Leveraging AI for Modeling MRD and Survival Outcomes in Multiple Myeloma

Abstract:

Introduction: Minimal residual disease (MRD) has been recently accepted by FDA as an endpoint for accelerated approval in Multiple Myeloma (MM) based on individual patient data collected from randomized trials. However, emerging data from recent trials were not included. Literature based meta-analysis on the correlation between MRD and survival outcomes requires extensive manual literature review. By leveraging AI with expert-in-the-loop, we can generate reliable evidence from a comprehensive list of clinical studies with up-to-date outcomes.

Method: This study presents an AI-assisted framework that identifies relevant studies and filters critical information to analyze published data via two independent tasks. In the first task, the associations between treatment effects on MRD and various clinical endpoints across different patient populations were modeled using weighted least squares. The strength of these associations was measured by the coefficients of determination (R^2) with their 95% confidence intervals (CIs). For the second task, we analyzed individual-level association via a digitization approach.

Results: The AI searched for eligible studies (> 50 patients per treatment arm) reporting progression-free survival (PFS), overall survival (OS), overall response rates (ORR), and MRD negative complete response rate (MRD- CR rate) data , using multi-parameter, next-generation flow cytometry or sequencing (MFC, NGC, NGF) methods with a minimum sensitivity threshold of 10^{-5} from January 1, 2010, to May 29, 2024. This enabled us to extend the list of studies from 15 in previous papers to 18 trials. The analysis results reported R^2 of 0.69 (95% CI 0.50 - 0.89) for log of hazard ratio (HR) regarding PFS versus log of odds ratio (OR) regarding MRD- CR rate, revealing a moderate to strong trial-level correlation between MRD- CR rates and PFS. Additionally, we utilized AI techniques to extract accurate individual-level evidence from published KM curves via digitization and scalable vector graphics. The results show that MRD- CR rate is usually associated with prolonged PFS.

Conclusion: This research introduces a novel AI-assisted method to conduct automatic, contemporary analyses of the association between MRD and different clinical endpoints based on published data in literatures.

1. **Introduction**

Significant advancements in the treatment of MM over recent decades have resulted in deeper responses. Recent reports indicate that Progression-free survival (PFS) in newly diagnosed MM now exceeds four years [1,2]. Additionally, nearly 100% ORR have been observed in both treatment and control groups for some studies [2,3,4]. The prolonged PFS duration and similar ORR between treatment and control groups may potentially delay approval and access to effective MM treatments. Minimal residual disease (MRD) has emerged as a candidate endpoint associated with improved survival outcomes, recently accepted by the FDA for accelerated approval in MM.

This study proposes an Artificial Intelligence (AI)-assisted, expert-in-the-loop framework, designed to efficiently identify relevant studies and filter critical information for complex medical analysis objectives. As an illustration, we applied our framework to two equally important tasks: 1. modeling the trial-level association between MRD and various clinical endpoints (PFS, OS, ORR); 2. proposing a benchmark model for estimating MRD negative complete response rates (MRD- CR rates) using mPFS, which is one of the most often reported arm level summary statistics in MM literature.

For the first task, AI tools were employed to identify eligible studies, followed by manual verification conducted by two independent investigators. Associations were quantified using trial-level coefficients of determination (R^2_{trial}) with 95% confidence intervals, using weighted least squares methods weighted by trial sample sizes. Alternative weighting strategies such as inverse variances were also considered. Subgroup analyses were stratified by patient type: newly diagnosed transplant eligible (NDTE), newly diagnosed transplant ineligible (NDTI), and relapsed or refractory (RRMM) populations. Criteria for interpreting R^2 were predefined as poor (R^2<0.4), moderate (0≤ R^2 < 0.8), and strong correlation ($R^2 \geq$ 0.8).

For the second task, we aimed to conduct a pooled analysis using IPD to evaluate the impact of MRD status on survival outcomes in multiple myeloma patients. This analysis focused on three distinct patient populations: newly diagnosed transplant-eligible (NDTE), newly diagnosed transplant-ineligible (NDTI), and relapsed/refractory multiple myeloma (RRMM). Second, we sought to perform a comprehensive subgroup analysis to assess the consistency of MRD's impact across various patient characteristics. To support these objectives, we utilized AI-powered tools and a novel algorithm for generating synthetic IPD (synthIPD), enabling a more robust and comprehensive analysis.

1. Methods

Study Selection

Studies published as papers and abstracts in PubMed and in major meetings such as ASCO, ASH, EHA, and IMS, spanning from January 1, 2010, to May 29, 2024 were systematically identified using AI. Eligible studies included in the analysis met the following criteria: reporting PFS HR, OS HR, ORR by two-treatment arms, and minimal residual disease negative rates at suspected complete response (MRD- CR rate) by treatment arms. Studies were required to have more than 50 patients per treatment arm and utilize Multi-parametric Flow Cytometry (MFC), Next-Generation Flow Cytometry (NGF), or Next-Generation Sequencing (NGS) with a minimum sensitivity threshold of 10^{-5}.

Data Preparation

Following the Oncologic Drugs Advisory Committee (ODAC) Meeting in April 2024, the MRD- CR rate was defined as the proportion of patients achieving MRD negativity after suspected complete response among those who had been randomized. Among the selected papers, the DETERMINATION trial (NCT01208662) [2] reported only the MRD negative rate. The PLEIADES trial (NCT03412565) [7] did not provide PFS/OS information. ICARIA-MM (NCT02990338) [8] and KarMMA-3 (NCT03651128) [9] reported MRD- CR rates in the control groups of 0\% and 0.8\% respectively, which were deemed as outliers and thus excluded from the analysis of MRD- rate versus PFS/OS/ORR in the primary objective.

Despite stricter selection criteria, the number of studies included in the primary analysis expanded from 15 trials as reported in [5] to 19 two-arm trials, encompassing 7 NDTE, 5 NDTI, and 7 RRMM trials. Study characteristics, including trial name, study population demographics, MRD- CR rates, mPFS, PFS/OS hazard ratios (HR) with 95\% confidence intervals, ORR, and details of MRD assessment methods, were systematically extracted and were partially presented in **Table 1**.

Trial	NCTID	Disease	Method	Drug	MRD- CR	mPFS	PFS HR
ATLAS [10]	NCT02659293	NDTE	NGS(10^{-5})	KRd vs R	53% vs 31%	59.1 vs 41.4	0.51
CASSIOPEA [3]	NCT02541383	NDTE	NGS(10^{-5})	D-VTd vs VTd	34% vs 20%	Inf vs Inf	0.47

GEM2012MENOS65 [11]	NCT01916252	NDTE	NGF(10^{-6})	VRD-BuMel vs VRD-Mel	58% vs 55%	Inf vs75.3	0.88
GRIFFIN [4]	NCT02874742	NDTE	NGS(10^{-5})	D-RVd vs RVd	62% vs 27%	Inf vs Inf	0.45
PERSEUS [12]	NCT03710603	NDTE	NGS(10^{-5})	D-VRd vs VRd	75% vs 48%	Inf vs Inf	0.42
TOURMALINE-MM3 [13]	NCT01850524	NDTE	MFC(10^{-5})	Ixazomib vs Placebo	26% vs 18%	26.5 vs 21.3	0.83
STAMINA[26]	NCT01109004	NDTE	NGS(10^{-5})	Single Auto vs Tandem Auto vs Auto +RVD	75% vs 85% vs 78%		0.695 0.985
CLARION [14]	NCT01818752	NDTI	NGF(10^{-6})	KMP vs VMP	5% vs 5%	22.3 vs 22.1	0.91
TOURMALINE-MM2 [15]	NCT01850524	NDTI	NGS(10^{-5})	Ixazomib-Rd vs Placebo-Rd	15% vs 7%	35.3 vs 21.8	0.83
ALCYONE [16]	NCT02195479	NDTI	NGF(10^{-5})	D-VMP vs VMP	28% vs 7%	36.5 vs 19.3	0.42
MAIA [17]	NCT02252172	NDTI	NGF(10^{-5})	DRd vs Rd	31% vs 10%	Inf vs 31.9	0.53
OCTANS[18]	NCT03217812	NDTI	MFC(10^{-5})	D-VMP vs VMP	30% vs 7%	Inf vs 18.2	0.43
POLLUX [19]	NCT02076009	RRMM	NGS(10^{-5})	DRd vs Rd	33% vs 7%	44.5 vs 17.3	0.44
CASTOR [20]	NCT02136134	RRMM	NGS(10^{-5})	D-Vd vs Vd	15% vs 2%	16.7 vs 7.1	0.31
BOSTON [21]	NCT03110562	RRMM	NGS(10^{-5})	SVd vs Vd	5% vs 4%	13.9 vs 9.5	0.70
CANDOR [22]	NCT03158688	RRMM	NGS(10^{-5})	KdD vs Kd	22% vs 8%	28.4 vs 15.2	0.64
APOLLO [23]	NCT03180736	RRMM	NGS(10^{-5})	DaraPD vs PD	9% vs 2%	12.4 vs 6.9	0.63
IKEMA [24]	NCT03275285	RRMM	NGS(10^{-5})	Isa-Kd vs Kd	26% vs 12%	35.7 vs 19.2	0.58

CARTITUDE-4 [25]	NCT04181827	RRMM	NGS(10^{-5})	Cilta-cel vs PVd/DPd	61% vs 16%	Inf vs 11.8	0.26

Table 1. Included trial information. KRd: Dexamethasone - Lenalidomide -Carfilzomib; D-VTd: Daratumumab - Bortezomib -Thalidomide - Dexamethasone; VRD-BuMel: Bortezomib - Lenalidomide - Dexamethasone -Busulfan - Melphalan; D-RVd: Daratumumab - Lenalidomide - Bortezomib - Dexamethasone; D-VRd: Daratumumab - Bortezomib - Lenalidomide - Dexamethasone; KMP: carfilzomib - melphalan - prednisone; VMP: bortezomib - melphalan - prednisone; SVd: Selinexor - Bortezomib +Dexamethasone; KdD: carfilzomib - dexamethasone - daratumumab; Kd: carfilzomib - dexamethasone; DaraPD: Daratumumab + Pomalidomide + Dexamethasone; Isa-Kd: Isatuximab + Carfilzomib + Dexamethasone; Cilta-cel: ciltacartagene autoleucel; MFC: multi-parametric flow cytometry; NGF: next-generation flow cytometry; NGS: next generation sequencing; Inf:infinity, not reached; NA: not available; *: Single-arm trial included in the secondary objective. The assessment times for MRD- CR rates are assumed to be at suspected/confirmed complete response (CR) or better at any time during the study.

Statistical Methods

In the first task, trial-level associations were evaluated using weighted least squares regression models. The dependent variable was the logarithm of HR for PFS, while the independent variable was the logarithm of the OR for MRD- CR. The strength of these trial-level associations was quantified using the coefficient of determination.

In the second task, the individual-level association was measured via individual patient data (IPD) approach. In updated studies that reported Kaplan-Meier (KM) curves stratified by MRD status, the plots were digitized via scalable vector graphics (SVG) and then individual patient data (IPD) were reconstructed using an algorithm specifically designed for SVG. Compared with traditional digitization + health technology assessment methods (IPDfromKM package reference), our method appears to be more robust and almost identical to reported curves, see *Figure S8*.

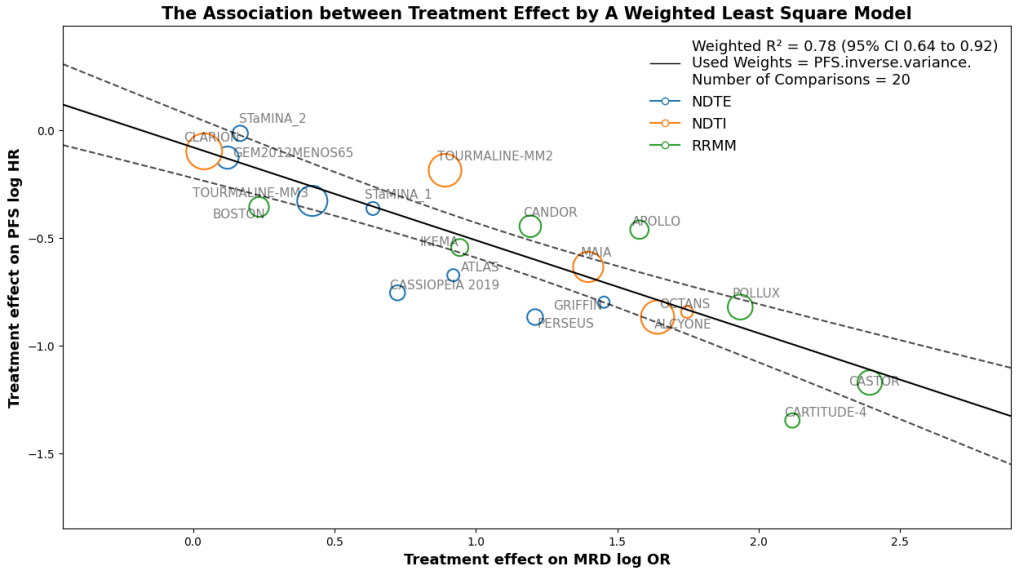
Specifically for POLLUX study, synthetic IPD data was generated for Age and Prior lenalidomide subgroups (SynthIPD Discover clinical insights and enhance statistical analyses with Synthetic individual patient data.pdf (hopeai.co). This method was shown to be robust and can capture the true subgroup KM curve even without knowing the IPD within that subgroup. This approach provides a comprehensive view of how MRD- CR impact survival outcomes within subgroups.

3. Results

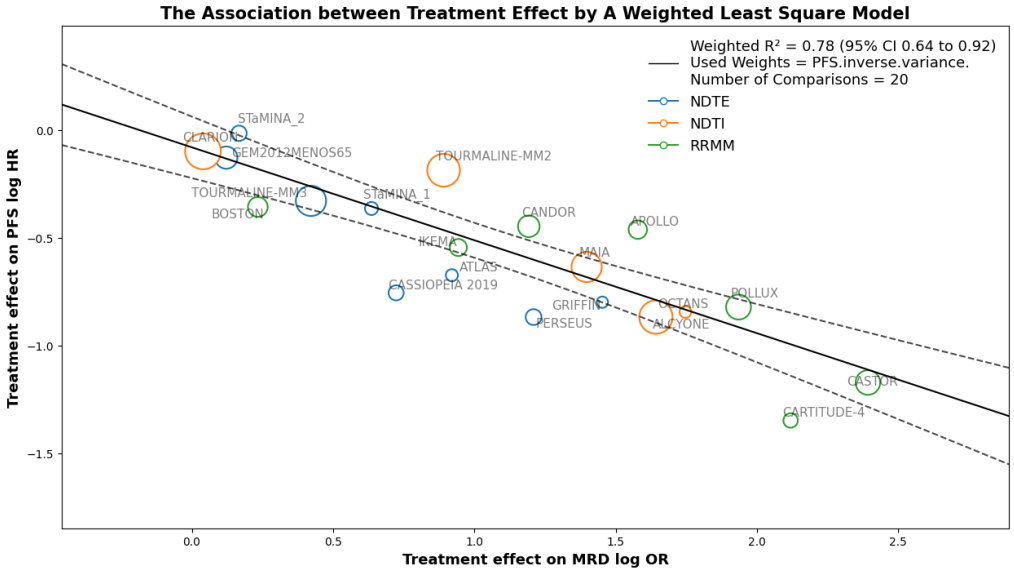
Taks 1: Trial-Level Association (The graphs need to be regenerated)

For the first task, trial-level coefficients of determination were calculated for (1) log(HR) PFS versus log(OR) MRD, (2)log(HR) OS versus log(OR) MRD *(see Supplementary materials Figure S3-S4)*, (3) log(OR) ORR versus log(OR) MRD main analysis *Figure S5*, (4)log(HR) PFS/OS versus log(OR) MRD for each subgroup stratified by disease type (N DTE, NDTI, RRMM), see Supplement \ref{supp: primary} *Figure S6*.

Results for (1) were reported in *Figure 1(a)* using sample sizes as weights, additionally, results using inverse variance of PFS HR were reported in *Figure 1(b*). The weighted trial-level coefficient of determination for (1) log(HR) PFS vs log(OR) MRD- CR observed in the aggregated analysis of the 19 clinical trials was $R^2_{trial} = 0.71$ (95% CI, 0.52-0.89), see *Figure 1(a)*. Sensitivity analyses using a leave-one-out approach showed R^2_{trial} ranging from 0.63 - 0.80 *Figure S7*. Another sensitivity analysis using 16 studies with sensitivity 10^{-5} only was conducted, see *Figures S1-S2*, showing moderate correlation R^2_{trial}=0.55, 0.66 respectively for two different weights. Summarizing the results, the odds ratio of MRD-negative rates in MM are moderately correlated with the hazard ratio in PFS. The result aligns with recent discoveries in [5] using 15 studies with reported R^2 = 0.70(0.41 - 0.98), and [27] using 13 studies with reported R^2 = 0.53 (0.21 - 0.77).



(a) PFS log(HR) versus MRD log(OR) weighted by sample size of each trial.



(b) PFS log(HR) versus MRD log(OR) weighted by inverse variance of PFS HR.

Figure 1. The weighted R^2_{trial} in the aggregated analysis of 19 clinical trials. PFS HR and MRD-CR odds ratio are natural log transformed. The black solid lines are the fitted regression lines and the black dotted lines are 95% confidence bands.

Result (2) was reported in *Figure S3* with sample sizes as weights. The weighted trial-level R^2_{trial} = 0.59 (0.35 - 0.84), showing a moderate correlation between MRD- CR rates odds ratio and OS hazard ratio. Similarly, the reported trial level R^2_{trial} = 0.53 (0.26-0.8) for (3) log(OR) ORR vs log(OR) MRD- CR. Results using inverse variance of OS log(HR) were provided in Supplementary materials *Figure S4*.

The subgroup analyses were conducted for NDTE, NDTI and RRMM populations. In NDTE, R^2_{trial} = 0.78 (0.61 - 0.95); in NDTI, R^2_{trial} = 0.85 (0.73 - 0.95); in RRMM, R^2_{trial} = 0.72 (0.50 - 0.93). Notably, the NDTI subgroup exhibited a strong correlation with R^2_{trial}> 0.8 and a lower CI bound greater than 0.6. The correlation, though very strong, included only 5 NDTI studies and should be interpreted with caution.

Task 2: A pooled analysis of MRD- CR rate using individual patient data

The second task aimed to investigate the role of MRD- CR rates in long-term survival outcomes based on individual patient data. AI searched for updated publications on MRD- CR rates with Kaplan-Meier (KM) curves stratified by MRD- CR status in the previous 18 studies. Updated versions of 9 studies are available, they are presented in *Figure 2* along with the $\log$ HR of MRD- versus MRD+ within each study. Additionally, the pooled log HR were reported by a stratified cox model of all studies within each disease population (NDTE, NDTI, RRMM).

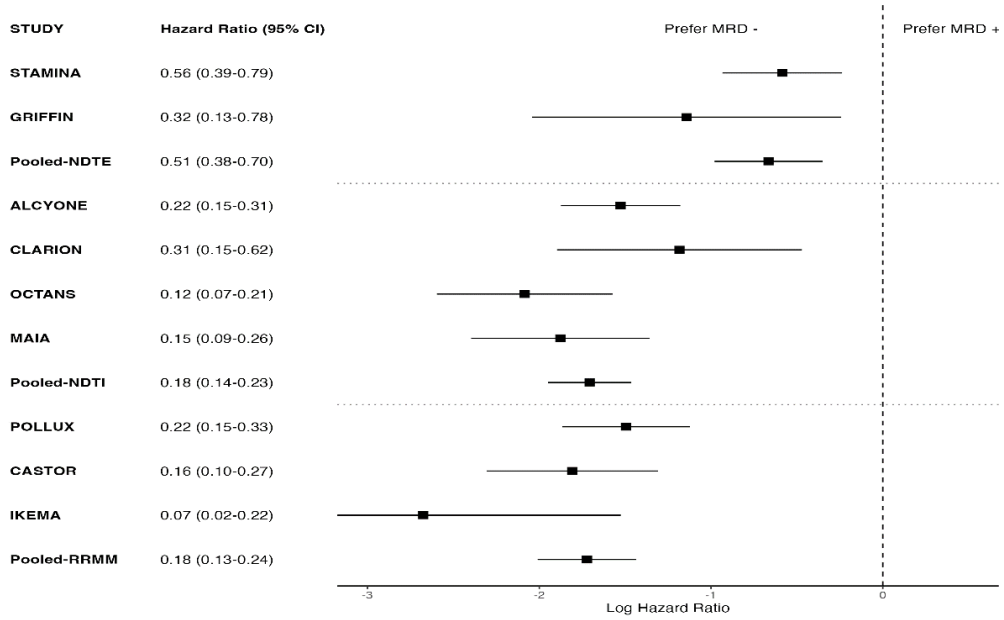


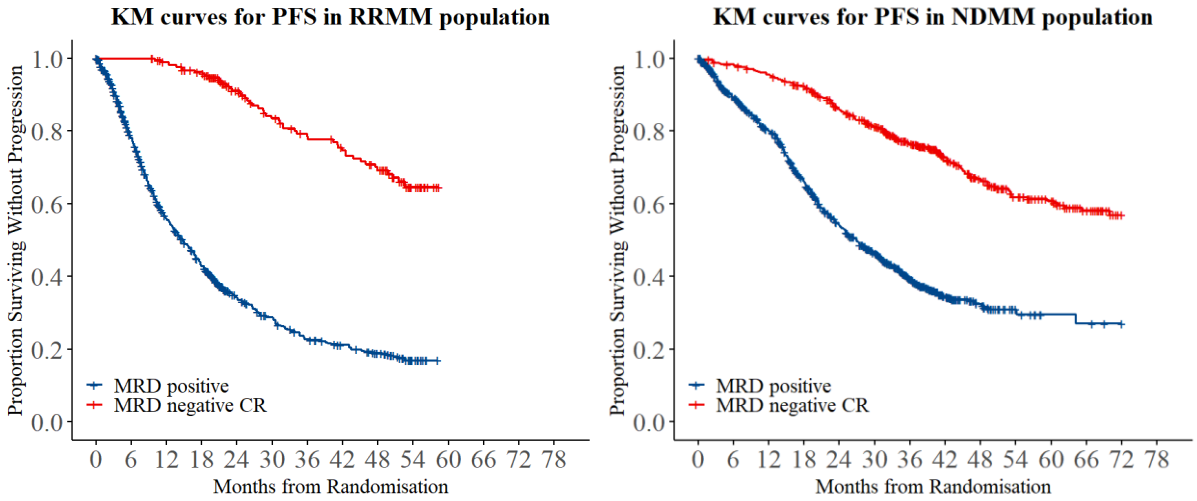
Figure 2. Base-case and pooled association analysis of MRD status. The updated versions of the studies are STAMINA [37], GRIFFIN [38], ALCYONE and MAIA [39], CLARION [14], OCTANS [40],POLLUX and CASTOR [19], IKEMA [41].

Survival outcomes (PFS) were reported with updated MRD- CR information for 9 studies, including 2 NDTE, 4 NDTI and 4 RRMM studies. PFS was significantly improved with MRD- CR vs MRD+ across all disease settings, see *Figure 3*. In NDTE group MRD- CR was associated with moderate improvement of PFS with HR 0.51(95% CI 0.38-0.70). Median PFS was Inf(95% CI 69.9-Inf) for MRD- CR patients and 64.1(95% CI 48.1-Inf) for MRD+ patients. Three-year PFS rates were 64.5% and 58.2% for patients who were MRD- and MRD+ respectively.

In NDTI group MRD- CR was associated with very significant improvement of PFS with HR 0.18(95% CI 0.14-0.23). Median PFS was Inf(95% CI 47.3-Inf) for MRD- CR patients and 25.0(95% CI 23.2-26.8) for MRD+ patients. Three-year PFS rates were 60.4% and 28.4% for patients who were MRD- and MRD+ respectively.

In NDMM group MRD- CR was associated with moderate to significant improvement of PFS with HR 0.26(95% CI 0.22-0.31). Median PFS was Inf(95% CI Inf-Inf) for MRD- CR patients and 26.8(95% CI 25.1-29.2) for MRD+ patients. Three-year PFS rates were 76.5% and 39.5% for patients who were MRD- and MRD+ respectively.

In RRMM group MRD- CR was associated with very significant improvement of PFS with HR 0.18(95% CI 0.13-0.24). Median PFS was Inf (95% CI Inf-Inf) for MRD- CR patients and 14.9(95% CI 13.2-16.4) for MRD+ patients. Three-year PFS rates were 69.3% and 19.0% for patients who were MRD- and MRD+ respectively.



(a) (b)

Figure 3. Kaplan-Meier curves for PFS for (a) RRMM (b)NDMM population stratified by MRD status.

Furthermore, SynthIPD method, a synthetic data solution powered by generative AI and an optimization algorithm, was used to generate IPD with covariates information using only published summary statistics and intention-to-treat Kaplan-Meier curves for the POLLUX study. Synthetic IPD produces high-quality data that mimics the statistical properties of original IPD. The errors of important statistical findings between generated and original IPD are typically within 2%.

As a simple illustration, the pooled analysis of hazard ratio within subgroups (age and prior lenalidomide exposure) are reported in *Figure 4*, the PFS outcomes summarized as KM curves were reported in *Figure 5*. The results suggested improved survival outcome is associated with MRD- CR status within both subgroups. The data used in these results are not publicly available and are generated via SynthIPD algorithm.

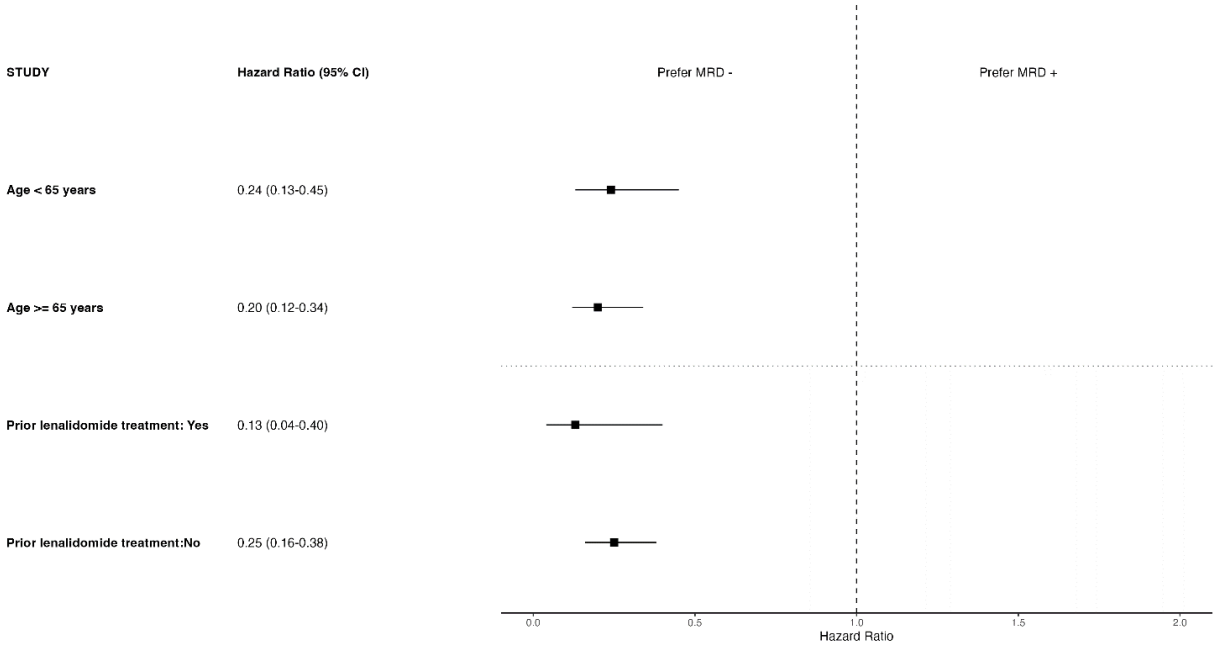
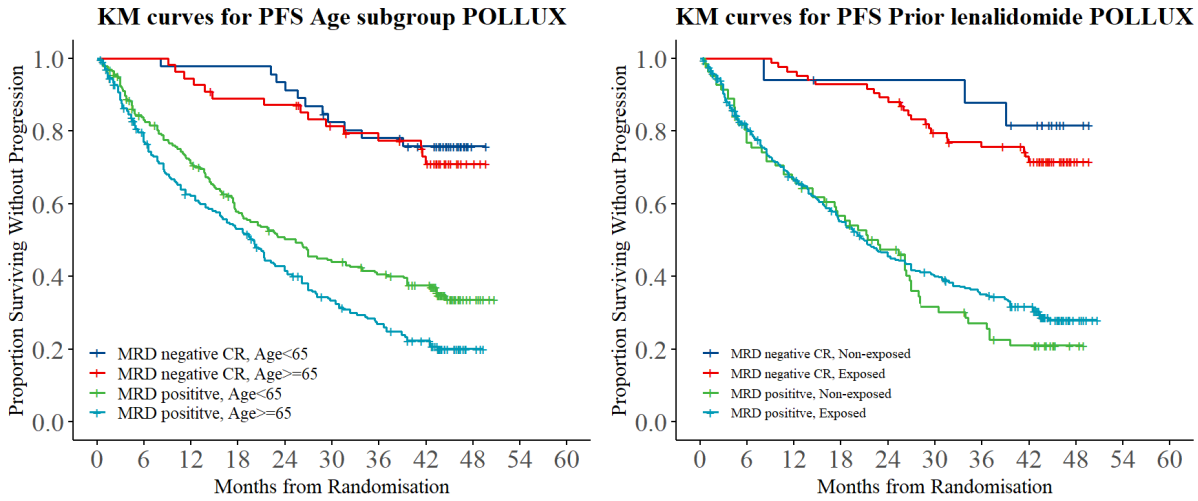


Figure 4. Association analysis of MRD status of POLLUX study stratified by Age and Prior Lenalidomide exposure subgroups.



(a) (b)

Figure 5. Kaplan-Meier curves for PFS for (a) Age subgroup (>=65 vs <65) (b) Previous Lenalidomide exposure subgroup (non-exposed vs exposed) stratified by MRD status in POLLUX study.

4. Discussion

Recent studies [32,5,33,34] have consistently shown a positive correlation between MRD- CR rates and mPFS in Multiple Myeloma (MM). Recognizing this, ODAC has recently accepted MRD- CR rates as an endpoint for accelerated approval in MM. However, previous integrated studies have limitations due to outdated clinical evidence, as well as significant time consumption.

This trial-level analysis of MRD- CR rates further underscore their potential role in improving survival outcomes across different MM patient populations. Our findings align with recent analyses [5] and ODAC recommendations, reporting R^2_{trial} = 0.70 (0.41 - 0.98) and R^2_{trial} = 0.69 (0.51 - 0.87) for the association between PFS log(HR) and MRD- CR log(OR). These results advocate the consideration of potential use of MRD- CR rates as a surrogate for mPFS.

Moreover, we assessed individual level association by digitizing the reported KM plots via a novel, SVG technique. The result of digitization is robust and accurate and the individual survival outcomes favored the MRD- CR group instead of MRD+ group, which is a further corroboration of our findings in the previous trial level analysis. Taking advantage of a newly proposed synthetic data generating algorithm, we were able to, using POLLUX study as an example, compare synthetic IPD for MRD- status on subgroup level as well. The statistical conclusions were very similar to the truth, the errors in summary statistics were typically within a range of 2% when comparing synthetic data versus the truth.

Our contribution is twofold: (1) we showed that MRD- CR rate is moderately correlated with mPFS at trial-level using up-to-date clinical evidence. Compared with existing literature [5], we removed 2 studies which failed our eligibility criteria and added 5 studies that were not considered previously. The MRD- CR rates were defined according to the latest ODAC document. Results obtained were consistent with previous results. (2) We evaluated the impact of MRD- CR rate on PFS via an IPD approach, we further proposed a way to analyze the MRD- CR rate behavior within each subgroup of a study. (3) In contrast to previous integrated studies and meta-analyses, which typically require up to six months to complete [6], our research achieved robust statistical analyses for MRD within an exceptionally short time frame (less than two weeks), leveraging the most current and comprehensive clinical evidence available. Additionally, the consistency of our results with previous studies proved the applicability and potential of our framework.

However, several limitations require attention. Variability in treatments, timing and methods of MRD assessment across studies introduces heterogeneity. Additionally, the study's reliance on aggregate trial data precluded individual-level analyses of the MRD- CR rate's relationship with PFS/OS/ORR. Lastly, our analysis was based on a limited set of 19 trials, underscoring the need for future studies to enhance reliability and generalizability.

In conclusion, our study supports using MRD- CR rate as an endpoint for accelerated approval for mPFS in MM. Both main and subgroup analyses demonstrate moderate correlations between MRD- CR rates and PFS/OS/ORR. Moreover, the proposed benchmark model can estimate MRD- CR rates from mPFS data with high interpretability and low prediction error. By leveraging AI tools, we are able to facilitate the achievement of complex research objectives efficiently and accurately, this, in turn, marks a significant advancement in medical research methodologies.

References

1. Michel Attal et al. “Lenalidomide, bortezomib, and dexamethasone with transplantation for myeloma”. In: New England Journal of Medicine 376.14 (2017), pp. 1311–1320.

2. Paul G Richardson et al. “Triplet therapy, transplantation, and maintenance until progression in myeloma”. In: New England Journal of Medicine 387.2 (2022), pp. 132–147.

3. Philippe Moreau et al. “Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (CASSIOPEIA): a randomised, open-label,phase 3 study”. In: The Lancet 394.10192 (2019), pp. 29–38.

4. Peter M Voorhees et al. “Addition of daratumumab to lenalidomide, bortezomib, and dexamethasone for transplantation-eligible patients with newly diagnosed multiple myeloma (GRIFFIN): final analysis of an open-label, randomised, phase 2 trial”. In: The Lancet Haematology 10.10 (2023), e825–e837.

5. Bruno Paiva et al. “Impact of treatment effect on MRD and PFS: an aggregate data analysis from randomized clinical trials in multiple myeloma”. In: Blood Advances 8.1 (2024), pp. 219–223.

6. I Elaine Allen and Ingram Olkin. “Estimating time to conduct a meta-analysis from number of citations retrieved”. In:Jama 282.7 (1999), pp. 634–635.

7. Philippe Moreau et al. “Daratumumab, carfilzomib, and dexamethasone in relapsed or refractory myeloma: final analysis of PLEIADES and EQUULEUS”. In: Blood cancer journal 13.1 (2023), p. 33.

8. Michel Attal et al. “Isatuximab plus pomalidomide and low-dose dexamethasone versus pomalidomide and low-dose dexamethasone in patients with relapsed and refractory multiple myeloma (ICARIA-MM): a randomised, multicentre, open-label, phase 3 study”. In:The Lancet 394.10214 (2019), pp. 2096–2107.

9. Paula Rodriguez-Otero et al. “Ide-cel or standard regimens in relapsed and refractory multiple myeloma”. In:New England Journal of Medicine 388.11 (2023), pp. 1002–1014.

10. Dominik Dytfeld et al. “Carfilzomib, lenalidomide, and dexamethasone or lenalidomide alone as maintenance therapy after autologous stem-cell transplantation in patients with multiple myeloma (ATLAS): interim analysis of a randomised, open-label, phase 3 trial”. In: The Lancet Oncology 24.2 (2023), pp. 139–150.

11.Laura Rosiñol et al. “Bortezomib, lenalidomide, and dexamethasone as induction therapy prior to autologous transplant in multiple myeloma”. In:Blood, The Journal of the American Society of Hematology 134.16 (2019), pp. 1337–1345.

12. Pieter Sonneveld et al. “Daratumumab, bortezomib, lenalidomide, and dexamethasone for multiple myeloma”. In: New England Journal of Medicine 390.4 (2024), pp. 301–313.

13. Meletios A Dimopoulos et al. “Oral ixazomib maintenance following autologous stem cell transplantation (TOURMALINE-MM3): a double-blind, randomised, placebo-controlled phase 3 trial”. In:The Lancet 393.10168 (2019), pp. 253–264.

14.Thierry Facon et al. “Carfilzomib or bortezomib with melphalan-prednisone for transplant-ineligible patients with newly diagnosed multiple myeloma”. In: Blood, The Journal of the American Society of Hematology 133.18 (2019), pp. 1953–1963.

15.Thierry Facon et al. “Oral ixazomib, lenalidomide, and dexamethasone for transplant-ineligible patients with newly diagnosed multiple myeloma”. In: Blood, The Journal of the American Society of Hematology 137.26 (2021), pp. 3616–3628.

16. Maria-Victoria Mateos et al. “Overall survival with daratumumab, bortezomib, melphalan, and prednisone in newly diagnosed multiple myeloma (ALCYONE): a randomised, open-label, phase 3 trial”. In:The Lancet 395.10218 (2020), pp. 132–141.

17.Thierry Facon et al. “Daratumumab, lenalidomide, and dexamethasone versus lenalidomide and dexamethasone alone in newly diagnosed multiple myeloma (MAIA): overall survival results from a randomised, open-label, phase 3 trial”. In: The Lancet Oncology 22.11 (2021), pp. 1582–1596.

18. Weijun Fu et al. “Bortezomib, melphalan, and prednisone with or without daratumumab in transplant-ineligible asian patients with newly diagnosed multiple myeloma: the phase 3 octans study”. In: Clinical Lymphoma Myeloma and Leukemia 23.6 (2023), pp. 446–455.

19. Hervé Avet-Loiseau et al. “Evaluation of sustained minimal residual disease negativity with daratumumab-combination regimens in relapsed and/or refractory multiple myeloma: analysis of POLLUX and CASTOR”. In: Journal of Clinical Oncology 39.10 (2021), p. 1139.

20. Pieter Sonneveld et al. “Overall survival with daratumumab, bortezomib, and dexamethasone in previously treated multiple myeloma (CASTOR): a randomized, open-label, phase III trial”. In: Journal of Clinical Oncology 41.8 (2023), pp. 1600–1609.

21. Sebastian Grosicki et al. “Once-per-week selinexor, bortezomib, and dexamethasone versus twice-per-week bortezomib and dexamethasone in patients with multiple myeloma

(BOSTON): a randomised, open-label, phase 3 trial". In: The Lancet 396.10262 (2020), pp. 1563–1573.

22. Saad Z Usmani et al. "Final analysis of carfilzomib, dexamethasone, and daratumumab vs carfilzomib and dexamethasone in the CANDOR study". In: Blood Advances 7.14 (2023), pp. 3739–3748.

23. Meletios A Dimopoulos et al. "Daratumumab plus pomalidomide and dexamethasone versus pomalidomide and dexamethasone alone in previously treated multiple myeloma (APOLLO): an open-label, randomised, phase 3 trial". In: The Lancet Oncology 22.6 (2021), pp. 801–812.

24.Thomas Martin et al. "Isatuximab, carfilzomib, and dexamethasone in patients with relapsed multiple myeloma: updated results from IKEMA, a randomized Phase 3 study". In: Blood cancer journal 13.1 (2023), p. 72.

25.Jesús San-Miguel et al. "Cilta-cel or standard care in lenalidomide-refractory multiple myeloma". In: New England journal of medicine 389.4 (2023), pp. 335–347.

26. Alexander M Lesokhin et al. "Elranatamab in relapsed or refractory multiple myeloma: phase 2 MagnetisMM-3 trial results". In: Nature medicine 29.9 (2023), pp. 2259–2267.

27. Joseph Ficek et al. "Association of minimal residual disease negativity rates with progression free survival in frontline therapy trials for newly diagnosed multiple myeloma: a meta-analysis". In: Clinical Lymphoma Myeloma and Leukemia 23.5 (2023), e213–e221.

28.Meletios A Dimopoulos et al. "Carfilzomib and dexamethasone versus bortezomib and dexamethasone for patients with relapsed or refractory multiple myeloma (ENDEAVOR): a randomised, phase 3, open-label, multicentre study". In: The Lancet Oncology 17.1 (2016), pp. 27–38.

29.Philippe Moreau et al. "Once weekly versus twice weekly carfilzomib dosing in patients with relapsed and refractory multiple myeloma (ARROW): interim analysis results of a randomised, phase 3 study". In: The Lancet Oncology 19.7 (2018), pp. 953–964.

30. Paul G Richardson et al. "Pomalidomide, bortezomib, and dexamethasone for patients with relapsed or refractory multiple myeloma previously treated with lenalidomide (OPTIMISMM): a randomised, open-label, phase 3 trial". In: The Lancet Oncology

20.6 (2019), pp. 781–794.

31. Meletios A Dimopoulos et al. "Elotuzumab plus pomalidomide and dexamethasone for multiple myeloma". In: New England Journal of Medicine 379.19 (2018), pp. 1811–1822.

32.Nikhil C Munshi et al. "A large meta-analysis establishes the role of MRD negativity in long-term survival outcomes in patients with multiple myeloma". In: Blood advances 4.23 (2020), pp. 5988–5999.

33. Ola Landgren et al. “Role of MRD status in relation to clinical outcomes in newly diagnosed multiple myeloma patients: a meta-analysis”. In: Bone marrow transplantation 51.12 (2016), pp. 1565–1568.

34. Kenneth C Anderson et al. “The role of minimal residual disease testing in myeloma treatment selection and drug development: current value and future applications”. In: Clinical Cancer Research 23.15 (2017), pp. 3980–3993.

35.Thierry Facon et al. “Oral ixazomib, lenalidomide, and dexamethasone for transplant-ineligible patients with newly diagnosed multiple myeloma”. In: Blood, The Journal of the American Society of Hematology 137.26 (2021), pp. 3616–3628.

36. Marcelo C. Pasquini et al., Minimal Residual Disease Status in Multiple Myeloma 1 Year After Autologous Hematopoietic Cell Transplantation and Lenalidomide Maintenance Are Associated With Long-Term Overall Survival. JCO 42, 2757-2768(2024).

37. Pasquini, M. C., Wallace, P. K., Logan, B., Kaur, M., Tario, J. D., Howard, A., ... & Hahn, T. (2024). Minimal Residual Disease Status in Multiple Myeloma 1 Year After Autologous Hematopoietic Cell Transplantation and Lenalidomide Maintenance Are Associated With Long-Term Overall Survival. *Journal of Clinical Oncology*, JCO-23.

38. Chari, A., Kaufman, J. L., Laubach, J., Sborov, D. W., Reeves, B., Rodriguez, C., ... & Richardson, P. G. (2024). Daratumumab in transplant-eligible patients with newly diagnosed multiple myeloma: final analysis of clinically relevant subgroups in GRIFFIN. Blood cancer journal, 14(1), 107.

39. San-Miguel, J., Avet-Loiseau, H., Paiva, B., Kumar, S., Dimopoulos, M. A., Facon, T., ... & Bahlis, N. J. (2022). Sustained minimal residual disease negativity in newly diagnosed multiple myeloma and the impact of daratumumab in MAIA and ALCYONE. Blood, The Journal of the American Society of Hematology, 139(4), 492-501.

40. Fu, W., Bang, S. M., Huang, H., Kim, K., Li, W., An, G., ... & Wang, J. (2023). Subgroup Analyses of Progression-Free Survival from the Phase 3 Octans and Alcyone Studies in Transplant-Ineligible Patients with Newly Diagnosed Multiple Myeloma Treated with Bortezomib, Melphalan, and Prednisone with or without Daratumumab. *Blood*, *142*, 2006.

41. Martin, T., Mikhael, J., Hajek, R., Kim, K., Suzuki, K., Hulin, C., ... & Moreau, P. (2022). Depth of response and response kinetics of isatuximab plus carfilzomib and dexamethasone in relapsed multiple myeloma. *Blood Advances*, *6*(15), 4506-4515.

Supplementary materials

Additional results and sensitivity analysis for the first task

Figure S3-S5 presented the association analysis of OS log(HR) and ORR log(OR) against MRD- CR log(OR). For OS analysis, only a total of 14 studies were included because the OS data were not mature for OCTANS, TOURMALINE-MM2, PLEIADES, TOURMALINE-MM3 [18,15,35,13] at the date of analysis. Similarly, GEM2012MENOS65 [11] was excluded when analyzing ORR because of missing information on ORR.

For trial-level association between PFS log(HR) versus MRD- CR log(OR), a sensitivity analysis using only 10^{-5} MRD measurement sensitivity was performed, using both sample sizes and PFS log(HR) variances as weights, see *Figure S1,S2*.

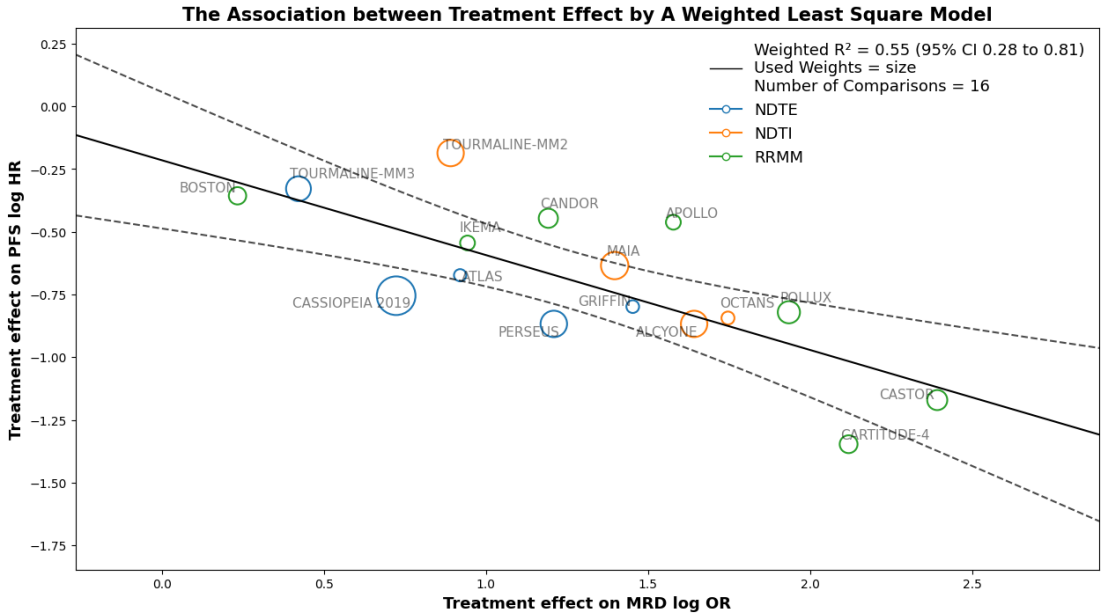


Figure S1: The weighted $R^2{}_{trial}$ in the aggregated analysis of 16 clinical trials with sensitivity level 10^{-5} only. PFS HR and MRD- CR odds ratio are natural log transformed. The weights equal sample sizes. The black solid lines are the fitted regression lines and the black dotted lines are 95% confidence bands.

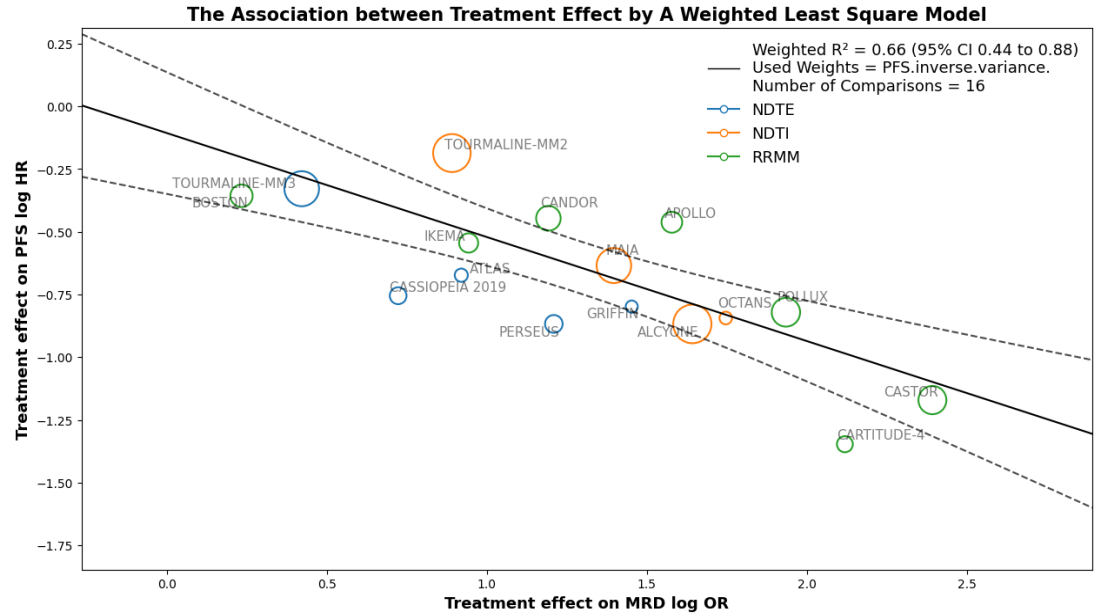


Figure S2: The weighted $R^2{}_{trial}$ in the aggregated analysis of 16 clinical trials with sensitivity level as 10^{-5} only. PFS HR and MRD- CR odds ratio are natural log transformed. The weights equal inverse variances of PFS log(HR). The black solid lines are the fitted regression lines and the black dotted lines are 95% confidence bands.

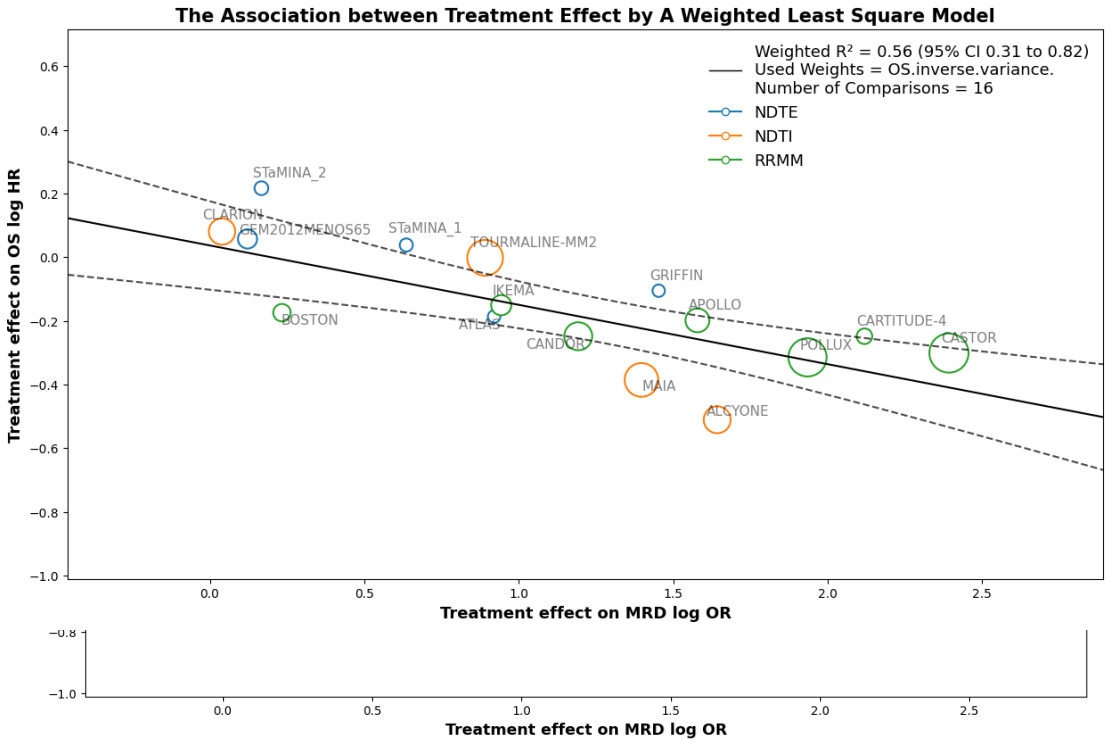


Figure S3. The weighted $R^2{}_{trial}$ in the aggregated analysis of 16 clinical trials reporting OS information. OS HR and MRD- CR odds ratio are natural log transformed. The weights equal sample sizes. The black solid lines are the fitted regression lines and the black dotted lines are 95% confidence bands.

Figure S4. The weighted $R^2{}_{trial}$ in the aggregated analysis of 16 clinical trials reporting OS information. OS HR and MRD- CR odds ratio are natural log transformed. The weights equal inverse variance of log(HR). The black solid lines are the fitted regression lines and the black dotted lines are 95% confidence bands.

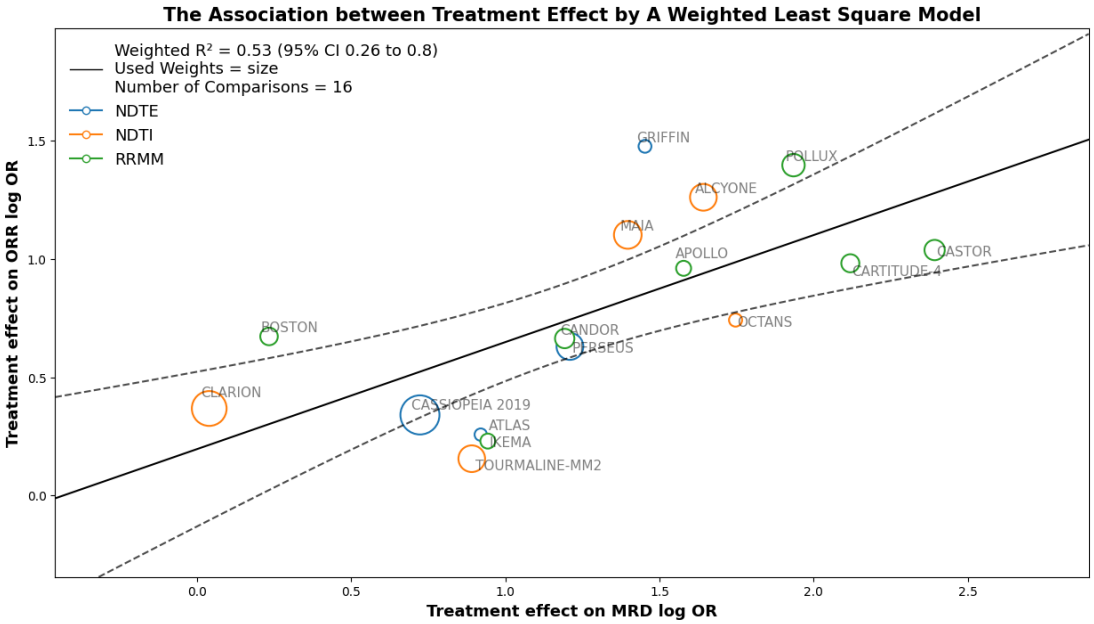
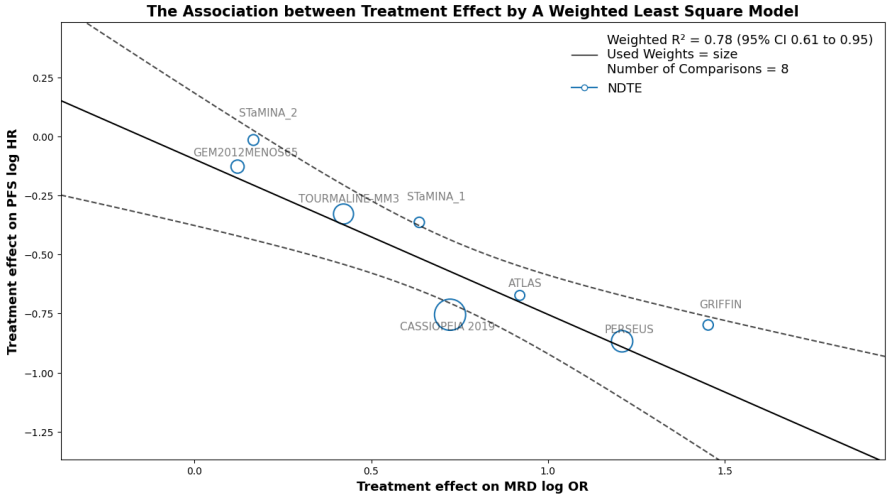
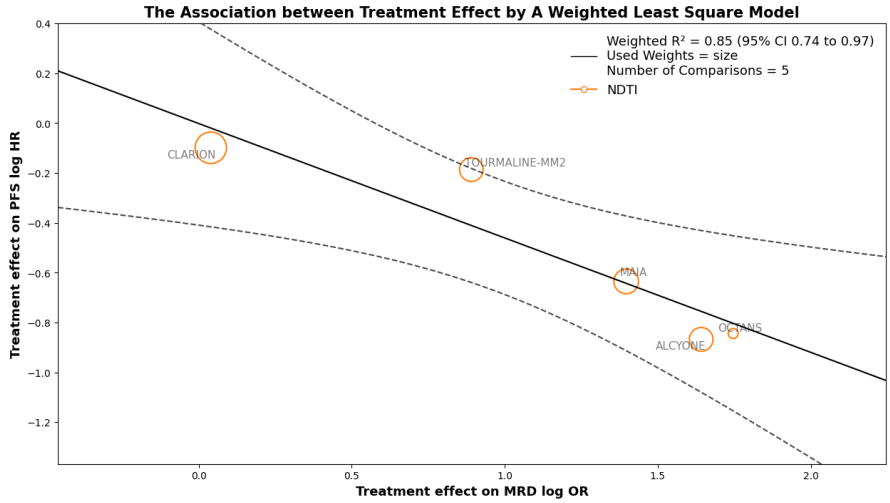


Figure S5. The weighted R^2_{trial} in the aggregated analysis of 16 clinical trials reporting ORR information. ORR OR and MRD- CR OR are natural log transformed. The weights equal sample sizes. The black solid lines are the fitted regression lines and the black dotted lines are 95% confidence bands.





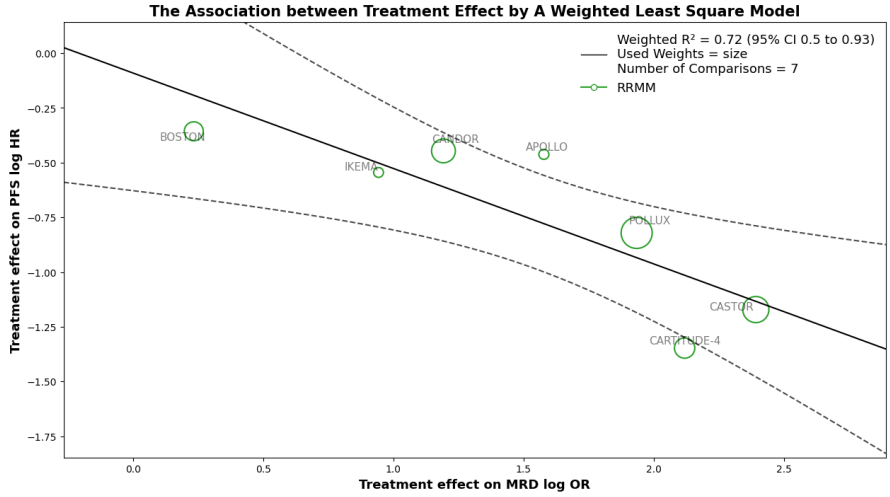


Figure S6. The weighted $R^2{}_{trial}$ in the subgroup analysis of 18 clinical trials reporting PFS information. PFS HR and MRD- CR odds ratio are natural log transformed. The weights equal sample sizes. The black solid lines are the fitted regression lines and the black dotted lines are 95% confidence bands.

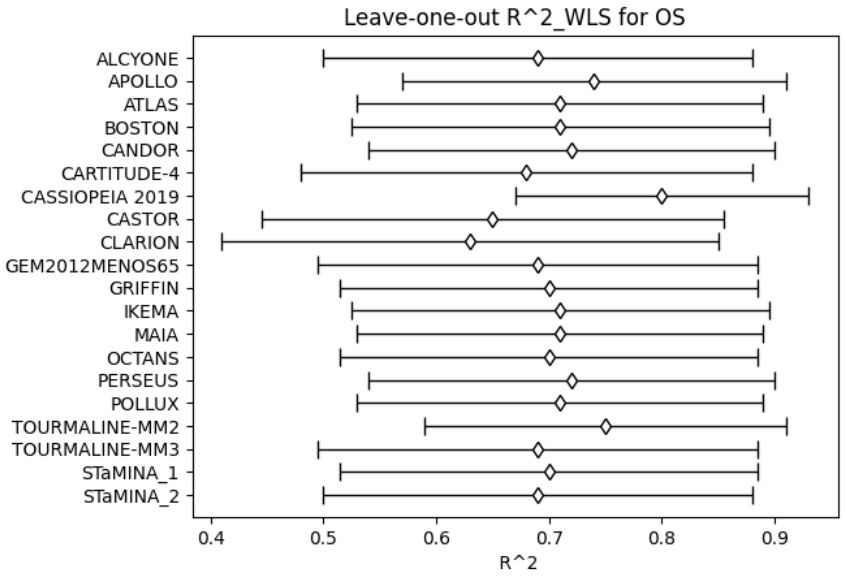
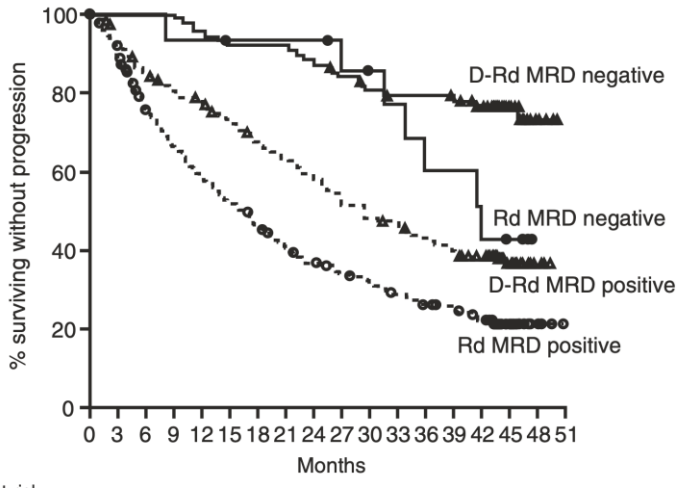
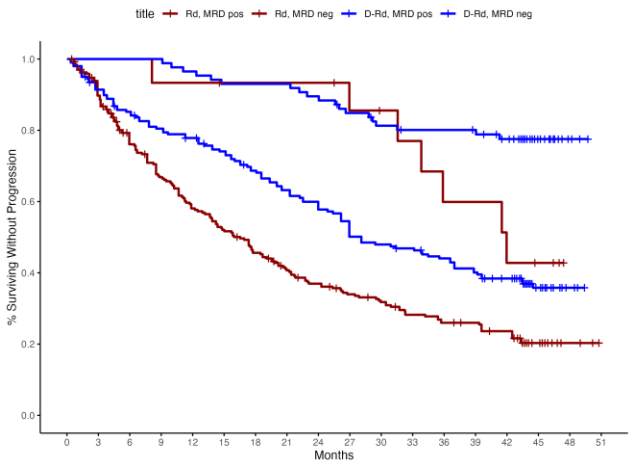


Figure S7. Leave-one out association plot. The y-axis are the names of excluded studies.

Additional results and sensitivity analysis for the second task



(b)



(a)

Figure S8. Comparison between reported KM curves and digitized KM curves for POLLUX study.