Development and evaluation of alternative tumor-response based metrics for immunotherapy clinical trials

Meihua Wang*, Chen Hu†, Iris Wu*, Heng Zhou*, Cong Chen*

*Merck & Co., Inc., Kenilworth, NJ, USA
† Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD

FDA-Industry Statistics Workshop, September 25, 2019

Outline

- Background
- Methods and Illustrations
- Simulation Setup and Results
- Remarks and Discussions

Background

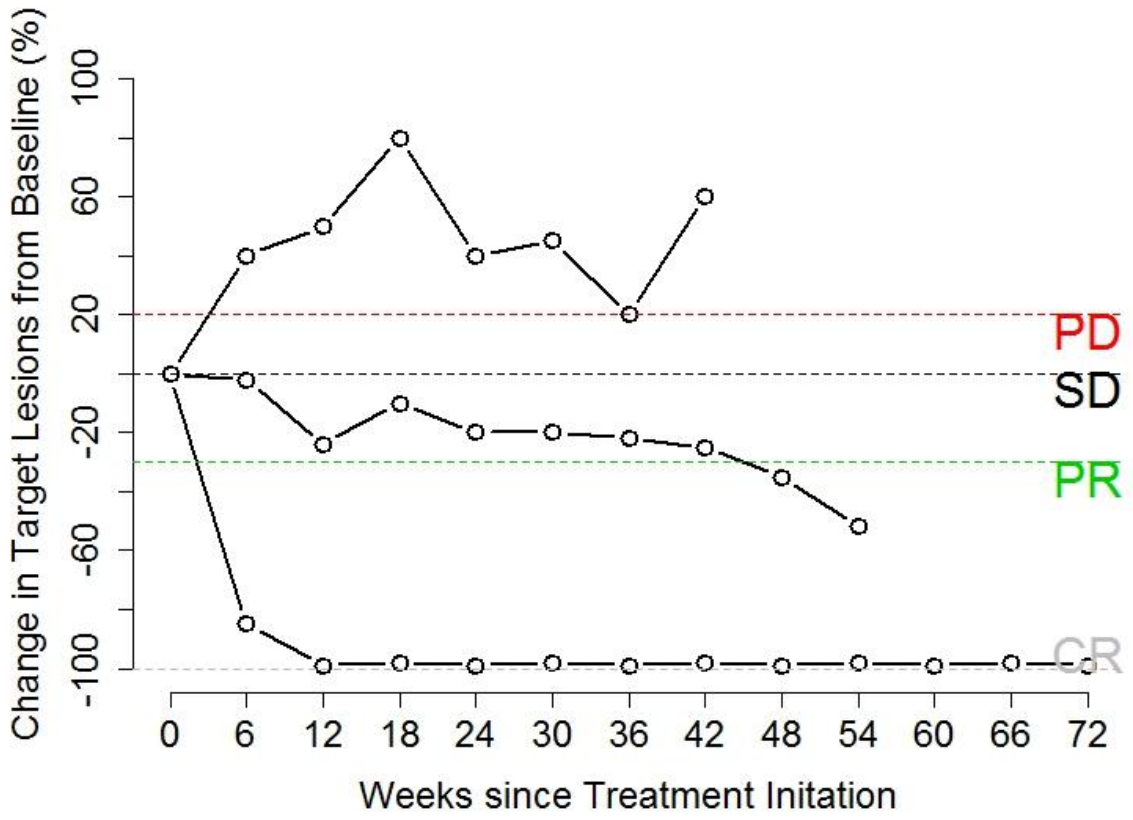
- Traditional endpoints:
  - Objective response rate (ORR)
  - Progression-free survival (PFS)
  - Overall survival (OS)
- Some observations in immuno-oncology trials
  - Low/intermediate ORR and prolonged duration of response
  - ORR, PFS and OS benefit could be seen
  - However,
    - Survival benefit seen in the absence of PFS benefit
    - Survival benefit seen in the absence of PFS and ORR benefit

Background

- Alternative efficacy metrics are needed to characterize the clinical benefit, and help go/no-go decision making process
- Can we incorporate Tumor Response and Duration into a single summary measure?

Methods

Examples of tumor burden trajectories over time



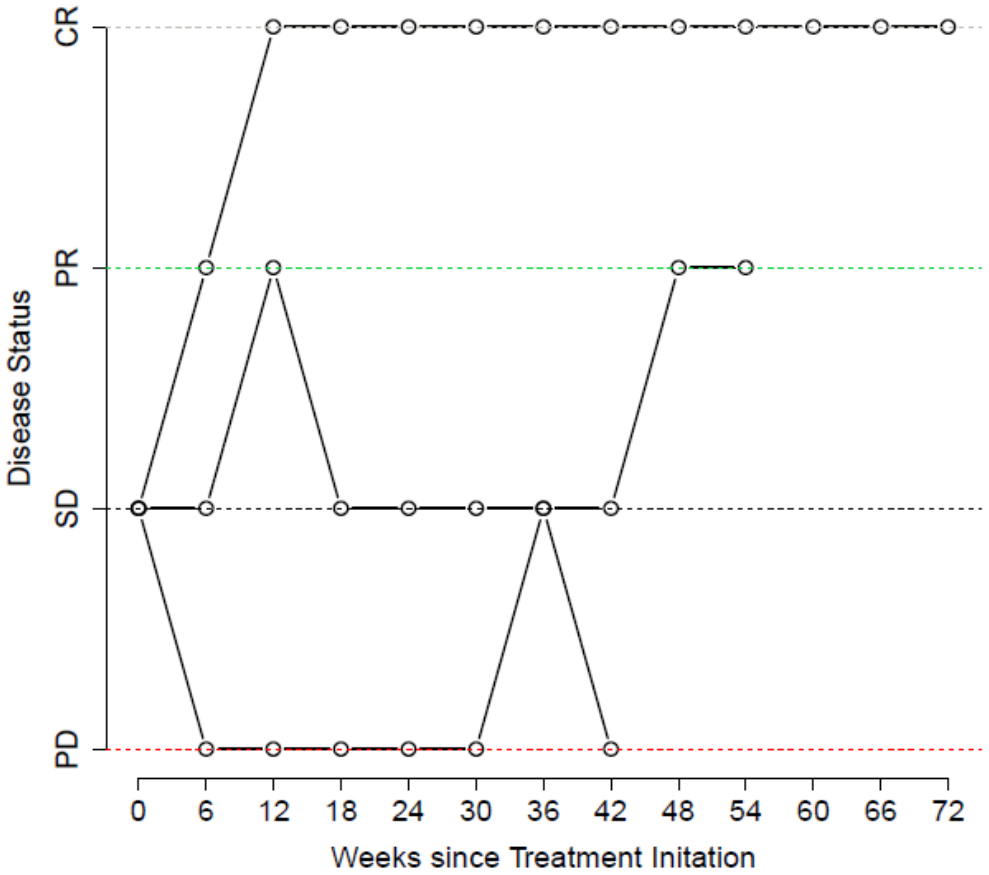
Methods

- Preferred regimen should both prolong survival and maintain a favorable tumor disease status
- Area Under Curve (AUC) of tumor kinetics (y-axis) could potentially carry more information than either alone (when cutoff timepoint is properly chosen)
- Without loss of generality, consider
  - Maintain better tumor disease status over time is more favorable outcome
  - Regimen(s) with larger AUC (on average) is preferred

Methods

- This efficacy metric is called “Cumulative Tumor Response Index (cTRI)”

$$\text{Cumulative Mean } \mu(t) = E\{M(t)\} = \int_0^t E\{Y^*(u)I(D \geq u)du\}$$

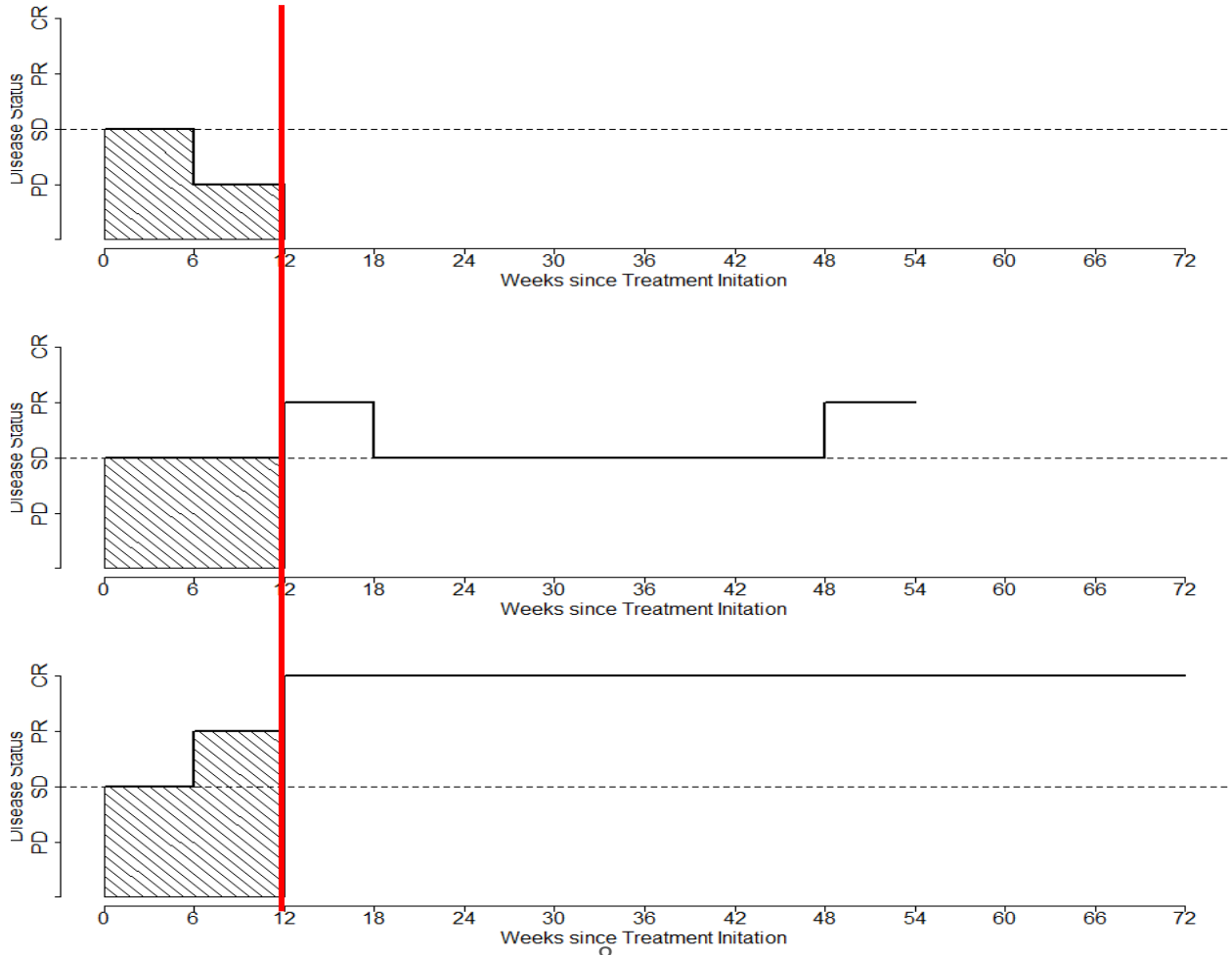


- Referred as the area under a tumor response trajectory up to a pre-specified timepoint
- A univariate summary measure integrating both temporal information (e.g., response duration) and tumor response status information
- Utility (weight) could be based on response categories, individual tumor size changes, etc.

MERCK

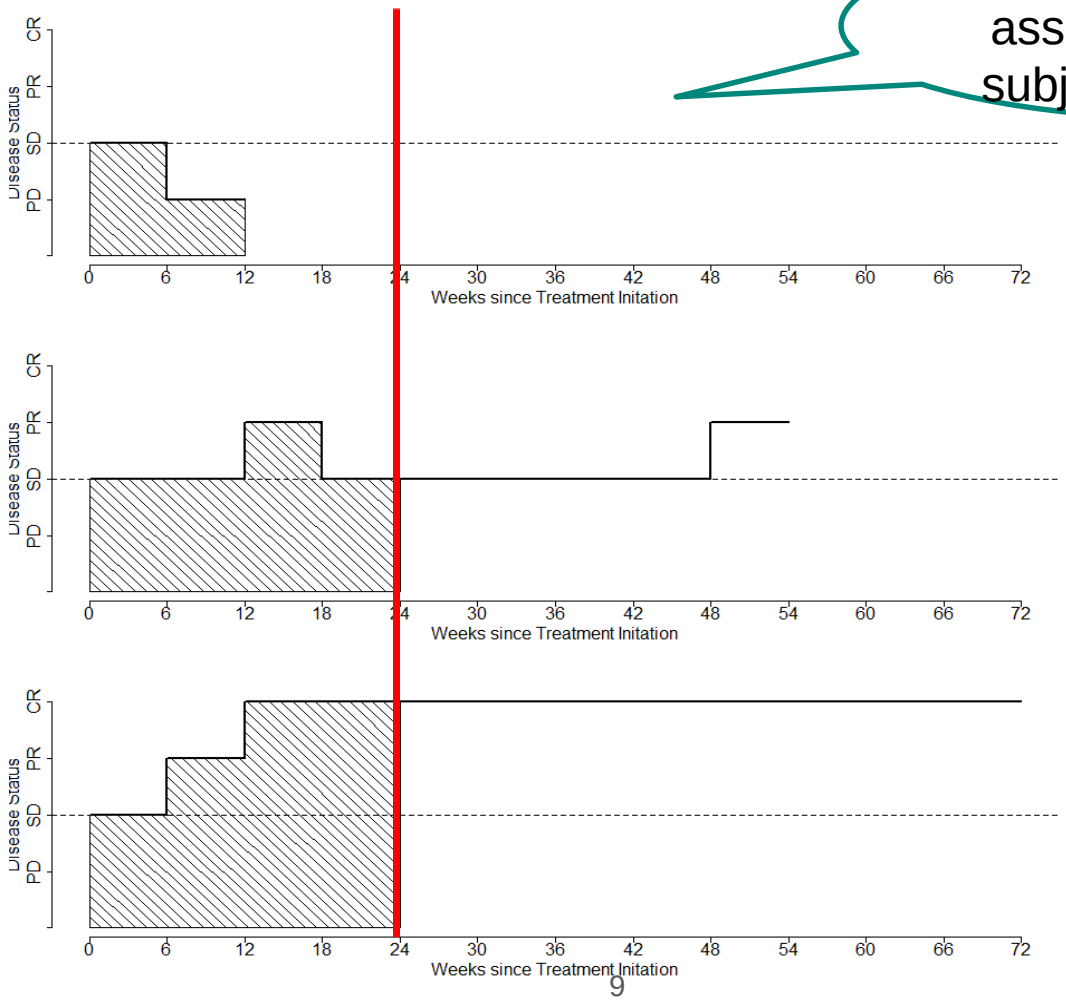
Methods

Examples: cTRI at week 12 is the shaded area



Methods

Examples: cTRI at week 24 is the shaded area



Method

- Calculation of cTRI
  - Subject to censoring as co-observed as OS
  - Naïve approach
    - Average of all study subjects
      - biased downward as responses after censoring are not accounted for
    - Average of all uncensored subjects
      - biased toward to patients with shorter survival times because larger survival times are more likely to be censored
    - Standard survival analysis methods do NOT apply either
      - cTRI at survival time is positively correlated with cTRI at censoring time, violating the independent censoring assumption

Method

- Calculation of cTRI
  - Proper approach
    - Consider the following inverse-probability of censoring weighting (IPCW) estimator

$$\hat{\mu}_L = \int_0^L \frac{1/n \sum_{i=1}^n \Delta_i(s) U_{is}}{\hat{K}(s)} ds$$

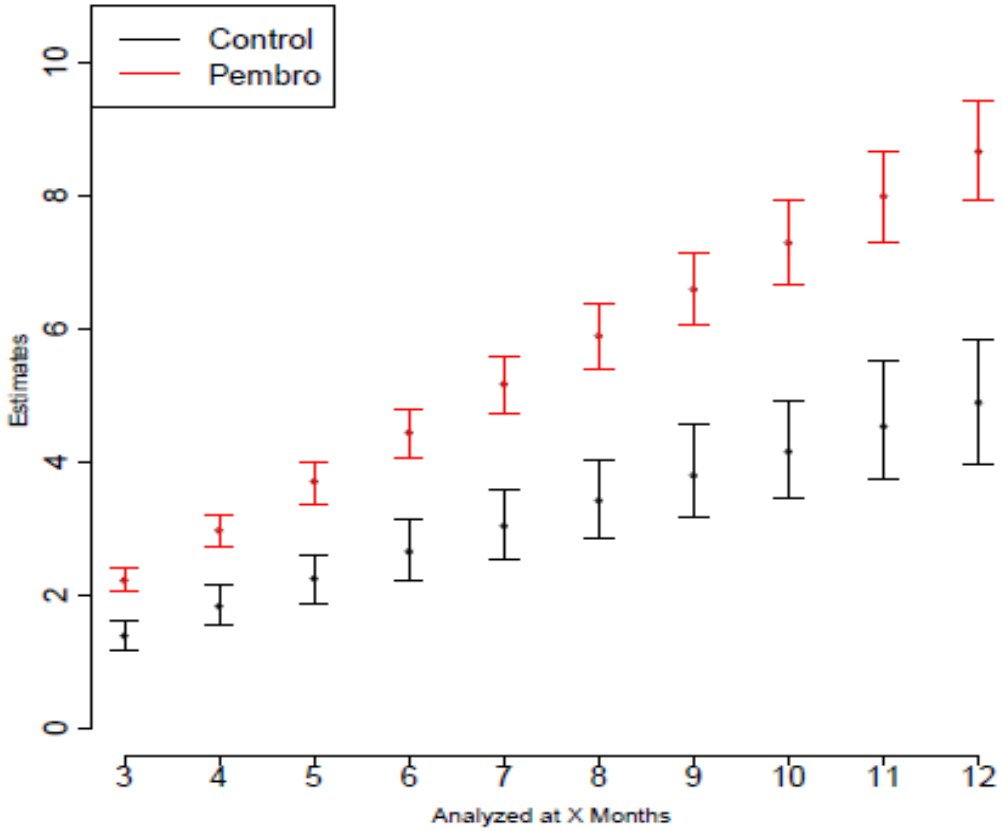
where i is the subject index, $\Delta_i(s)$ is the indicator if i-th subject dies prior to s in presence of censoring time C, and $\hat{K}(s)$ is the KM estimator of censoring time

Method

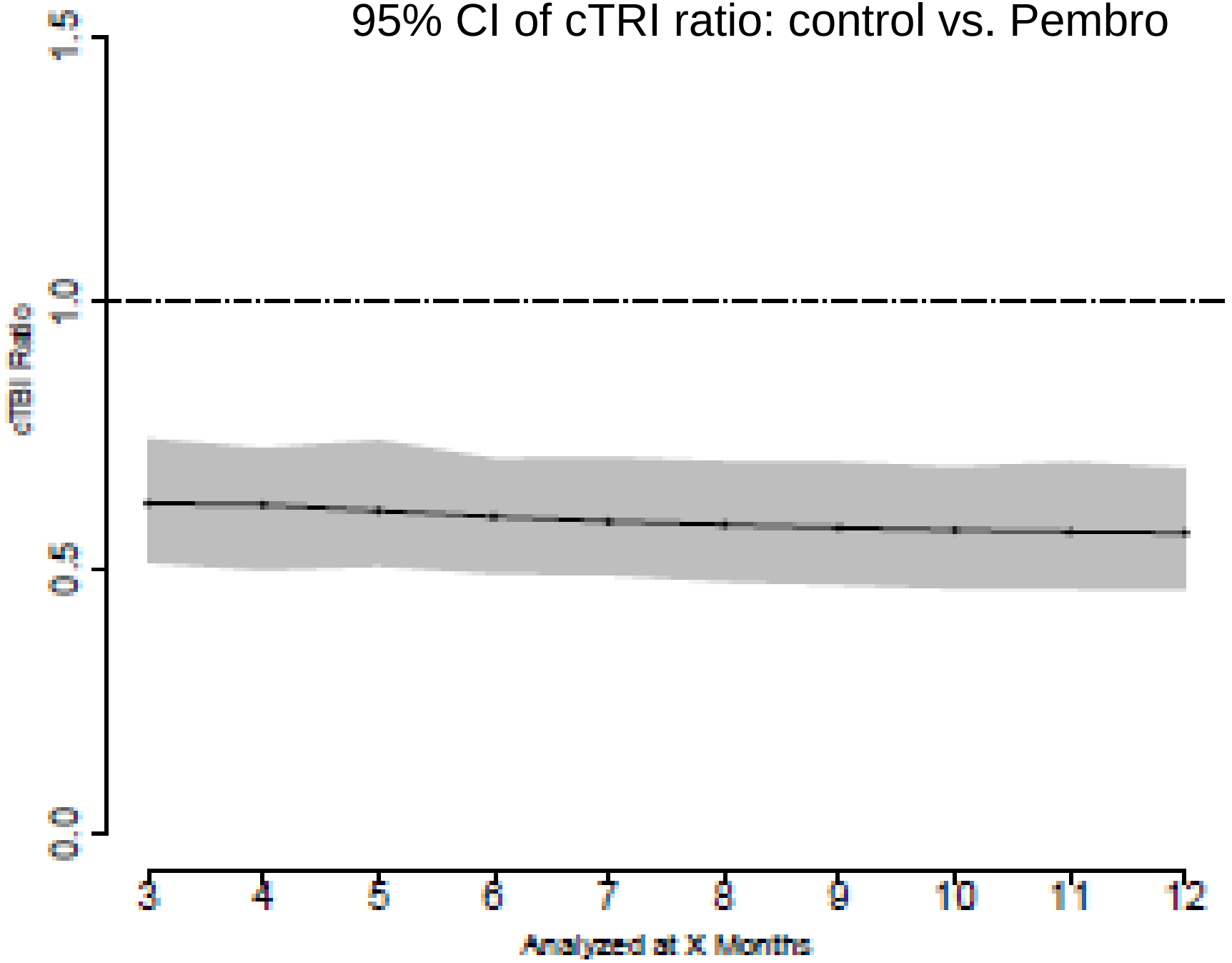
- Implications of cTRI:
  - cTRI (t) is larger when CR/PR occur more frequently and duration is longer
  - cTRI (t) is smaller when PD occur more frequently and faster
  - Choice of Utility (weight) may be arbitrary
  - "Ratio of cTRI" is used for between arm comparison
    - "Ratio" may be less dependent on utility(weight) instead of "difference"
    - Bootstrapping is use for statistical inference (95% CI, p value)

Illustration

- Ordinal "weights"
- PFS + Duration of CR + 1/2 Duration of PR, on RMST scale
  - CR=2.0, PR=1.5, SD=1.0, PD=0



95% CI of cTRI ratio: control vs. Pembro



K

Illustration

- Alternative weight(utility) is evaluated; the 95% CIs and p values are very similar.

For example:

- CR=2.0, PR=2, SD=1, PD=0
  - PFS+ Duration of CR/PR
- CR=3, PR=2, SD=1, PD=0
  - PFS+2X Duration of CR + Duration of PR

Re-sampling simulation

- Simulate randomized phase II trials based on complete phase III trials
- N: 100 (50 per am) or 200 (100 per arm) patients
- Limited follow up time with right censoring
  - Censoring uniform (6m, 12m), (6m, 9m), (9m, 15m), (9m, 12m)
- Compare the following metrics:
  - Ratio of cTRI at 6m (or 9m)
    - Fixed utility(weight) based on RECIST response categories
      - CR=1, PR=1, SD=0, PD=0
      - CR=2, PR=1.5, SD=1, PD=0
      - CR=1, PR=1, SD=0.5, PD=0
    - Individual utility (weight) based on relative tumor size changes (TSC)
  - PFS
  - ORR
- Evaluate the Power (based on type I error of 0.025) and Type I error

Simulation Results

- "Study 1" trial outcome (all positive):
  - PFS: HR 0.61
  - ORR: 36% vs.13%
  - OS: HR 0.68

- **True positive rate (power)** based on resampling from "Study 1" data

N		PFS	ORR	cTRI metric (at 6m)			
				CR=1, PR=1, SD=0, PD=0	CR=2, PR=1.5, SD=1, PD=0	CR=1, PR=1, SD=0.5, PD=0	TSC
100	cen unif(6, 12)	0.607	0.701	0.805	0.626	0.695	0.668
	cen unif(6, 9)	0.568	0.721	0.807	0.636	0.7	0.662
200	cen unif(6, 12)	0.865	0.951	0.97	0.9	0.937	0.905
	cen unif(6, 9)	0.846	0.953	0.971	0.902	0.937	0.909

  - The cTRI metric with utility of CR/PR of 1 has higher power than other utility scenarios.
  - It has higher power than PFS and ORR metric

Simulation Results

- "Study 2" trial outcome:
  - PFS: HR 0.5 (positive)
  - ORR: 45% vs. 28% (some difference)
  - OS: HR 0.6 (positive)

- **True positive rate (power)** based on resampling from "Study 2" data

				cTRI metric (at 6 m)			
N		PFS	ORR	CR=1,PR=1, SD=0, PD=0	CR=2,PR=1.5, SD=1, PD=0	CR=1, PR=1, SD=0.5, PD=0	TSC
100	cen unif(6, 12)	0.75	0.21	0.44	0.42	0.44	0.12
	cen unif(6, 9)	0.69	0.22	0.44	0.4	0.43	0.12
200	cen unif(6, 12)	0.96	0.41	0.76	0.72	0.74	0.16
	cen unif(6, 9)	0.92	0.46	0.78	0.76	0.78	0.21

- The cTRI metric with utility of CR/PR of 1 has higher power than other utilities
- The cTRI metric has higher power than ORR, but lower power than PFS

Simulation Results

- "Study 3" trial outcome:
  - PFS: HR 1 (no difference)
  - ORR: 21% vs. 11% (some difference)
  - OS: HR 0.7 (positive)

- **True positive rate (power)** based on resampling from "Study 3" data

				cTRI metric (at 6 months)			
N		PFS	ORR	CR=1,PR=1, SD=0, PD=0	CR=2, PR=1.5, SD=1, PD=0	CR=1, PR=1, SD=0.5, PD=0	TSC
100	cen unif(6, 12)	0.05	0.11	0.26	0.04	0.04	0.03
	cen unif(6, 9)	0.05	0.11	0.26	0.04	0.04	0.03
200	cen unif(6, 12)	0.07	0.29	0.47	0.05	0.07	0.06
	cen unif(6, 9)	0.06	0.29	0.47	0.05	0.07	0.06

- The cTRI metric with utility of CR/PR of 1 has higher power than other utility scenarios.
- It has higher power than ORR

Simulation Results

- **Type I error** based on resampling from the two dosing schedules without efficacy difference in “Study 1”:

				cTRI metric (at 6 months)			
N		PFS	ORR	CR=1,PR=1, SD=0, PD=0	CR=2, PR=1.5, SD=1, PD=0	CR=1, PR=1, SD=0.5, PD=0	TSC
100	cen unif(6, 12)	0.03	0.05	0.02	0.03	0.03	0.04
	cen unif(6, 9)	0.03	0.04	0.02	0.03	0.03	0.04
200	cen unif(6, 12)	0.11	0.04	0.05	0.1	0.08	0.1
	cen unif(6, 9)	0.1	0.03	0.05	0.1	0.08	0.1

- The type I error rates are comparable across all metrics

Remarks and Discussions

- The proposed cTRI metric (with incorporating Response status and Time) appears to have higher power than ORR
- The cTRI metric with utilities only on CR/PR, corresponds to the Restricted Mean Survival Time (RMST) for Duration of Response (DOR), and has higher power than with other utility scenarios
- The cTRI metric uniquely combines “ITT-based ORR ” with “subset-based DOR ” into a single metric applicable to ITT population
- The cTRI with utilities only on CR/PR may be better than ORR and PFS in some highly immunogenetic settings