The added value of Bayesian interim analysis in randomized phase II and phase III clinical trials in hemato-oncology

Prospective phase III randomized controlled trials (RCT) are the gold standard for evaluating the benefit-risk ratio of new therapies.[1,2] RCT frequently include interim analyses, primarily to ensure patient safety and monitor treatment efficacy.[3] These interim analyses may potentially result in early trial termination or modification if there is evidence of harm, overwhelming efficacy, or futility. The time between initiation and obtaining final results of RCT is generally very long, which delays access to new therapies for patients. Conventional frequentist statistical approaches apply conservative stopping rules at interim analyses and require long-term follow-up. Bayesian statistical methods may address these challenges and may potentially reduce the time of RCT and the exposure of patients to ineffective or harmful treatment regimens, as recently suggested.[4-7] However, the use of Bayesian interim analyses in phase II/III hemato-oncology trials, particularly in varying situations involving superiority, no superiority and inferiority has not been well explored. Therefore, the current study applied Bayesian interim analysis to three phase II/III randomized controlled trials to evaluate the added value of Bayesian statistical models next to conventional interim designs.

Clinical and outcome data from three randomized trials were used, including the phase III HOVON95/EMN02 trial and the phase III HOVON87/NMSG18 trial for patients with multiple myeloma (MM) and the phase II HOVON103/SAKK30/10 trial for patients with AML, which were previously reported.[8-10] The HOVON95/EMN02 trial is registered at the *EU Clinical Trials Register: EudraCT 2009-017903-28* and *clinicaltrials gov. Identifier: NCT01208766*. The HOVON87/NMSG18 trial is registered at *www.trialregister.nl* as *NTR1630 (EudraCT 2007-004007-34)*. The HOVON103/SAKK30/10 trial is registered at *www.trialregister.nl* as *NL5748 (NTR5902)*.The study designs are summarized in *Online Supplementary Figure S1*. The trials were approved by the institutional review boards of each participating site, as previously reported.[8-10] Three interim analyses were retrospectively mimicked based on actual data after the inclusion of 50%, 75% and 100% of the total included patients in all three trials for early benefit-risk assessment. This corresponded to three simulated interim analyses after the inclusion of 599, 898, and 1,197 patients in the HOVON95/EMN02 trial, after the inclusion of 319, 478, and 637 patients in the HOVON87/NMSG18 trial and after the inclusion of 51, 77 and 102 patients in the HOVON103/SAKK30/10 trial (Figures 1A, 2A and 3A). The primary endpoint in the HOVON95/EMN02 trial was progression-free survival (PFS) and the anticipated median

PFS improvement in the experimental arm corresponded to a target hazard ratio (HR)=0.78.[8] The primary endpoint in the HOVON87/NMSG18 trial was PFS with a projected median PFS improvement of 8 months for the experimental arm corresponding to a target HR=0.714.[9] The primary endpoint in the HOVON103/SAKK30/10 trial was complete response (CR) and the anticipated difference in CR rates between the two treatment arms was set at 20%.[10] The HOVON95/EMN02 trial showed superiority for PFS in the experimental arm, whereas in the HOVON87/NMSG18 trial no superiority for PFS was observed, and CR rates were lower with the experimental arm in the HOVON103/SAKK30/10 trial.

Bayesian inference allows for the evaluation of the probability that a treatment provides a benefit exceeding various clinically meaningful thresholds. In contrast to traditional frequentist approaches, which rely on null hypothesis significance testing and *P* values, Bayesian inference distributes probability across a spectrum of parameter values - such as a treatment effect - updating these as new data emerge, designated as the posterior probability. Bayesian posterior probability distributions of the treatment difference were calculated for the primary endpoints of the trial: PFS (HOVON95/EMN02 trial; HOVON87/NMGSG18 trial) and CR (HOVON103/SAKK30/10 trial), respectively. The probability distributions were summarized to provide estimates of the probability distribution median and 95% Bayesian credible intervals (95% CrI) of the HR, whereas frequencies were calculated using the highest density interval method. Non-informative priors were used for the interim analyses as no prior data was available. Four Markov Monte Carlo chains were run with 50,000 iterations and chain convergence was evaluated by quantile plots and Gelman-Rubin diagnostics. PFS was evaluated using a Bayesian Cox proportional hazards survival model. The probabilities of HR were estimated based on the assumed effect size for PFS. CR, adverse events (AE) grade 3-5 and death within 100 days after the start of treatment (early death) were evaluated using a Bayesian β-binomial model. The probability of any benefit (treatment difference >0%) in CR, AE or early death was estimated for the experimental treatment arm compared with the control treatment arm. Bayesian statistical methods were retrospectively applied and the results were aligned with a conventional group sequential design at the same interim time points to offer additional context. Therefore, no stopping boundaries were applied nor applicable. Analyses were performed in R (version 4.2.2.) using the software of JAGS version 4.3.0 in the package "rjags" and the "brms" package and using

EAST statistical software (version 6).[11,12] The R-script of all the main analyses are published online (https://github.com/niekvandermaas/Analysis-of-The-Added-Value-of-Bayesian-Interim-Analysis).

In the HOVON95/EMN02 trial, Bayesian interim analyses showed that the probability of a HR below 0.78 (target HR<0.78) for the HSCT arm compared with the VMP arm was 76.2% at interim analysis 1, 77.1% at interim analysis 2 and 93.7% at interim analysis 3 (Figure 1B). Collectively,

Bayesian analysis provided early signs of efficacy upon all three successive interim analyses. Nevertheless, the conventional group sequential design showed that the HR of PFS crossed the efficacy boundary only at the third interim analysis (*Online Supplementary Figure S2A*). Grade 3-5 AE excluding hematological AE were more frequently observed in the HSCT arm compared with the VMP arm with a treatment difference of 8-10% across the three interim analyses (Figure 1C).
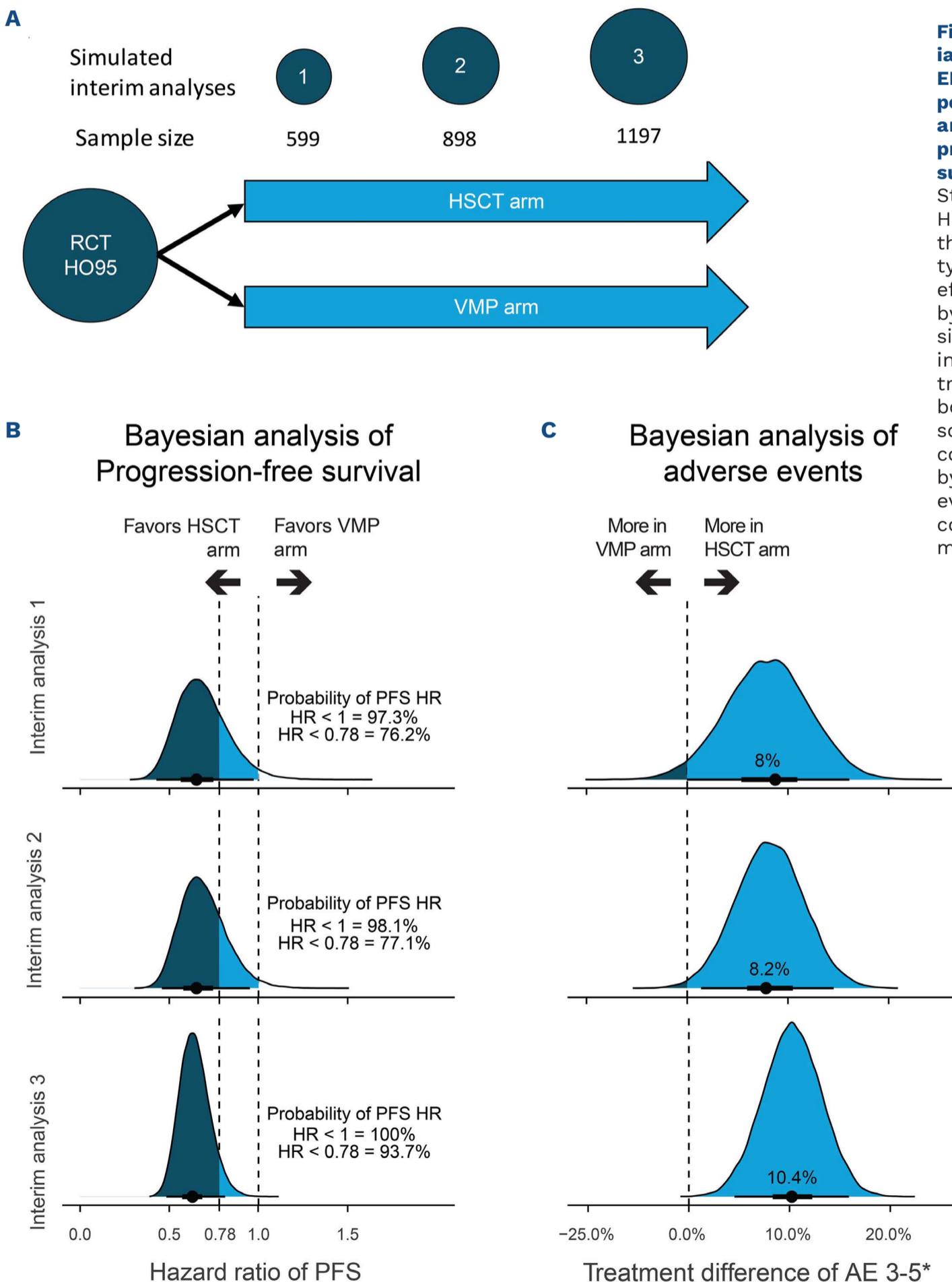

Figure 1. Study design and Bayesian analysis of the HOVON95/EMN02 trial comparing hematopoietic stem cell transplantation and bortezomib, melphalan, prednisone in progression-free survival and adverse events. (A) Study design flow diagram of the HOVON95/EMN02 trial with (B) the estimation of the probability for the assumed treatment effect (hazard ratio [HR]<0.78) by Bayesian analysis of progression-free survival (PFS) comparing hematopoietic stem cell transplantation (HSCT) *versus* bortezomib, melphalan, prednisone (VMP) (control arm) and (C) comparison of HSCT *versus* VMP by Bayesian analysis of adverse events (AE). RCT: randomized controlled trial; *excluding hematological AE.

In the HOVON87/NMSG18 trial, Bayesian interim analyses showed low posterior probabilities of the target benefit (target HR<0.714) for the MPR-R arm compared with the MPT-T arm being 76.2% at interim analysis 1, 37.7% at interim analysis 2 and 9.1% at interim analysis 3 (Figure 2). Collectively, from the second interim analysis onward, the likelihood of achieving the targeted relative benefit corresponding to at least an 8-month increase in median PFS became highly unlikely, and even elusive at the last interim analysis. However, the conventional group sequential design did not show signs of futility given that the HR of PFS did not cross the futility boundary at interim analyses 1, 2 and 3, respectively (*Online Supplementary Figure S2*). AE grade 3-5 were more frequently observed in the MPR-R arm compared with the MPT-T arm with a treatment difference of ~7% across interim analyses (Figure 2C).
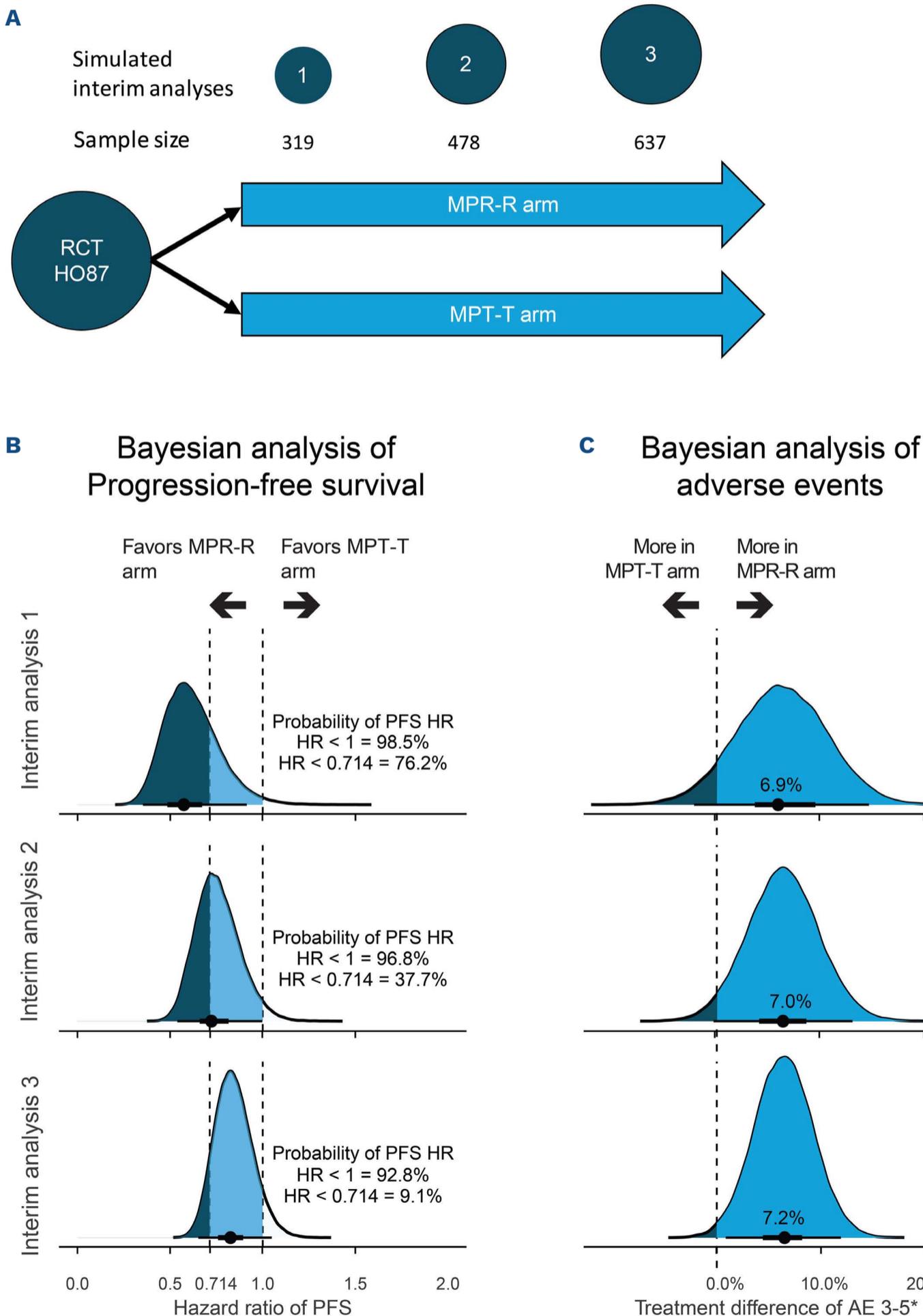

Figure 2. Study design and Bayesian analysis of the HOVON87/NMSG18 trial comparing melphalan, prednisone, and lenalidomide (MPR-R) and melphalan, prednisone, and thalidomide (MPT-T) in progression-free survival and adverse events. (A) Study design flow diagram of the HOVON87/NMSG18 trial (B) with the estimation of the probability for the assumed treatment effect (hazard ratio [HR]<0.714) by Bayesian analysis of progression-free survival (PFS) comparing MPR-R treatment *versus* control (MPT-T) and (C) comparison of MPR-R *versus* control by Bayesian analysis of adverse events (AE). RCT: randomized controlled trial; *including hematological AE.

In the HOVON103/SAKK30/10 trial, both Bayesian interim analyses and conventional interim methods consistently indicated futility. The probability of a higher CR rate in the selinexor treatment arm compared with the control treatment arm was 2.9% at interim analysis 1, 0.5% at interim analysis 2 and 0.8% at interim analysis 3 (Figure 3B). Similarly, the conventional group sequential design showed that the treatment difference of CR crossed the futility boundary at all three interim analyses (*Online Supplementary Figure S2*). Additionally, death within 100 days was more frequently observed in the selinexor treatment arm compared with the control treatment arm at all three interim analyses (Figure 3C). In summary, this study showed that Bayesian interim analyses provided early and informative insight into the efficacy and harm of treatments, which compared favorably with the conventional group sequential design in two of three
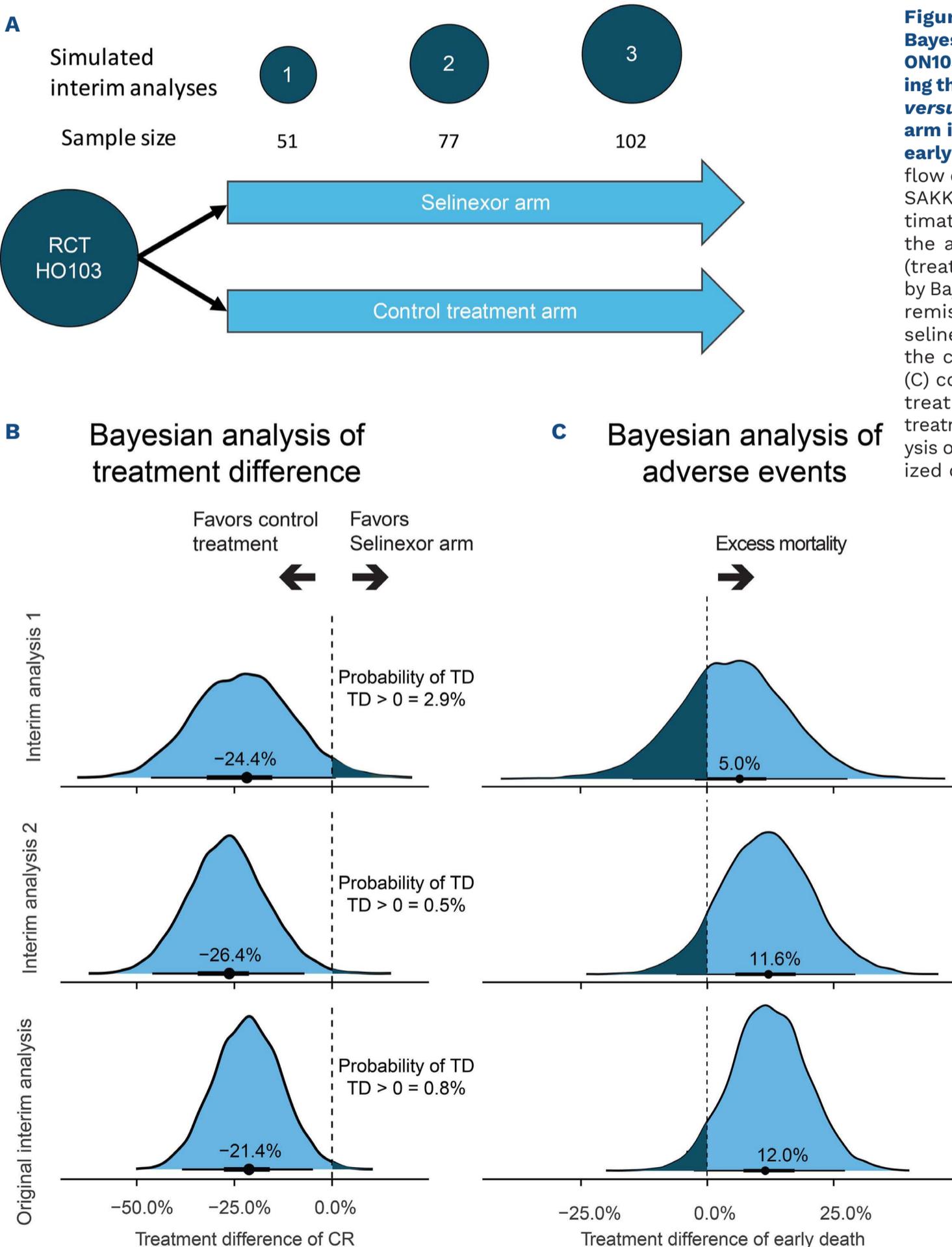

Figure 3. Study design and Bayesian analysis of the HOVON103/SAKK30/10 trial comparing the selinexor treatment arm *versus* the control treatment arm in complete remission and early deaths. (A) Study design flow diagram of the HOVON103/SAKK30/10 trial with (B) the estimation of the probability for the assumed treatment effect (treatment difference [TD] >0) by Bayesian analysis of complete remission (CR) comparing the selinexor treatment arm *versus* the control treatment arm and (C) comparison of the selinexor treatment arm *versus* control treatment arm by Bayesian analysis of early death. RCT: randomized controlled trial.

trials (HOVON95/EMN02 trial and HOVON87/NMSG18 trial). It aligns with previous findings from a Bayesian interim analysis study of the HOVON132 AML trial by van der Maas *et al.*, in which an early lack of efficacy in all four simulated interim analyses was detected, while the frequentist group sequential design only showed futility at the third interim time point.[4]

The paucity of events early in the trial conduct may pose a challenge in early interim analyses as conducted in our study. This is especially a challenge in MM, where novel agents have substantially improved survival outcomes and delayed disease progression, resulting in prolonged time-to-event times. In contrast to MM, AML is still characterized by a more rapid disease progression and time-to-event times in AML trials are shorter compared with other hemato-oncological diseases. This shorter timeline enhances the potential of Bayesian interim analyses and might enable a more rapid data and safety monitoring board (DSMB) advice, in order to limit patient exposure to ineffective or harmful treatments. On the other hand, short timelines necessitate careful consideration and in the case of overwhelming efficacy, early termination of a trial is not always feasible and depends on the strength of evidence and the time to develop adverse events. Incorporating external control data from historical MM or AML trials as a Bayesian prior in our model may further enhance the utility of Bayesian analyses, which data were not available for the trial settings used in this analysis.[5,13,14] Successful implementation of Bayesian interim monitoring will, however, depend on timely data delivery, robust data infrastructure, and defining standardized, prespecified Bayesian decision thresholds to ensure consistency across trials.

In conclusion, the current study supports a broader use of Bayesian analysis in early interim analyses in phase III randomized controlled trials in hemato-oncology. While the overall study design may remain consistent with conventional RCT practices, Bayesian analyses show added value for interim analyses that may inform DSMB members for their meetings in addition to conventional interim analyses. Bayesian interim analysis may allow for identifying a clinically relevant treatment difference and/or clinically relevant toxicities at early interim time points. By providing informative, additional insights during a trial, this approach holds promise for optimizing methodologies in future prospective comparative studies, ultimately enhancing trial efficiency and decision-making.

Authors

Anna J.T. Smit,[1]* Niek G. van der Maas,[1]* Bronno van der Holt,[1,2] Yvette van Norden,[1,2] Erik van Werkhoven,[1,2] Annette Juul Vangsted,[3] Paula F. Ypma,[4] Heleen A. Visser-Wisselaar,[2] Rossella Troia,[5] Mario Boccadoro,[5] Bob Löwenberg,[1] Jeroen J. W. M. Janssen,[6] Gert J. Ossenkoppele,[7] Fredrik Schjesvold,[8] Pieter Sonneveld,[1] Sonja Zweegman,[9] Jurjen Versluis[1] and Jan J. Cornelissen[1]

[1]Department of Hematology, Erasmus University Medical Center Cancer Institute, Rotterdam, the Netherlands; [2]HOVON Foundation, Rotterdam, the Netherlands; [3]Department of Hematology, Rigshospitalet, Copenhagen, Denmark; [4]Department of Hematology, HagaZiekenhuis, Den Haag, the Netherlands; [5]European Myeloma Network, Turin, Italy; [6]Department of Hematology, Radboud UMC, Nijmegen, the Netherlands; [7]Amsterdam UMC, location VUMC, Cancer Center Amsterdam, Amsterdam, the Netherlands; [8]Oslo Myeloma Center, Department of Hematology, Oslo University Hospital, Oslo, Norway and [9]Department of Hematology, Amsterdam UMC, Cancer Center Amsterdam, Amsterdam, the Netherlands

AJTS and NGvdM contributed equally as first authors.

Correspondence:
J.J. CORNELISSEN - j.cornelissen@erasmusmc.nl

Disclosures

MB has received honoraria from Sanofi, Celgene, Amgen, Janssen, Novartis, Bristol Myers Squibb, and AbbVie; has served on the advisory boards for Janssen and GlaxoSmithKline; has received research funding from Sanofi, Celgene, Amgen, Janssen, Novartis, Bristol Myers Squibb, and Mundipharma. All other authors have no conflicts of interest to disclose.

Contributions

AJTS, NGvdM, JV, and JC contributed to the conceptualization and writing of the original draft. AJTS and NGvdM participated in formal analysis. JV accessed and verified the data. AJTS, NGvdM, JV, EvW, and JC contributed to data interpretation, manuscript revision, and approval of this manuscript for submission. All other authors provided study materials or enrolled patients, were involved in the collection and assembly of clinical data, contributed to writing, reviewing and editing, and approved the manuscript for submission.

Data-sharing statement

The data that support the findings of this study are available from HOVON but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of HOVON.

References

1. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet. 2002;359(9300):57-61.
2. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.
3. Ciolino JD, Kaizer AM, Bonner LB. Guidance on interim analysis methods in clinical trials. J Clin Transl Sci. 2023;7(1):e124.
4. van der Maas NG, Versluis J, Nasserinejad K, et al. Bayesian interim analysis for prospective randomized studies: reanalysis of the acute myeloid leukemia HOVON 132 clinical trial. Blood Cancer J. 2024;14(1):56.
5. Ruberg SJ, Beckers F, Hemmings R, et al. Application of Bayesian approaches in drug development: starting a virtuous cycle. Nat Rev Drug Discov. 2023;22(3):235-250.
6. Sherry AD, Msaouel P, Miller AM, et al. Bayesian interim analysis and efficiency of phase III randomized trials. Br J Cancer. 2025;133(8):1145-1151.
7. Kidwell KM, Roychoudhury S, Wendelberger B, et al. Application of Bayesian methods to accelerate rare disease drug development: scopes and hurdles. Orphanet J Rare Dis. 2022;17(1):186.
8. Cavo M, Gay F, Beksac M, et al. Autologous haematopoietic stem-cell transplantation versus bortezomib-melphalan-prednisone, with or without bortezomib-lenalidomide-dexamethasone consolidation therapy, and lenalidomide maintenance for newly diagnosed multiple myeloma (EMN02/HO95): a multicentre, randomised, open-label, phase 3 study. Lancet Haematol. 2020;7(6):e456-e468.
9. Zweegman S, van der Holt B, Mellqvist U-H, et al. Melphalan, prednisone, and lenalidomide versus melphalan, prednisone, and thalidomide in untreated multiple myeloma. Blood. 2016;127(9):1109-1116.
10. Janssen JJWM, Löwenberg B, Manz M, et al. Addition of the nuclear export inhibitor selinexor to standard intensive treatment for elderly patients with acute myeloid leukemia and high risk myelodysplastic syndrome. Leukemia. 2022;36(9):2189-2195.
11. Plummer M. Bayesian Graphical Models using MCMC [R package rjags version 4-16]. 2024; published online August 19, 2024. https://CRAN.R-project.org/package=rjags. Accessed November 7, 2024.
12. Kieser M. Methods and Applications of sample size calculation and recalculation in clinical trials. Springer Nature, 2020.
13. Viele K, Berry S, Neuenschwander B, et al. Use of historical control data for assessing treatment effects in clinical trials. Pharm Stat. 2014;13(1):41-54.
14. van Rosmalen J, Dejardin D, van Norden Y, Löwenberg B, Lesaffre E. Including historical data in the analysis of clinical trials: Is it worth the effort? Stat Methods Med Res. 2018;27(10):3167-3182.