Onco Estimand WG Knowledge Sharing of all task forces

June 2022

www.oncoestimand.org

Poster with more details about the WG

Oncology estimand working group: newsletter 23 May 2022 (oncoestimand.github.io)

Publications

Published:

- Lawrence, R., Degtyarev, E., Griffiths, P., Trask, P., Lau, H., D'Alessio, D., Griebsch, I., Wallenstein, G., Cocks, K., Rufibach, K. *What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials* (2020). Journal of Patient-Reported Outcomes, **4(1)**, 68. doi.
- Degtyarev, E., Rufibach, K., Shentu, Y., Yung, G., Casey, M., Englert, S., Liu, F., Liu, Y., Sailer, O., Siegel, J., Sun, S., Tang, R., Zhou, J. *Assessing the impact of COVID-19 on the objective and analysis of oncology clinical trials – application of the estimand framework* (2020). Statistics in Biopharmaceutical Research, **12(4)**, 427-437. doi | arxiv
- Casey M., Degtyarev E., Lechuga M.J., Aimone P., Ravaud A., Motzer R., Liu F., Stalbovskaya V., Tang R., Butler E., Sailer O., Halabi S., George D. *Estimand framework: Are we asking the right question? A case study in the solid tumor setting* (2020). Pharmaceutical Statistics, **20**, 324–334. doi
- Sun, S., Weber, J., Butler, E., Rufibach, K., Roychoudhury, S. *Estimands in Hematology Trials* (2021). Pharmaceutical Statistics, 20, 793-805. doi | arxiv
- Bornkamp, B., Rufibach, K., Lin, J., Liu, Y., Mehrotra, D., Roychoudhury, S., Schmidli, H., Shentu, Y., Wolbers, M. *Principal Stratum Strategy: Potential Role in Drug Development* (2021). Pharmaceutical Statistics, **20**, 737-751. doi | arxiv | github | markdown
- Manitz, J., Kan-Dobrosky, N., Buchner, H., Casadebaig, M.L., Degtyarev, E., Dey, J., Haddad, V., Fei, J., Martin, E., Mo, M., Rufibach, K., Shentu, Y., Stalbovskaya, V., Tang, R., Yung, G., Zhu, J. *Estimands for Overall Survival in Clinical Trials with Treatment Switching in Oncology* (2021). Pharmaceutical Statistics, accepted. doi.
- Lisa V. Hampson, Evgeny Degtyarev, Rui (Sammi) Tang, Jianchang Lin, Kaspar Rufibach & Cheng Zheng (2021) Comment on "Biostatistical Considerations When Using RWD and RWE in Clinical Studies for Regulatory Purposes: A Landscape Assessment", Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2021.1994459

Submitted:

- Siegel, J.M., Grinsted, L., Liu, F., Weber, J., Englert, S., Casey, M. *Censoring and censoring mechanisms in oncology in light of the estimands framework* (2022).
- Siegel, J.M. Weber, J., Englert, S. *The Role of Occlusion: Potential Extension of the ICH E9 (R1) Addendum on Estimands and Sensitivity Analysis for Time-to-Event Oncology Studies* (2022).
- Rufibach, K., Grinsted, L., Li, J., Weber, H.J., Zheng, C., and Zhou, J.. *Quantification of Follow-up Time in Oncology Clinical Trials with a Time-to-Event Endpoint: Asking the Right Questions.* (2022). Submitted. | arXiv | github | markdown.
- Liu, Y., Yang, M., Kil, S., Li, J., Mondal, S., Shentu, Y., Tian, H., Wang, L., Yung, G., From logic-respecting efficacy estimands to logic-ensuring analysis principle for time-to-event endpoint in randomized clinical trials with subgroups. (2022). Submitted.

NEW

Past and Upcoming Events

- Events with contributions from the working group (oncoestimand.github.io)

- An overview includes all past events (conferences, webinars etc) including slide decks and links to videos as well as upcoming events with confirmed WG participation
  - Use the opportunity to meet F2F if you attend any of these conferences

Outline for today

Short presentation by the following task forces:

- DoR
- Follow-up Quantification
- Principal Stratification
- EDEN
- Biomarker subgroup

Written updates in the word document from these task forces:

- PRO
- Safety
- Conditional vs Marginal
- Engagement
- RWD

Task force DOR/TTR

Task force: duration of response, time to response

What are the questions the group is working on?

- What are the relevant clinical questions to be addressed by DOR and TTR?
- Discussion of the clinical context on what DOR and TTR is based: ORR
- What are approaches to quantify DOR and TTR in the context of the corresponding clinical question of interest?
- When and how should we present DOR and TTR in clincal trials?

What has been achieved so far?

- Review of relevant literature
- Review of labels based on accelerated approvals
- Illustration of approaches based on a case study in mantle cell lypmphoma
- Presentations at SCT and PSI (2022)
- Paper draft

What are plans for the coming months?

- Present at conferences (ASA BIOP and webinars in 2nd half 2022)
- Finalize paper

Task force: duration of response, time to response

- ORR typical primary endpoint in oncology ph2 studies
  - ORR estimand implies several clinical assumptions and is agnostic to onset and durability: relevant to provide further characteristics like duration of response (DOR) and time to response (TTR)
  - We find DOR in many product labels of accelerated approvals as supportive information to ORR to illustrate that responses are not just transient
- DOR and TTR estimands are not described in protocols or publications
  - From re-engineering: typically conditional DOR and TTR are assessed (principal stratum of responders)
  - What are we interested in? In line with the intention of interpreting cDOR and cTTR as supportive information of ORR
    - “In the experimental treatment the ORR was 62% and the median DOR amongst responders was 9.7 months”
    - “... and the DOR lasted at least 6 months for the majority of responders”

Task force: duration of response, time to response

- Dealing with intercurrent events: follow considerations that are relevant for intercurrent event strategies for ORR
  - E.g., we may consider a while on treatment strategy for start of new therapies for ORR since we might want to avoid counting responses attributable to the new therapy. Accordingly, for DOR treatment policy might not be consistent and would rather choose a hypothetical strategy
- There are approaches for DOR estimands on the full population
  - Valid estimands but not very common
  - Respective questions might be more appropriately addressed in controlled ph3 studies
- Recommendations
  - Although secondary endpoints, we recommend to make the DOR and TTR estimand definitions transparent
  - We propose to stay with current practice and provide cDOR and cTTR as supportive information to ORR
  - If one wishes to compare DOR across treatment groups then valid estimands should be taken into account that also consider ORR
    - E.g. time in response (EMA 2017), probability of being in response at month x (Ellis 2008, Garnett 2013)

Follow-up Quantification

Median follow-up was 30 months.

- Sentence featured in virtually every paper reporting on clinical trial with time-to-event endpoint.
- But:
  - What can we conclude from it?
  - What is the question we are answering with this statement?
  - What do two trials with 30 months share in terms of properties?

Median follow-up was 30 months.

- Sentence featured in virtually every paper reporting on clinical trial with time-to-event endpoint.
- But:
  - What can we conclude from it? **Virtually nothing!**
  - What is the question we are answering with this statement? **Not clear!**
  - What do two trials with 30 months share in terms of properties? **Not much!**

Contributions of paper

- List of relevant scientific questions trialists have.
- Illustrate that none of them can be answered using some unclearly defined «follow-up» quantifier.
- Show how to answer these questions instead.
- 1- and 2-group case.
- Define and illustrate routinely used follow-up quantifiers. Take inspiration from estimand framework.
- Special features: Non-proportional hazards, delayed separation, cure, etc.
- Two case-studies: PH and NPH.

Conclusions

- Be clear on the scientific question(s) you want to answer!
- No hope that single number, however defined,
  - can say everything about “follow-up",
  - answers relevant questions trialists have,
  - allows to compare relevant aspects of trials.
- **Just do not give any measure of follow-up! If you must, at least define what you compute.**
- Discuss precision, stability, information (all defined in paper), and potential assumptions separately for any quantity of interest:
  - estimates of survival function(s),
  - effect measures.

Quantification of follow-up time in oncology clinical trials with a time-to-event endpoint: Asking the right questions

Kaspar Rufibach*
Lynda Grinsted†
Jiang Li‡
Hans Jochen Weber§
Cheng Zheng¶
Jiangxiu Zhou‖

June 10, 2022

https://doi.org/10.48550/arXiv.2206.05216

https://oncoestimand.github.io/quantFU/quantFU.html

Task force Treatment Switching & Principal Stratification

Intro Based on Mattei (2022)

Motivation

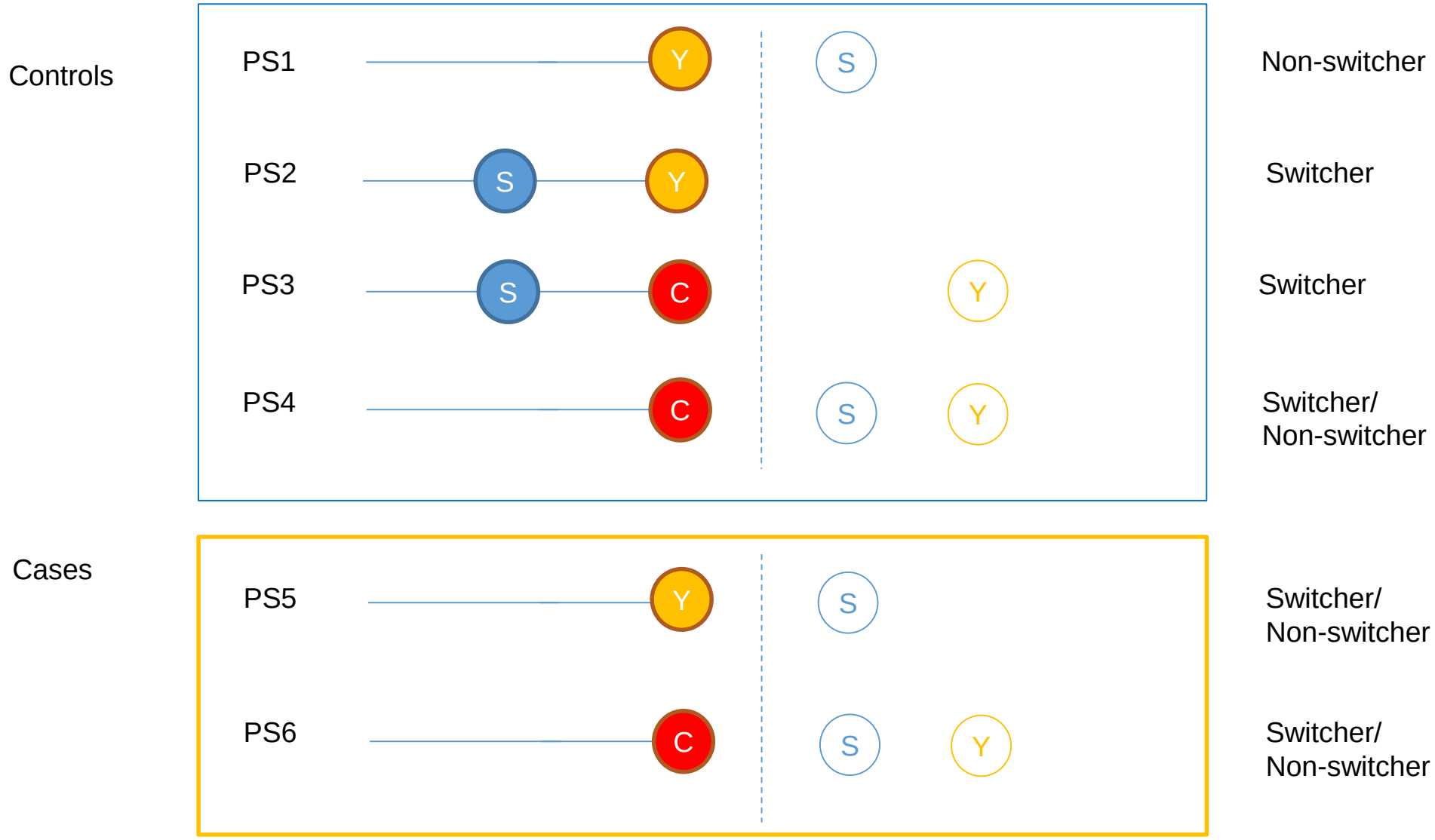
- Clinical trials focusing on survival outcomes often allow patients in the control arm to switch to the treatment arm if their physical conditions are worse than certain tolerance levels.

- The intention-to-treat analysis ignores the information of treatment switching.

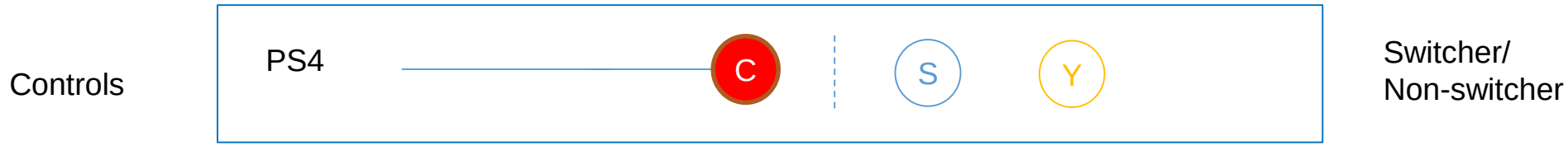
- Other existing methods propose to reconstruct the outcome a subject would have had if he or she had not switched under strong assumptions.

- The proposed method focuses on principal causal effects for patients belonging to subpopulations defined by the switching behavior under control.

Principal Strata Setup for Treatment Switching



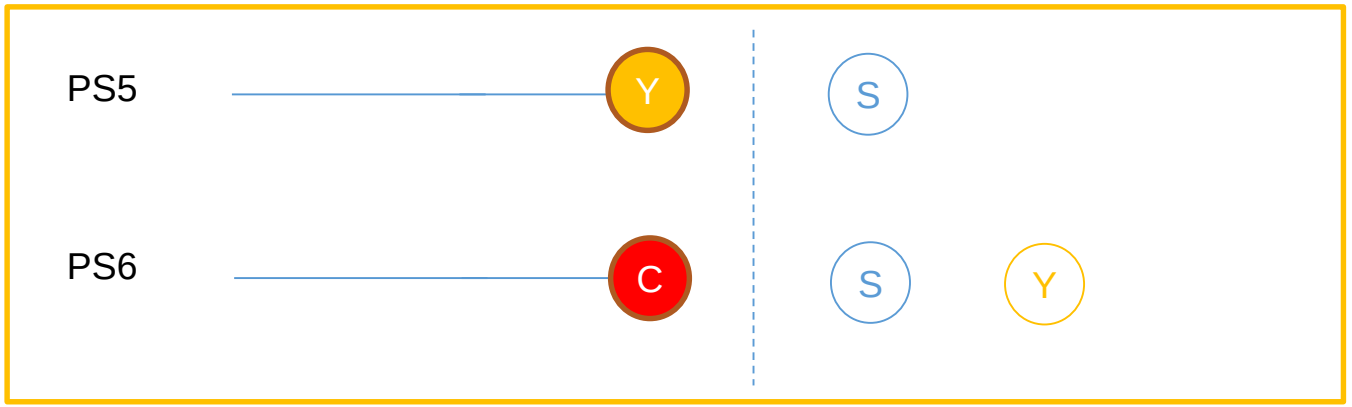
Assumptions on Switching Behaviour



$$\pi_{NS} = \frac{\pi G^{\bar{S}}_{Y(0)}(C_i)}{\pi G^{\bar{S}}_{Y(0)}(C_i) + (1-\pi) G_{S(0)}(C_i) \times 1}$$

Assumptions on Switching Behaviour

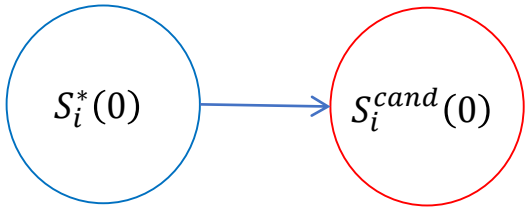
Cases



Switcher/
Non-switcher

Switcher/
Non-switcher

Metropolis-Hastings

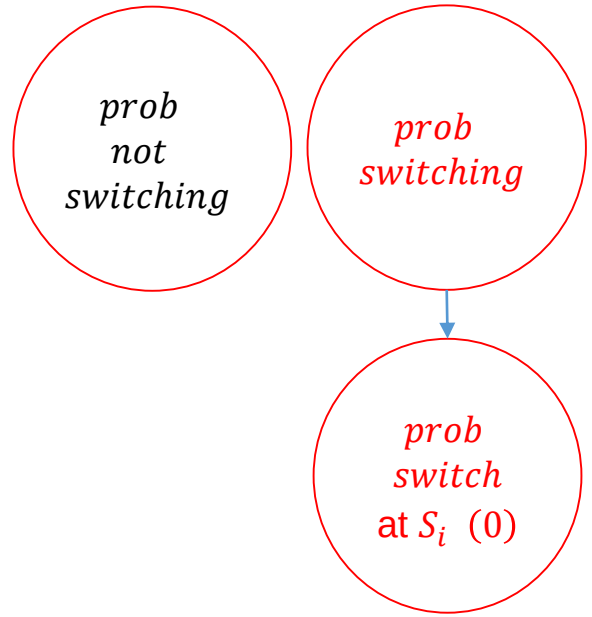


Previous draw

Currentdraw

$$Accept\ current\ draw\ with\ P = \min\left(P\left(S_i\ (0)\right), 1\right)$$

$$P\left(S_i\ (0)\right) = \frac{P\{S_i^{cand}(0)|\boldsymbol{\theta}, D_i^{obs}\}}{P\{S_i^*(0)|\boldsymbol{\theta}, D_i^{obs}\}} \times \frac{g\left(S_i^*(0)\right)}{g\left(S_i^{cand}(0)\right)}$$



Assumptions on Outcome conditioned on Switching

Controls

- $Y_i(0)|S_i(0) = \bar{S} \sim \text{Weibull}\left(\bar{\alpha}_Y, \bar{\beta}_Y\right)$
- $Y_i(0)|S_i(0) \in R_+ \sim S_i(0) + \text{Weibull}\left(\alpha_Y, \beta_Y + \boldsymbol{\lambda} \log\left(S_i(0)\right)\right)$

Cases

- $Y_i(1)|Y_i(0), S_i(0) = \bar{S} \sim \boldsymbol{\kappa} Y_i(0) + \text{Weibull}(\bar{\nu}_Y, \bar{\gamma}_Y)$
- $Y_i(1)|Y_i(0), S_i(0) \in R_+ \sim \boldsymbol{\kappa} Y_i(0) + \text{Weibull}\left(\nu_Y, \gamma_Y + \boldsymbol{\lambda} \log\left(S_i(0)\right)\right)$

Upcoming Working

- Simulate studies with survival outcomes suitable for assessing strandard treatment switching techniques (IPTW/RPSFT/Two-stage etc) and pricinpal stratification

- Benchmarking pricinpal stratification with other treatment switching methods

Biomarker subgroup task force

Task force: Time to event/binary endpoints with prognostic or predictive biomarker subgroups

- What are the questions the group is working on?
  - Among commonly used efficacy estimands, which ones are logic respecting - meaning efficacy in overall population is in between the efficacies in subgroups at population level?
  - How to ensure such logical relationships hold when analyzing real clinical trial data for time-to-event endpoint with pre-specified subgroups
- What has been achieved so far?
  - 1 JSM presentation; 1 session with 4 presentations in ASA BIOP Regulatory/Industry workshop in 2021
  - Draft paper submitted to SBR in Jan 2022; comments received in June 2022; paper revision in progress
  - ASA Biopharm section webinar presentation on May 26, 2022
- What are plans for the coming months?
  - 1 JSM, 1 ICSA, and 1 MCP conference presentation in 2022
  - Brainstorming additional topics to work on for the next chapter of the TF

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

Yi Liu
Nektar Therapeutics, San Francisco, California, USA
Miao Yang
Nektar Therapeutics, San Francisco, California, USA
Siyoen Kil
LSK Global Pharma Services Co, Seoul, Korea
Jiang Li
BeiGene, Ridgefield park, New Jersey, USA
Shoubhik Mondal
AstraZeneca, Gaithersburg, Maryland, USA
Yue Shentu
Daiichi Sankyo Inc., Basking ridge, New Jersey, USA
Hong Tian
BeiGene, Ridgefield park, New Jersey, USA
Liwei Wang
Genmab US, Inc, Princeton, New Jersey, USA
and
Godwin Yung
Genentech, South San Francisco, California, USA

January 12, 2022

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

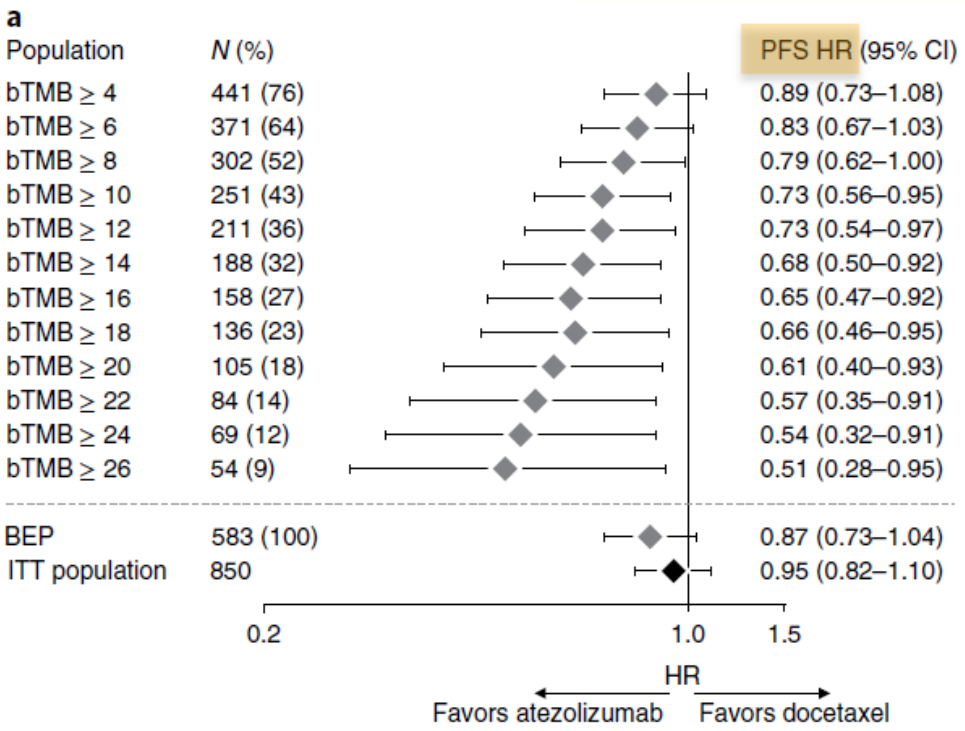
Background Atezolizumab is a humanised antiprogrammed death-ligand 1 (PD-L1) monoclonal antibody that *Lancet* 2017; 389: 255-65

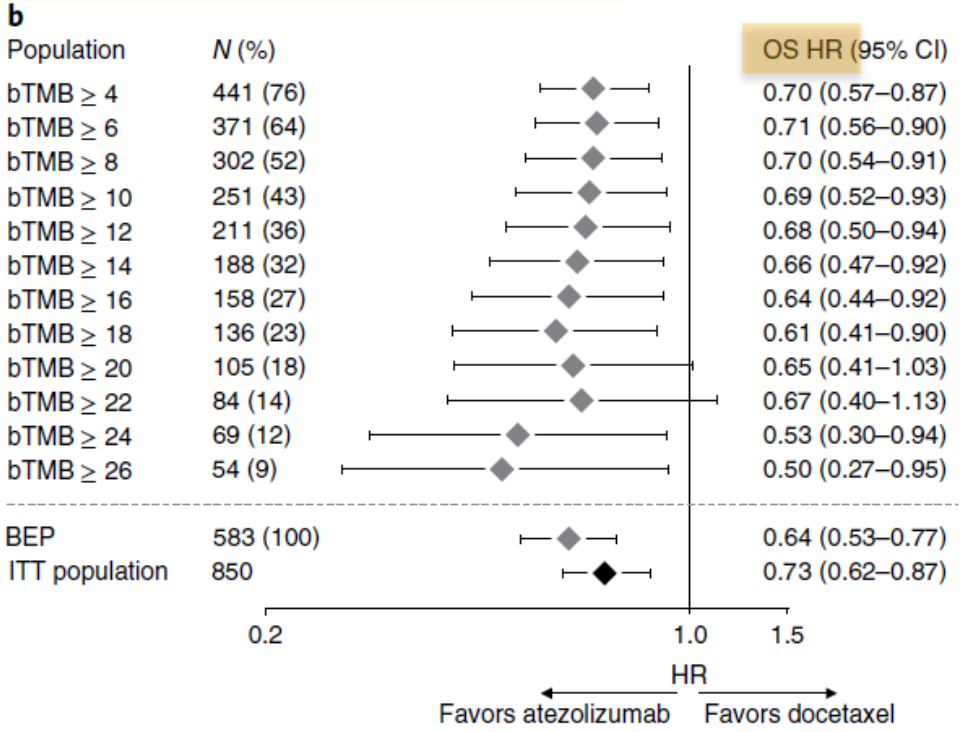
Is bTMB a predictor of clinical benefit in NSCLC patients treated with atezolizumab in OAK study?

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

PFS

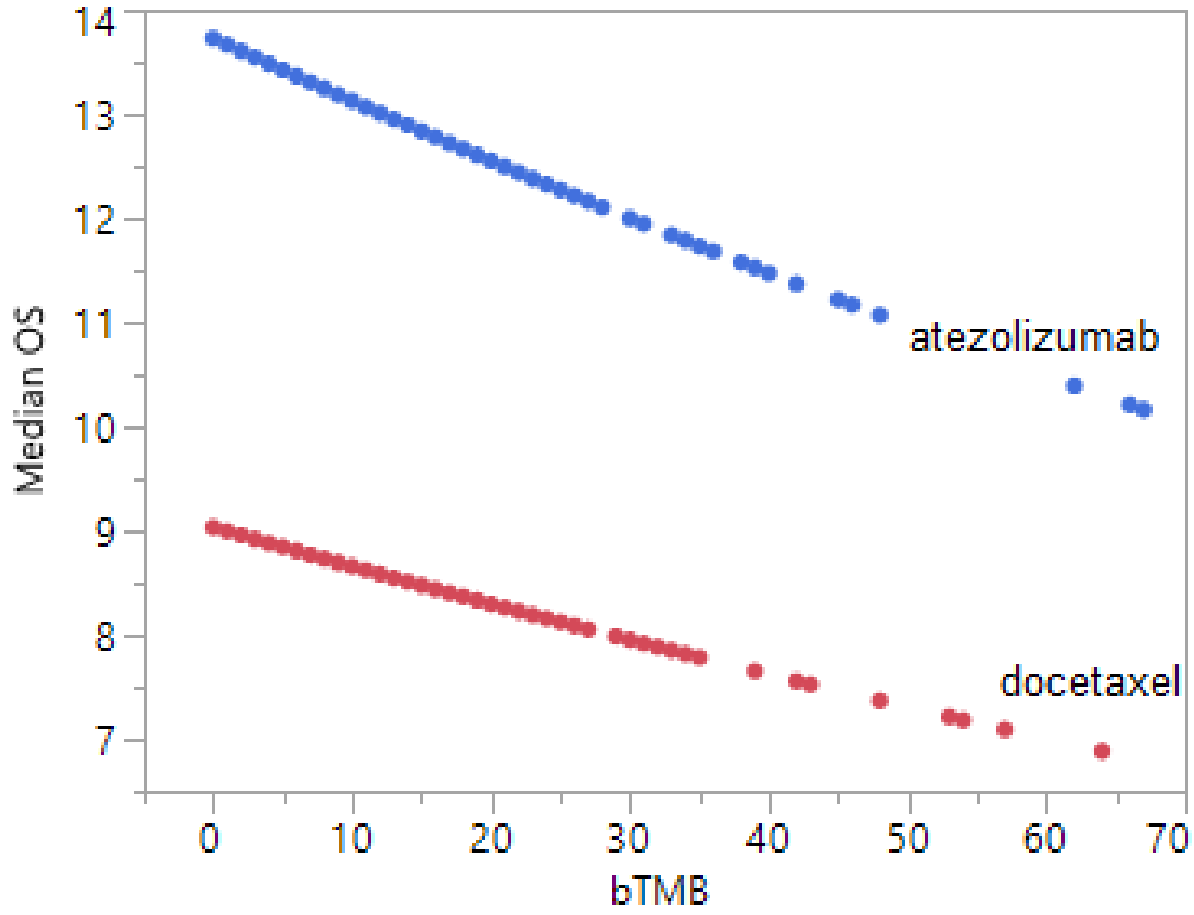
OS



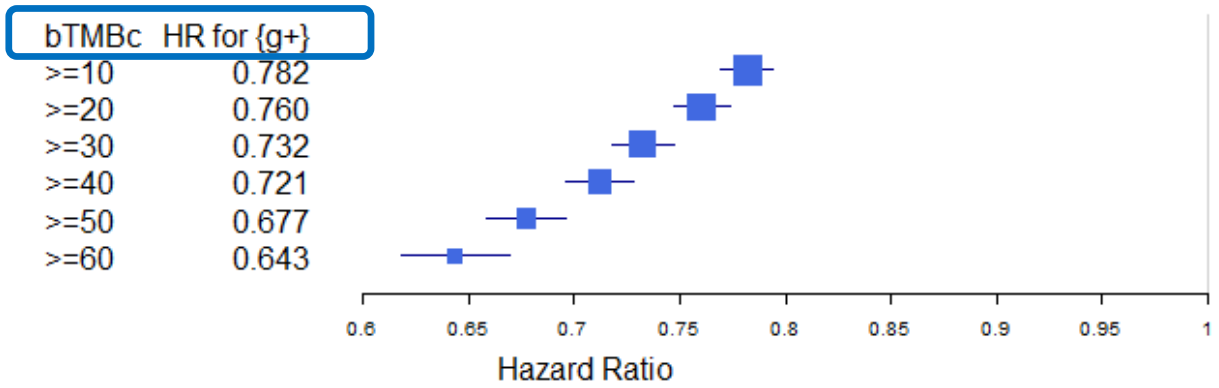


Rerun of the OAK trial data shows that bTMB is mostly a prognostic (instead of predictive) biomarker in terms of OS

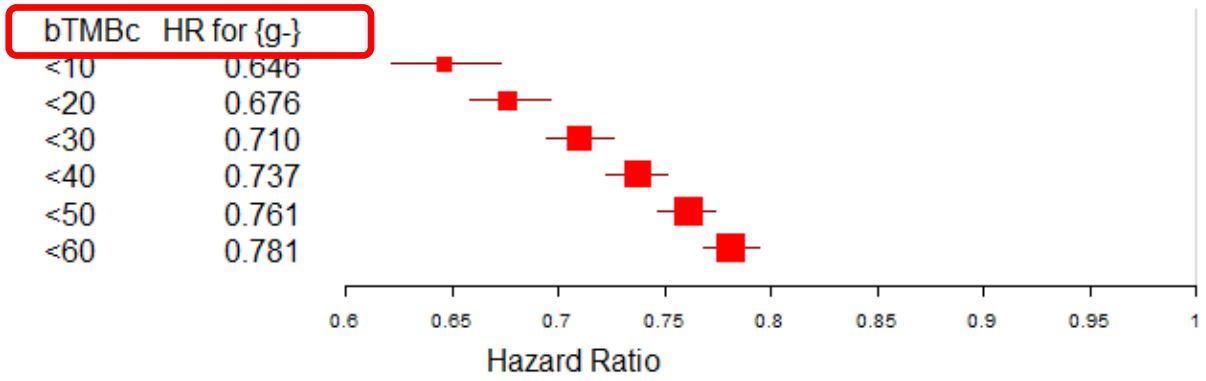
Estimated median OS from Weibull fit with bTMB, Trt and the interaction term



HR behavior for purely prognostic biomarker based on simulation



Replicated the pattern observed in OAK trial



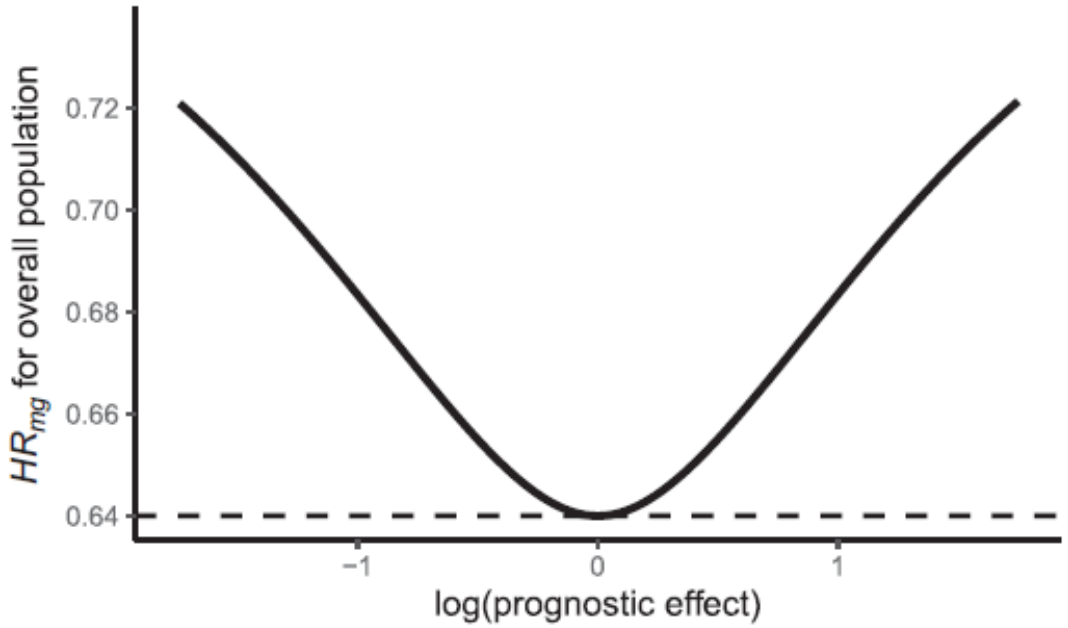
Conflicting message in terms which pt subgroup benefits most

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).

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)

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*

*Ding et al (2016); Lin 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

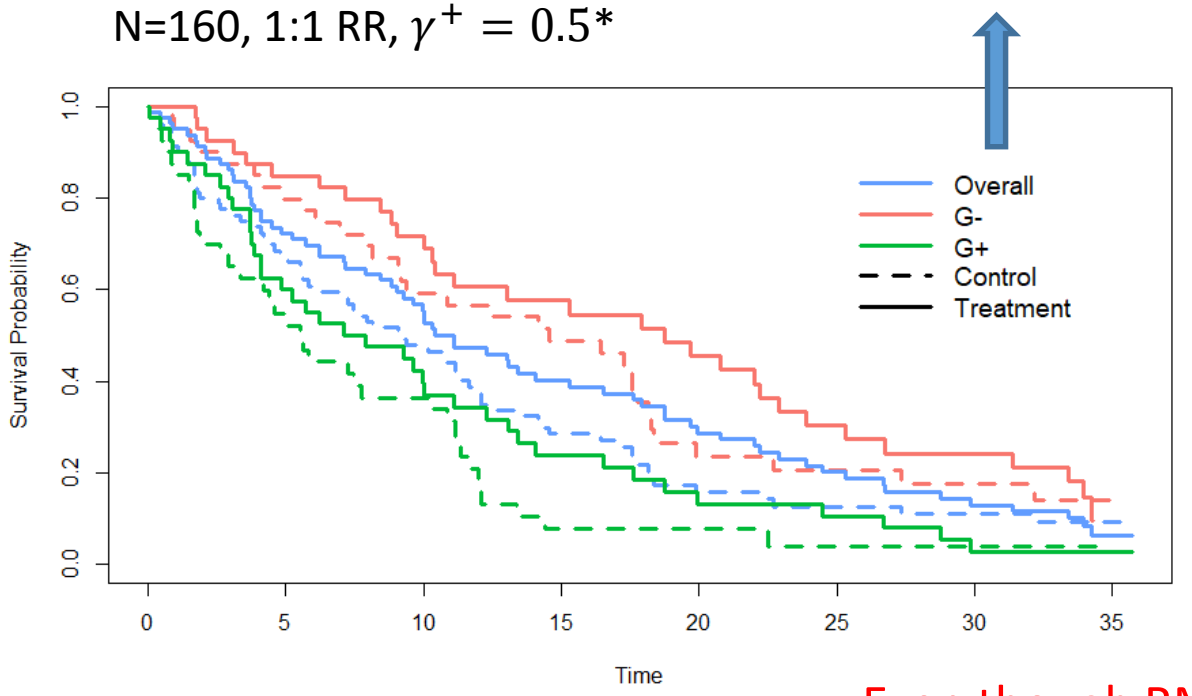
    True marginal HR ≠ $HR_m = exp\{\gamma^+(logHR_+) + \gamma^-(logHR_-)\}$
  - So it appears that “marginal HR” is always in between subgroup HRs
  - However, this is not the real marginal HR
- For logic-respecting efficacy measures in the form of DOE
  - Marginal models/analysis can lead to illogical behavior in estimates such as difference of means for continuous endpoint, and difference of RMST for TTE

Sample space

RMST difference based on marginal KM curves may disrespect logic

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

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



τ	RMST difference			
	g^-	g^+	Overall	
			marginal	SME
6	0.2203	0.4949	0.3539	0.3576
12	0.7039	1.0099	0.8422	0.8569
18	1.1301	1.9452	1.4951	1.5376
24	2.2333	2.3975	2.2331 (illogical)	2.3154
30	2.6531	2.6861	2.5506 (illogical)	2.6696

Even though RMST difference is logic respecting at population level, estimated RMST difference by the marginal method is not between those from the subgroups for tau=24 or 30. But using SME always ensures such logical relationships in the estimates.

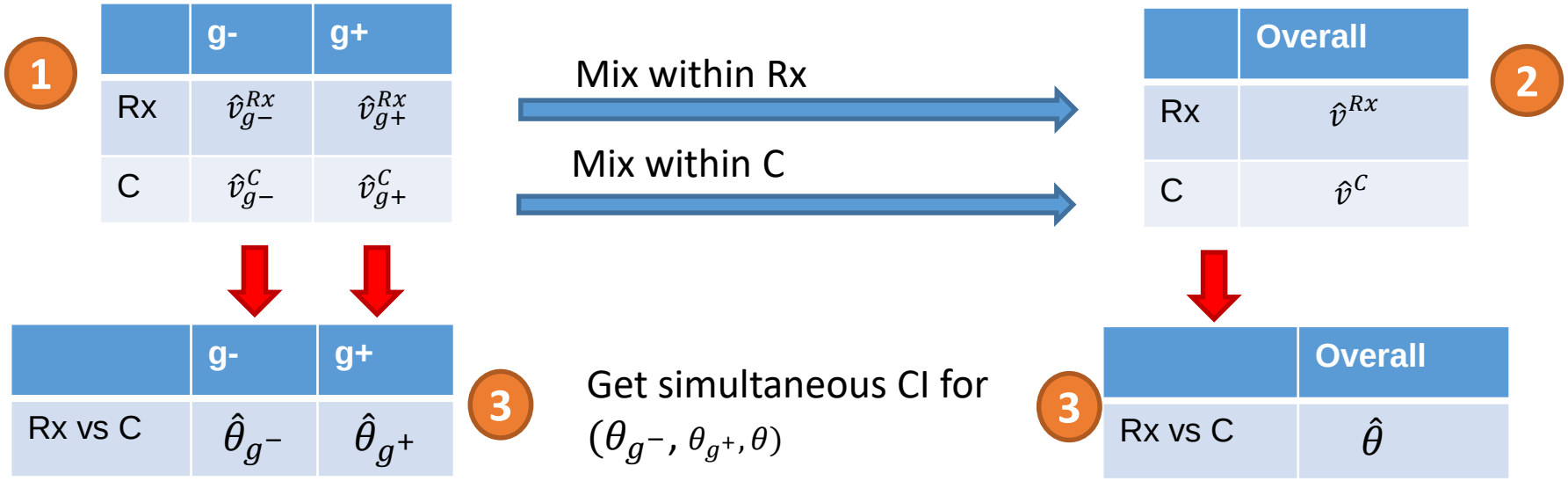
*Data generated with exponential distribution, median for C arm is 6, 10 for g+, g- and HR=0.7 for both subgroups

Sample space

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

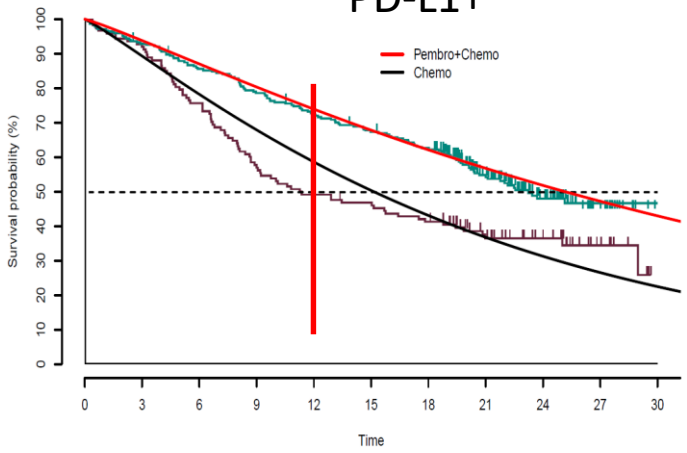
Principle of Subgroup Mixable Estimation (SME)

1. Get estimated treatment effect for (g+,Rx), (g-, Rx), (g+,C), (g-,C) and associated variance matrix estimates using **either parametric, semi-parametric or non-parametric method**
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

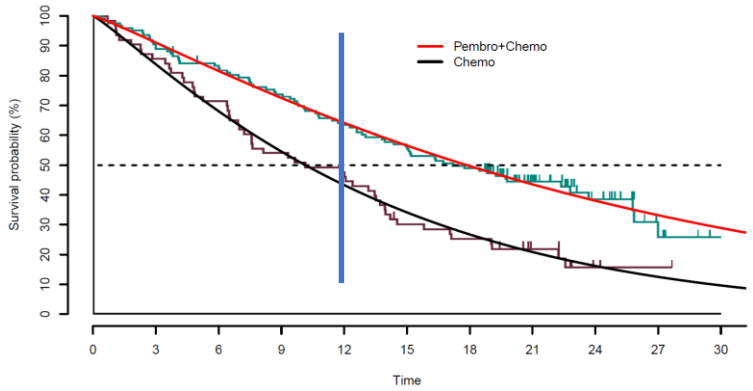


Applying SME to Keynote189 OS

PD-L1+



PD-L1-



Sample space

Weibull model results

Efficacy Measure	Group	Estimates		Ratio (95% Simultaneous CI)	Difference (95% Simultaneous CI)
		Rx	C		
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% Simultaneous CI)	Difference (95% Simultaneous CI)
		Rx	C		
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

<u>Our recommendation:</u>

- 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

BACKUP SLIDES for PS/Treatment Switching Presentation

Potential outcomes

- Administrative censoring: $C_i = c - E_i$, where c is the calendar time of study end, and E_i is the entry time for subject i with staggered entry
- Randomized treatment assignment: $Z_i = 1$ (experimental), $Z_i = 0$ (control)
- Potential survival outcomes: $Y_i(1)$ is the survival time if subject i is assigned to treatment, $Y_i(0)$ is the survival time if subject i is assigned to control. Of note, for switchers, $Y_i(0)$ is the value of survival if switchers were initially assigned to the control treatment, and thus they were exposed to the control treatment up to the time of switching and exposed to the active treatment from the time of switching onward.
- Switching status under control: $S_i(0)$, which is a potential outcome if subject i was randomized to the active treatment arm, $S_i(0) \in \bar{S} \cup R_+$, where $\bar{S}$ denotes non-switchers, and R_+ denotes the set of switchers.
- For switchers, $S_i(0) \leq Y_i(0)$

Causal estimands

- Intention-to-treat causal effects
  - Average causal effect:
$$ACE = E[Y_i(1)] - E[Y_i(0)]$$
  - Distributional causal effect:
$$DCE(y) = P\{Y_i(1) > y\} - P\{Y_i(0) > y\}$$
- Principal causal effects
  - Principal average causal effects:
$$ACE(s) = E[Y_i(1)|S_i(0) = s] - E[Y_i(0)|S_i(0) = s]$$
  - Principal distributional causal effects:
$$DCE(y|s) = P\{Y_i(1) > y|S_i(0) = s\} - P\{Y_i(0) > y|S_i(0) = s\}$$
  - Conditional principal distributional causal effects for switchers:

$$cDCE(y|s)$$
$$= P\{Y_i(1) > y|Y_i(1) \geq S_i(0), S_i(0) = s\} - P\{Y_i(0) > y|Y_i(1) \geq S_i(0), S_i(0) = s\}$$

Observed data

- Without censoring:
  - $Y_i^{obs} = Z_i Y_i(1) + (1 -$

Table 1: Observed data pattern and possible latent principal strata

Z_i	$\tilde{S}_i^{\text{obs}}$	$\tilde{Y}_i^{\text{obs}}$	Principal strata	Principal stratum label
0	C_i	$Y_i^{\text{obs}} \in [0, C_i]$	$\{i : S_i(0) = \overline{\mathbb{S}}\}$	Non-switchers
0	$S_i^{\text{obs}} \leq C_i$	$Y_i^{\text{obs}} \in [S_i^{\text{obs}}, C_i]$	$\{i : S_i(0) = S_i^{\text{obs}}\}$ $(S_i^{\text{obs}} \in \mathbb{R}_+)$	Switchers at time $S_i^{\text{obs}} \in \mathbb{R}_+$
0	$S_i^{\text{obs}} \leq C_i$	C_i	$\{i : S_i(0) = S_i^{\text{obs}}\}$ $(S_i^{\text{obs}} \in \mathbb{R}_+)$	Switchers at time $S_i^{\text{obs}} \in \mathbb{R}_+$
0	C_i	C_i	$\{i : S_i(0) = \overline{\mathbb{S}}\}$ or $\{i : S_i(0) = s \in (C_i, +\infty)\}$	Non-switchers or Switchers at some time $s > C_i$
1	$\overline{\mathbb{S}}$	$Y_i^{\text{obs}} \in [0, C_i]$	$\{i : S_i(0) = \overline{\mathbb{S}}\}$ or $\{S_i(0) \in \mathbb{R}_+\}$	Non-switchers or Switchers
1	$\overline{\mathbb{S}}$	C_i	$\{i : S_i(0) = \overline{\mathbb{S}}\}$ or $\{S_i(0) \in \mathbb{R}_+\}$	Non-switchers or Switchers

Missing data

- Switching status:
  $S_i^*(0) = (1 - Z_i)\left[\tilde{S}_i^{obs} I\{S_i(0) \in R_+\} + \bar{S} I\{S_i(0) = \bar{S}\}\right] + Z_i S_i(0)$
- Potential survival outcome:
  $$Y_i^*(0) = (1 - Z_i)\tilde{Y}_i^{obs} + Z_i Y_i(0)$$

Assumptions

- Assumption 1 (Unconfounded Treatment Assignment):

$$P\{Z_i|S_i(0), Y_i(0), Y_i(1), C_i, X_i\} = P\{Z_i|X_i\}$$

- Assumption 2 (Conditional Ignorability of the Censoring Mechanism):

$$P\{C_i|S_i(0), Y_i(0), Y_i(1), X_i\} = P\{C_i|X_i\}$$

- For each treatment arm:

$$P\{C, S(0), Y(0), Y(1), X\}$$

$$= \int \prod_{i=1}^{n} P\{X_i|\theta\}\boldsymbol{P\{S_i(0)|X_i;\theta\}P\{Y_i(0)|S_i(0), X_i, \theta\}}$$

$$\boldsymbol{P\{Y_i(1)|Y_i(0), S_i(0), X_i;\theta\}}P\{C_i|S_i(0), Y_i(0), Y_i(1), X_i;\theta\}P(\theta)d\theta$$

Parametric models

- Sub-model for the switching behavior $S_i(0)$
  - $\pi = P\{S_i(0) = \bar{S}\}$
  - $S_i(0)|S_i(0) \in R_+ \sim \text{Weibull}(\alpha_S, \beta_S)$ with survival function $\exp\left(-e^{\beta_S} t^{\alpha_S}\right)$
- Sub-model for $Y_i(0)|S_i(0)$
  - $Y_i(0)|S_i(0) = \bar{S} \sim \text{Weibull}\left(\bar{\alpha}_Y, \bar{\beta}_Y\right)$
  - $Y_i(0)|S_i(0) \in R_+ \sim S_i(0) + \text{Weibull}\left(\alpha_Y, \beta_Y + \boldsymbol{\lambda} \log\left(S_i(0)\right)\right)$
- Sub-model for $Y_i(1)|Y_i(0), S_i(0)$
  - $Y_i(1)|Y_i(0), S_i(0) = \bar{S} \sim \boldsymbol{\kappa} Y_i(0) + \text{Weibull}(\bar{\nu}_Y, \bar{\gamma}_Y)$
  - $Y_i(1)|Y_i(0), S_i(0) \in R_+ \sim \boldsymbol{\kappa} Y_i(0) + \text{Weibull}\left(\nu_Y, \gamma_Y + \boldsymbol{\lambda} \log\left(S_i(0)\right)\right)$

Prior distributions

- $\pi \sim \text{Beta}(a, b)$, $a = b = 1$
- $\alpha_S \sim \text{Gamma}(a_S, b_S)$, $a_S = 1, b_S = 10$
- $\beta_S \sim N(\mu_S, \sigma_S^2)$, $\mu_S = 0, \sigma_S = 100$
- $\bar{\alpha}_Y \sim \text{Gamma}(\bar{a}_Y, \bar{b}_Y)$, $\bar{a}_Y = 1$, $\bar{b}_Y = 10$
- $\bar{\beta}_Y \sim N(\bar{\mu}_Y, \bar{\sigma}_Y^2)$, $\bar{\mu}_Y = 0, \bar{\sigma}_Y = 100$
- $\alpha_Y \sim \text{Gamma}(a_Y, b_Y)$, $a_Y = 1, b_Y = 10$
- $\beta_Y \sim N(\mu_Y, \sigma_Y^2)$, $\mu_Y = 0, \sigma_Y = 100$
- $\bar{\nu}_Y \sim \text{Gamma}\left(\bar{d}_Y, \bar{s}_Y\right)$, $\boldsymbol{\bar{d}_Y = 125, \bar{s}_Y = 0.01}$
- $\bar{\gamma}_Y \sim N(\bar{m}_Y, \bar{\tau}_Y^2)$, $\bar{m}_Y = 0, \boldsymbol{\bar{\tau}_Y = 0.5}$
- $\nu_Y \sim \text{Gamma}(d_Y, s_Y)$, $\boldsymbol{d_Y = 125, s_Y = 0.01}$
- $\gamma_Y \sim N(m_Y, \tau_Y^2)$, $m_Y = 0, \boldsymbol{\tau_Y = 0.5}$
- $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$, $\mu_\lambda = 0, \sigma_\lambda = 100$
- κ fixed at $0, 0.25, 0.5, 0.75, 1$

Posterior sampling algorithm

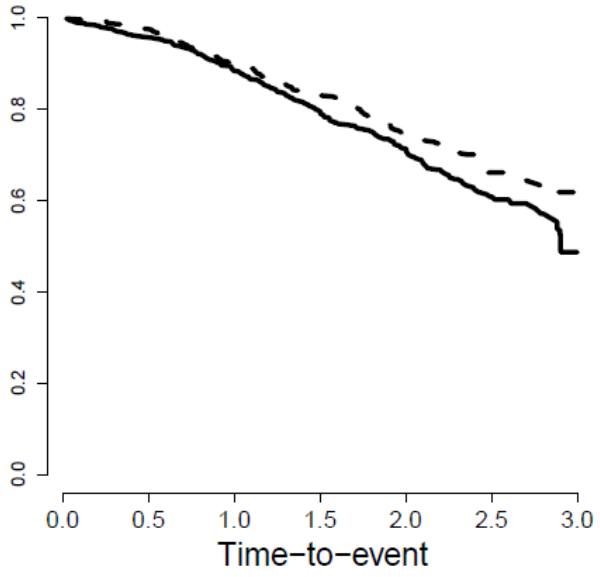
- Data augmentation to obtain complete-data log posterior
- Metropolis-Hastings steps for
  - imputing missing data, $Y_i^*(0)$, given $\theta, D^{obs}, S^*(0)$
  - imputing switching status and time, $S_i^*(0)$, given $\theta, D^{obs}, Y^*(0)$
  - drawing $\alpha_S, \beta_S, \bar{\alpha}_Y, \bar{\beta}_Y, \alpha_Y, \beta_Y, \bar{\nu}_Y, \bar{\gamma}_Y, \nu_Y, \gamma_Y, \lambda$, with proposal distributions centered at the current draw, and scaling parameters for the proposal distributions calibrated to have good acceptance rates of the MH steps
- For $\kappa = 0$, $Y_i(1)$ and $Y_i(0)$ are independent, hence impution of $Y^*(0)$ can be omitted to simplify the algorithm
- See Web appendix for details

Application to the synthetic Concorde dataset

- Effect of immediate versus deferred treatment with zidovudine in symptom-free individuals infected with HIV.
- In principle, patients in the deferred arm should not receive zidovudine until they progress to AIDS-related complex (ARC) or AIDS.
- Nevertheless, some patients in the deferred arm are allowed to switch to the active treatment arm starting zidovudine before the onset of ARC or symptoms of HIV based on persistently low CD4 cell counts.
- The outcome is time to disease progression or death, subject to censoring.
- The trial lasted 3 years, with staggered entry over the first 1.5 years.

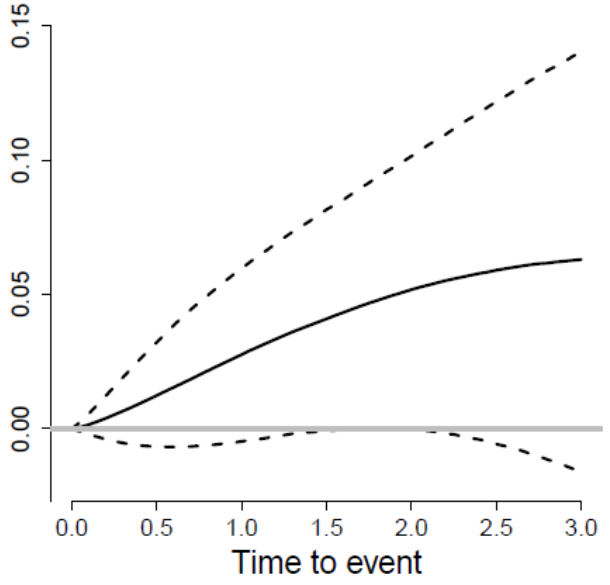
Intention-to-treat analysis

$$P\left\{\tilde{Y}_i^{\text{obs}} > y \mid Z_i\right\}$$



(a) Kaplan–Meier estimates of the survival functions

$$\text{DCE}(y) = P\{Y(1) > y\} - P\{Y(0) > y\}$$



(b) Bayesian analysis using Weibull models

ACE using Weibull models

$$ACE = E[Y_i(1)] - E[Y_i(0)]$$
$$= \exp\left(-\frac{\gamma_y}{\nu_y}\right)\Gamma\left(1+\frac{1}{\nu_y}\right)$$
$$-\exp\left(-\frac{\beta_y}{\alpha_y}\right)\Gamma\left(1+\frac{1}{\alpha_y}\right)$$

- Posterior mean 0.43
- 95% posterior credible interval: (-0.48, 1.43)
- $P\left(ACE > 0 \middle| D^{obs}\right) = 0.82$

Principal stratification analysis

- Consider $\kappa = 0$, i.e., independence of $Y_i(1)$ and $Y_i(0)$ given $S_i(0)$
- 38% non-switchers matches that observed in the control arm

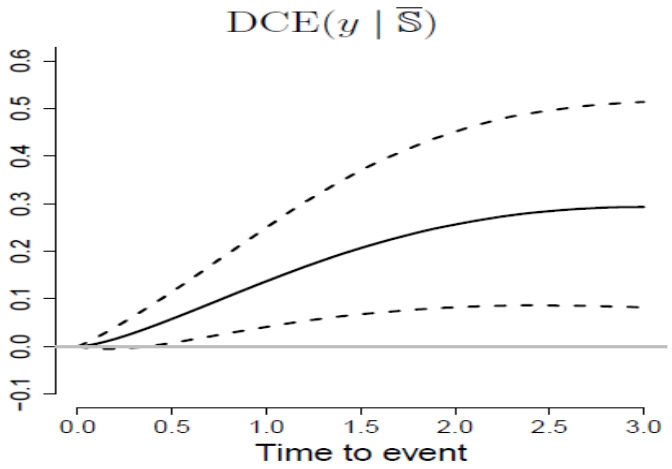
Table 2: Summary statistics of the posterior distributions

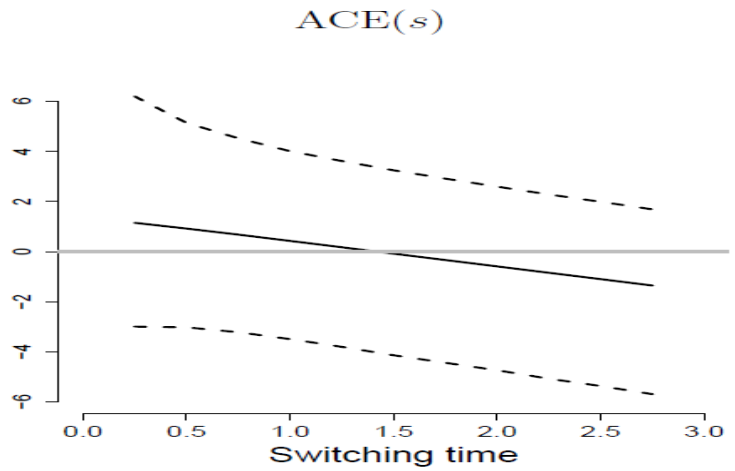
			Percentiles					
Parameter	*Mean*	*sd*	2.5%	25%	50%	75%	97.5%	$\widehat{R}$
π	0.38	0.06	0.28	0.33	0.37	0.42	0.50	1.001
α_S	1.56	0.11	1.35	1.48	1.56	1.63	1.79	1.001
β_S	−1.29	0.15	−1.55	−1.39	−1.29	−1.19	−0.98	1.001
$\overline{\alpha}_Y$	1.38	0.13	1.14	1.29	1.38	1.47	1.67	1.000
$\overline{\beta}_Y$	−1.09	0.21	−1.49	−1.24	−1.09	−0.94	−0.68	1.000
α_Y	0.94	0.12	0.72	0.86	0.94	1.02	1.19	1.000
β_Y	−1.21	0.15	−1.51	−1.31	−1.20	−1.11	−0.92	1.000
$\overline{\nu}.y$	1.29	0.11	1.09	1.22	1.29	1.36	1.51	1.000
$\overline{\gamma}_Y$	−1.85	0.29	−2.43	−2.04	−1.84	−1.65	−1.30	1.000
ν_Y	1.30	0.10	1.11	1.23	1.30	1.36	1.50	1.000
γ_Y	−2.24	0.23	−2.70	−2.39	−2.24	−2.09	−1.81	1.000
λ	0.10	0.17	−0.22	−0.02	0.09	0.21	0.44	1.000

Principal causal effects

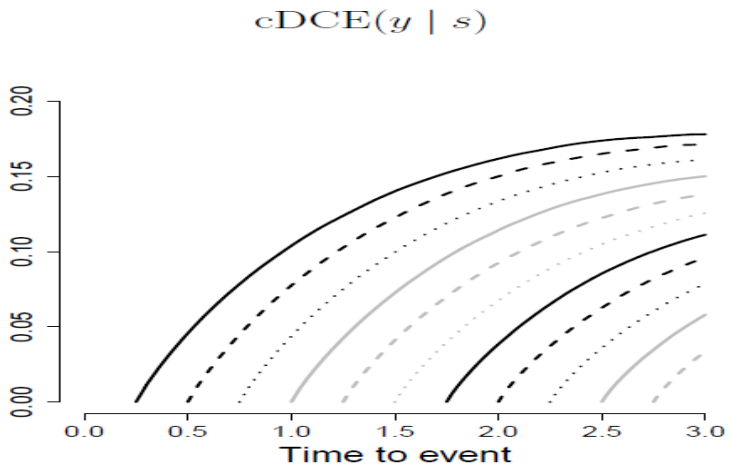
Table 3: Principal stratification analysis: Posterior median and 95% posterior credil interval for causal estimands for non-switchers for different values of κ

κ	$\mathbb{E}[Y_i(0) \mid S_i(0) = \overline{\mathbb{S}}]$	$\mathbb{E}[Y_i(1) \mid S_i(0) = \overline{\mathbb{S}}]$	$\mathrm{ACE}(\overline{\mathbb{S}})$
$\kappa = 0$	2.02 (1.44; 2.97)	3.85 (2.41; 6.94)	1.78 (0.39; 4.78)





(a) $\mathrm{ACE}(s)$, $s \in \mathbb{R}_+$ with 95% credible interval



(b) $\mathrm{cDCE}(y \mid s)$ for $s = 0.25, 0.50, \ldots, 2.50, 2.75$

Figure 4: Posterior medians of principal causal effects for switchers

Sensitivity analyses

- Mattei et al demonstrated that the results are robust with respect to the prior distribution of λ
- Preliminary results for $\kappa > 0$ showed sensitivity to the value of κ

Conclusions

- The proposed method targets the principal causal effects for subpopulations defined by switching status and time
- The Bayesian parametric modelling is flexible; however, the results are sensitive to the assumed relationship between potential survival outcomes within a principal stratum
- The method may be extended to handle two-way switching and informative censoring