CO91

Association between progression-free survival and overall survival in patients with relapsed/refractory multiple myeloma

Liza Mastikhina,[1] Shannon Cope,[1] Thomas S. Marshall,[2] Dylan Maciel,[1] Ali Mojebi,[1] Korinna Karampampa,[3] Devender Dhanda[2]

[1]PRECISIONheor, Evidence Synthesis and Decision Modeling, Vancouver, BC, Canada; [2]Bristol Myers Squibb, Princeton, NJ, USA; [3]Bristol Myers Squibb, Uxbridge, UK

Introduction

- Overall survival (OS) is the most clinically relevant endpoint in oncology; however, emerging treatments for relapsed/refractory multiple myeloma (RRMM), such as anti-CD38 monoclonal antibodies (mAbs), bispecific antibodies, and chimeric antigen receptor T cell therapy,[1] do not have the mature OS data necessary for regulatory and reimbursement decision-making[2,3]
- In the absence of mature OS data, use of surrogate endpoints such as progression-free survival (PFS) and time to progression (TTP) potentially allow for more timely evaluation of the benefit-to-risk profile of experimental treatments
  - Trial-level surrogacy in the form of meta-analyses of several randomized controlled trials (RCTs) represents the highest level of evidence for validating surrogate endpoints[4-6]
  - Previous analyses in RRMM[7,8] have evaluated surrogacy at the individual level, which represents a lower level of evidence for validating surrogate endpoints[4]
- Etekal et al.[9] recently explored trial-level surrogacy in RRMM; however, the evidence base was restricted to phase 3 RCTs published between 2005 and 2019 that focused on PFS and regression coefficients were not reported, making it challenging to predict OS hazard ratios (HRs) from this analysis or to integrate surrogacy estimates into a cost-effectiveness analysis
  - Given the evolution of the treatment landscape for patients who are double-class refractory (DCR) to an immunomodulatory drug (IMiD) and a proteasome inhibitor (PI); or triple-class exposed (TCE) to IMiDs, PIs, and anti-CD38 mAbs, it is of interest to account for class exposure and refractoriness in the context of the surrogacy relationship

Objective

- To investigate trial-level surrogacy of PFS and TTP for OS in the RRMM population

Methods

Identification of studies

- RCTs of systemic treatments in RRMM that reported PFS or TTP and OS HRs were identified on July 7, 2022 via an umbrella review of systematic literature reviews (SLRs) using predefined study eligibility criteria (**Table 1**); additional RCTs were identified from bibliographies of published surrogacy analyses[7,9] and other SLRs identified through hand searching in January 2023. Supplemental materials are available in the ISPOR Europe 2023 conference app

Table 1. Study eligibility criteria of the literature review

Criteria	Inclusion criteria	Exclusion criteria
Population	Patients with RRMM[a]	Patients without RRMM or with MM being treated with first-line therapy
Intervention/ comparators	Any systemic intervention	Non-systemic interventions (eg, stem cell transplant)
Outcomes	Published HRs for both endpoints of interest:[b] • OS • PFS or TTP	NA
Study design	Phase 2 or phase 3 RCTs	Phase 1 RCTs, single-arm studies, observational studies, case studies, case series, narrative reviews, SLRs, meta-analyses, indirect treatment comparisons, editorials, commentaries
Language	English only	NA
Time	Inception to January 31, 2023	NA

[a]No restrictions were placed on patient age, comorbidities, treatment history, or disease severity; [b]HRs were not derived from Kaplan-Meier curves if not directly reported.
MM, multiple myeloma; NA, not available.

- A feasibility assessment was performed to identify any imbalances across the unique identified studies with respect to characteristics that could potentially impact treatment effects, including study design (trial phase), baseline patient characteristics (class exposure and refractoriness), treatments, and outcomes (definitions, PFS vs TTP, follow-up duration)
- To maximize the number of data points included, study quality was not prespecified as a determinant of study eligibility for the analysis; therefore, a formal risk-of-bias assessment was not conducted

Statistical analysis

- Linear regression weighted by sample size of the reported log PFS or TTP and OS HRs (with the same data cut within each trial where possible) was performed (frequentist framework) and summarized in terms of the parameters and the weighted Pearson correlation coefficient (R)
- A bivariate meta-analysis (Bayesian framework) was carried out and summarized in terms of the corresponding parameters
- A leave-one-out cross-validation was also performed (Bayesian framework)
- Results were translated into the percentage decrease (ie, improvement) in log (OS HR) associated with a 10% decrease in log (PFS or TTP HR), which was calculated for each model and framework using the formula: % decrease in log (OS HR) = $\lambda_1 \times 10$, where λ_1 represents the model-specific slope value analysis
  - The 95% confidence interval (CI) for the slope term was used to calculate the associated 95% CI for the percentage decrease in log (OS HR)

Sensitivity/scenario analyses

- The criteria used to evaluate the surrogacy relationship for PFS or TTP and OS are summarized in **Table 2**

Table 2. Criteria used to evaluate surrogacy

Framework	Criteria	Term	Definition	Classification
Bayesian	NICE[4]	Leave-one-out cross-validation	95% of reported HRs are captured by the prediction interval	Valid surrogate
	--	Intercept, slope, and conditional variance	Intercept: 95% CrI includes 0; Slope: 95% CrI does not include 0; Conditional variance: close to 0	Strong surrogate
Frequentist	IQWiG[5]	R	Lower limit of the 95% CI of R is ≥ 0.85	High correlation
			R is > 0.70 but < 0.85	Medium correlation
			Upper limit of the 95% CI of R is ≤ 0.70	Low correlation
	BSES[10]	R^2	$R^2 \geq 0.60$	Excellent surrogate
			$R^2 \geq 0.40$	Good surrogate
			$R^2 \geq 0.20$	Fair surrogate
			$R^2 < 0.20$	Poor surrogate
	--	Intercept and slope	Intercept: statistically not significant; Slope: statistically significant	Strong surrogate

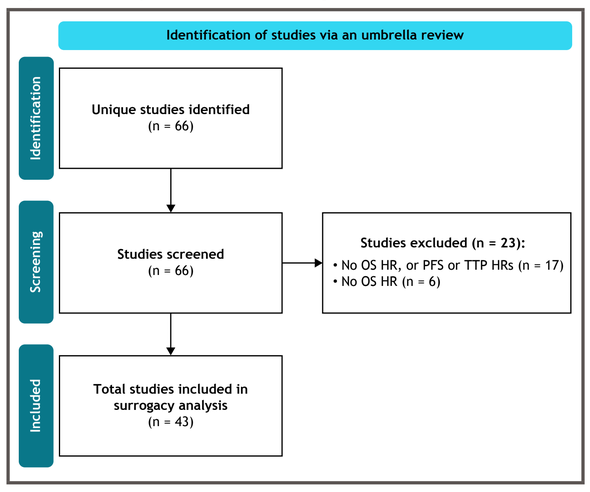
BSES, Biomarker-Surrogacy Evaluation Schema; CrI, credible interval; IQWiG, Institute for Quality and Efficiency in Health Care; NICE, National Institute for Health and Care Excellence.

- Several sensitivity analyses were explored, including:
  - Scenario restricted to studies including any DCR patients (used to approximate the TCE population due to an insufficient number of trials conducted in a predominantly TCE population)
  - Scenario including OS HRs reported based on the longest available follow-up per trial (possibly with a different data cut than PFS or TTP HR)
  - Scenario restricted to trials reporting PFS
  - Scenario restricted to phase 3 trials

Results

- A total of 66 phase 2 and 3 RCTs were identified, of which 43 were included in the surrogacy analysis (see **Figure 1** for study selection)

Figure 1. Study selection flow diagram

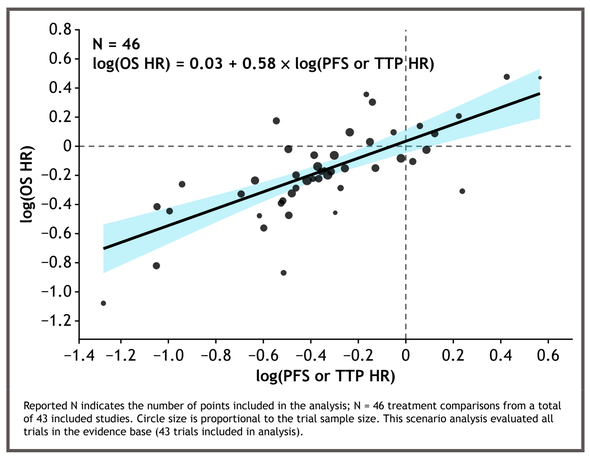


- **Table 3** provides an overview of trial-level surrogacy estimates across the scenarios
- Models across all scenarios in both Bayesian and frequentist frameworks met the criteria for a strong surrogate relationship wherein the intercepts were close to 0 and the slopes were statistically significant
  - All models in the Bayesian framework had conditional variances that were close to 0
- In the leave-one-out cross-validation, prediction intervals captured 98%-100% of the reported OS HRs in the bivariate meta-analyses, indicating a valid surrogate relationship

All-evidence analysis

- This analysis included all 43 RCTs with a total of 46 treatment comparisons; OS and PFS or TTP HRs from the same data cut within each trial were used where possible
- In the Bayesian framework (**Figure 2**), the slope was 0.58 (95% CrI, 0.41-0.75)
  - Based on the estimated correlations between PFS or TTP and OS, a 5.8 (95% CrI, 4.1-7.5) decrease in log (OS HR), in the bivariate meta-analysis could occur given a 10% decrease in log (PFS or TTP HR)

Figure 2. Bayesian framework fit for the all-evidence analysis (log HR for PFS or TTP vs log HR for OS)



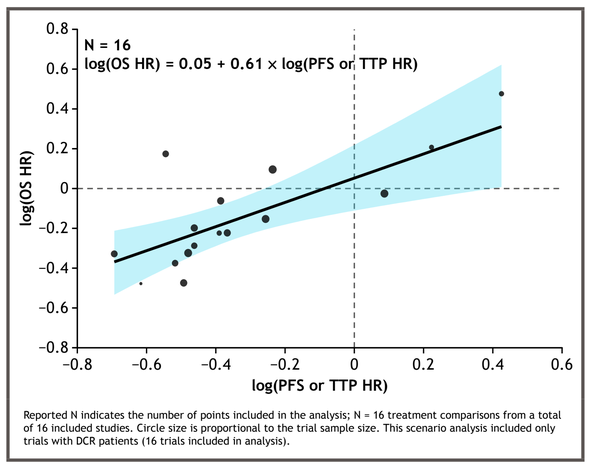
Reported N indicates the number of points included in the analysis; N = 46 treatment comparisons from a total of 43 included studies. Circle size is proportional to the trial sample size. This scenario analysis evaluated all trials in the evidence base (43 trials included in analysis).

- Results were similar in the frequentist framework, where in the slope was 0.59 (95% CI, 0.43-0.75)
  - The correlation term (R = 0.74) was classified as medium correlation per IQWiG criteria, and the correlation coefficient (R^2 = 0.54) was classified as a good surrogate per BSES criteria
  - Based on the estimated correlations between PFS or TTP and OS, a 5.9 (95% CI, 4.3-7.5) decrease in log (OS HR) in the frequentist regression could occur given a 10% decrease in the log (PFS or TTP) HR

Scenario analysis 1: trials including any DCR patients

- This scenario was used to approximate the TCE population because none of the included RCTs were conducted in a predominantly TCE population
- This scenario resulted in a higher correlation term (R = 0.75)
  - The estimated slope was also higher (**Figure 3**), which resulted in a higher predicted percentage decrease in the log (OS HRs) associated with a 10% decrease in log (PFS or TTP HRs) than the all-evidence analysis

Figure 3. Bayesian framework fit for trials including any DCR patients (log HR for PFS or TTP vs log HR for OS)

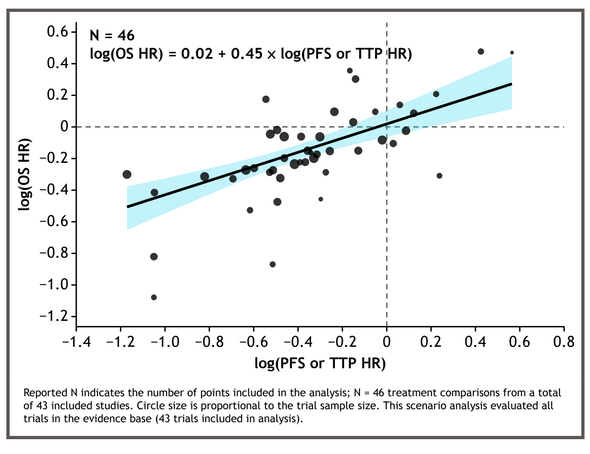


Reported N indicates the number of points included in the analysis; N = 16 treatment comparisons from a total of 16 included studies. Circle size is proportional to the trial sample size. This scenario analysis included only trials with DCR patients (16 trials included in analysis).

Scenario analysis 2: longest OS follow-up

- This scenario resulted in a lower correlation term (R = 0.68)
  - The estimated slope was also lower (**Figure 4**), which resulted in a lower predicted decrease in the log (OS HRs) associated with a 10% decrease in log (PFS or TTP HRs) than the all-evidence analysis

Figure 4. Bayesian framework fit for longest OS follow-up (log HR for PFS or TTP vs log HR for OS)



Reported N indicates the number of points included in the analysis; N = 46 treatment comparisons from a total of 43 included studies. Circle size is proportional to the trial sample size. This scenario analysis evaluated all trials in the evidence base (43 trials included in analysis).

Table 3. Overview of surrogacy analysis results (Bayesian and frequentist frameworks)

Analysis	No. of trials	No. of comparisons	Bayesian framework			Frequentist framework				% decrease (95% CrI or 95% CI) in log (OS HR) given 10% decrease in log (PFS or TTP) HR	
			Intercept (95% CrI)	Slope (95% CrI)	Conditional variance (95% CrI)	Intercept (95% CI)	Slope (95% CI)	R (95% CI)	R^2 (95% CI)	Bayesian models	Frequentist models
All evidence	43	46	0.03 (−0.05 to 0.12)	0.58 (0.41-0.75)	0.00 (0.00-0.02)	0.04 (−0.04 to 0.13)	0.59 (0.43-0.75)	0.74 (0.57-0.85)	0.54 (0.32-0.72)	**5.8 (4.1-7.5)**	**5.9 (4.3-7.5)**
Trials including any DCR patients	16	16	0.05 (−0.11 to 0.22)	0.61 (0.24-0.99)	0.01 (0.00-0.06)	0.08 (−0.07 to 0.22)	0.65 (0.32-0.99)	0.75 (0.40-0.91)	0.56 (0.16-0.82)	**6.1 (2.4-9.9)**	**6.5 (3.2-9.9)**
Using longest follow-up for OS and PFS	43	46	0.02 (−0.06 to 0.10)	0.45 (0.29-0.62)	0.00 (0.00-0.02)	0.05 (−0.04 to 0.14)	0.52 (0.35-0.69)	0.68 (0.49-0.81)[a]	0.47 (0.24-0.66)	**4.5 (2.9-6.2)**	**5.2 (3.5-6.9)**
Restricted to trials reporting PFS	39	40	0.03 (−0.06 to 0.12)	0.54 (0.33-0.75)	0.00 (0.00-0.02)	0.04 (−0.05 to 0.13)	0.55 (0.37-0.73)	0.71 (0.50-0.83)	0.50 (0.25-0.69)	**5.4 (3.3-7.5)**	**5.5 (3.7-7.3)**
Restricted to phase 3 trials	31	33	0.00 (−0.10 to 0.11)	0.54 (0.33-0.75)	0.00 (0.00-0.02)	0.02 (−0.09 to 0.13)	0.56 (0.36-0.75)	0.72 (0.50-0.85)	0.51 (0.25-0.72)	**5.4 (3.3-7.5)**	**5.6 (3.6-7.5)**

Outputs are in green for scenarios that met various surrogacy criteria indicative of a surrogate relationship: intercept with a 95% CI including 0, a slope with a 95% CI that does not include 0, a conditional variance value close to 0, an R value ≥ 0.85 (IQWiG high correlation), or an R^2 value ≥ 0.40 (BSES good surrogate). Outputs are in yellow for scenarios indicative of a medium correlation per IQWiG criteria (0.70 < R < 0.85). Outputs in **black** in the columns reporting % decrease in log (OS HR) associated with decrease in log (PFS or TTP HR) are estimates calculated based on slope values and are not evaluated against surrogacy criteria. [a]The correlation value falls between the IQWiG criteria for low and medium correlation.

Scenario analysis 3 (restricting to trials reporting PFS) and scenario analysis 4 (restricting to phase 3 trials)

- Both of these scenarios resulted in lower correlation terms (R = 0.71 for trials reporting PFS only and R = 0.72 for phase 3 trials only)
  - The estimated slopes were also lower, which resulted in lower predicted percentage decreases in the log (OS HRs) associated with a 10% decrease in log (PFS or TTP HRs) than the all-evidence analysis

Discussion

- All criteria used to assess the strength of the relationship between PFS or TTP and OS in RRMM, except for the IQWiG high-correlation criteria, were met for the all-evidence scenario suggesting that PFS or TTP is a valid surrogate for OS in RRMM
- Sensitivity analyses were conducted to investigate factors that may affect the association between PFS or TTP and OS HRs:
  - No RCTs evaluating triple-class refractory or TCE patients were identified; therefore, a scenario analysis restricting to trials including any DCR patients was conducted to approximate this population, which resulted in a higher slope and stronger correlation term compared with the all-evidence scenario, suggesting that surrogacy of PFS or TTP for OS may be stronger in patients with more refractory disease
  - The analysis using longer follow-up data resulted in a lower slope and lower correlation term compared with the analysis based on the same data cuts, suggesting that longer follow-up duration may affect the surrogacy relationship
  - Restricting to phase 3 RCTs and to trials reporting PFS yielded similar results to the all-evidence scenario
  - Results of the scenario restricted to 31 phase 3 RCTs (R = 0.72) were also similar to those obtained in Etekal et al.,[9] which was restricted to 16 phase 3 RCTs (R = 0.76)

Strengths and limitations

- This identified evidence base included 43 trials, being more robust than previous surrogacy analyses in RRMM[7,8,11] that assessed the association between PFS or TTP and OS
- The current analysis included fixed-effect bivariate meta-analysis models (Bayesian framework), recommended by NICE, and simple linear-regression models (frequentist framework); results were consistent across frameworks
- Both random- and fixed-effects models were considered for the analyses; however, convergence was not achieved with the former; therefore, estimates were only presented for the fixed-effect models, which did not account for the observed between-study differences
- The models used for the current analysis are based on the reported HRs in intention-to-treat populations and rely on the proportional hazards assumption, which is implausible if the hazard functions of competing interventions cross
- This analysis leveraged data from other published SLRs rather than a de novo SLR
  - Additionally, HRs were not derived from Kaplan-Meier curves if not directly reported
- Scenario analyses could not be conducted for RCTs evaluating treatment regimens with similar methods of action, due to heterogeneity across the included studies, or for RCTs reporting crossover-adjusted OS due to an insufficient number of studies reporting crossover adjustment
- A scenario analysis restricting to trials including any DCR patients was explored to approximate the surrogacy relationship in a TCE population; however, results of this scenario may not be generalizable to the TCE population

Conclusions

- Based on a trial-level surrogacy analysis of reported HRs from 43 published RCTs in RRMM, either PFS or TTP appear to be a valid surrogate for OS among patients with RRMM
- Additionally, scenario analyses suggest this surrogacy relationship may be stronger among more exposed and refractory patient populations
- Regression coefficients from these analyses can be used to predict OS HRs or to integrate surrogacy estimates into a cost-effectiveness analysis for regulatory and decision-making purposes

References

1. Jeryczynski G, et al. *Cancers* 2021;13:5886.
2. Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics. Published Dec 2018. Updated May 7, 2020. Accessed October 10, 2023.
3. Kemp R, Prasad V. *BMC Med* 2017;15:134.
4. Bujkiewicz S, et al. NICE DSU Technical Support Document 20. Version 2. https://www.sheffield.ac.uk/nice-dsu/tsds/full-list. Published Oct 22, 2019. Accessed October 10, 2023.
5. IQWiG. IQWiG Reports – Commission No. A10-05: Validity of surrogate endpoints in oncology. Updated Nov 23, 2011. Accessed October 10, 2023.
6. Murphy P, et al. *Health Technol Assess* 2021;25:1-228.
7. Félix J, et al. *BMC Cancer* 2013;13:122.
8. Dimopoulos M, et al. *Value Health* 2017;20:A408.
9. Etekal T, et al. *Br J Haematol* 2023;200:587-594.
10. Lassere MN, et al. *BMC Med Res Methodol* 2012;12:27.
11. Cartier S, et al. *Oncol Res Treat* 2015;38:88-94.

Acknowledgments

- The study was supported by Bristol Myers Squibb
- All authors contributed to and approved the presentation; writing support was provided by PRECISIONheor; writing and editorial assistance were provided by Ana Almeida, MSc, of Excerpta Medica, funded by Bristol Myers Squibb