From Logic-respecting Efficacy Estimands to Logic-ensuring Analysis Principle for Time-to-event Endpoint in Randomized Clinical
Trials with Biomarker Subgroups

Liwei Wang (Genmab)

LIDS presentations

June 1st, 2023

STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2023, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/19466315.2023.2186945

Taylor & Francis
Taylor & Francis Group

Check for updates

From Logic-Respecting Efficacy Estimands to Logic-Ensuring Analysis Principle for Time-to-Event Endpoint in Randomized Clinical Trials with Subgroups

Yi Liu[a], Miao Yang[a,*], Siyoen Kil[b], Jiang Li[c], Shoubhik Mondal[d], Yue Shentu[e], Hong Tian[f], Liwei Wang[g], and Godwin Yung[h]

[a]Nektar Therapeutics, San Francisco, CA; [b]LSK Global Pharma Services Co, Seoul, Korea; [c]BeiGene, Ridgefield Park, NJ; [d]AstraZeneca, Gaithersburg, MD; [e]Daiichi Sankyo Inc., Basking Ridge, NJ; [f]BeiGene, Ridgefield Park, NJ; [g]Genmab US, Inc, Princeton, NJ; [h]Genentech, South San Francisco, CA

Acknowledgement

Oncology Estimand WG TF8

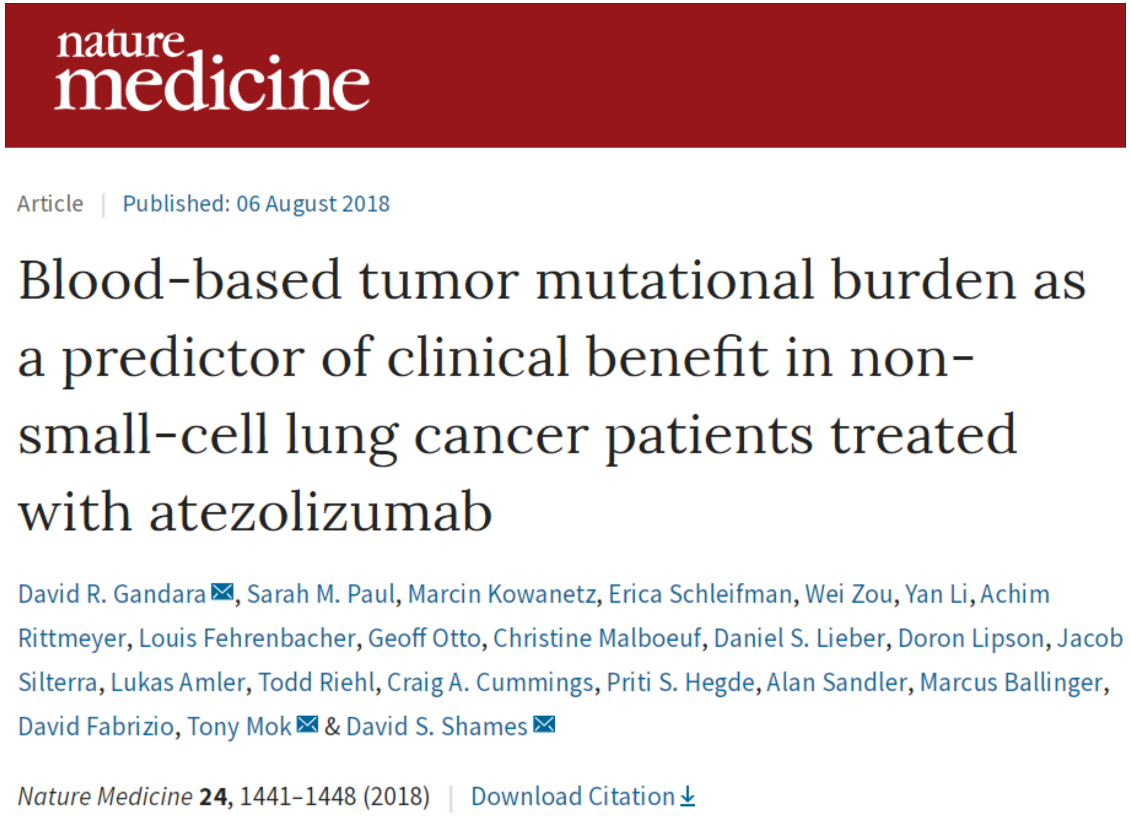
- Yi Liu (Nektar)
- Yue Shentu (Daiichi)
- Miao Yang (Seagen)
- Shoubhik Mondal (AstraZeneca)
- Hong Tian (Beigene)
- Siyoen Kil (LSK global PS)
- Jiang Li (Beigene)
- Godwin Yung (Genentech)
- Jonathan Siegel (Bayer)

Additional collaborators

- Jason Hsu
- Ying Ding
- Hui-Min Lin
- Szu-Yu Tang
- Bushi Wang
- Haiyan Xu

Outline

- Puzzling behavior of HR in real Clinical trials with subgroups
  - HR can make a purely prognostic biomarker seem predictive
- Two issues:
  - Efficacy measure such as HR and OR are not logic respecting and non-collapsible at the population level
  - Current computer software and common analysis methods help mask the problem
- Our proposal: logic respecting estimands at population level and SME for data analysis
  - Steps to implement SME using either parametric or non-parametric approach
  - Simultaneous CI for biomarker subgroups and overall population based on real clinical trials
- Summary



nature medicine

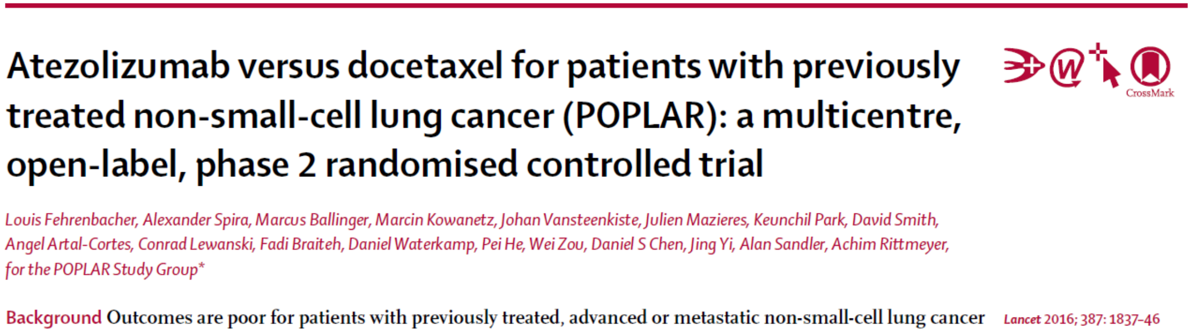
Article | Published: 06 August 2018

Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab

David R. Gandara, Sarah M. Paul, Marcin Kowanetz, Erica Schleifman, Wei Zou, Yan Li, Achim Rittmeyer, Louis Fehrenbacher, Geoff Otto, Christine Malboeuf, Daniel S. Lieber, Doron Lipson, Jacob Silterra, Lukas Amler, Todd Riehl, Craig A. Cummings, Priti S. Hegde, Alan Sandler, Marcus Ballinger, David Fabrizio, Tony Mok & David S. Shames

Nature Medicine **24**, 1441–1448 (2018) | Download Citation

- POPLAR data demonstrated proof of principle for bTMB as a predictor of PFS clinical outcome



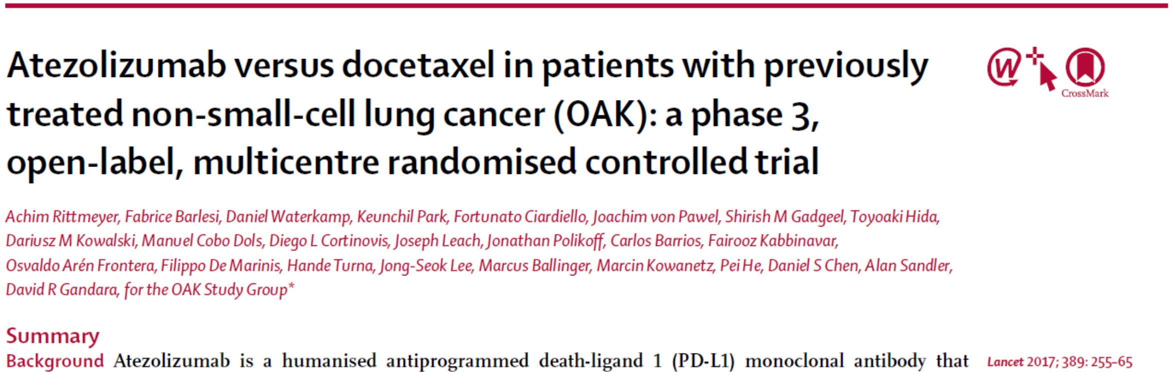
Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial

*Louis Fehrenbacher, Alexander Spira, Marcus Ballinger, Marcin Kowanetz, Johan Vansteenkiste, Julien Mazieres, Keunchil Park, David Smith, Angel Artal-Cortes, Conrad Lewanski, Fadi Braiteh, Daniel Waterkamp, Pei He, Wei Zou, Daniel S Chen, Jing Yi, Alan Sandler, Achim Rittmeyer, for the POPLAR Study Group**

Background Outcomes are poor for patients with previously treated, advanced or metastatic non-small-cell lung cancer

Lancet 2016; 387: 1837–46

- OAK data confirm bTMB as a potential non-invasive biomarker of PD-L1-directed immunotherapy.



Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial

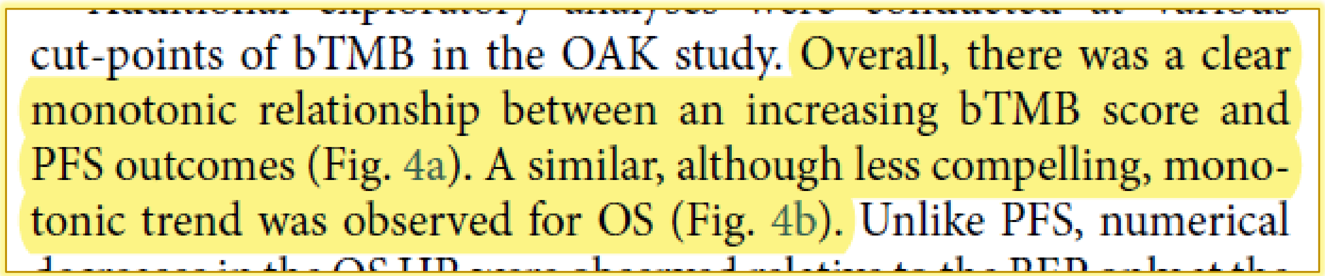
*Achim Rittmeyer, Fabrice Barlesi, Daniel Waterkamp, Keunchil Park, Fortunato Ciardiello, Joachim von Pawel, Shirish M Gadgeel, Toyoaki Hida, Dariusz M Kowalski, Manuel Cobo Dols, Diego L Cortinovis, Joseph Leach, Jonathan Polikoff, Carlos Barrios, Fairooz Kabbinavar, Osvaldo Arén Frontera, Filippo De Marinis, Hande Turna, Jong-Seok Lee, Marcus Ballinger, Marcin Kowanetz, Pei He, Daniel S Chen, Alan Sandler, David R Gandara, for the OAK Study Group**

Summary

Background Atezolizumab is a humanised antiprogrammed death-ligand 1 (PD-L1) monoclonal antibody that

Lancet 2017; 389: 255–65

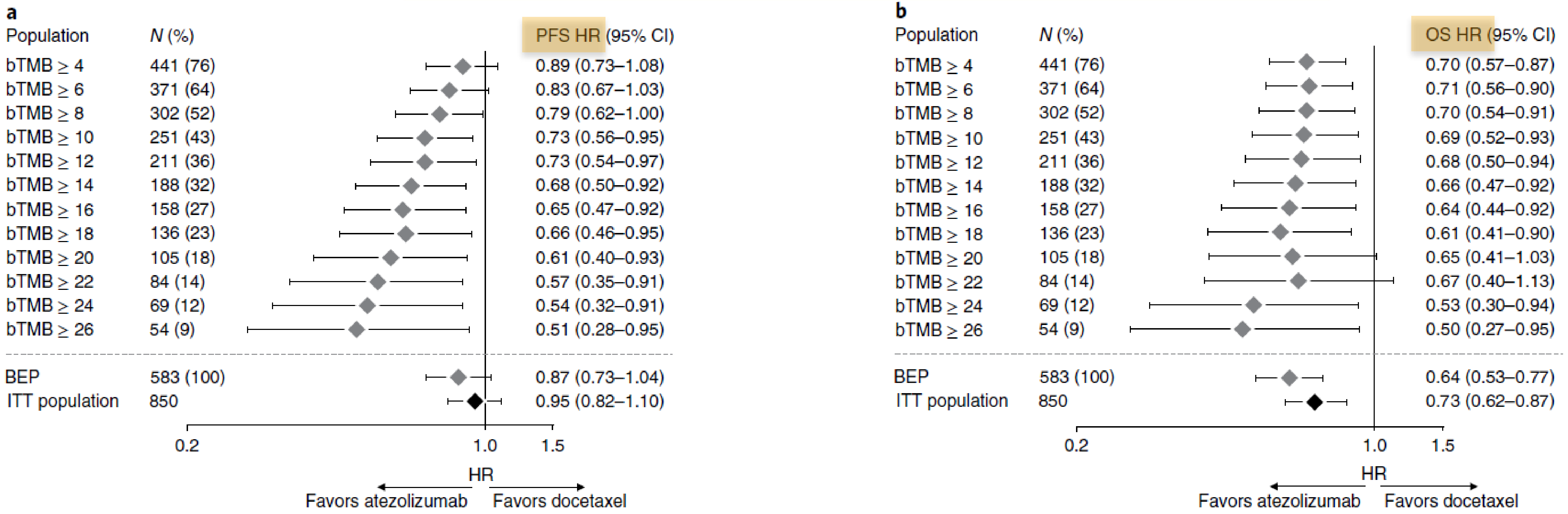
Is bTMB a predictor of clinical benefit in NSCLC patients treated with atezolizumab in OAK study?



Additional exploratory analyses were conducted at various cut-points of bTMB in the OAK study. Overall, there was a clear monotonic relationship between an increasing bTMB score and PFS outcomes (Fig. 4a). A similar, although less compelling, monotonic trend was observed for OS (Fig. 4b). Unlike PFS, numerical decreases in the OS HR were observed relative to the BEP only at the

PFS

OS



a

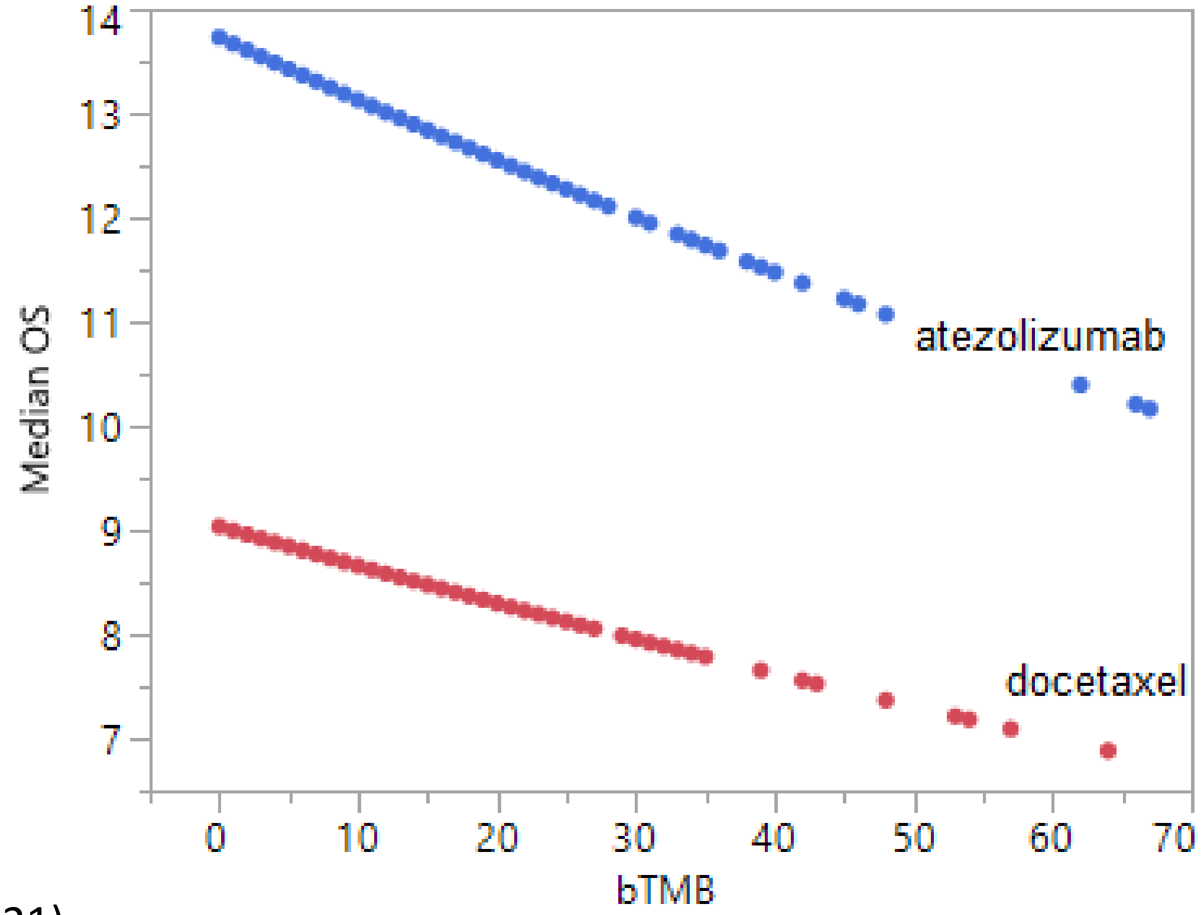
Population	N (%)	PFS HR (95% CI)
bTMB ≥ 4	441 (76)	0.89 (0.73–1.08)
bTMB ≥ 6	371 (64)	0.83 (0.67–1.03)
bTMB ≥ 8	302 (52)	0.79 (0.62–1.00)
bTMB ≥ 10	251 (43)	0.73 (0.56–0.95)
bTMB ≥ 12	211 (36)	0.73 (0.54–0.97)
bTMB ≥ 14	188 (32)	0.68 (0.50–0.92)
bTMB ≥ 16	158 (27)	0.65 (0.47–0.92)
bTMB ≥ 18	136 (23)	0.66 (0.46–0.95)
bTMB ≥ 20	105 (18)	0.61 (0.40–0.93)
bTMB ≥ 22	84 (14)	0.57 (0.35–0.91)
bTMB ≥ 24	69 (12)	0.54 (0.32–0.91)
bTMB ≥ 26	54 (9)	0.51 (0.28–0.95)
BEP	583 (100)	0.87 (0.73–1.04)
ITT population	850	0.95 (0.82–1.10)

b

Population	N (%)	OS HR (95% CI)
bTMB ≥ 4	441 (76)	0.70 (0.57–0.87)
bTMB ≥ 6	371 (64)	0.71 (0.56–0.90)
bTMB ≥ 8	302 (52)	0.70 (0.54–0.91)
bTMB ≥ 10	251 (43)	0.69 (0.52–0.93)
bTMB ≥ 12	211 (36)	0.68 (0.50–0.94)
bTMB ≥ 14	188 (32)	0.66 (0.47–0.92)
bTMB ≥ 16	158 (27)	0.64 (0.44–0.92)
bTMB ≥ 18	136 (23)	0.61 (0.41–0.90)
bTMB ≥ 20	105 (18)	0.65 (0.41–1.03)
bTMB ≥ 22	84 (14)	0.67 (0.40–1.13)
bTMB ≥ 24	69 (12)	0.53 (0.30–0.94)
bTMB ≥ 26	54 (9)	0.50 (0.27–0.95)
BEP	583 (100)	0.64 (0.53–0.77)
ITT population	850	0.73 (0.62–0.87)

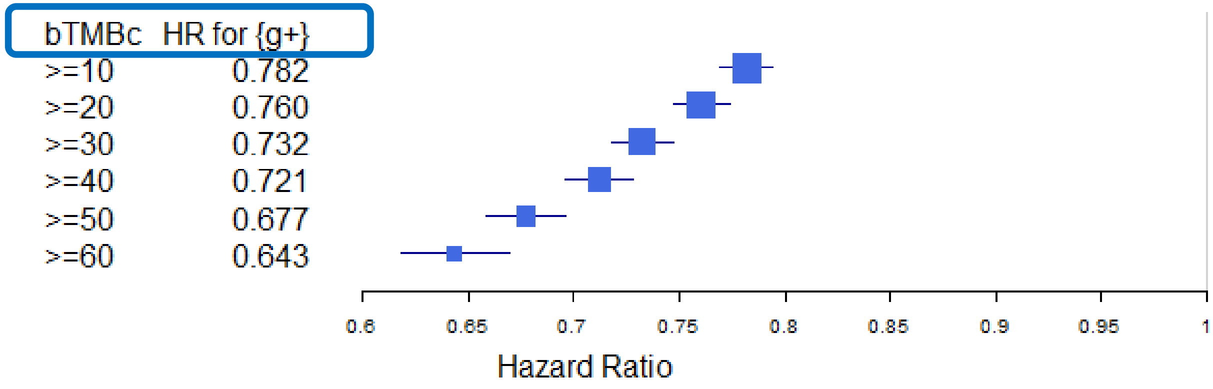
Rerun of the OAK trial data* shows that bTMB is mostly a prognostic (instead of predictive) biomarker in terms of OS

Estimated median OS from Weibull fit with bTMB, Trt and the interaction term



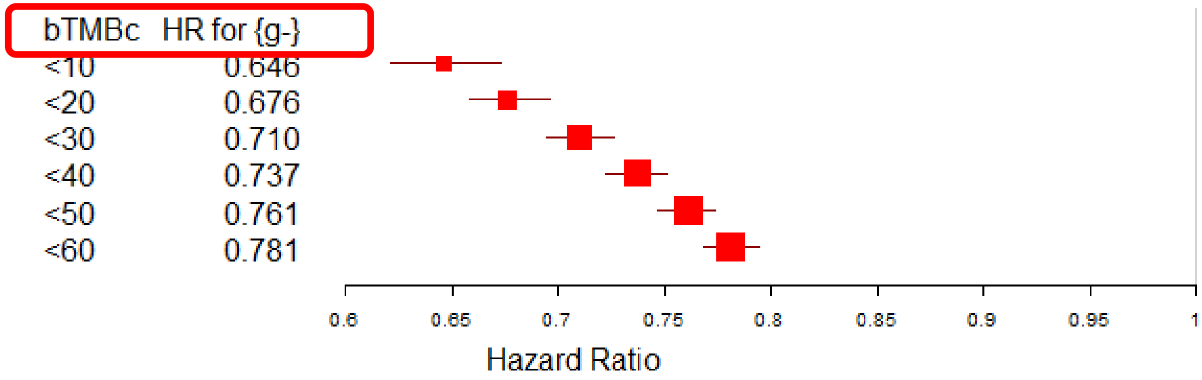
*Liu et al (2021)

HR behavior for purely prognostic biomarker based on simulation



bTMBc	HR for {g+}
>=10	0.782
>=20	0.760
>=30	0.732
>=40	0.721
>=50	0.677
>=60	0.643

Replicated the pattern observed in OAK trial

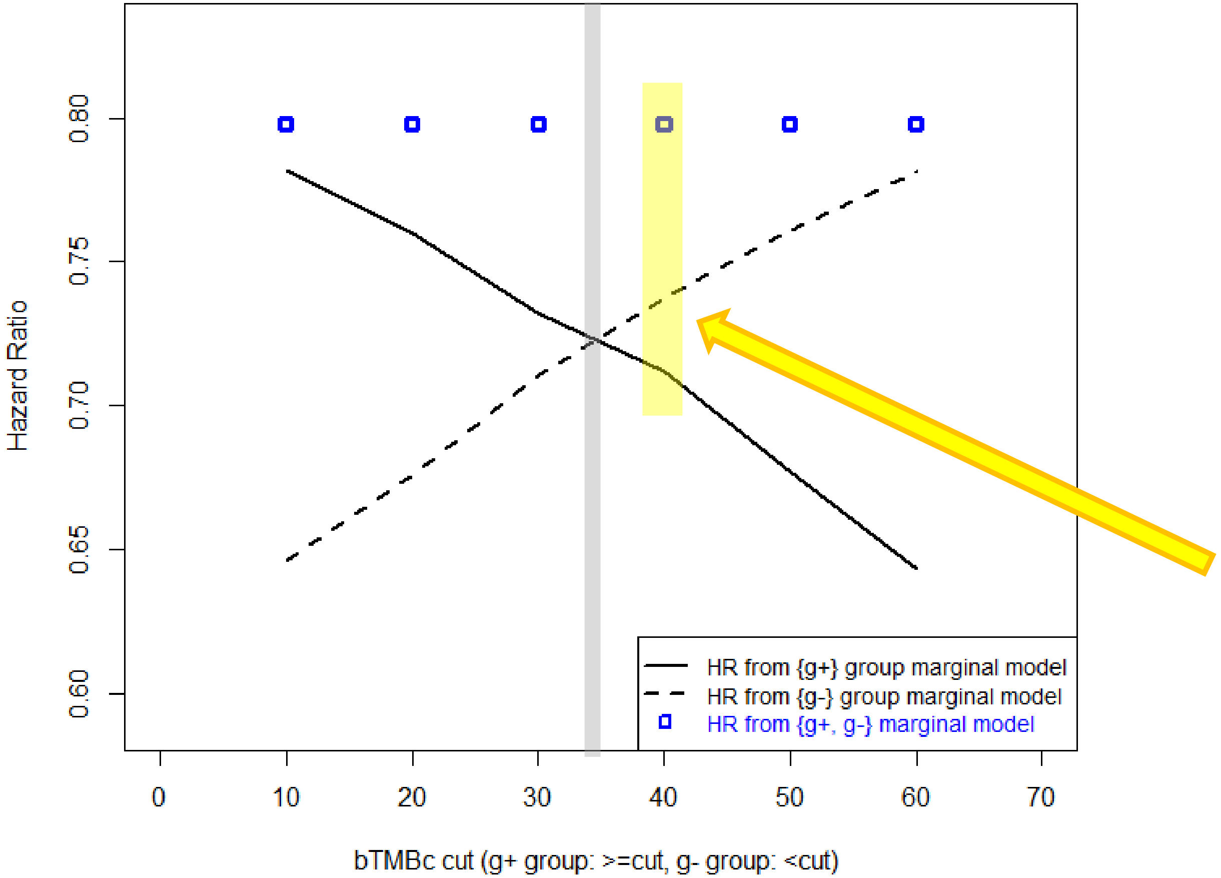


bTMBc	HR for {g-}
<10	0.646
<20	0.676
<30	0.710
<40	0.737
<50	0.761
<60	0.781

Conflicting message in terms which pt subgroup benefits most

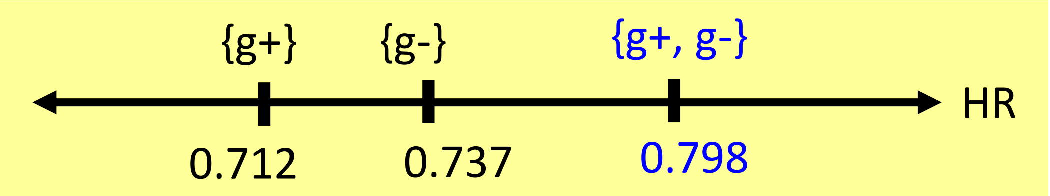
Per disjoint biomarker subgroup, generated 10,000 (total 70,000) time-to-event random variable that follows Weibull distribution. Simulated data present purely prognostic biomarker (i.e. constant HR within each disjoint biomarker subgroup but with increasing baseline hazard across different subgroups).

HR behavior for purely prognostic biomarker based on simulation



For any cut point of the bTMBc value, the marginal HR for whole data {g+, g-} is always outside range of the HRs of bTMBc subgroups.

Ex) bTMBc cut = 40



Marginal HR: HR for the overall population using Trt as only covariate in the cox model

Our proposal

In population space

- *logic respecting Estimands**:
  - $\theta \in [\theta_{g^-}, \theta_{g^+}]$
  - θ is efficacy in {g-, g+}
  - θ_{g^-} is efficacy in {g-}
  - θ_{g^+} is efficacy in {g+}

In sample space

- *Logic-ensuring Estimation:*
  - Analysis principles that ensures logical relationships in the estimates
  - $\hat{\theta} \in [\hat{\theta}_{g^-}, \hat{\theta}_{g^+}]$
  - Subgroup Mixable Estimation (SME)*

*Ding et al (2016); Lin et al (2019)

population space

Logic-respecting vs collapsible Estimands

Logic-respecting

$$\theta \in [\theta_{g^-}, \theta_{g^+}]$$

- No requirement on weights

Collapsible*

$$\theta = (w_{g^-}\theta_{g^-} + w_{g^+}\theta_{g^+})/(w_{g^-} + w_{g^+})$$

- Introduced in general setting, not specific to subgroups
- Require specification of weights $w_{g^-}, w_{g^+} \geq 0$

Commonalities:

- Population level definition
- Not tied to specific models
- Non-logic-respecting and non-collapsible behavior are different from confounding and can occur despite randomization and large sample size

*Huitfeldt et al. (2019)

population space

Logic respecting efficacy estimands for all endpoint types

Endpoint type	Efficacy Estimand	Logic-respecting?
Continuous	Difference of means	Yes
Binary	Difference of props	Yes
	Relative risk (RR)	Yes
	Odds ratio (OR)	No
Time-to-event (TTE)	HR	No
	Difference of medians	No
	Ratio of medians (RoM)	Yes*
	Difference of RMSTs/milestone probabilities	Yes
	Ratio of RMSTs/milestone probabilities	Yes

* When there is proportional hazards within each subgroup under Weibull model

Sample space

Incorrect analysis methods in analyzing real clinical trial data

- For non-logic-respecting efficacy measures such as HR
  - LSMEANS in PROC PHREG produces marginal HR that is between the subgroup HRs by

$$\boxed{HR_m = exp\{ \gamma^+(logHR_+) + \gamma^-(logHR_-)\}}$$

  - So it appears that marginal HR is always in between subgroup HRs
  - However, this is not the real marginal HR
- For logic-respecting efficacy measures in the form of difference of expectation
  - Marginal models/analysis can lead to illogical behavior in estimates

Sample space

Marginal model estimates can lead to illogical behavior even for logic respecting efficacy measure

Two models to estimate difference of means (DoM): $\theta = E(Y_i|T_i = Rx) - E(Y_i|T_i = C)$

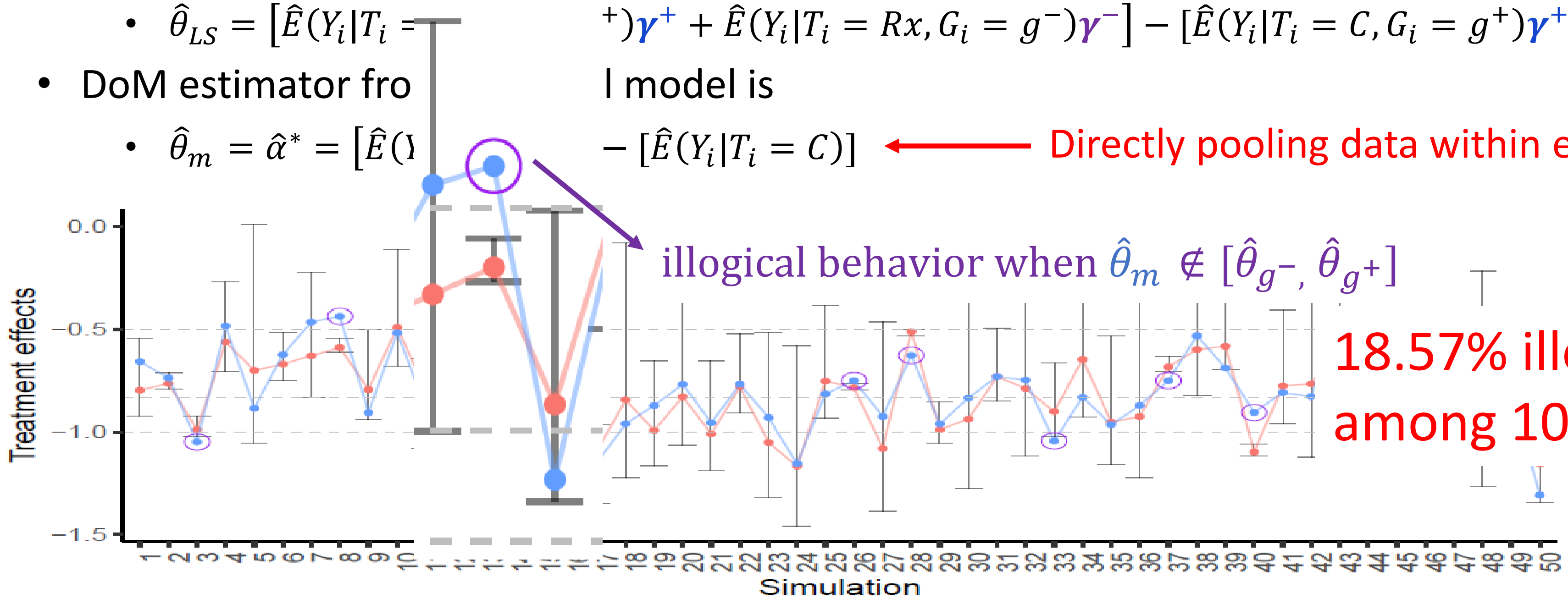
Conditional model : $Y_i = \mu + \alpha T_i + \beta G_i + \delta T_i G_i + \varepsilon_i$ **with** $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Marginal model : $Y_i = \mu^* + \alpha^* T_i + \varepsilon_i$ **with** $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Mix within each Rx and C using $\boldsymbol{\gamma^+}, \boldsymbol{\gamma^-}$

- DoM estimator fro nal model is LS-means estimator
  - $\hat{\theta}_{LS} = [\hat{E}(Y_i|T_i =$ $^+)\boldsymbol{\gamma^+} + \hat{E}(Y_i|T_i = Rx, G_i = g^-)\boldsymbol{\gamma^-}] - [\hat{E}(Y_i|T_i = C, G_i = g^+)\boldsymbol{\gamma^+} + \hat{E}(Y_i|T_i = C, G_i = g^-)\boldsymbol{\gamma^-}]$
- DoM estimator fro l model is
  - $\hat{\theta}_m = \hat{\alpha}^* = [\hat{E}($ $- [\hat{E}(Y_i|T_i = C)]$

Directly pooling data within each treatment arm



illogical behavior when $\hat{\theta}_m \notin [\hat{\theta}_{g^-}, \hat{\theta}_{g^+}]$

18.57% illogical behavior among 10,000 simulations*

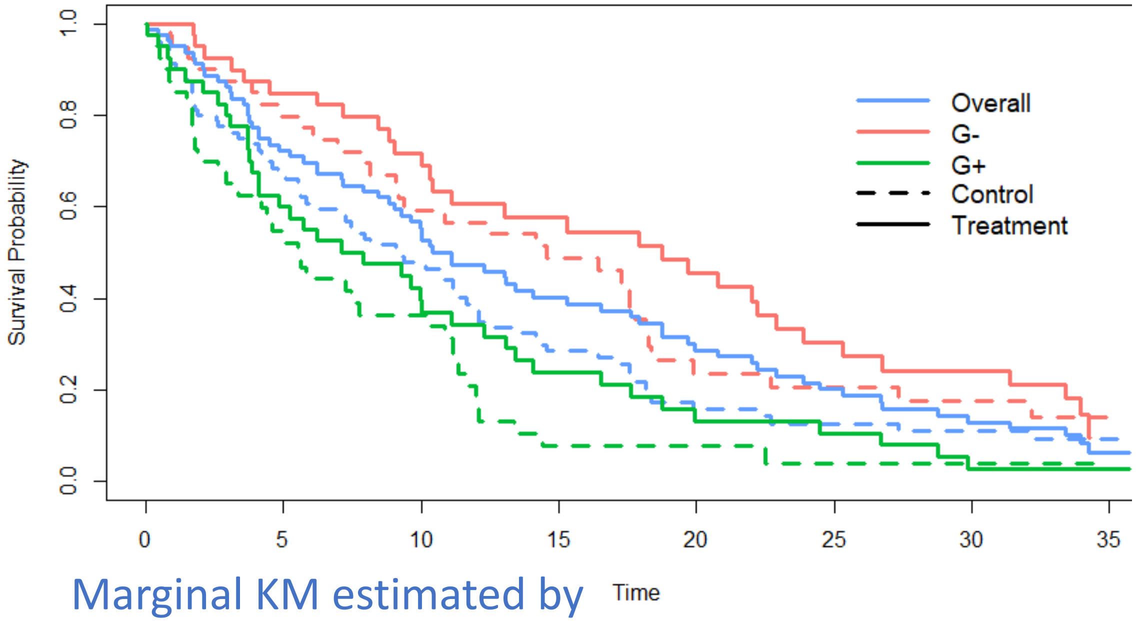
● $\hat{\theta}_{LS}$, ● $\hat{\theta}_m$, — $\hat{\theta}_{g^+}, \hat{\theta}_{g^-}$,

*$\mu = 1, \alpha = -1, \beta = 1, \delta = 0.5, \gamma^+ = 1/3, \sigma = 1$, 1:1 allocation with N=120

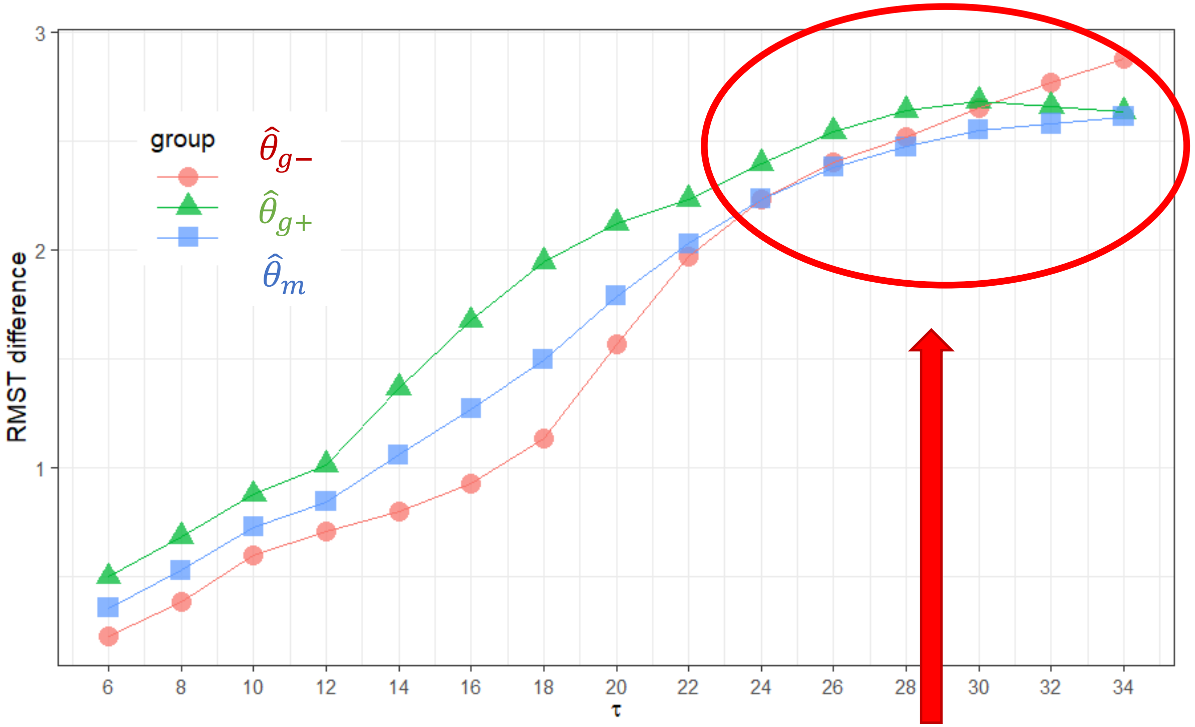
Sample space

RMST difference based on marginal KM curves may disrespect logic

N=160, 1:1 RR, $\gamma^{+} = 0.5$*



Marginal KM estimated by pooling g-, g+ pts in Rx and C arm separately



Even though RMST difference is logic respecting at population level, estimated RMST difference by the pooled KM estimate for Rx and C is not always in between those from the subgroups

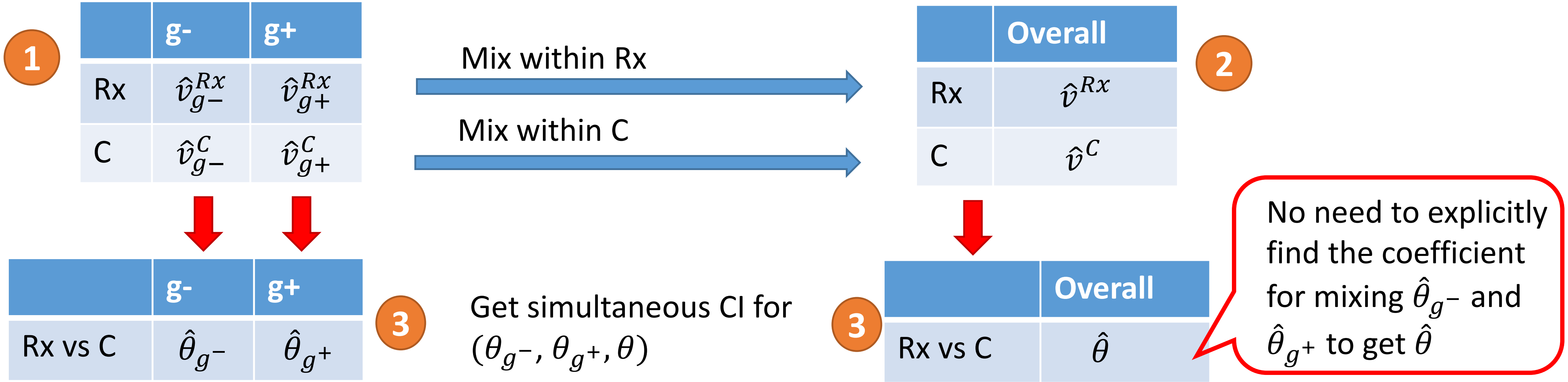
*Data generated with exponential distribution, median for C arm is 6, 10 for g+, g- and HR=0.7 for both subgroups

Sample space

Correct analysis methods for logic respecting efficacy measures for all endpoint types

Principle of Subgroup Mixable Estimation (SME)

1. Get estimated treatment effect for (g+,Rx), (g-, Rx), (g+,C), (g-,C) and associated variance matrix estimates
2. Get estimates of Rx and C treatment effect for overall pop:
   - mix within *Rx* and *C* on the probability scale by population or pooled sample prevalence
3. Calculate estimates of efficacy (Rx vs C) in g+ and g- and overall pop and associated simultaneous CI



1

	g-	g+
Rx	$\hat{\nu}_{g-}^{Rx}$	$\hat{\nu}_{g+}^{Rx}$
C	$\hat{\nu}_{g-}^{C}$	$\hat{\nu}_{g+}^{C}$

Mix within Rx

Mix within C

2

	Overall
Rx	$\hat{\nu}^{Rx}$
C	$\hat{\nu}^{C}$

3

	g-	g+
Rx vs C	$\hat{\theta}_{g^-}$	$\hat{\theta}_{g^+}$

Get simultaneous CI for $(\theta_{g^-}, \theta_{g^+}, \theta)$

3

	Overall
Rx vs C	$\hat{\theta}$

No need to explicitly find the coefficient for mixing $\hat{\theta}_{g^-}$ and $\hat{\theta}_{g^+}$ to get $\hat{\theta}$

Sample space

Following SME to produce simultaneous CI for RoM under Weibull model

- Fit Weibull model $h(t) = h_0(t) exp\{\beta_1 T + \beta_2 G + \beta_3 TG\}$
  - $h_0(t) = \kappa\lambda^{\kappa}t^{\kappa-1}$, κ and λ are the shape and rate parameters, respectively

0. Estimate all parameters and covariance matrix $\hat{\phi} = (\hat{\kappa}, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ and $\hat{\Sigma}$.

1. Within Rx or C, compute median for g+/g- and overall population $v = (v_{g^-}^{C}, v_{g^-}^{Rx}, v_{g^+}^{C}, v_{g^+}^{Rx}, v^{C}, v^{Rx}) = g(\phi, t)$ where g(.) is implicit function by solving the following equations

$$\begin{aligned} t &= v_{g^-}^{C} : g_1(\phi, t) = \exp\left(-\lambda^{\kappa}t^{\kappa}\right) - 0.5 = 0 \\ t &= v_{g^-}^{Rx} : g_2(\phi, t) = \exp\left(-e^{\beta_1}\lambda^{\kappa}t^{\kappa}\right) - 0.5 = 0 \\ t &= v_{g^+}^{C} : g_3(\phi, t) = \exp\left(-e^{\beta_2}\lambda^{\kappa}t^{\kappa}\right) - 0.5 = 0 \\ t &= v_{g^+}^{Rx} : g_4(\phi, t) = \exp\left(-e^{\beta_1+\beta_2+\beta_3}\lambda^{\kappa}t^{\kappa}\right) - 0.5 = 0 \end{aligned}$$

Mix on the probability scale within Rx and C

$$\begin{aligned} t &= v^{C} : g_5(\phi, t) = \gamma^{-}\exp\left(-\lambda^{\kappa}t^{\kappa}\right) + \gamma^{+}\exp\left(-e^{\beta_2}\lambda^{\kappa}t^{\kappa}\right) - 0.5 = 0 \\ t &= v^{Rx} : g_6(\phi, t) = \gamma^{-}\exp\left(-e^{\beta_1}\lambda^{\kappa}t^{\kappa}\right) + \gamma^{+}\exp\left(-e^{\beta_1+\beta_2+\beta_3}\lambda^{\kappa}t^{\kappa}\right) - 0.5 = 0 \end{aligned}$$

Replacing estimator $\hat{\boldsymbol{\phi}}$ with $\boldsymbol{\phi}$ above to get the estimator $\hat{\boldsymbol{v}}$

Sample space

Following SME to produce simultaneous CI for RoM under Weibull model

2. Compute the estimated variance and covariance matrix of $\hat{v}$ by the implicit delta method (Benichou & Gail 1989)

- We know $\widehat{\boldsymbol{\phi}} \sim N(\phi, \Sigma)$ from Weibull model fitting, then $\widehat{\boldsymbol{\nu}} \sim N(\boldsymbol{\nu}, \Sigma_\nu)$ where
  - $\Sigma_\nu = J^{-1}\mathrm{H}\Sigma\mathrm{H}'(J^{-1})'$
  - $J = \frac{\partial g_i}{\partial v_j}$ for i,j=1,…,6 should be a diagonal 6X6 matrix
  - $\mathrm{H} = \frac{\partial g_i}{\partial \phi_j}$ for i=1,…,6; j=1,…,5 is a 6X5 matrix
- Covariance matrix of $\hat{v}$ can be estimated as $\hat{\Sigma}_\nu$evaluated at $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\nu}})$

Sample space

Following SME to produce simultaneous CI for RoM under Weibull model

3. Calculate ratio of median for g+/g-, overall and estimated variance and covariance matrix based on multivariate Delta method
   - Let $\hat{\eta} = \log(\hat{\boldsymbol{\nu}})$ then $\hat{\eta} \sim N(log(\boldsymbol{\nu}), \Sigma_{l\nu} = D\Sigma_{\nu}D')$ where D is the diagonal matrix with $(1/\nu_i)_{i=1,\ldots,6}$
   - Let $u_1 = \eta_2 - \eta_1; u_2 = \eta_4 - \eta_3; u_3 = \eta_6 - \eta_5$ then
     - $\boldsymbol{u} = (\log(\nu_{g-}^{Rx}/\nu_{g-}^{C}), \log(\nu_{g+}^{Rx}/\nu_{g+}^{C}), \log(\nu^{Rx}/\nu^{C}))$
     - $\hat{\boldsymbol{u}} \sim N(\boldsymbol{u}, \Sigma_u = \mathrm{M}\Sigma_{l\nu}\mathrm{M}')$ where $\mathrm{M} = \partial u_i/\partial \eta_j$ i=1,2,3;j=1,…,6 is a 3X6 matrix
     - Calculate the critical value q using the multivariate normal distribution of $\hat{\boldsymbol{u}}$ as follows

$$P\left(\left|\frac{\hat{u}_1 - u_1}{se(\hat{u}_1)}\right| < q, \left|\frac{\hat{u}_2 - u_2}{se(\hat{u}_2)}\right| < q, \left|\frac{\hat{u}_3 - u_3}{se(\hat{u}_3)}\right| < q\right) = 1 - \alpha$$

     - Simultaneous CI for $\boldsymbol{u}$ is then $\boldsymbol{I_u} = I_{u_1} \times I_{u_2} \times I_{u_3}$where $I_{u_i} = \hat{u}_i \pm q \times se(\hat{u}_i)$
   - Point estimator for $(\nu_{g-}^{Rx}/\nu_{g-}^{C}, \nu_{g+}^{Rx}/\nu_{g+}^{C}, \nu^{Rx}/\nu^{C})$ is $\exp(\hat{\boldsymbol{u}})$ with simultaneous CI $\exp(\boldsymbol{I_u})$

Sample space

Following SME to produce simultaneous CI for RMST difference using non-parametric KM estimates

1. Let us use $v = (v_{g^-}^{C}, v_{g^-}^{Rx}, v_{g^+}^{C}, v_{g^+}^{Rx}, v^{C}, v^{Rx})$ to denote the RMST for g-/g+ and overall population within each treatment arm

$$\hat{v}_{g-}^{C} = \int_0^{\tau} \hat{S}_{g-}^{C}(t)\,dt, \hat{v}_{g-}^{Rx} = \int_0^{\tau} \hat{S}_{g-}^{Rx}(t)\,dt, \hat{v}_{g+}^{C} = \int_0^{\tau} \hat{S}_{g+}^{C}(t)\,dt, \hat{v}_{g+}^{Rx} = \int_0^{\tau} \hat{S}_{g+}^{Rx}(t)\,dt.$$

2. Obtain the overall Rx and C RMST estimates and claim $\hat{\boldsymbol{v}} \sim N(\boldsymbol{v}, \Sigma_{\boldsymbol{v}})$

$$\hat{v}^{C} = \gamma^{-}\hat{v}_{g-}^{C} + \gamma^{+}\hat{v}_{g+}^{C}, \hat{v}^{Rx} = \gamma^{-}\hat{v}_{g-}^{Rx} + \gamma^{+}\hat{v}_{g+}^{Rx}.$$

3. RMST difference u can be written as $\hat{\boldsymbol{u}} = \hat{H}\hat{\boldsymbol{v}} \sim N\left(H\boldsymbol{v}, \boldsymbol{\Sigma}_{\boldsymbol{u}} = H\Sigma_{v}H^{\top}\right)$ and simultaneous CI can be calculated using $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{u}}}$

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} v_{g-}^{Rx} - v_{g-}^{C} \\ v_{g+}^{Rx} - v_{g+}^{C} \\ v^{Rx} - v^{C} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix} \times \boldsymbol{v} := H \times \boldsymbol{v}.$$

Sample space

Applying SME to Keynote189 OS

Fit following Weibull model:

$$h(t) = h_0(t)exp\{\beta_1 T + \beta_2 G + \beta_3 TG\}$$

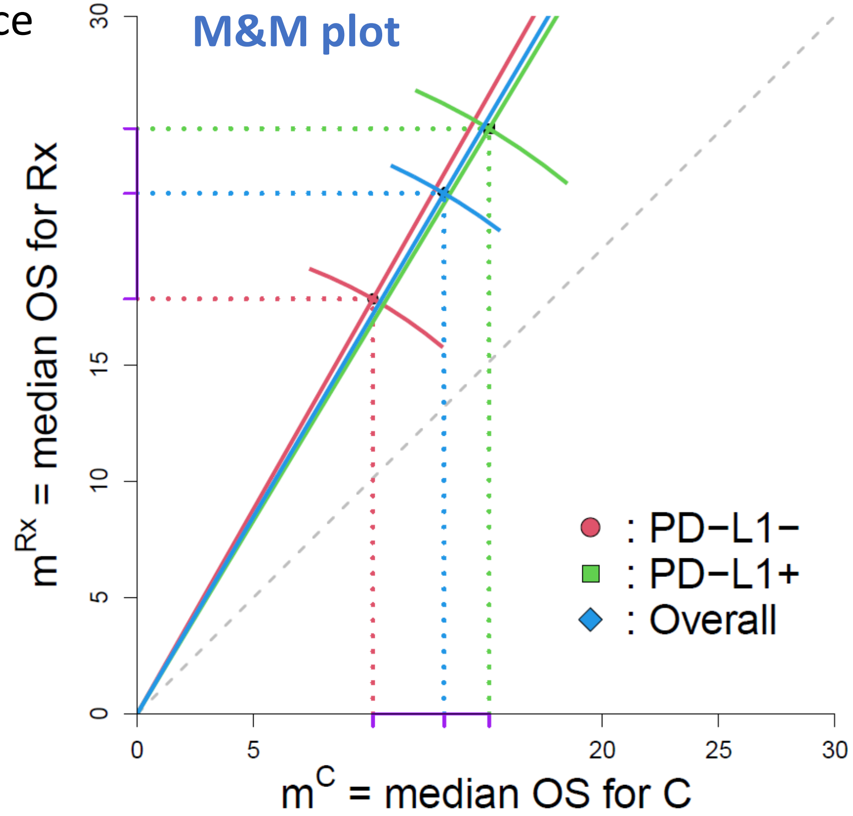
where $h_0(t) = \kappa\lambda^\kappa t^{\kappa-1}$

95% sim. CIs for RoM (right) and ratio/difference of RMST and 1-year OS rate (below)

Efficacy Measure	Group	Weibull model	
		Ratio	Difference
RMST	PD-L1-	1.393 (1.101,1.762)	4.726 (1.624,7.827)
	PD-L1+	1.245 (1.088,1.424)	3.777 (1.579,5.976)
	Overall	**1.286** **(1.143,1.446)**	**4.089** **(2.292,5.887)**
1-year survival rate	PD-L1-	1.482 (1.102,1.993)	20.8% (6.8%,34.8%)
	PD-L1+	1.261 (1.088,1.463)	15.3% (6.2%,24.4%)
	Overall	**1.320** **(1.154,1.510)**	**17.1%** **(9.5%,24.8%)**

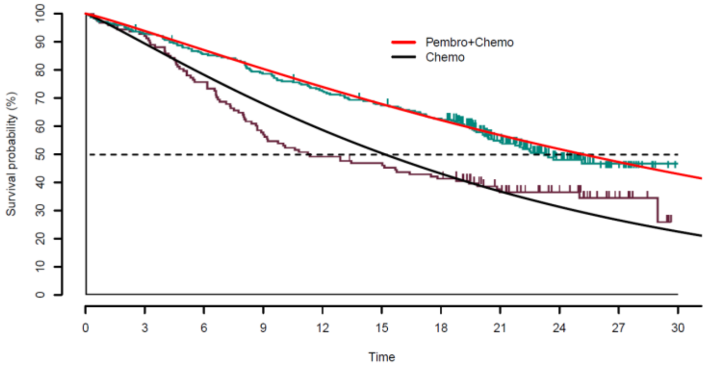
RoM estimate=1.76, 1.66, 1.70

95% sim. CI are the arcs in M&M plot

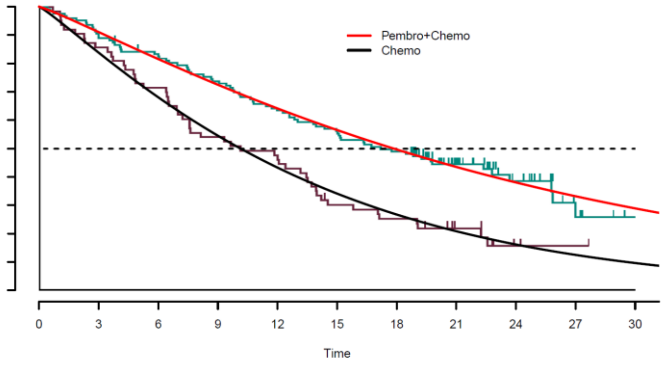


M&M plot

PD-L1+

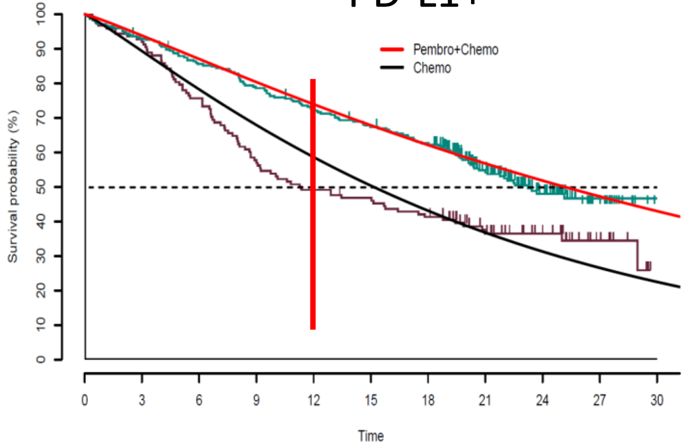


PD-L1-

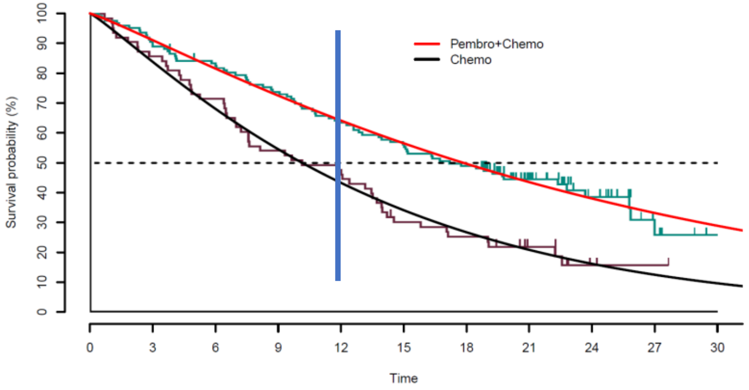


Applying SME to Keynote189 OS

PD-L1+



PD-L1-



Weibull model results

Sample space

Efficacy Measure	Group	Estimates		Ratio (95%	Difference (95%
		Rx	C	Simultaneous CI)	Simultaneous CI)
RMST (months)	PD-L1-	16.8	12.0	1.393 (1.101,1.762)	4.726 (1.624,7.827)
	PD-L1+	19.2	15.4	1.245 (1.088,1.424)	3.777 (1.579,5.976)
	Overall	18.4	14.3	1.286 (1.143,1.446)	4.089 (2.292,5.887)
1-year OS rate (%)	PD-L1-	64.1	43.3	1.482 (1.102,1.993)	20.8 (6.8,34.8)
	PD-L1+	73.9	58.6	1.261 (1.088,1.463)	15.3 (6.2,24.4)
	Overall	70.7	53.6	1.320 (1.154,1.510)	17.1 (9.5,24.8)

Non-parametric KM results

Efficacy Measure	Group	Estimates		Ratio (95%	Difference (95%
		Rx	C	Simultaneous CI)	Simultaneous CI)
RMST (months)	PD-L1-	16.9	12.1	1.392 (1.103,1.756)	4.746 (1.638,7.855)
	PD-L1+	19.1	15.0	1.275 (1.098,1.481)	4.110 (1.736,6.483)
	Overall	18.3	14.0	1.308 (1.153,1.484)	4.319 (2.426,6.212)
1-year OS rate (%)	PD-L1-	64.0	47.6	1.344 (0.954,1.895)	16.4 (-1.2,34.0)
	PD-L1+	73.1	49.2	1.485 (1.185,1.861)	23.9 (11.8,35.9)
	Overall	70.1	48.7	1.440 (1.261,1.644)	21.4 (11.5,31.4)

Summary

- Using non-logic respecting efficacy measures such as HR can potentially harm patients due to incorrect treatment benefit assessment
- Explaining to clinicians that "*HR in the overall pop and HR in the subgroups are apples and oranges and should not be compared*" is not the right message

Our recommendation:

- Summarize clinical trial results with logic respecting efficacy measure
- Use SME to correctly analyze clinical trial results using either parametric or non-parametric approaches to guarantee logical behavior (thus marginal agreeing with conditional)
    - Shiny app and R codes available for implementation

References

- Liu, Y, Wang, B, Yang, M, Hui, J, Xu, H, Kil, S, Hsu, JC. Correct and logical causal inference for binary and time-to-event outcomes in randomized controlled trials. Biometrical Journal. 2022; 64: 198– 224. https://doi.org/10.1002/bimj.202000202
- Liu, Y., Wang, B., Tian, H., & Hsu, J. C. Rejoinder for discussions on correct and logical causal inference for binary and time-to-event outcomes in randomized controlled trials. Biometrical Journal, 2022; 64: 246– 255. https://doi.org/10.1002/bimj.202100089
- Xi, D., Bretz, F. Discussion on 'correct and logical causal inference for binary and time-to-event outcomes in randomized controlled trials'. Biometrical Journal. 2022; 64: 243– 245. https://doi.org/10.1002/bimj.202100060
- Pennello, G, Xu, D. Discussion on "Correct and logical causal inference for binary and time-to-event outcomes in randomized controlled trials" by Yi Liu, Bushi Wang, Miao Yang, Jianan Hui, Heng Xu, Siyoen Kil, and Jason C. Hsu. Biometrical Journal. 2022; 64: 225– 234. https://doi.org/10.1002/bimj.202000320
- Didelez, V, Stensrud, MJ. On the logic of collapsibility for causal effect measures. Biometrical Journal. 2022; 64: 235– 242. https://doi.org/10.1002/bimj.202000305
- Rubin, D. B. (1978). *Annals* of *Statistics* 6, 34–58.
- Holland, P. (1986). *J. Amer*. *Statist*. *Assoc*. 81, 945–970.
- Huitfeldt, A., Stensrud, M.J. & Suzuki, E. On the collapsibility of measures of effect in the counterfactual causal framework. Emerg Themes Epidemiol 16, 1 (2019). https://doi.org/10.1186/s12982-018-0083-9
- Greenland, Sander, James M. Robins, and Judea Pearl (1999). Confounding and Collapsibility in Causal Inference. *Statistical Science* 14, 29–46.
- Ding, Ying, Hui-Min Lin, and Jason C. Hsu (2016). Subgroup Mixable Inference on treatment efficacy in mixture populations, with an application to time-to-event outcomes. *Statistics in Medicine* 35, 1580–1594.
- Lin, Hui-Min, Haiyan Xu, Ying Ding, and Jason C. Hsu (2019). Correct and Logical Inference on efficacy in subgroups and their mixture for binary outcomes. *Biometrical Journal* 61, 8–26.
- Ding, Peng and Fan Li (2018). *Statistical Science* 33, 214–237.

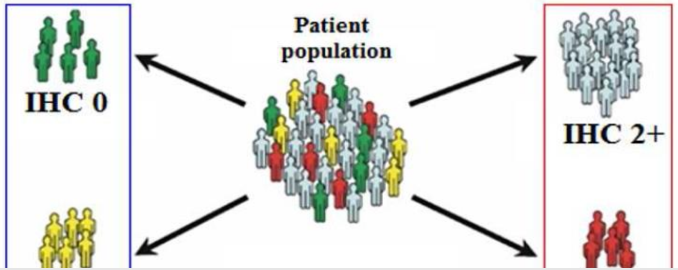
References

- Gandara et al (2018). Blood-based tumor mutational burden as predictor … NSCLC … atezolizumab. *Nature Medicine* 24, 1441–1448.
- Spigel et. al. (2013). Randomized phase II trial of onartuzumab in combination with erlotinib in patients with advanced non-small-cell lung cancer. *Journal of Clinical Oncology:* 31(32): 4105-4114.
- Paz-Ares, L., Ciuleanu, T.E., Cobo, M., Schenker, M., Zurawski, B., Menezes, J., Richardet, E., Bennouna, J., Felip, E., Juan-Vidal, O. and Alexandru, A., 2021. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international, randomised, open-label, phase 3 trial. The Lancet Oncology, 22(2), pp.198-211.
- Powles, T., Plimack, E.R., Soulières, D., Waddell, T., Stus, V., Gafanov, R., Nosov, D., Pouliot, F., Melichar, B., Vynnychenko, I. and Azevedo, S.J., 2020. Pembrolizumab plus axitinib versus sunitinib monotherapy as first-line treatment of advanced renal cell carcinoma (KEYNOTE-426): extended follow-up from a randomised, open-label, phase 3 trial. *The Lancet Oncology*, *21*(12), pp.1563-1573.
- Daniel, R, Zhang, J, Farewell, D. Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal*. 2021; 63: 528– 557. https://doi.org/10.1002/bimj.201900297

Back up

Clinical Trials with two subgroups where HR is not logic respecting

MET study: Ph2 NSCLC[1]

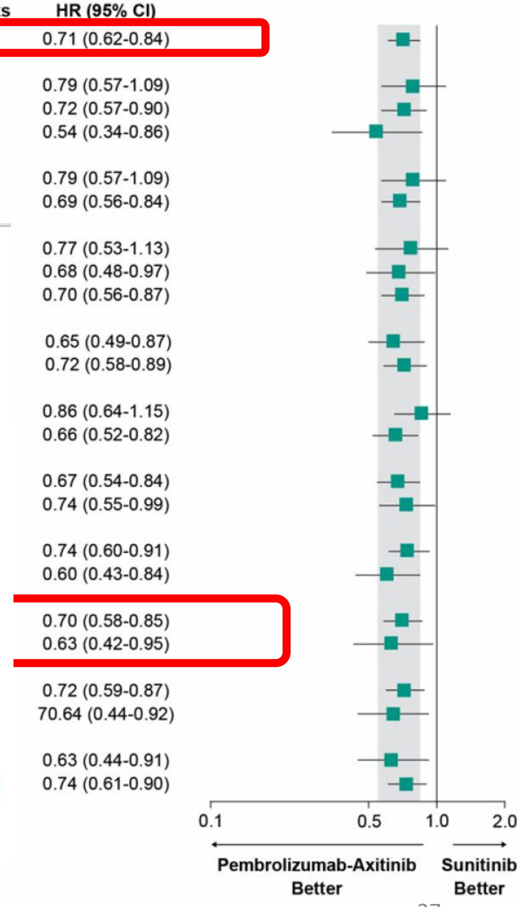


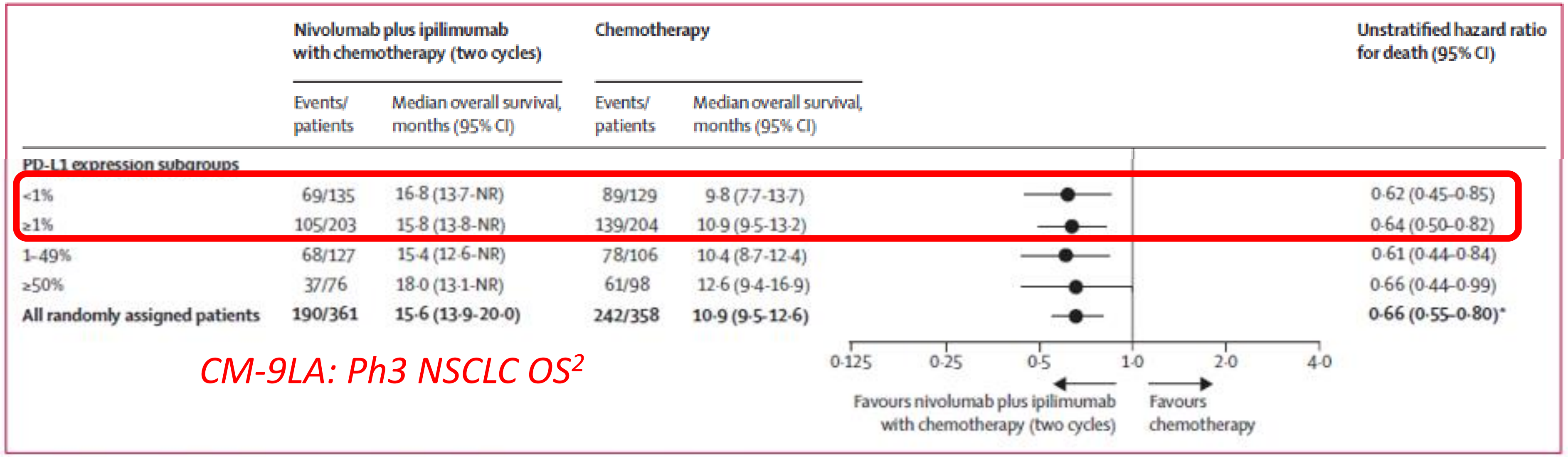
HR=0.47

HR=0.34

KN-426: Ph3 RCC PFS[3]

	Pembrolizumab + Axitinib N/No. Events	Sunitinib N/No. Events	HR (95% CI)
Overall	**432/264**	**429/281**	0.71 (0.62-0.84)
IMDC category 1			
Favorable	138/77	131/75	0.79 (0.57-1.09)
Intermediate	238/145	246/163	0.72 (0.57-0.90)
Poor	56/42	52/43	0.54 (0.34-0.86)
IMDC risk category 2			
Favorable	138/77	131/75	0.79 (0.57-1.09)
Intermediate/Poor	294/187	298/206	0.69 (0.56-0.84)
Geographic region			
			0.77 (0.53-1.13)
			0.68 (0.48-0.97)
			0.70 (0.56-0.87)
			0.65 (0.49-0.87)
			0.72 (0.58-0.89)
			0.86 (0.64-1.15)
			0.66 (0.52-0.82)
			0.67 (0.54-0.84)
			0.74 (0.55-0.99)
			0.74 (0.60-0.91)
			0.60 (0.43-0.84)
			0.70 (0.58-0.85)
			0.63 (0.42-0.95)
			0.72 (0.59-0.87)
			70.64 (0.44-0.92)
			0.63 (0.44-0.91)
			0.74 (0.61-0.90)





	Nivolumab plus ipilimumab with chemotherapy (two cycles)		Chemotherapy		Unstratified hazard ratio for death (95% CI)
	Events/ patients	Median overall survival, months (95% CI)	Events/ patients	Median overall survival, months (95% CI)	
PD-L1 expression subgroups					
<1%	69/135	16·8 (13·7-NR)	89/129	9·8 (7·7-13·7)	0·62 (0·45-0·85)
≥1%	105/203	15·8 (13·8-NR)	139/204	10·9 (9·5-13·2)	0·64 (0·50-0·82)
1-49%	68/127	15·4 (12·6-NR)	78/106	10·4 (8·7-12·4)	0·61 (0·44-0·84)
≥50%	37/76	18·0 (13·1-NR)	61/98	12·6 (9·4-16·9)	0·66 (0·44-0·99)
All randomly assigned patients	**190/361**	**15·6 (13·9-20·0)**	**242/358**	**10·9 (9·5-12·6)**	**0·66 (0·55-0·80)***

0·125
0·25
0·5
1·0
2·0
4·0
Favours nivolumab plus ipilimumab with chemotherapy (two cycles)
Favours chemotherapy

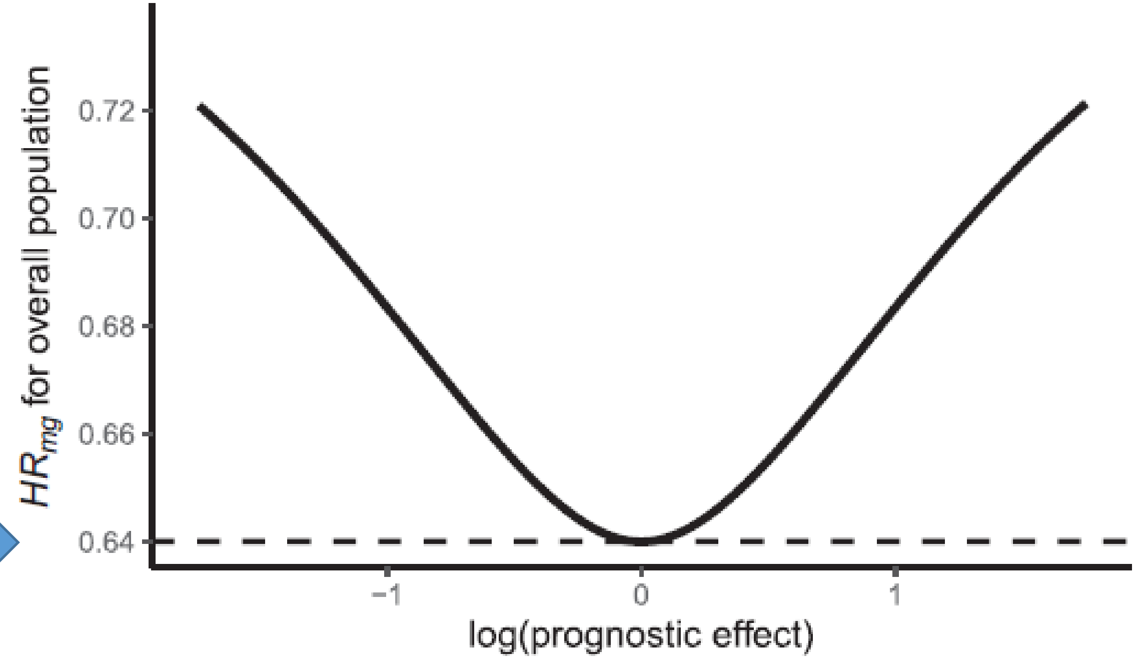
CM-9LA: Ph3 NSCLC OS[2]

Figure 3: Forest plot of overall survival based on longer follow-up in predefined patient subgroups
ECOG=Eastern Cooperative Oncology Group. NR=not reached. *Stratified hazard ratio. Unstratified hazard ratio was 0·67 (95% CI 0·55–0·81).

1.Spigel et. al. (2013). 2. Paz-Ares et. Al. (2021); 3. Powles et. Al. (2020)

Conditional and marginal HR disagree at both pop and sample level

- At population level:
  - With a purely prognostic subgroup G={g+,g-}, marginal HR gets closer to 1 than the common subgroup HR



Subgroup HR=0.64 for g+ and g-

50% prevalence; prognostic effect is the HR between g+ and g-; HR_mg is calculated as HR from the cox model with Trt as the only covariate – even though the theoretical HR for overall pop depends on time when prognostic effect is present; HR_mg is viewed as average HR (Xu and O’Quigley 2000)

population space

Logic-respecting vs collapsible Estimands

Logic-respecting

$$\theta \in [\theta_{g^-}, \theta_{g^+}]$$

- No requirement on weights

Collapsible*

$$\theta = (w_{g^-}\theta_{g^-} + w_{g^+}\theta_{g^+})/(w_{g^-} + w_{g^+})$$

- Introduced in general setting, not specific to subgroups
- Require specification of weights $w_{g^-}, w_{g^+} \geq 0$

Commonalities:

- Population level definition
- Not tied to specific models
- Non-logic-respecting and non-collapsible behavior are different from confounding and can occur despite randomization and large sample size

*Huitfeldt et al. (2019)

population space

Causal interpretations

*“How the outcome of treatment compares to what would have happened to the same subjects under different treatment conditions…”**

Difference of expectations:

- $\boxed{E(Y_i(Rx) - Y_i(C))} = \boxed{E\left(Y_i^{Rx}\right) - E\left(Y_j^{C}\right)}$

Population average of the difference in potential outcomes when the same person takes Rx vs C

Difference in population average of observed outcomes from pts taking Rx vs other pts taking C

*ICH E9 (R1) guidance

population space

Difference of expectation (DOE)

In the setting of RCT, let G_i=g+ or g- denote subgroup and T_i=Rx or C denote randomization assignment, we have

- Ignorability: $T_i \perp (Y_i(Rx), Y_i(C)) | G_i$ ★
- $T_i \perp G_i$ ☆

- $E(Y_i(Rx) - Y_i(C)) = E(Y_i(Rx)) - E(Y_i(C)) = E\left(Y_i^{Rx}\right) - E\left(Y_j^{C}\right)$
  - $E(Y_i(Rx)) = E_G[E(Y_i(Rx)|G_i)] \overset{★}{=} E_G[E(Y_i(Rx)|T_i = Rx, G_i)] \overset{☆}{=} E_{G|T=Rx}\,[E(Y_i(Rx)|T_i = Rx, G_i)] = E(Y_i(Rx)|T_i = Rx) = E\left(Y_i^{Rx}\right)$
  - Similarly $E(Y_i(C)) = E\left(Y_i^{C}\right) = E\left(Y_j^{C}\right)$
- DoE in the overall is a weighted ave of DoE in the subgroups by prevalence
  - $E\left(Y_i^{Rx}\right) - E\left(Y_i^{C}\right)$= $[E\left(Y_i^{Rx}|G_i = g^+\right)\gamma^+ + E\left(Y_i^{Rx}|G_i = g^-\right)\gamma^-] - [E\left(Y_i^{C}|G_i = g^+\right)\gamma^+ + E\left(Y_i^{C}|G_i =\right.$

DOE for g+

DOE for g-

population space

Efficacy estimand in the form of ratio

- Following similar ideas for difference, ideally one is interested in

$$E(Y_i(Rx)/Y_i(C))$$

  - Population average of the ratio in potential outcomes when the same person i takes Rx vs C, but can't be estimated using observed data Y_i^{Rx}, Y_j^{C}
  - $E(Y_i(Rx)/Y_i(C)) \neq E(Y_i^{Rx}/Y_j^{C})$ as i and j are from different pts
- <u>Alternative 1:</u> $E(Y_i(Rx))/E\big(Y_i(C)\big)$=$E(Y_i^{Rx})/E(Y_j^{C})$
  - Example: Relative Risk for binary endpoint
  - Note: $$\frac{E(Y_i(Rx))}{E(Y_i(C))} \neq E_G\left[\frac{E(Y_i(Rx)|G_i)}{E(Y_i(C)|G_i)}\right]$$

population space

Efficacy estimand in the form of ratio

- Ideally, we want $\mathrm{E}\left[\frac{Y_i(Rx)}{Y_i(C)}\right]$
- <u>Alternative 2</u>:
  - $\mathrm{E}\left[\log\left(\frac{Y_i(Rx)}{Y_i(C)}\right)\right] = \mathrm{E}[\log(Y_i(Rx)) - \log(Y_i(C))] = \mathrm{E}[\log(Y_i^{Rx}) - \log(Y_j^{C})] = \mathrm{E}\left[\log\left(\frac{Y_i^{Rx}}{Y_j^{C}}\right)\right]$
  - This is a different estimand from $\log\left[\mathrm{E}\left(\frac{Y_i(Rx)}{Y_i(C)}\right)\right]$
  - Under log normal, common variance (e.g. bioequivalence), it relates to “Alternative 1” – assumption doesn’t hold with subgroup effect
    - $\mathrm{E}\left[\log\left(\frac{Y_i(Rx)}{Y_i(C)}\right)\right] = \mathrm{E}\left[\log\left(\frac{Y_i^{Rx}}{Y_j^{C}}\right)\right] = \log\left[\frac{E[Y_i^{Rx}]}{E[Y_j^{C}]}\right]$

population space

Hypothetical example 1- prognostic & predictive subgroup effect

Same pt taking Rx and C, 10^6 SS each cell, Total N=2*10^6, $Y_i(Rx)/Y_i(C)$ can't be observed

Survival time (months)	g+	g-
$Y_i(Rx)$	2	5
$Y_i(C)$	1	10
$Y_i(Rx)/Y_i(C)$	2	1/2

Different pts taking Rx and C, 10^6 SS each cell, Total N=4*10^6
Can be observed in clinical trials

Survival time (months)	g+	g-	g+ and g- combined
Y_i^{Rx}	2	5	0.5*2+0.5*5=3.5
Y_j^{C}	1	10	0.5*1+0.5*10=5.5

$$\mathrm{E}\left[\frac{Y_i(Rx)}{Y_i(C)}\right] = 0.5*2+0.5*\frac{1}{2} = 1.25$$

$$\not\longleftarrow$$

$$E\left(\frac{Y_i^{Rx}}{Y_j^C}\right) = 0.25*\frac{2}{1}+0.25*\frac{2}{10}+0.25*\frac{5}{1}+0.25*\frac{5}{10} = 1.925$$

$$\frac{E[Y_i(Rx)]}{E[Y_i(C)]} = \frac{0.5*2+0.5*5}{0.5*1+0.5*10} = \frac{3.5}{5.5} = 0.64$$

$$\longleftarrow$$

$$\frac{E[Y_i^{Rx}]}{E[Y_j^C]} = \frac{0.5*2+0.5*5}{0.5*1+0.5*10} = \frac{3.5}{5.5} = 0.64$$

$\mathrm{E}\left[\log\left(\frac{Y_i(Rx)}{Y_i(C)}\right)\right] = 0.5*\log(2)+0.5*\log\left(\frac{1}{2}\right) =$
$0 \quad \Rightarrow \quad 1$ (after exponentiation)

$$\longleftarrow$$

$$\mathrm{E}\left[\log\left(\frac{Y_i^{Rx}}{Y_j^C}\right)\right] = \mathrm{E}\left[\log(Y_i^{Rx})-\log(Y_j^C)\right] = \mathrm{E}\left[\log(Y_i^{Rx})\right]-\mathrm{E}\left[\log(Y_j^C)\right]$$
$$= [0.5*\log(2)+0.5*\log(5)]-[0.5*\log(1)+0.5*\log(10)] = 0$$

Q: Is there a need to define treatment effect for the **overall population?**

population space

Hypothetical example 2 - purely prognostic subgroup effect

Same pt taking Rx and C, 10^6 SS each cell, Total N=$2*10^6$, $Y_i(Rx)/Y_i(C)$ can't be observed

Survival time (months)	g+	g-
$Y_i(Rx)$	2	10
$Y_i(C)$	1	5
$Y_i(Rx)/Y_i(C)$	2	2

$$\mathrm{E}\left[\frac{Y_i(Rx)}{Y_i(C)}\right] = 0.5*2+0.5*2 = 2$$

$$\frac{E[Y_i(Rx)]}{E[Y_i(C)]} = \frac{0.5*2+0.5*10}{0.5*1+0.5*5} = \frac{6}{3} = 2$$

$\mathrm{E}\left[\log\left(\frac{Y_i(Rx)}{Y_i(C)}\right)\right] = 0.5*\log(2)+0.5*\log(2) = \log(2) \Rightarrow$ 2 (after exponentiation)

Different pts taking Rx and C, 10^6 SS each cell, Total N=$4*10^6$

Can be observed in clinical trials

Survival time (months)	g+	g-	g+ and g- combined
Y_i^{Rx}	2	10	0.5*2+0.5*10=6
Y_j^C	1	5	0.5*1+0.5*5=3

$$E\left(\frac{Y_i^{Rx}}{Y_j^C}\right) = 0.25*\frac{2}{1}+0.25*\frac{2}{5}+0.25*\frac{10}{1}+0.25*\frac{10}{5} = 3.6$$

$$\frac{E[Y_i^{Rx}]}{E[Y_j^C]} = \frac{0.5*2+0.5*10}{0.5*1+0.5*5} = \frac{6}{3} = 2$$

$$\mathrm{E}\left[\log\left(\frac{Y_i^{Rx}}{Y_j^C}\right)\right] = \mathrm{E}[\log(Y_i^{Rx})-\log(Y_j^C)] = \mathrm{E}[\log(Y_i^{Rx})]-\mathrm{E}[\log(Y_j^C)]$$
$$= [0.5*\log(2)+0.5*\log(10)]-[0.5*\log(1)+0.5*\log(5)]$$
$$= \log(2)$$

Two alternatives are consistent with the ideal estimand of expectation of ratios in this case

population space

Hypothetical example 3 - purely predictive subgroup effect

Same pt taking Rx and C, 10^6 SS each cell, Total N=$2*10^6$, $Y_i(Rx)/Y_i(C)$ can't be observed

Survival time (months)	g+	g-
$Y_i(Rx)$	2	10
$Y_i(C)$	1	1
$Y_i(Rx)/Y_i(C)$	2	10

$$\mathrm{E}\left[\frac{Y_i(Rx)}{Y_i(C)}\right] = 0.5*2+0.5*10=6$$

$$\frac{E[Y_i(Rx)]}{E[Y_i(C)]} = \frac{0.5*2+0.5*10}{0.5*1+0.5*1} = \frac{6}{1} = 6$$

$$\mathrm{E}\left[\log\left(\frac{Y_i(Rx)}{Y_i(C)}\right)\right] = 0.5*\log(2)+0.5*\log(10) = \log(\sqrt{20}) = \log(4.5)$$

⇒ 4.5 (after exponentiation)

Different pts taking Rx and C, 10^6 SS each cell, Total N=$4*10^6$
Can be observed in clinical trials

Survival time (months)	g+	g-	g+ and g- combined
Y_i^{Rx}	2	10	0.5*2+0.5*10=6
Y_j^C	1	1	0.5*1+0.5*1=1

$$E\left(\frac{Y_i^{Rx}}{Y_j^C}\right) = 0.25*\frac{2}{1}+0.25*\frac{2}{1}+0.25*\frac{10}{1}+0.25*\frac{10}{1} = 6$$

$$\frac{E[Y_i^{Rx}]}{E[Y_j^C]} = \frac{0.5*2+0.5*10}{0.5*1+0.5*1} = \frac{6}{1} = 6$$

$$\mathrm{E}\left[\log\left(\frac{Y_i^{Rx}}{Y_j^C}\right)\right] = \mathrm{E}\left[\log(Y_i^{Rx}) - \log(Y_j^C)\right] = \mathrm{E}\left[\log(Y_i^{Rx})\right] - \mathrm{E}\left[\log(Y_j^C)\right]$$
$$= [0.5*\log(2)+0.5*\log(10)] - [0.5*\log(1)+0.5*\log(1)]$$
$$= \log(4.5)$$

Ratio of expectations is the same as the expectation of ratios in this case, but not exponential of expectation of log ratios

population space

Summary on Efficacy Estimand in the form of ratio

- Among the three different causal estimands:

$$\mathrm{A} = \mathrm{E}\left[\frac{Y_i(Rx)}{Y_i(C)}\right], \qquad \mathrm{B} = \frac{E[Y_i(Rx)]}{E[Y_i(C)]}, \qquad \mathrm{C} = \mathrm{E}\left[\log\left(\frac{Y_i(Rx)}{Y_i(C)}\right)\right]$$

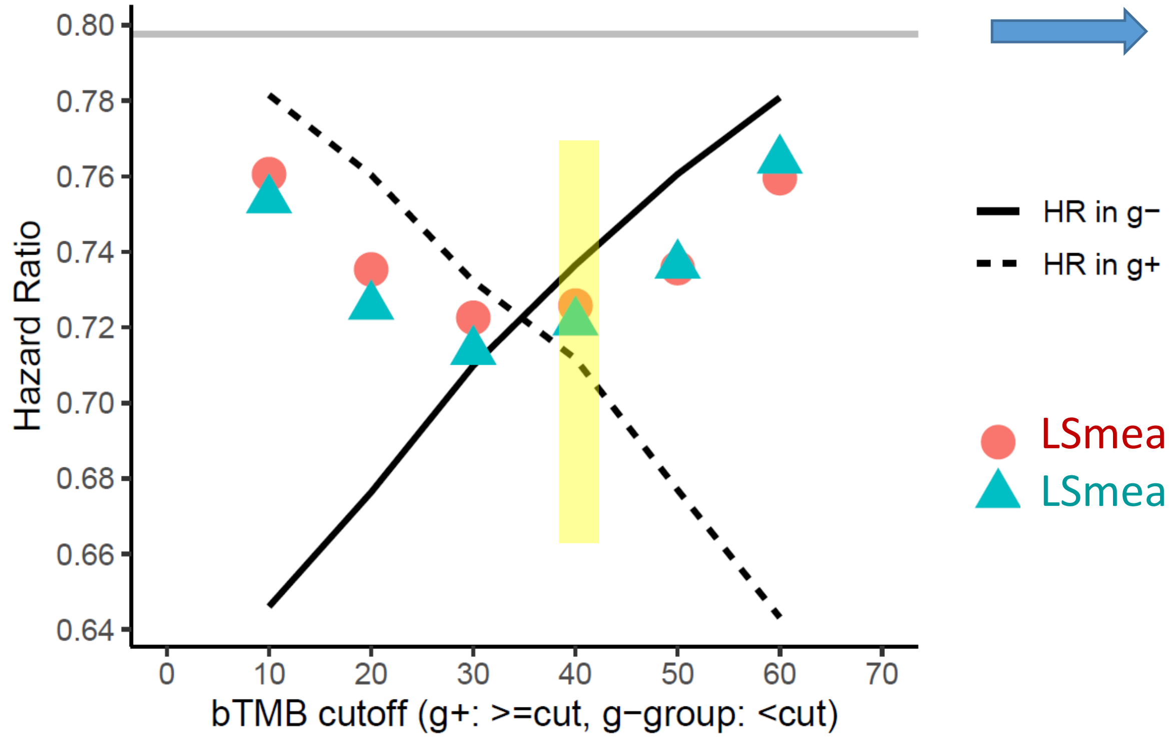
- A can’t be estimated using observed data Y_i^{Rx}, Y_j^{C} while B and C can
- C is not the same as A after exponentiation in most cases except when the subgroup effect is purely prognostic

⇒ B seems to represent treatment effect reasonably well and are the same as A in purely prognostic and purely predictive subgroup effect cases

  - Examples: RR for binary endpoint, ratio of RMSTs/Milestone probabilities for TTE endpoint

Sample space

Incorrect estimate of marginal HR in SAS LSMEANS that masks illogical behavior of HR



True marginal HR

```
PROC PHREG DATA=DA2;
CLASS TRT01P(REF="CTL") BTMB40(REF="g-") /PARAM=GLM;
MODEL OS*OSCNSR(1)=TRT01P;
HAZARDRATIO 'H1' TRT01P/DIFF=ALL CL=BOTH;
LSMEANS TRT01P;
RUN;
```

● LSmeans estimate of marginal HR from stratified model

▲ LSmeans estimate of marginal HR from unstratified model

```
PROC PHREG DATA=DA2;
CLASS TRT01P(REF="CTL") BTMB40(REF="g-") /PARAM=GLM;
MODEL OS*OSCNSR(1)=TRT01P BTMB40 TRT01P*BTMB40;
~~STRATA BTMB40;~~
HAZARDRATIO 'H1' TRT01P/DIFF=ALL CL=BOTH;
LSMEANS TRT01P/EXP BYLEVEL;
RUN;
```

$$\boxed{exp\{\,\gamma^{+}(logHR_{+}) + \gamma^{-}\,(logHR_{-})\}}$$

LSMEANS estimate of marginal HR always stays between subgroup HRs and changes depending on the cutoff value!

Sample space

Following SME to produce simultaneous CI for RMST difference using non-parametric KM estimates

1. Let us use $v = (v_{g^-}^{C}, v_{g^-}^{Rx}, v_{g^+}^{C}, v_{g^+}^{Rx}, v^{C}, v^{Rx})$ to denote the RMST for g-/g+ and overall population within each treatment arm

$$\hat{v}_{g^-}^{C} = \int_0^{\tau} \hat{S}_{g^-}^{C}(t)\,dt, \hat{v}_{g^-}^{Rx} = \int_0^{\tau} \hat{S}_{g^-}^{Rx}(t)\,dt, \hat{v}_{g^+}^{C} = \int_0^{\tau} \hat{S}_{g^+}^{C}(t)\,dt, \hat{v}_{g^+}^{Rx} = \int_0^{\tau} \hat{S}_{g^+}^{Rx}(t)\,dt.$$

2. Obtain the overall Rx and C RMST estimates and claim $\hat{\boldsymbol{v}} \sim N(\boldsymbol{v}, \Sigma_{\boldsymbol{v}})$

$$\hat{v}^{C} = \gamma^{-}\hat{v}_{g^-}^{C} + \gamma^{+}\hat{v}_{g^+}^{C}, \hat{v}^{Rx} = \gamma^{-}\hat{v}_{g^-}^{Rx} + \gamma^{+}\hat{v}_{g^+}^{Rx}.$$

3. RMST difference u can be written as $\hat{\boldsymbol{u}} = \hat{H}\hat{\boldsymbol{v}} \sim N\left(H\boldsymbol{v}, \boldsymbol{\Sigma}_{\boldsymbol{u}} = H\Sigma_{v}H^{\top}\right)$ and simultaneous CI can be calculated using $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{u}}}$

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} v_{g^-}^{Rx} - v_{g^-}^{C} \\ v_{g^+}^{Rx} - v_{g^+}^{C} \\ v^{Rx} - v^{C} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix} \times \boldsymbol{v} := H \times \boldsymbol{v}.$$

Sample space

Applying SME to Checkmate-9LA OS

HR for overall: 0.67 ∉(0.62, 0.64)

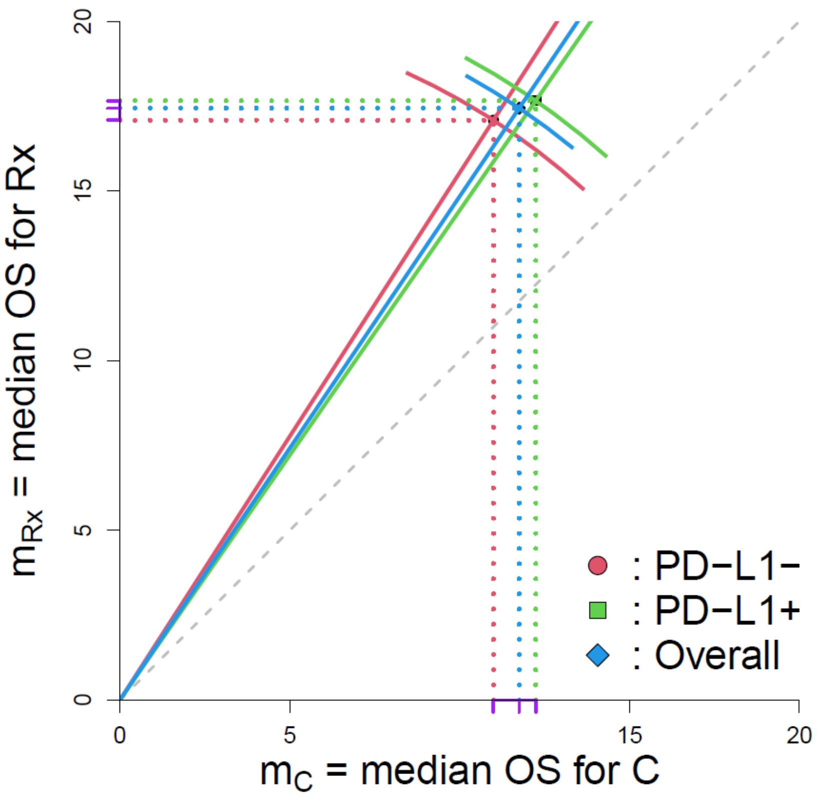
Fit following separate Weibull models:

$h_G(t) = h_{0,G}(t)exp\{\beta_G T\}$ for G={g^+, g^+}

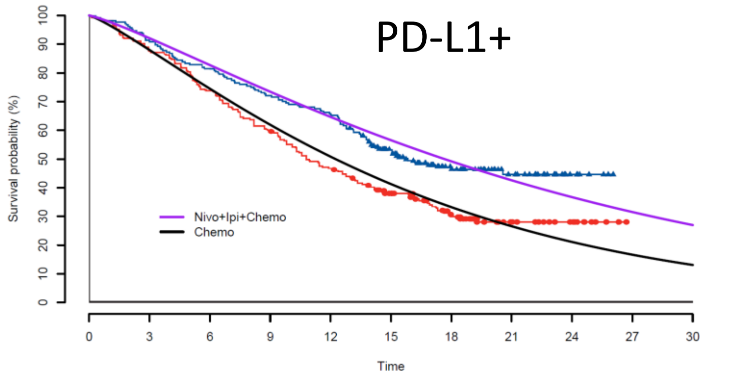
95% sim. CIs for RoM (right) and ratio/differenc of RMST and 1-year OS rate (below)

Efficacy Measure	Group	Weibull model	
		Ratio	Difference
RMST	PD-L1-	1.265 (1.058,1.512)	3.270 (0.814,5.725)
	PD-L1+	1.216 (1.065,1.388)	2.840 (0.930,4.749)
	Overall	**1.234 (1.109,1.373)**	**3.009 (1.500,4.517)**
1-year survival rate	PD-L1-	1.343 (1.068,1.688)	16.0% (4.0%,28.0%)
	PD-L1+	1.272 (1.078,1.502)	13.8% (4.5%,23.2%)
	Overall	**1.299 (1.135,1.486)**	**14.7% (7.3%,22.0%)**

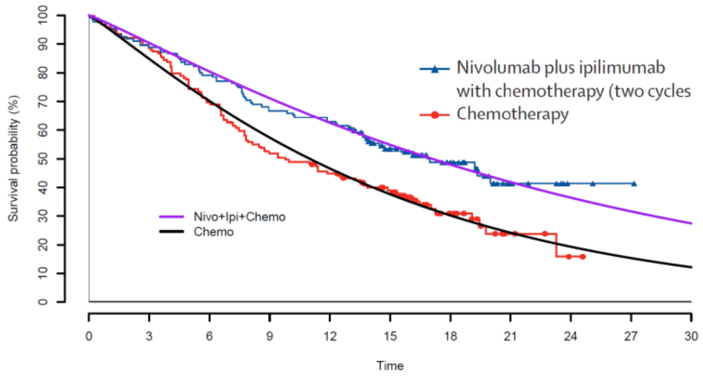
M&M plot



PD-L1+



PD-L1-


Survival probability (%)
Nivolumab plus ipilimumab with chemotherapy (two cycles
Chemotherapy
Nivo+Ipi+Chemo
Chemo
Time