When Trial Context Shapes Value: Same Effect, Different Outcomes

HTA4

David Dai-Wee Lee,[1] Wong Yoke Fui,[2] Nathorn Chaiyakunapruk,[3] Junice Ng,[4] Foo Chee Yoong[5]

[1]Sunway Medical Centre, Bandar Sunway, Selangor, Malaysia; [2]National Cancer Institute, Putrajaya, Malaysia; [3]Department of Pharmacotherapy, College of Pharmacy, University of Utah, Salt Lake City, UT, USA; [4]BeOne Medicines, Ltd, Singapore; [5]BeOne Medicines, Ltd, Petaling Jaya, Selangor, Malaysia

CONCLUSIONS

- The phase 3 trials JUPITER-02 and RATIONALE-309 reported nearly identical PFS HR but markedly different active arm median PFS
- RATIONALE-309 showed higher control arm progression hazards; after recalibration, the active arm curves converged and the 18-month RMST gap decreased from 2.57 to 0.17 months
- These findings suggest that absolute PFS differences across trials may reflect the progression risk environment in which the treatment effect was observed, rather than differences in treatment effect
- Relative measures and absolute outcomes should therefore be interpreted together, particularly when trial evidence is compared or translated into decision-making settings

Similar relative effects can produce very different absolute outcomes when the underlying trial risk environments differ

INTRODUCTION

- Clinical trial outcomes are shaped by more than treatment effect alone. They also reflect who was enrolled, how the study was conducted, and the underlying risk of progression in that trial
- Relative measures, such as hazard ratios (HRs), describe the treatment effect within a trial. Absolute outcomes, such as median progression-free survival (PFS), also depend on the trial's baseline progression risk environment
- As a result, similar relative treatment effects can produce very different absolute outcomes across trials. Without considering trial context, these differences may be mistaken for differences in treatment potency
- When trials evaluate the same disease setting, use a comparable control regimen, and assess the same endpoint in a broadly similar way, the control arms can help reveal how much of the observed outcome difference may be related to trial context

Real-World Example

- JUPITER-02 and RATIONALE-309 illustrate this challenge.[1,2] Both were phase 3 randomized studies evaluating a programmed cell death protein-1 (PD-1) inhibitor plus gemcitabine-cisplatin in the first-line treatment of recurrent or metastatic nasopharyngeal carcinoma. PFS data were derived from the published final analysis of both trials. Their PFS HRs were similar, but their active arm median PFS differed markedly.

Similar relative effect	Very different absolute outcome
PFS HR: 0.53 (tislelizumab) vs 0.52 (toripalimab)	Median PFS: 9.6 (tislelizumab) vs 21.4 (toripalimab) months

RATIONALE-309 overall median study follow-up time was 27.5 months (range: 0.1-53.0 months); JUPITER-02 median follow-up time at final PFS analysis cutoff was 22.1 months in the toripalimab arm and 21.4 months in the placebo arm

Objective

- To assess whether differences in control arm progression risk could help explain the divergence in active arm PFS despite similar relative treatment effects

Trial Context

- Although JUPITER-02 and RATIONALE-309 were similar in disease setting, chemotherapy backbone, and early PFS assessment, differences in patient mix, eligibility criteria, and trial conduct could still influence the progression risk environment observed in each study (**Table 1**)
- Because these factors operate together, the control arms provide a way to capture their combined influence on the progression risk observed in each trial

Table 1. Selected Baseline Characteristics, Eligibility Criteria, and Trial Design Differences Between RATIONALE-309 and JUPITER-02

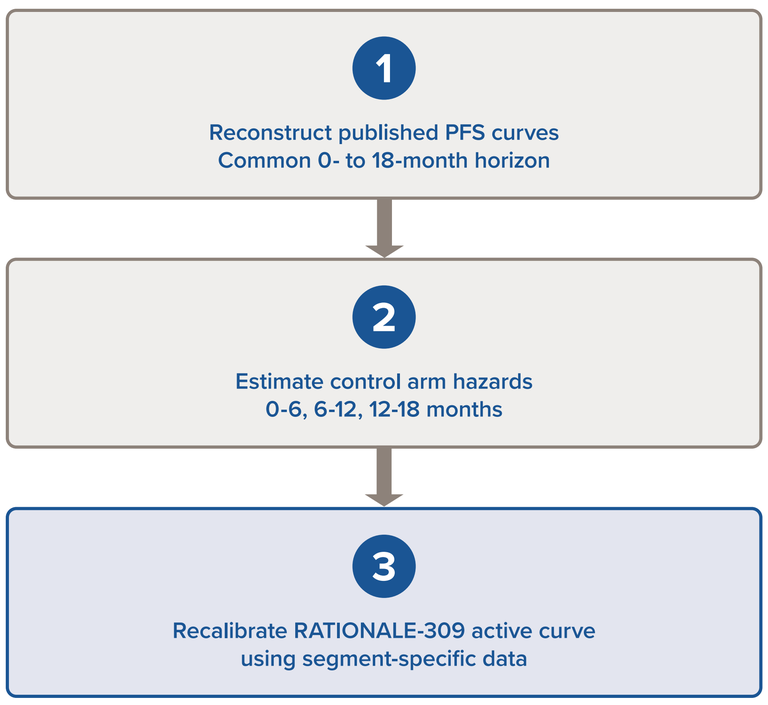
Domain	RATIONALE-309	JUPITER-02	Potential relevance to trial context
Population fitness	ECOG PS 1: 63.1%	ECOG PS 1: 43.3%	Higher proportion of less functionally fit patients in RATIONALE-309
Marrow reserve	ANC ≥1.5 × 10^9/L; no WBC minimum specified	ANC ≥2.0 × 10^9/L; WBC ≥4.0 × 10^9/L	More restrictive hematologic eligibility in JUPITER-02
Nutritional/ hepatic reserve	No explicit minimum albumin or ALP entry threshold; grade ≥3 hypoalbuminemia excluded	Albumin ≥3.0 g/dL; ALP ≤3 × ULN, or ≤5 × ULN with liver or bone metastases	Additional reserve-based selection in JUPITER-02
Cardiovascular exclusions	Explicit exclusions including recent cardiac chest pain or pulmonary embolism, MI within 6 months, NYHA class III-IV heart failure, and grade ≥2 ventricular arrhythmia	Explicit exclusions including NYHA class II or greater cardiac disease, MI within 3 months, unstable arrhythmia, and unstable angina	Criteria differed in scope and threshold; neither trial can be characterized simply as having universally stricter cardiovascular eligibility
Metastatic burden	Lung 46.0%; bone 44.5%; liver 43.0%	Lung 39.8%; bone 39.8%; liver 40.8%	Visceral burden broadly similar, with modestly higher lung/ bone involvement in RATIONALE-309
GP cycles	4-6 cycles, investigator discretion	Up to 6 cycles	Potential difference in chemotherapy exposure and patient selection
Assessment schedule	q6w x 6 months, q9w remainder of Year 1, then q12w	q6w in Year 1, then q9w	Broadly similar early imaging; later schedules differed
Control arm crossover	Permitted after IRC-confirmed progression	No explicit patient-level crossover	Relevant mainly to post-progression and OS interpretation

Abbreviations: ALP, alkaline phosphatase; ANC, absolute neutrophil count; ECOG PS, Eastern Cooperative Oncology Group performance status; GP, gemcitabine-platinum; IRC, independent review committee; MI, myocardial infarction; NYHA, New York Heart Association; OS, overall survival; q12w, every 12 weeks; q6w, every 6 weeks; q9w, every 9 weeks; ULN, upper limit of normal; WBC, white blood cell count.

METHODS

- Published PFS Kaplan–Meier curves and numbers-at-risk tables from JUPITER-02 and RATIONALE-309 were digitized and used to reconstruct pseudo-individual patient data over a common 18-month horizon[3]
- Follow-up was divided into 0- to 6-, 6- to 12-, and 12- to 18-month intervals. Average progression hazards were estimated within each interval using a piecewise constant-hazard approximation[4]
- For each interval, a control arm progression risk ratio, δ, was calculated as the JUPITER-02 control arm hazard divided by the RATIONALE-309 control arm hazard
- The interval-specific δ values were applied to the RATIONALE-309 active arm curve to estimate its PFS trajectory under the JUPITER-02 control arm risk environment
- Outcomes included reconstructed median PFS, 0- to 18-month restricted mean survival time (RMST), and interval-specific RMST
- This exploratory analysis was intended to contextualize absolute PFS differences, not to estimate head-to-head comparative efficacy
- If trial context accounted for much of the observed difference, the active arm curves would be expected to converge after recalibration

Figure 1. Control Arm Hazard-Anchored Recalibration Framework



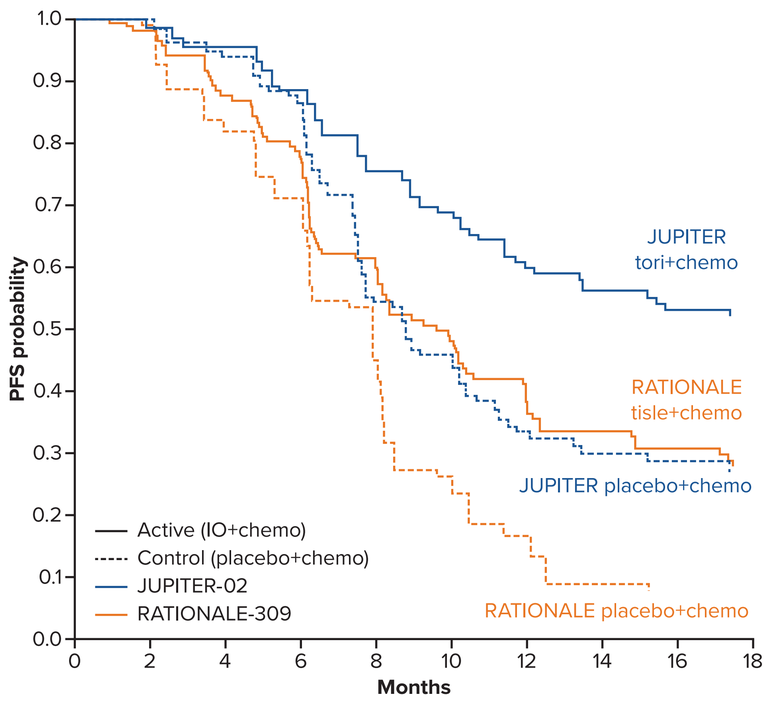
Abbreviation: PFS, progression-free survival.

RESULTS

Control Arm Progression Hazards Differed Across Trials

- The RATIONALE-309 control arm showed higher estimated progression hazards than the JUPITER-02 control arm in all three intervals (**Figure 2** and **Table 2**)
- The corresponding control arm baseline risk ratios were 0.571, 0.640, and 0.330 for 0-6, 6-12, and 12-18 months, respectively
- These findings indicate a consistently less favorable observed control arm risk environment in RATIONALE-309

Figure 2. Reconstructed PFS Curves Over a Standardized 18-Month Horizon



RATIONALE-309 showed a faster decline in the control arm, indicating a less favorable observed baseline risk environment. Data were derived from the published final analysis of both trials (RATIONALE-309 data cutoff: December 8, 2023; JUPITER-02 data cutoff: June 8, 2021).
Abbreviations: chemo, chemotherapy; IO, immuno-oncology; PFS, progression-free survival; tisle, tislelizumab; tori, toripalimab.

Table 2. Segment-Specific Control Arm Progression Hazards and Baseline Risk Ratios

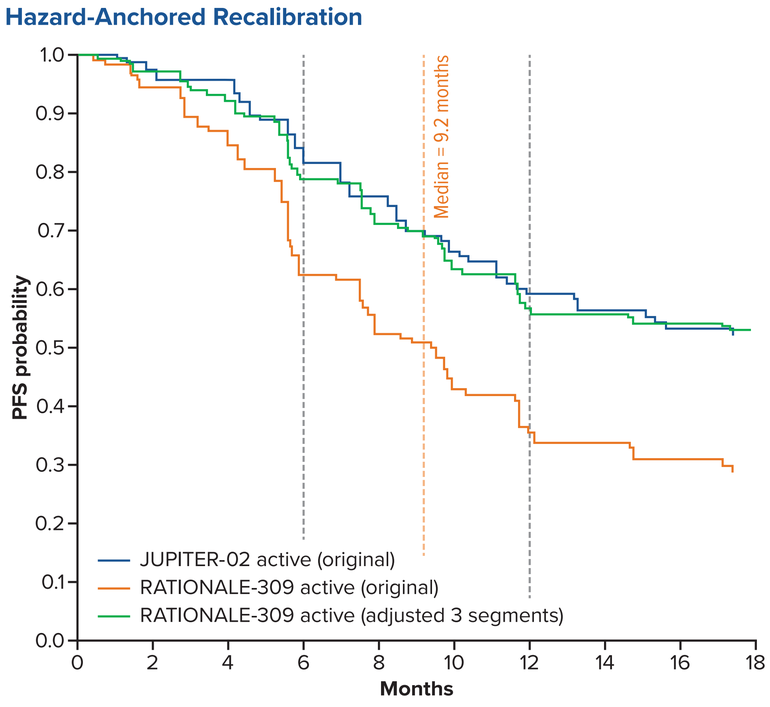
Interval	JUPITER-02 control hazard	RATIONALE-309 control hazard	Baseline risk ratio, δ
0-6 months	0.058	0.101	0.571
6-12 months	0.150	0.235	0.640
12-18 months	0.030	0.090	0.330

δ values below 1 indicate higher observed control arm progression risk in RATIONALE-309.

After Recalibration, the Active Arm Curves Converged

- Before recalibration, the reconstructed active arm PFS trajectories differed substantially between trials (**Figure 3**)
- After applying the interval-specific control arm baseline risk ratios, the recalibrated RATIONALE-309 active arm curve closely aligned with the JUPITER-02 active arm curve over the common 18-month horizon (**Figure 3**)

Figure 3. Active Arm PFS Before and After Control Arm Hazard-Anchored Recalibration



The recalibrated RATIONALE-309 active arm curve closely tracked the JUPITER-02 active arm curve over 18 months.
Abbreviation: PFS, progression-free survival.

The Gap Reduction Occurred Mainly in the Middle and Late Intervals

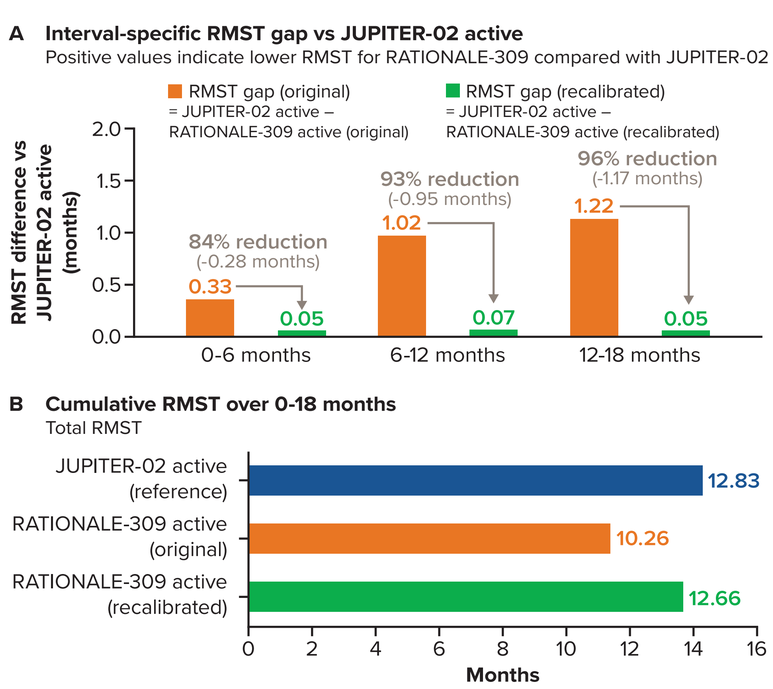
- The original 18-month RMST was 12.83 months for JUPITER-02 and 10.26 months for RATIONALE-309, representing a difference of 2.57 months (**Table 3**)
- After recalibration, the RATIONALE-309 18-month RMST increased to 12.66 months, leaving a residual difference of 0.17 months
- The approximate reduction was 93%
- The reduction in RMST difference occurred mainly during the 6- to 12- and 12- to 18-month intervals (**Figure 4**)

Table 3. PFS Outcomes Before and After Recalibration

PFS curve	Reconstructed median PFS	RMST 0-6	RMST 6-12	RMST 12-18	RMST 0-18
JUPITER-02 active, original	15.59	5.64	4.06	3.13	12.83
RATIONALE-309 active, original	9.19	5.31	3.04	1.91	10.26
RATIONALE-309 active, recalibrated	17.35	5.59	3.99	3.08	12.66

Median PFS values are reconstructed estimates within the standardized 0- to 18-month analysis and are not the published full-follow-up medians.
Abbreviations: PFS, progression-free survival; RMST, restricted mean survival time.

Figure 4. (A) Interval-Specific and (B) Cumulative RMST Differences Before and After Recalibration



Difference before recalibration	Difference after recalibration	Approximate reduction
2.57 months (12.83-10.26)	0.17 months (12.83-12.66)	93%

The RMST gap versus JUPITER-02 active decreased from 0.33 to 0.05 months at 0-6 months, from 1.02 to 0.07 months at 6-12 months, and from 1.22 to 0.05 months at 12-18 months. Recalibration applied control arm hazard ratios estimated in each interval to the RATIONALE-309 active curve. Positive gap values indicate lower RMST for RATIONALE-309 compared with JUPITER-02.
Abbreviation: RMST, restricted mean survival time.

REFERENCES

1. Mai HQ, et al. *JAMA*. 2023;330:1961-1970.
2. Yang Y, et al. *JAMA Oncol*. 2026;12:384-393.
3. Guyot P, et al. *BMC Med Res Methodol*. 2012;12:9.
4. Michael F. *Ann Stat*. 1982;10:101-113.

ACKNOWLEDGMENTS

The authors thank the patients and their families, investigators, co-investigators, and study teams at each of the participating centers. This study was sponsored by BeOne Medicines, Ltd.

CORRESPONDENCE: Foo Chee Yoong, cheeyoong.foo@beonemed.com

Presented at ISPOR Asia Pacific Summit 2026, Bangkok, Thailand, September 6-8, 2026