Assessing the impact of integrating extrapolated long-term outcomes into elicitation of unreported subgroup-specific survival in randomized controlled trials (RCTs): insights from advanced stage gastrointestinal cancers

MSR104

Oğuzhan Alagoz, PhD[1]; Swetha Srinivasan, PhD[2]; Hong Xiao, PhD[2]; Prianka Singh, PharmD[2]; Joseph Gricar, MS[2]; Matthew Dixon, PharmD, PhD[2]; Inkyu Kim, PhD[2]; Murat Kurt, PhD[2]

[1]University of Wisconsin-Madison, Madison, WI, USA; [2]Bristol Myers Squibb, Lawrenceville, NJ, USA

Introduction

- Analyses of randomized controlled trials (RCTs) typically report
  - Kaplan-Meier (KM) survival curves for intervention and comparator arms for the overall trial population
  - Subgroup-specific survival information provided only in a summarized form: forest plots displaying hazard ratios (HRs) across selected subgroups
- Lack of subgroup-specific KM curves prevents a robust comparison of the treatments under consideration for subsequent comparative effectiveness and cost-effectiveness analyses

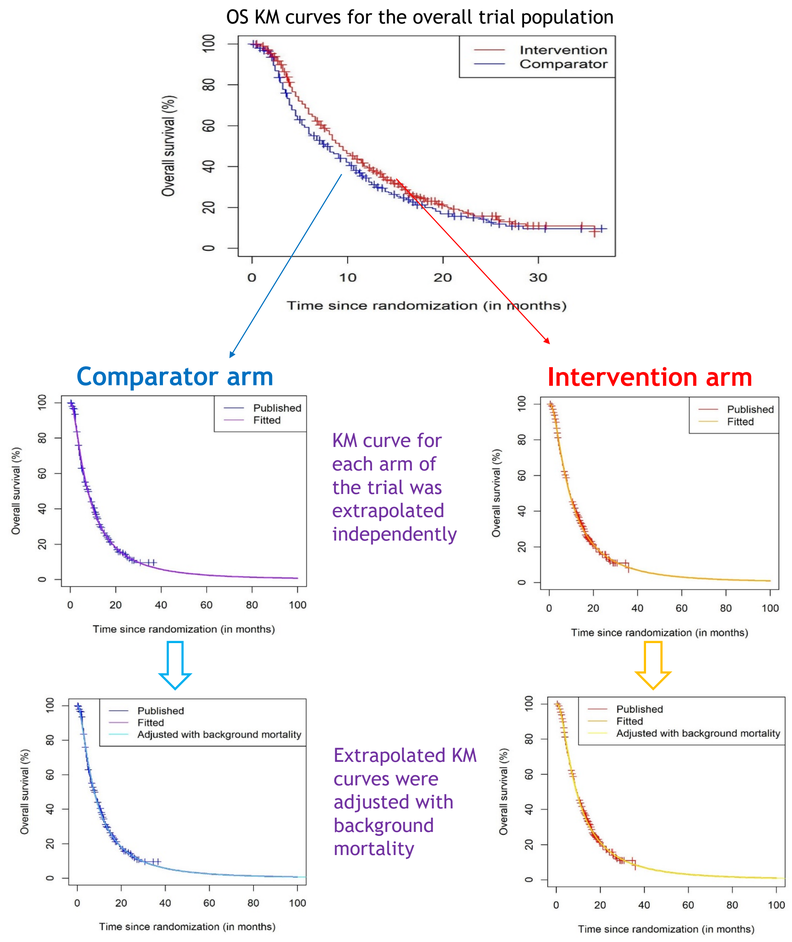
Objectives

- Develop an analytical framework to elicit unreported subgroup specific survival from reported aggregate RCT results using lifetime mean survival (LTMS) as an objective.
- Compare this approach to a previously developed alternative approach that uses restricted mean survival time (RMST) based on reported KM curves as an objective.

Methods

- For a given RCT, reconstructed KM curves of the intervention and comparator arms for overall survival (OS) were extrapolated using parametric survival models over a lifetime while adjusting with background mortality rates (**Figure 1**)
  - Seven commonly used parametric distributions suggested by the UK's National Institute for Health and Care Excellence (NICE)[1,2] were fit to each arm's OS separately
  - Extrapolated OS curves were adjusted with background mortality rates that account for country of enrollment, patients' median baseline age, and sex distribution in the trial.

Figure 1. Example for OS extrapolations with background mortality adjustment: REACH trial from second-line treatment of advanced hepatocellullar carcinoma

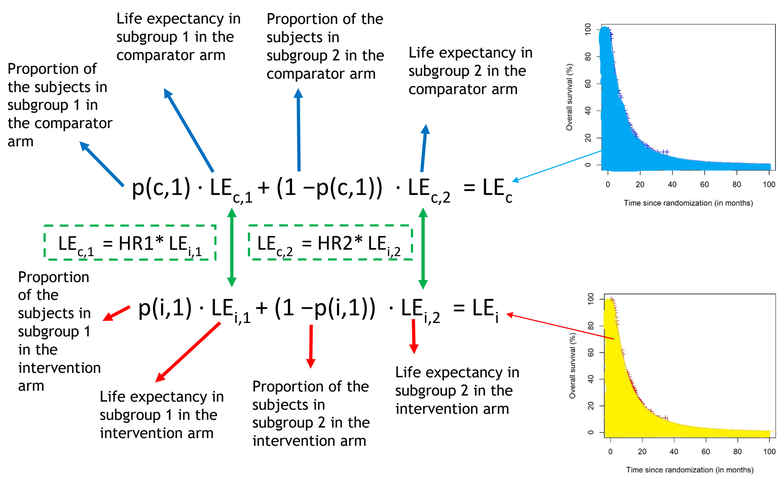


*Plot is adapted from Zhu AX, et al. *Lancet Oncol* 2015;16:859–870. Note: Vertical markers on KM-plots represent censoring information

- Theoretically, while LTMS is defined as the area under the extrapolated survival curves with no limit on time, for practical purposes and to be able to integrate background mortality adjustments into calculations, in our study they are (approximately) calculated as the area under the extrapolated curves until age 100 for both intervention and comparator arms (**Figure 2**)

- Using subgroup-specific HRs that are reported from the forest plots and assuming subgroup-specific survival times follow exponential distribution, a 2 x 2 system (2 equations in 2 unknowns) of linear equations was solved to calculate the underlying hazard rates of the distributions for each subgroup (**Figure 2**)

Figure 2. Overview of LTMS approach to estimate the underlying hazard rates for each subgroup



*Plots are adapted from Zhu AX, et al. *Lancet Oncol* 2015;16:859–870.

Note: When subgroup-specific survival times are exponentially distributed, for each subgroup, LTMS in comparator arm ($LE_{c,1}$ or $LE_{c,2}$) can be expressed by a product of the LTMS in the intervention arm ($LE_{i,1}$ or $LE_{i,2}$) and the corresponding HR (HR 1 or HR 2) between intervention and comparator arms for the same subgroup as reported by forest plots.

Methods (continued)

Synthesizing Estimates of Subgroup Specific Survival

- For each combination of parametric fits (7 x7 = 49 in total) used in survival extrapolations for intervention and comparator arms, corresponding hazard rates of the exponential distributions for subgroup-specific survival were estimated

- A relative likelihood (RL) of a model (specified by x and y) being the best candidate (i.e. probability that it minimizes the (estimated) information loss) was calculated based on Akaike information criterion (AIC) for both comparator (c) and intervention (i) arms:

$$RL_x(c) = \exp((AIC_{min}(c) - AIC_x)/2),\ RL_y(i) = \exp((AIC_{min}(i) - AIC_y)/2)$$

Then, for each combination of extrapolations (i.e. x and y), a combined relative likelihood score was computed as $RL_{x,y} = RL_x(c)^* RL_y(i)$

- Combinations that led to infeasible solutions were eliminated and $RL_{x,y}$ values corresponding to filtered feasible solutions were normalized to define the weights of the remaining combinations. Normalized likelihood ratios were utilized to calculate weighted average of the feasible solutions to generate a single hazard rate estimate for each subgroup. They were also used to estimate the probability that predicted medians would be within the 95% CI of the reported median survival times

Testing Model's Predictive Performance

- Model performance was tested in a case study of 18 distinct RCTs from advanced stage gastrointestinal tumors reporting KM-curves for validation in 96 subgroups in total (**Table 1**)

- Predictive accuracy of the method was compared to an alternative approach using an objective relying on restricted mean survival time (RMST) limited by the trial follow-up or a sufficiently long pre-specified threshold time to allow enough number of events in each arm (**Figure 3**)

- Sensitivity analyses considered spline-based and dependent models between the arms for survival extrapolations

Table 1. List of studies included in the case study

Test case	Tumor type	Trial	Subgroup 1	Subgroup 2
1	HCC	REACH	α-fetoprotein < 400 ng/mL	α-fetoprotein ≥ 400 ng/mL
2		REACH	East-Asian	Non-East Asian
3		REACH-2	REACH-2 trial patients	REACH trial patients with α-fetoprotein ≥ 400 ng/mL
4		RESORCE	Last sorafenib dose 800 mg/day	Last sorafenib dose < 800 mg/day
5		REFLECT	Patient body weight: <60 kg	Patient body weight: >60 kg
6	EC/GC	KEYNOTE-181	PD-L1 CPS ≥ 10%	PD-L1 CPS < 10%*
7		KEYNOTE-181	SCC	Adenocarcinoma*
8		KEYNOTE-061	ECOG Status 0	ECOG Status 1
9		REGARD	Age < 65	Age ≥ 65
10		RAINBOW	Age < 65	Age ≥ 65
11		TAGS	With gastrectomy	Without gastrectomy
12		KEYNOTE-590	PD-L1 CPS >10%	PD-L1 CPS <10%
13		KEYNOTE-590	SCC	Adenocarcinoma
14		GATSBY	HER2 IHC3+	No HER2 IHC3+
15		SOLAR	Japan	South Korea
16	CRC	RAISE	KRAS status wild-type	KRAS status mutant
17		RAISE	Tumor side (left)	Tumor side (right)
18		CORRECT	Japanese	Non-Japanese
19		ASPECCT	Prior bevacizumab	No Prior bevacizumab
20		CRYSTAL	KRAS codon 12/13 wt Metastasis LLD	KRAS codon 12/13 wt Metastasis non-LLD
21		CRYSTAL	RAS wt Metastasis LLD	RAS wt Metastasis non-LLD
22		FRESCO	Prior targeted therapy (anti-VEGF/EGFR)	No Prior targeted therapy (anti-VEGF/EGFR)
23		FRESCO	Prior VEGFi	No Prior VEGFi
24		FRESCO	Liver metastasis	No liver metastasis
25		XELAVIRI	Males	Females

HCC: Hepatocellular carcinoma, EC: Esophageal cancer, GC: Gastric cancer, CRC: Colorectal cancer

Performance Measures

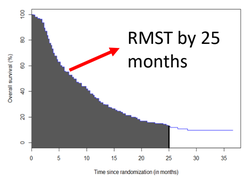
- For each subgroup, quality of the predictions from the model was assessed visually and evaluated statistically to the data estimated/reported from the RCTs through following performance metrics:

1. *Survival prediction accuracy*: % of times in which model-predicted survival rate fell within the 95% CI of the reported survival rate
2. *Median alignment*: predicted median survival vs. 95% CI of the reported median survival
3. *RMST gap (ΔRMST):* % gap between the model-predicted RMST and the RMST estimated from the reported KM-curve
4. *RMST alignment*: model-predicted RMST vs. 95% CI of the RMST estimated from the reported KM-curve
5. *Average survival rate gap [ΔS(t)]:* Average absolute difference between model-predicted and reported survival rates

- Calculation of all performance metrics except 2 was based on the reported follow-up durations of the RCTs, whereas metrics 1 and 5 were estimated on a monthly basis

Overview of the RMST-based approach

RMST for the overall population by a pre-specified threshold time (t^*) in each arm can be expressed as a weighted average of the RMSTs (defined by the corresponding t^*) of a list of mutually exclusive and exhaustive subgroups. Example graph below depicts the RMST for the comparator arm by month 25 (i.e. $t^* = 25$) in the REACH trial



Unlike the LTMS-based approach, since the RMST for an exponential distribution does not attain a linear closed form of its hazard rate, the RMST-based approach requires the solution of a nonlinear optimization problem for the derivation of subgroup specific hazard rates.

RMST for the intervention arm for the overall population by t^* | Prevalence of subgroup 1 in the intervention arm | RMST of subgroup 1 by t^* in the intervention arm | RMST of subgroup 2 by t^* in the intervention arm

$$RMST_i(t^*) = p(i,1) \times RMST_{i,1}(t^*) + (1 - p(i,1)) \times RMST_{i,2}(t^*)$$

HRs from the forest plots connect the hazard rates of the subgroups

$$RMST_c(t^*) = p(c,1) \times RMST_{c,1}(t^*) + (1 - p(c,1)) \times RMST_{c,2}(t^*)$$

RMST of the comparator arm for the overall population by t^* | Prevalence of subgroup 1 in the comparator arm | RMST of subgroup 1 by t^* in the comparator arm | RMST of subgroup 2 by t^* in the comparator arm

The nonlinear optimization model minimizing sum-of-square of errors defined over RMST

Predicted RMST for the intervention arm | Reported RMST for the intervention arm

$$\min \left[p(i,1)RMST_{i,1}(t^*) + (1 - p(i,1)) \times RMST_{i,2}(t^*) - RMST_i(t^*) \right]^2$$

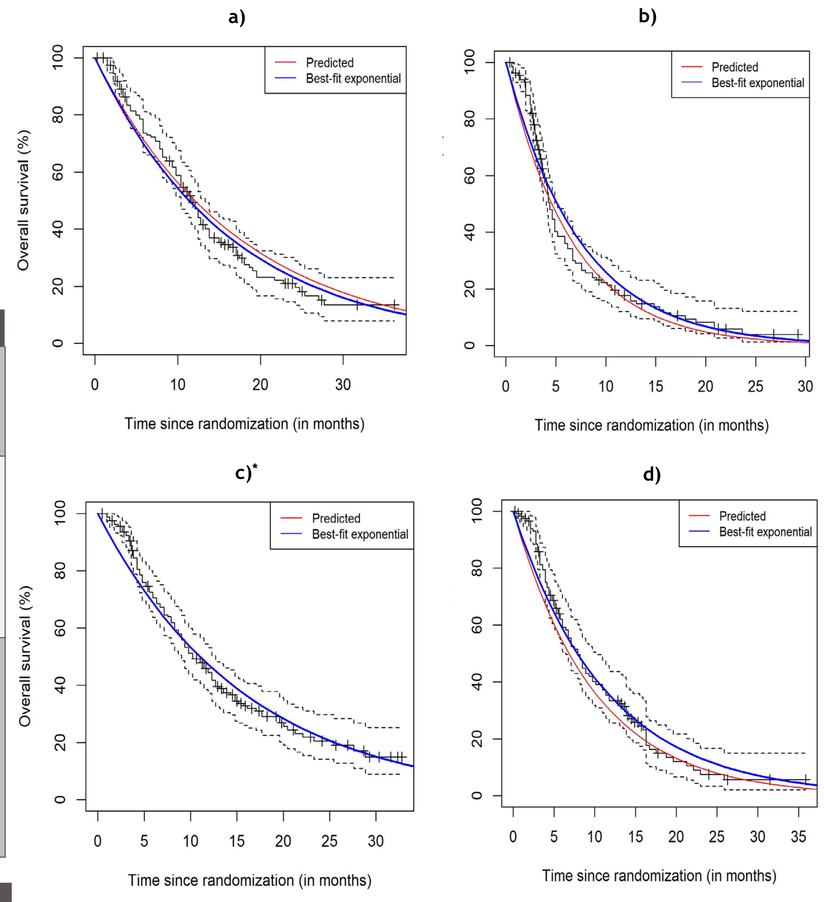
$$+ \left[p(c,1) \times RMST_{c,1}(t^*) + (1 - p(c,1)) \times RMST_{c,2}(t^*) - RMST_c(t^*) \right]^2$$

Predicted RMST for the comparator arm | Reported RMST for the comparator arm

Results

- LTMS-based approach worked visually and statistically well for some cases as depicted in **Figure 4.**
- LTMS-based approach showed median alignment in 23 subgroups, which was 27 subgroups less than the alignment achieved by RMST-based approach (**Table 2**)
- RMST-based approach had median alignment in 32 subgroups where LTMS-based approach showed none.
- In 4 RCTs where RMST-based approach showed median alignment in all subgroups, LTMS-based approach showed no alignment in any subgroups
- When RMST-based approach allowed subgroup-specific survival times to follow Weibull or loglogistic distribution, median alignment improved by 54 subgroups over LTMS-based approach
- Considering splines and dependent models between the arms of the trials for survival extrapolations did not improve the performance of LTMS-based approach

Figure 4. Visual comparison of LTMS-based approach to reported KM curves: Example from the REACH trial, a) α-fetoprotein < 400 ng/mL subgroup in the placebo arm; b) α-fetoprotein ≥ 400 ng/mL subgroup in the placebo arm; c) α-fetoprotein < 400 ng/mL subgroup in the ramucirumab arm; d) α-fetoprotein ≥ 400 ng/mL subgroup in the ramucirumab arm



*Predicted hazard rate was negligibly different than the hazard rate of the best-fitting exponential distribution for this subgroup. Therefore, red line is not visible.

Table 2. Performance summary of the LTMS-based approach and its comparison to RMST-based approach in the case study

Performance metrics						
Modeling approach	Total number of subgroups	Survival prediction accuracy, %	Median alignment, n (%)	ΔRMST*, %	RMST alignment, n (%)	ΔS(t)*, %
Exponential†, 4	96	50	50 (52)	20	45 (47)	10.3
Weibull†,5	96	73	70 (73)	10	66 (69)	5.4
Loglogistic†, 5	96	80	80 (83)	10	70 (73)	4.6
LTMS-based approach	96	34	23 (24)	33	23 (24)	15.6

*: Calculated as average across all 96 subgroups. †Indicates the assumed distribution of subgroup-specific survival from previously published work using RMST as an objective criterion.

Conclusions

- LTMS-based approach did not work very accurately for majority of the cases
- LTMS-based approach was not sensitive to the quality of the fits by parametric distributions to the data as statistically better fits did not necessarily lead to better predictions
- Eliciting subgroup-specific survival using RMST rather than LTMS can be statistically more accurate, and free of the subjectivity and uncertainty introduced by long-term survival extrapolations
- While RMST-based approach led to more accurate results than the LTMS-based approach, the nonlinear optimization model solved to implement the RMST-based approach is highly complex. Therefore, novel computational approaches are needed to improve the solution quality for the RMST-based approach

References

1. Latimer N. NICE DSU Technical Support Document 14. http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf
2. Rutherford M et al: NICE DSU Technical Support Document 21. http://nicedsu.org.uk/wpcontent/uploads/2020/11/NICE-DSU-Flex-Surv-TSD-21_Final_alt_text.pdf
3. Zhu AX, et al. *Lancet Oncology* 2015;16:859-870.
4. Alagoz O, et al. *Value in Health 24 (2021):S3.*
5. Alagoz O, et al. *Value in Health 25 (2022):S10.*

Acknowledgments

This study was supported by Bristol Myers Squibb
All authors contributed to and approved the presentation

Presented at ISPOR Europe 2022; November 6-9, 2022; Vienna, Austria, and virtual

Email: alagoz@engr.wisc.edu