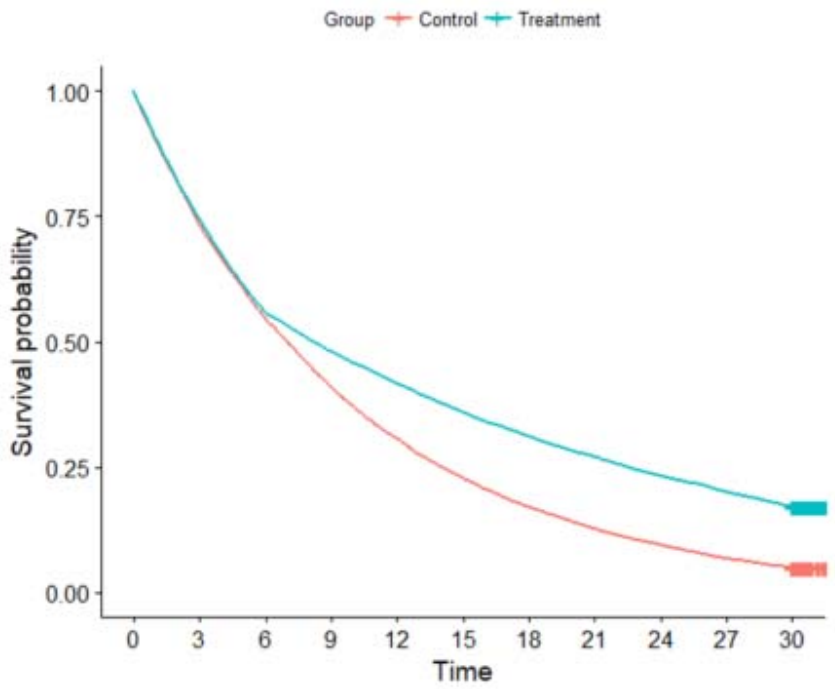
Max Combo vs Logrank for 2 Scenarios

Enroll 300 patients over 12 months; 7 month median for control arm; one interim analysis with $\gamma(-5)$ spending; final analysis at month 30

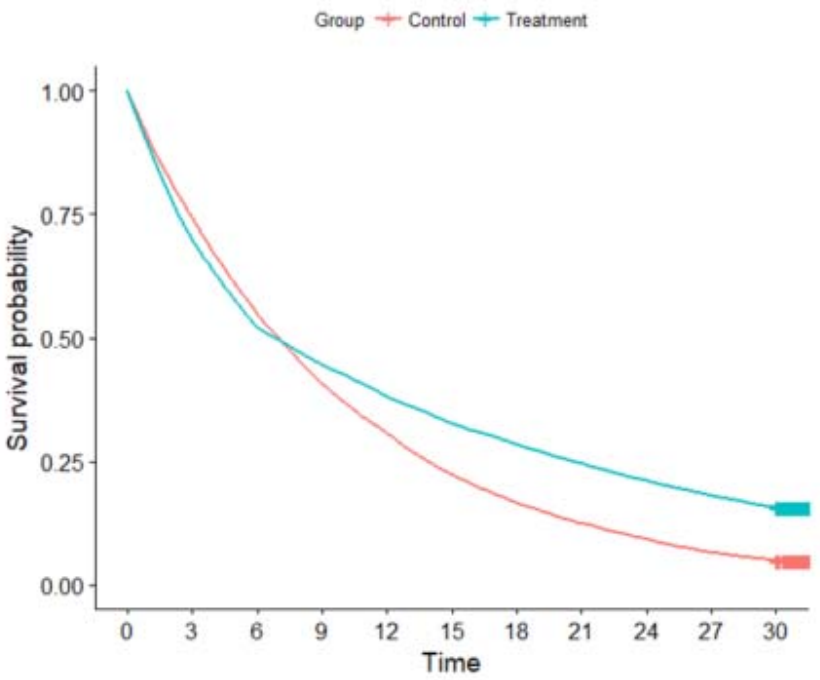
Late Separation **Hazard ratio is 1 for first 6 months and becomes 0.5 from 6 month onwards.**

Crossing Hazards **Hazard ratio is 1.2 for first 3 months, 1 for next 3 months and 0.5 thereafter,**

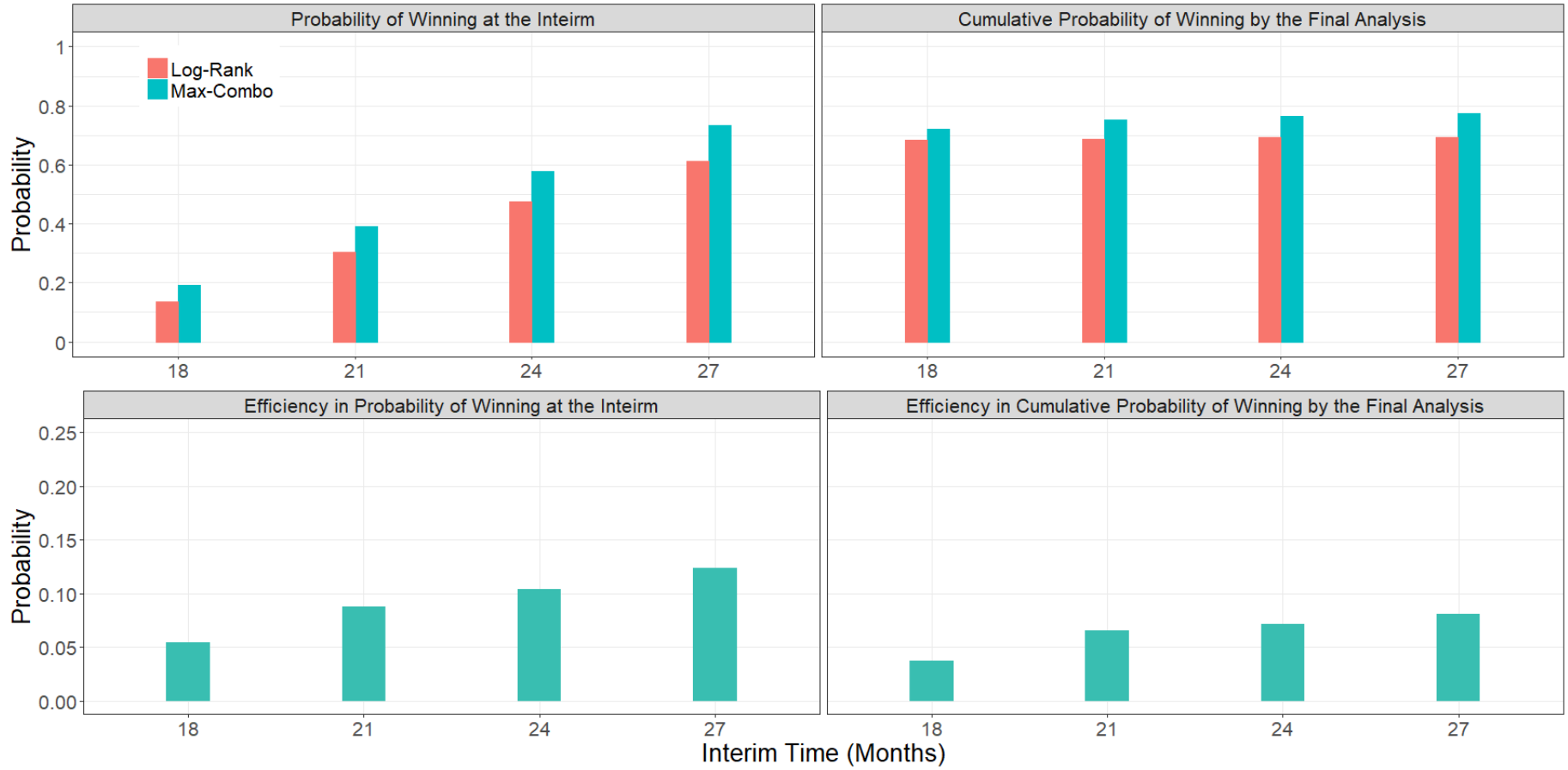
Late Separation



Crossing Hazards

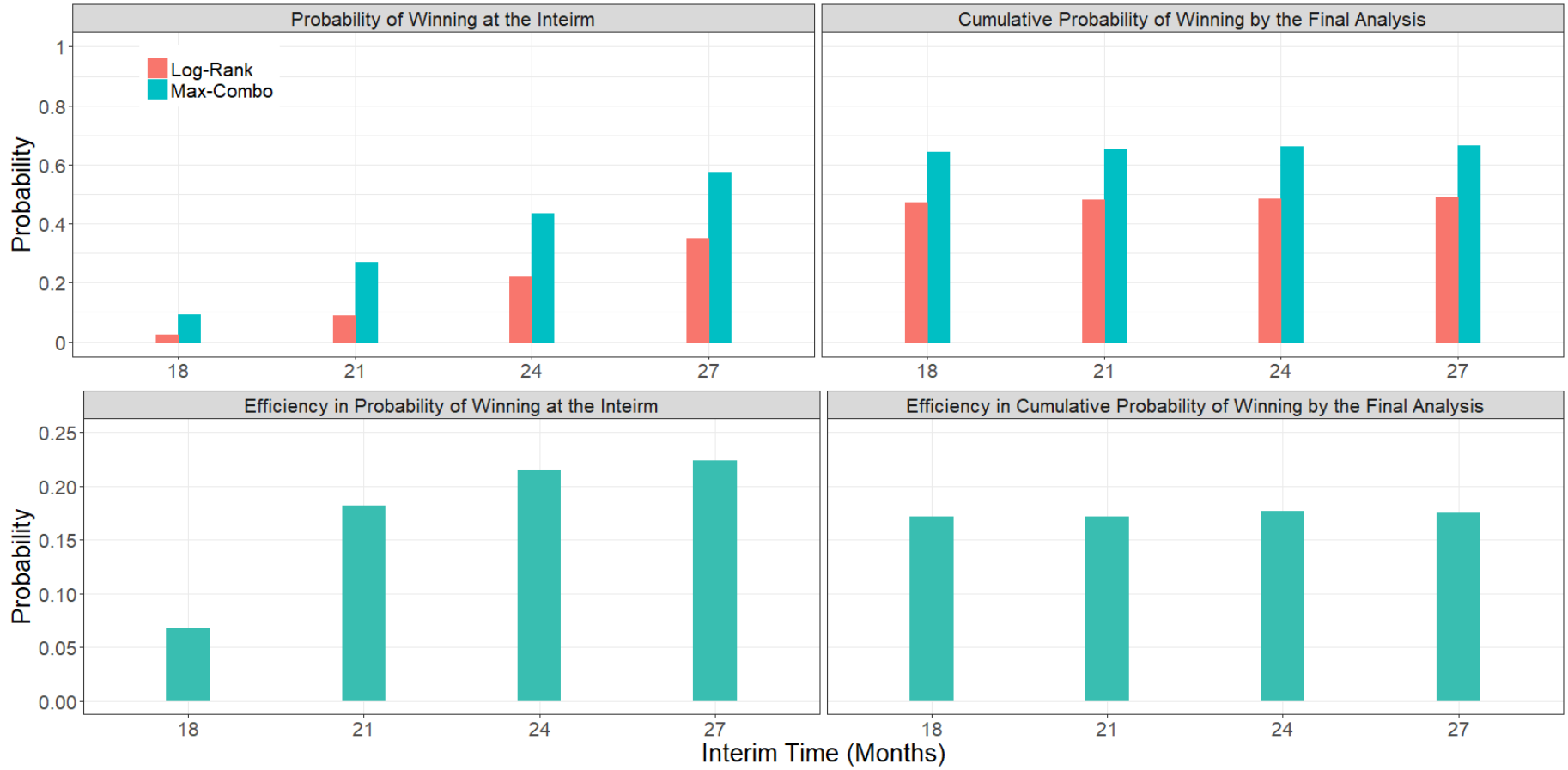


Late Separation: Logrank vs Max Combo



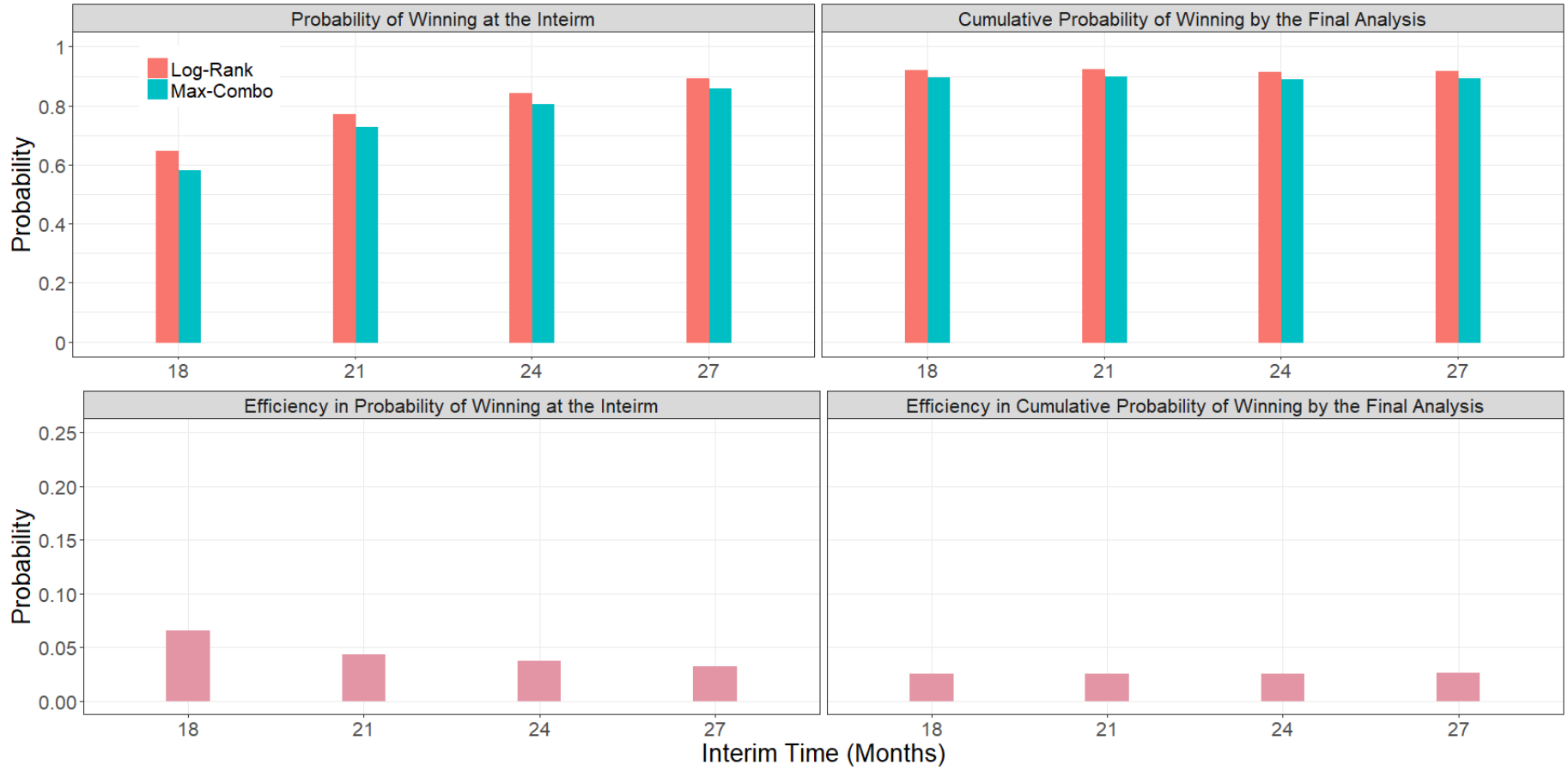
- **Up to 13% power gain at interim**
- **Up to 6% power gain at final**

Crossing Hazards: Logrank vs Max Combo



- **Up to 20% power gain at interim**
- **Up to 16% power gain at final**

Prop Hazards: Logrank vs Max Combo



- **Could lose up to 6% power at interim and 3% power at final with Max Combo if proportional hazards holds**

Cost-Benefit of Single Look vs 2-Stage

Scenario	Interim Analysis	Power		Power Loss of 2-Stage	Prob of Early Stop
		Single Look	2-Stage		
Late	18 mths	0.775	0.7223	0.0527	0.1916
	21 mths	0.775	0.7522	0.0228	0.391
	24 mths	0.775	0.7633	0.0117	0.5789
	27 mths	0.775	0.7735	0.0015	0.7347
Crossing	18 mths	0.671	0.6446	0.0264	0.0924
	21 mths	0.671	0.6527	0.0183	0.2709
	24 mths	0.671	0.6619	0.0091	0.4345
	27 mths	0.671	0.6644	0.006	0.5752

Concluding Remarks

- Logrank test is not the most powerful test if the proportional hazards assumption is violated
- We derived group sequential boundaries for the Max-Combo statistic by combining results from Tsiatis (1982), Karrison (2016), and Ghosh et al (2018). Industry consortium only considered single-look case
- Max-Combo two-stage test dominates Logrank two-stage test in terms of early stopping; up to 20% higher prob of early stopping; up to 12% gain in overall power
- Max-Combo two-stage has about 3% less power than Max-Combo single-stage, but can stop early