Quantifying the (long-term) treatment benefit with cancer immunotherapies

Bo Huang, PhD
Pfizer Inc.

Joint work with Dr. Lu Tian (Stanford) and Dr. LJ Wei (Harvard)

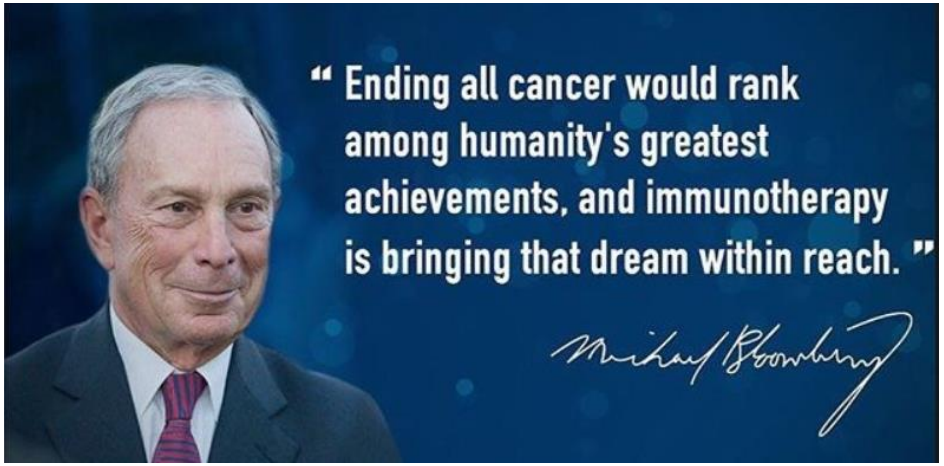
NISS-Merck Virtual Meet-Up, Jan 22, 2019

Pfizer Oncology

Acknowledgement

- Pfizer colleagues for useful input and example data
  - Enayet Talukder
  - Margarida Geraldes
  - Mace Rothenberg
  - XALKORI® (crizotinib) team: Paulina Selaru , Tiziana Usari and Timothy Kluge
  - BESPONSA® (inotuzumab ozogamicin) team: Jane Liang White, Tao Wang

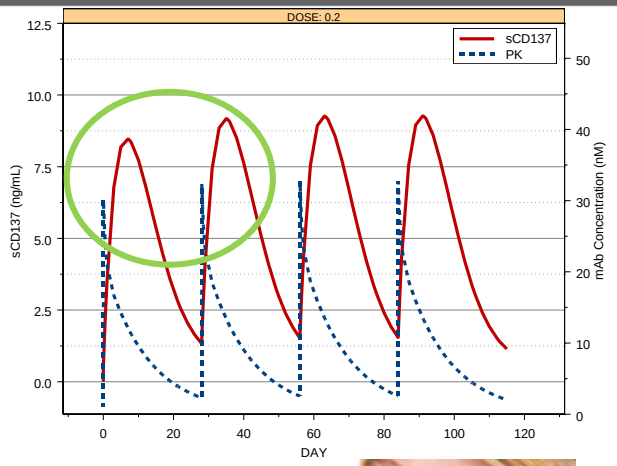
Immunotherapy: frontier of (cancer) drug development

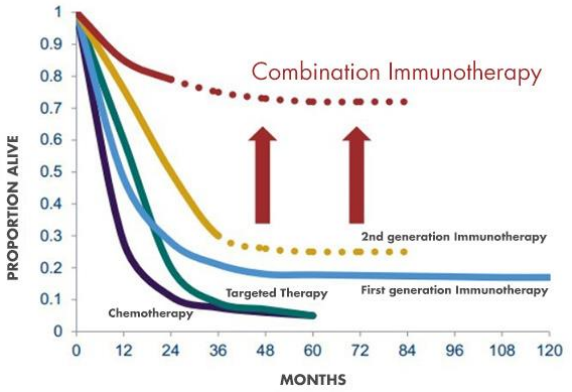


Challenges in IO dose finding: Manage toxicity, dose optimization

Due to the life-threatening nature of cancer, a high degree of dose limiting toxicity (DLT) is generally considered acceptable.

- Late/cumulative effect of IO: Late-onset toxicity not observed in traditional 1-cycle DLT window
- Select the optimal dose & schedule
  - Higher dose -> higher toxicity, but no necessarily higher activity → More is not necessarily better
  - How about dosing schedule?
- How about dose finding for combination agents?
  - Complexity increases on 2-dimensional space
  - Partial ordering

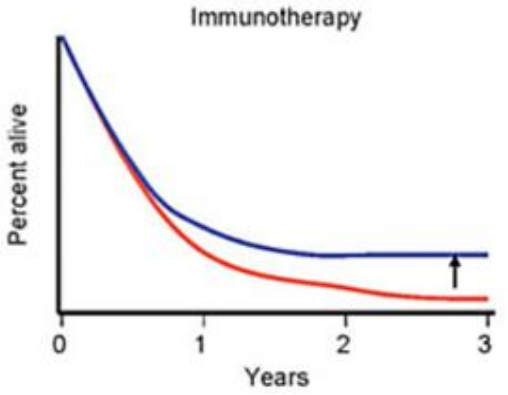


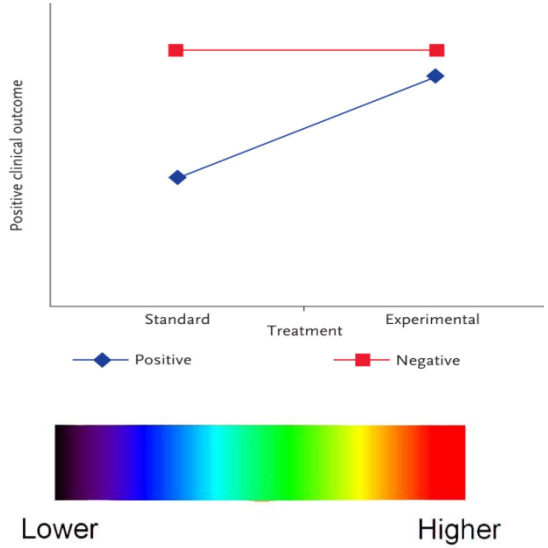


Pfizer Oncology

Challenges in IO efficacy evaluation

- Non-proportional hazards and late separation in curves due to delayed (late) anti-tumor effect; Durable responses lead to long-term benefit (cured effect)
  - LR test and HR still optimal?
  - Timing of IA?
- Weak/negative correlation between PFS and OS
  - Careful with futility IA based on PFS
- Predictive biomarker
  - Ignoring may lead to trial failure
  - PD-L1 expression continuous (no perfect dichotomization), TMB (how many mut/mb?), gene signatures etc.





Pfizer Oncology

Ref: Huang B. Some statistical considerations in the clinical development of cancer immunotherapies. Pharmaceutical statistics. 2018 Feb;17(1):49-60.

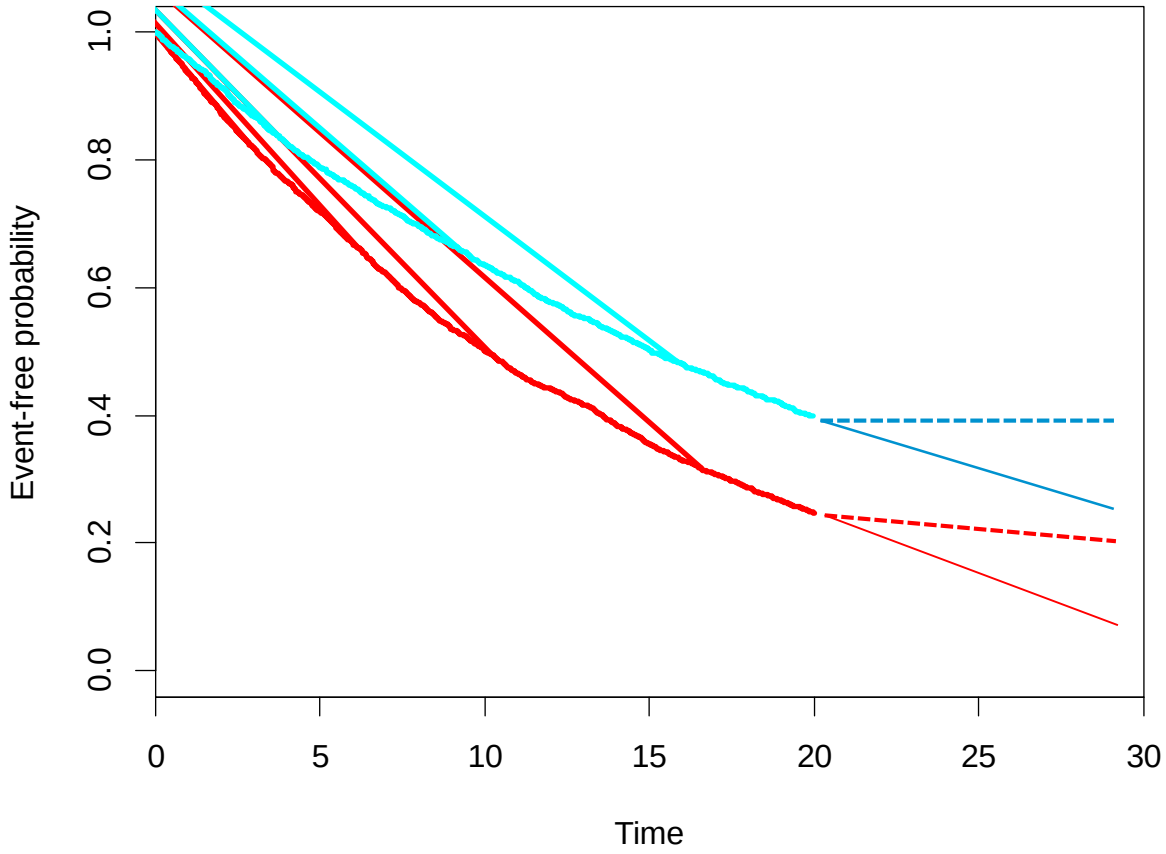
Limitation of Log-rank test, HR (from Cox model) and the medians

- The log-rank test is optimal under the proportional hazard (PH) assumption, but may lose power under NPH
  - Equivalent to the score test in the Cox-PH model
- The power of log-rank test is event driven
  - Low power with small events; not sensitive to separated flat tails

- Interpretation of HR from Cox model is a problem under NPH
- Only a relative measure of effect

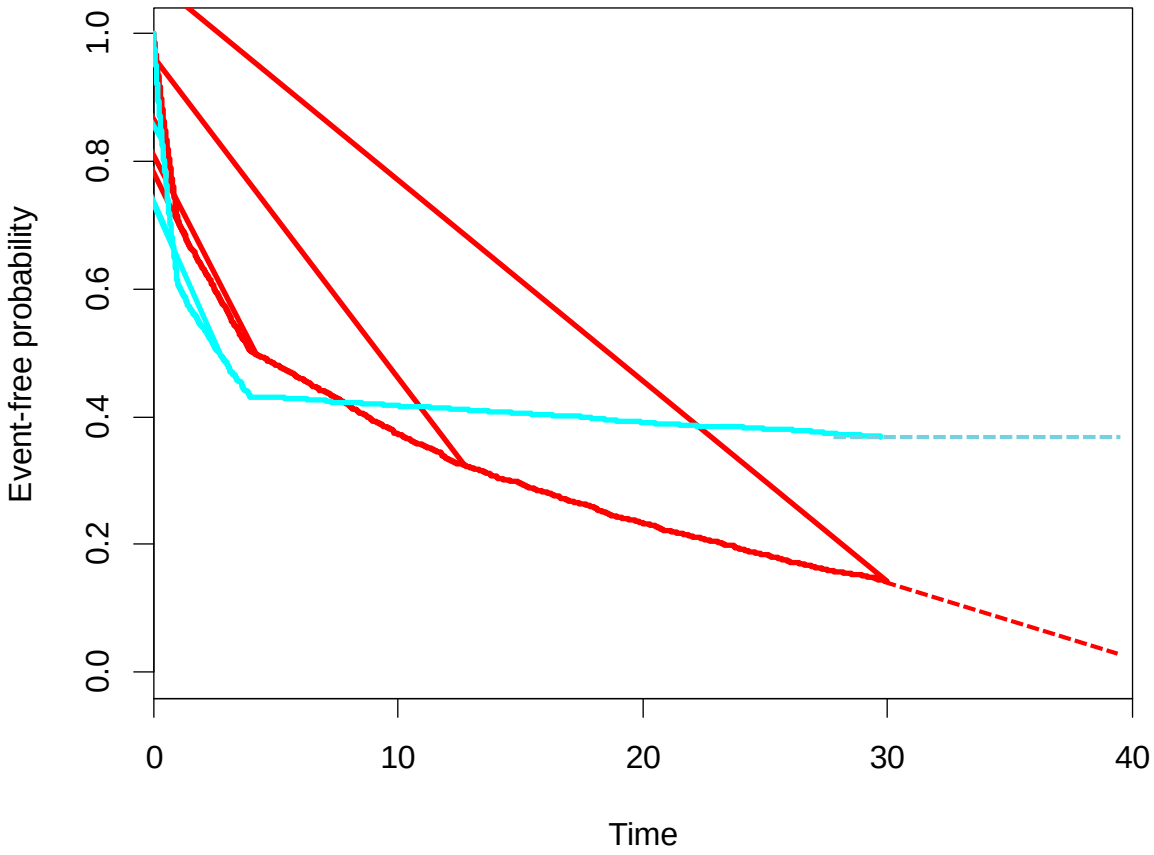
- The median is an arbitrary percentile (despite good clinical interpretation)
  - Not a global summary measure – does not capture what's before and after

Graphical examples



Pfizer Oncology

Graphical examples



Pfizer Oncology

Alternative methods to the rank-based methods

- t-time event-free probability
- Win-ratio (in particular for prioritized multiple endpoints)
- Net-benefit
- Generalized pairwise comparison
- Kaplan-Meier based method
  - RMST
  - Weighted KM based test

Restricted mean survival time (RMST)

- Mean survival time (life expectancy) truncated by time τ: area under the survival curve $S(t)$ from 0 to τ

$$\mathrm{RMST}(\tau) = \int_0^{\tau} S(t)dt$$

- $S(t)$ can be estimated by the KM estimator $\hat{S}(t)$. The variance of the observed RMST estimator is (Klein and Moeschberger, 2005)

$$\sum_{i=1}^{D} \left[\int_{t_i}^{\tau} \hat{S}(t)\, dt \right]^2 \frac{d_i}{Y_i(Y_i - d_i)}$$

# of events at t_i

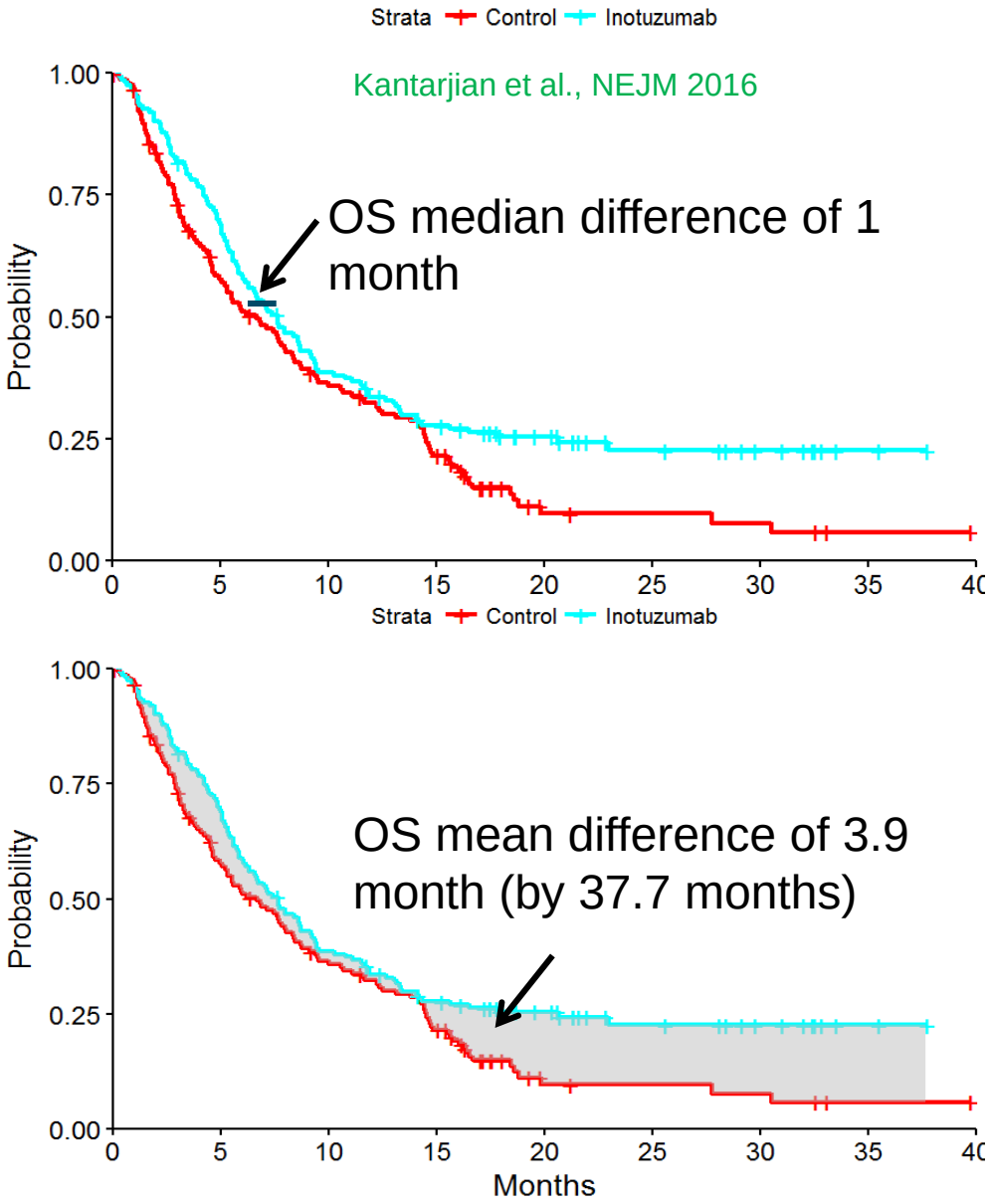
# of pts at risk at t_i

- Measure treatment effect by difference and ratio in the RMSTs of different drugs

Pfizer Oncology

Versus the HR, the median, t-time event-free probability

- **Versus the HR**
  - Clinical interpretation (whether or not PH holds)
  - Non-parametric
  - Dual presentation of relative and absolute effect
- **Versus the median/t-time event-free probability**
  - Informative global summary, no arbitrary percentile/cutoff
  - RMST curve: temporal profile by different truncation points



Pfizer Oncology

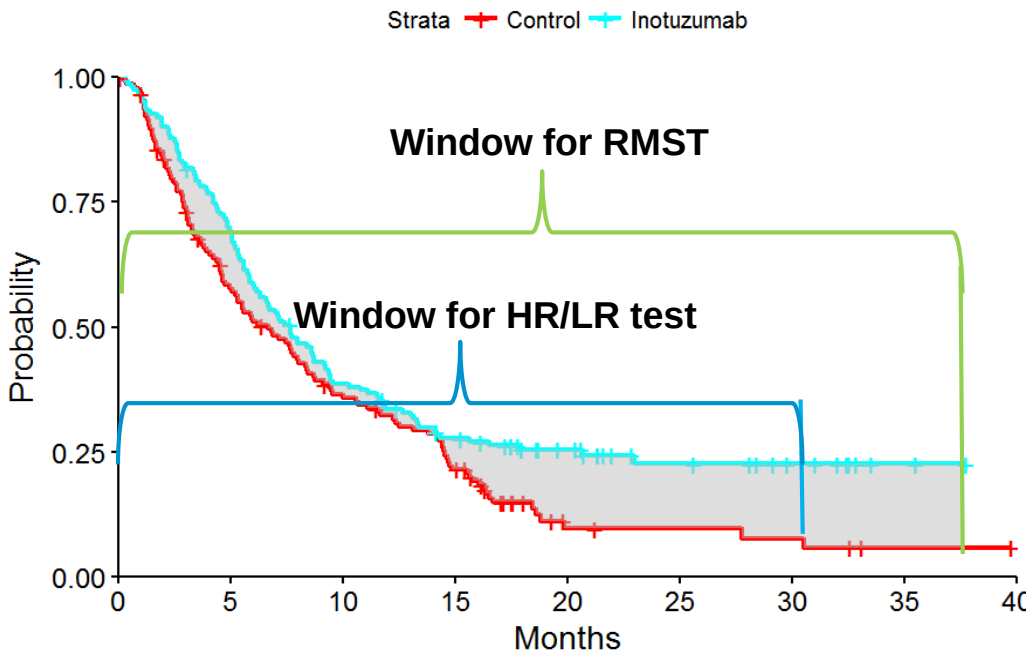
Time-window of RMST vs Time-window of HR/LR test

Under mild conditions on the censoring distribution:

- One can make inference on RMST up to **the last follow-up time (either event or censored) for the arm with shorter follow-up**

In contrast,

- For HR/LR test, one can only use data up to **the minimum of (last event time for any arm, the last follow-up time of the arm with the shorter follow-up)**



Reference:

- Tian L, Jin H, Uno H, Lu Y, Huang B, Anderson K, Wei LJ. On the Empirical Choice of the Time Window for Restricted Mean Survival Time, submitted (under review).
- Huang B, Kuan P. (2018). Comparison of the Restricted Mean Survival Time with the Hazard Ratio in Superiority Trials with a Time-to-Event Endpoint, *Pharmaceutical Statistics* 17(3):202-13

Pfizer Oncology

A composite endpoint to measure the effect of cancer treatment including immunotherapy

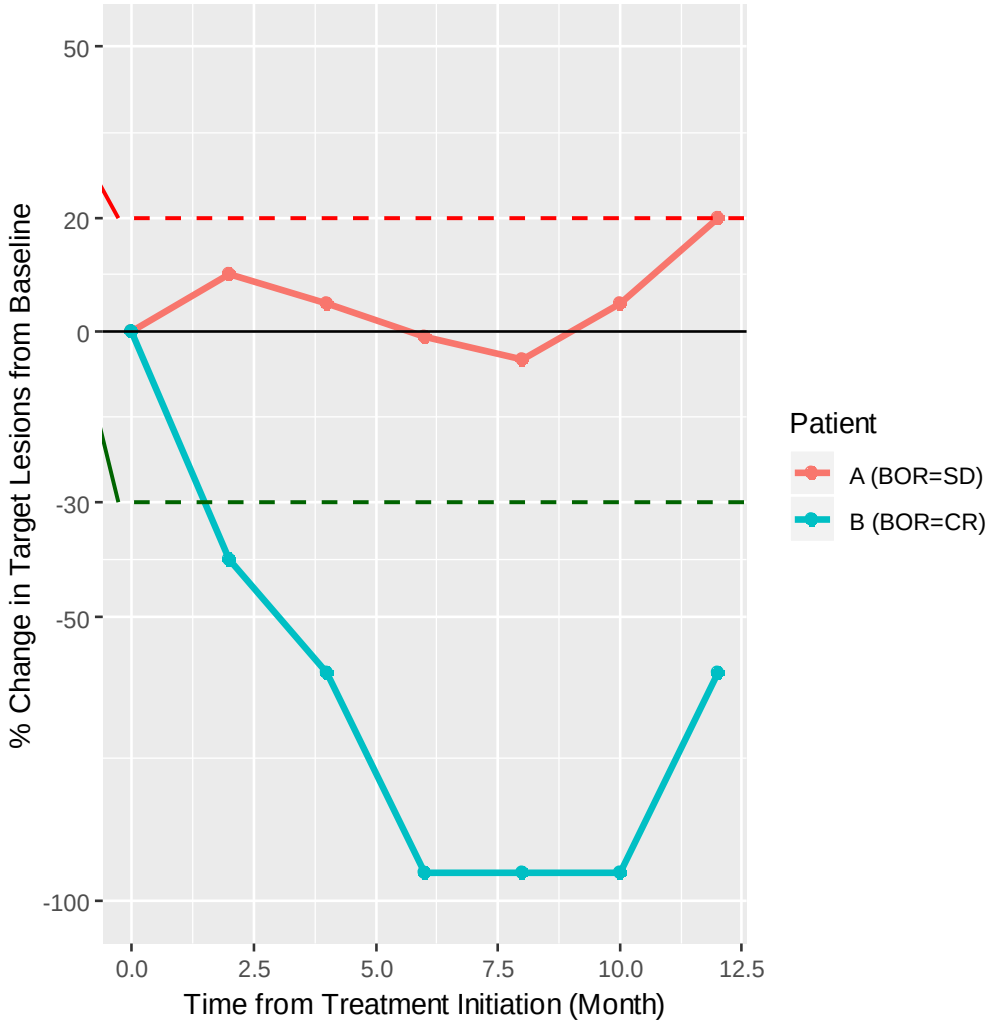
What's a desirable future cancer treatment?

- (future) cancer treatment may not be a "cure", but can effectively control the disease, and patients can live a "normal" life with the disease for a long period of time
- What are the characteristics of a desirable cancer treatment for patients and doctors?
  - ✓ **Life extending (OS benefit)**
  - ✓ **High likelihood of tumor response (reduction in size)**
  - ✓ **Fast time to response**
  - ✓ **Long duration of being in response (durable response)**
  - ✓ **Manageable side effect**
  - ✓ **Improved/not-worsening HRQOL**

Pfizer Oncology

Limitation of PFS

- “Disease stabilization” may not translate to long-term survival benefit
- Cannot distinguish tumor reduction from no change/slight increase
- With cancer becoming a chronic disease, we will lose the capability to design faster and smaller trials with PFS endpoint



Pfizer Oncology

A composite endpoint of duration of response in the ITT population

Research Letter

June 2018

Evaluating Treatment Effect Based on Duration of Response for a Comparative Oncology Study

Bo Huang, PhD[1]; Lu Tian, ScD[2]; Enayet Talukder, PhD[1]; Mace Rothenberg, MD[3]; Dae Hyun Kim, MD, ScD[4]; Lee-Jen Wei, PhD[5]

Method

To evaluate treatment effect of a drug, there are three states:

- 0: Time zero (randomization or first dose)
- 1: Time zero to response, progression or death (whichever is earlier)
- 2: Time from response(R) (or state 1) to progression (P) or death (D)

Parameter of interest:

The mean duration a patient expected to spend at state 2 (from response to progression or death) before time τ, for responders, this is equivalent to the duration of response (DOR) by traditional definition (but conditional on responders only).

Pfizer Oncology

Method

The mean duration of response (by time τ) then is:

$$E\,[\min(T[P/D], \tau) - \min(T[R], T[P/D], \tau)\,]$$

$$= E\,[\min(T[P/D], \tau)] - E[\min(T[R], T[P/D], \tau)\,]$$

$$= (\text{AUC of PFS curve from } 0 \text{ to } \tau) - (\text{AUC of P/D/R-free curve from } 0 \text{ to } \tau)$$

$$= \text{area between the PFS curve and P/D/R-free curve from } 0 \text{ to } \tau$$

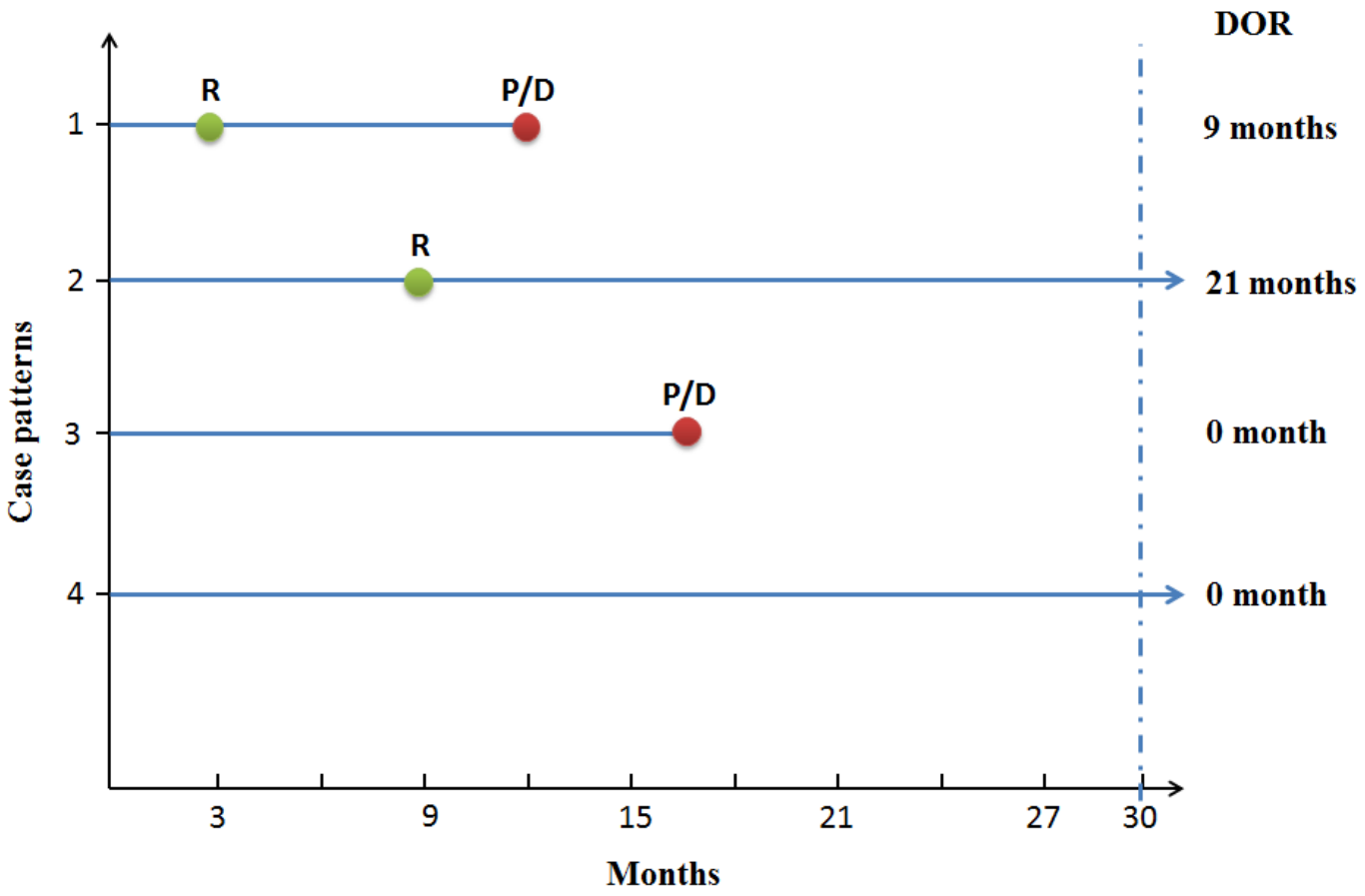
which can be interpreted as the expected duration of response up to time τ for a patient receiving treatment.

Variance term can be derived analytically, or by bootstrapping

- Programming code available at https://web.stanford.edu/~lutian/Software.HTML

Pfizer Oncology

Possible patterns of times to response (R) and progression/death (P/D) up to Month-30

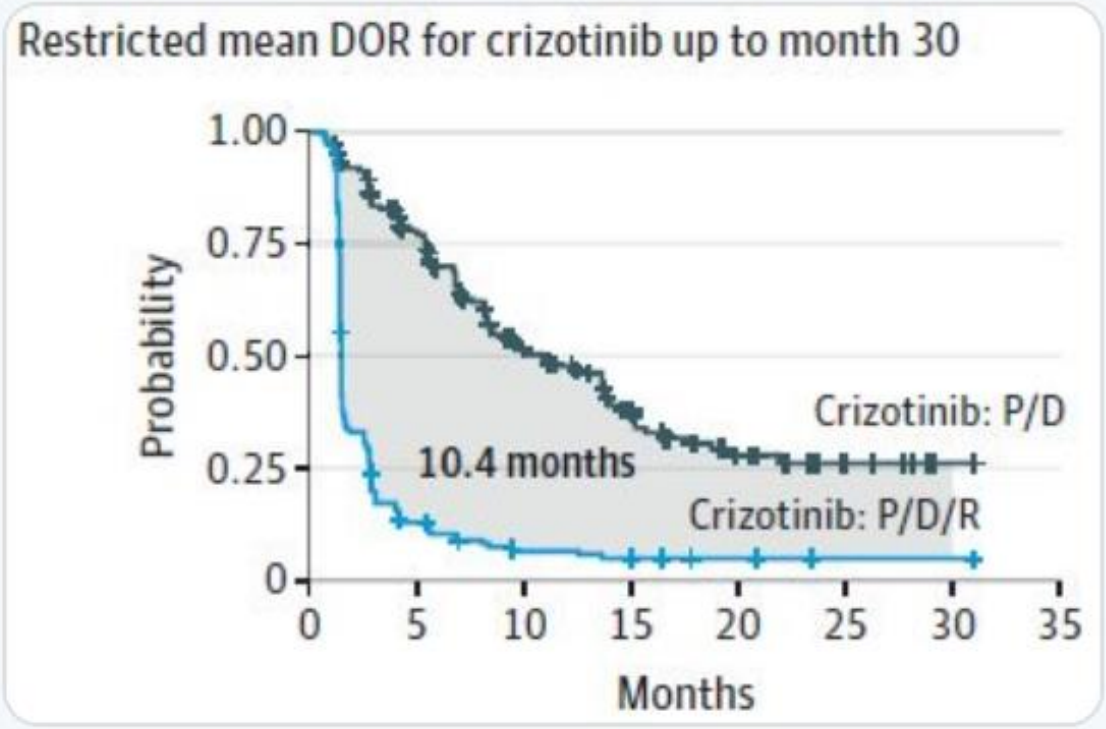


Pfizer Oncology

Restricted mean DOR for Crizotinib up to Month-30

JAMA Oncology @JAMAOnc · Apr 21

Read this specific proposal and explanation for how to define duration of response in #clinicaltrials **of** #cancer treatments ja.ma/2HzKV6j



33 45

shaded area between

- P/D event-free curve
- P/D/R event-free curve

A simulated study to compare endpoints and analysis methods

Scenario 1 (no "cured" effect)

- N=400 pts (1:1)
- ORR 50% in Arm A (new drug) and 25% in Arm B (SOC)
- Exponential distribution assumed for TTR & DOR for responders, PFS
  - Arm A: mPFS 10 months for non-responders, for responders, mTTR 2 months, mDOR 12 months
  - Arm B: mPFS 10 months for non-responders, for responders, mTTR 4 months, mDOR 8 months
- Uniform accrual in 12 months
- Data cut = 25, 50, 75 months from start of accural
- tau=minimax follow-up time

Scenario 2 (with "cured" effect)

- N=400 pts (1:1)
- ORR 50% in Arm A and 25% in Arm B
- Exponential distribution assumed for TTR & DOR for responders, PFS
  - Arm A: mPFS 10 months for non-responders, for 80% of responders, mTTR 2 months, mDOR 12 months, for 20% of responders, mTTR 2 months, mDOR 60 months
  - Arm B: mPFS 10 months for non-responders, for 80% of responders, mTTR 4 months, mDOR 8 months, for 20% of responders, mTTR 4 months, mDOR 60 months
- Others specifications same as Scenario 1

Pfizer Oncology

Result (Scenario 1 – no “cured” effect)

	PFS					DOR (ITT)	
	HR	Median Diff	Power (LR)	RMST Diff	Power (RMST)	Mean Diff	Power (RMST)
25 mos	0.88	1.37m	19.2%	1.00m	20.0%	3.92m	99.8%
50 mos	0.87	1.38m	26.6%	1.74m	26.2%	5.21m	99.9%
75 mos	0.87	1.38m	27.8%	2.00m	28.0%	5.57m	99.9%

Pfizer Oncology

Result (Scenario 2 – with “cured” effect)

	PFS					DOR (ITT)	
	HR	Median Diff	Power (LR)	RMST Diff	Power (RMST)	Mean Diff	Power (RMST)
25 mos	0.84	2.06m	30.2%	1.31m	30.8%	4.32m	99.8%
50 mos	0.83	2.07m	43.2%	2.57m	43.8%	6.12m	99.9%
75 mos	0.83	2.07m	45.0%	3.35m	45.2%	7.01m	99.9%

Pfizer Oncology

Summary and Discussion

- Conventional methods for TTE endpoints such as HR/median/LR tests have limitations
- OS endpoint will be harder to meet with the advance of cancer treatment
- PFS endpoint has limitations as a surrogate endpoint
- With the advances in cancer treatment (including immunotherapies), it is expected that a higher proportion of patients will respond to treatment (higher ORR, higher CR rate)
- Mean is the gold standard in many therapeutic areas
- A composite endpoint (DOR) and measure (mean) was proposed as an alternative method for statistical inference to assess TTR, ORR, DOR simultaneously
  - Mean duration of complete response in the ITT population could be potentially useful and efficient for next generation cancer therapies

Pfizer Oncology


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
obrigado
merci
dziękuję
bedankt
terima kasih
sukriya
kop khun krap
go raibh maith agat
dakujem
arigatô
grazie
tapadh leat
mochchakkeram
спасибо
рахмат
Баярлалаа
faafetai lava
kiitos
dankie
dhanyavad
nanni
nandri
bayarlalaa
gràcie
hvala
mauruuru
köszönöm
enkosi
sobodi
dēkuji
mèsi
didi madloba
sagolun
najis tuke
kam sah hamnida
তোমাকে ধন্যবাদ
rahmat
감사합니다
xièxie
ευχαριστώ
tanemirt
rahmet
diolch
dhanyavadagalu
shukriya
mercè
мерси
trugarez
takk
tenki
chokrane
murakoze
obrigada
asante
manana
хвала
mamnun
дякую
djiere dieuf
tau
paldies
grazzi
mahalo
matondo
misaotra
dank je
blagodaram
спасибі
vinaka
mersi
kia ora
barka
welalin
tack
spas
akun
dankon
ačiū
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
ありがとう
tak
faafetai lava