Oncologic Drugs Advisory Committee (ODAC) Meeting

April 12, 2024

Drug Topic: Use of Minimal Residual Disease (MRD) as an Endpoint in Multiple Myeloma Clinical Trials

Applicants:

Sylvester Comprehensive Cancer Center, University of Miami and,

i2TEAMM

Combined FDA and Applicants ODAC Briefing Document

DISCLAIMER STATEMENT

The attached package contains background information prepared by the Applicants and the Food and Drug Administration (FDA) for the panel members of the advisory committee. We have brought this topic to this Advisory Committee in order to gain the Committee's insights and opinions, and the background package may not include all issues relevant to the final regulatory recommendation and instead is intended to focus on issues identified by the Agency for discussion by the advisory committee. The FDA will not issue a final determination on the issues at hand until input from the advisory committee process has been considered and all reviews have been finalized. The final determination may be affected by issues not discussed at the advisory committee meeting.

Advisory Committee Briefing Materials: Available For Public Release

1 Table of Contents

2 Table of Tables

University of Miami Tables

i2TEAMM Tables

FDA Tables

3 Table of Figures

University of Miami Figures

i2TEAMM Figures

FDA Figures

Glossary

AA	Accelerated approval
ASCO	American Society of Clinical Oncology
ASCT	Autologous stem-cell transplantation
ASH	American Society of Hematology
ASO-qPCR	allele-specific oligonucleotide polymerase chain reaction
CI	confidence interval
CR	complete response
ECOG PS	Eastern Cooperative Oncology Group Performance Status
EFS	Event-free survival
FDA	Food and Drug Administration
FDASIA	Food and Drug Administration Safety and Innovation Act
FOIA	Freedom of Information Act
HR	hazard ratio
I2TEAMM	International Independent Team for Endpoint Approval of Myeloma Minimal Residual Disease
IMiD	Immunomodulatory drug
IMM	irreversible morbidity and mortality
IMWG	International Myeloma Working Group
IPD	Individual patient data
ISS	International Staging System
ITT	Intention-to-treat
MeSH	Medical Subject Heading
MFC	multiparameter-flow cytometry
MM	Multiple Myeloma
MRD	minimal residual disease
NCCN	National Comprehensive Cancer Network
NDMM	Newly diagnosed multiple myeloma
NDTE	Newly diagnosed transplant-eligible
NDTinE	Newly diagnosed transplant-ineligible
NGF	next-generation flow cytometry
NGS	next-generation sequencing
OR	Odds ratio
OS	overall survival
pCR	Pathological complete response
PFS	progression-free survival
PI	Proteasome inhibitor
RRMM	relapsed/refractory multiple myeloma
SAP	Statistical Analysis Plan
SCCC	Sylvester Comprehensive Cancer Center
sCR	stringent complete response
US	United States
WLS	weighted least squares

Representatives of FDA:	Nicole J Gormley, MD, Director
	Bindu Kanapuru, MD, Associate Director
	Rachel Ershler, MD, Clinical Reviewer
	Jing Zhang, PhD, Statistical Reviewer
	Jonathan Vallejo, PhD, Supervisory Mathematical Statistician
	Xiaofeng (Tina) Wang, MPH, Statistical Analyst
	Lisa Rodriguez, PhD, Deputy Director
	Theresa Carioti, MPH, Chief, Project Management Staff
	Denise Felluca, PharmD, MBA, Regulatory Project Manager
Representatives of Applicant:	Professor Ola Landgren, MD, PhD, Chief, Myeloma Division Sylvester Comprehensive Cancer Center, University of Miami
	Sean Devlin, PhD Associate Attending Biostatistician Memorial Sloan Kettering Cancer Center
	Representatives of i2TEAMM:
	Brian G.M. Durie, MD Cedars-Sinai Outpatient Cancer Center 8700 Beverly Blvd Los Angeles CA 90048
	Nikhil Munshi, MD Professor of Medicine Harvard Medical School VA Boston Healthcare System Kraft Family Chair and Director of Basic and Correlative Sciences Jerome Lipper Myeloma Center

Dana Farber Cancer Institute

Prof. Jesús San Miguel, MD, PhD
Professor of Hematology
Senior Consultant and Strategic Advisor
Clinica Universidad de Navarra, SPAIN

Bruno Paiva, PhD
Director of Flow Cytometry
Department of Hematology and Immunology
CIMA Laboratory Diagnostics
University of Navara, SPAIN

Qian Shi, PhD,
Professor of Biostatistics and Oncology
Department of Quantitative Health Sciences
Mayo Clinic, Rochester MN

1 Introduction

1.1 Purpose of the Meeting

The FDA is convening this ODAC to discuss the adequacy of available data to support the use of minimal residual disease (MRD) as an endpoint to support accelerated approval of new therapies for patients with multiple myeloma (MM). While overall response rate (ORR) has generally supported accelerated approval of MM therapies, recent advances, and improved understanding of the impact of minimal residual disease on long term outcomes has increased interest in evaluating MRD as an endpoint to support drug approval for patients with multiple myeloma.

No specific products will be presented or discussed at this ODAC. Rather, the committee will be asked to evaluate the totality of the data and to discuss the adequacy of the data to support the use of MRD as an endpoint to support accelerated approval in MM clinical trials.

2 Background and Context for the Meeting

2.1 Multiple Myeloma Clinical Setting

Multiple myeloma (MM) is a malignant plasma cell disorder characterized by clonal expansion of plasma cells in the bone marrow and over-production of monoclonal immunoglobulins leading to impaired hematopoiesis, bone destruction, and renal dysfunction.[1] MM is the second most common hematologic malignancy and accounts for approximately 17% of all hematologic malignancies and 1.8% of all new cancer diagnoses annually.[2] The median age at diagnosis is 69 years.

The treatment for MM depends on several disease-specific factors. In the newly diagnosed setting, treatment is often based on whether the patient is eligible for autologous stem cell transplant (ASCT), and in the relapsed or refractory setting treatment takes into account the type of prior therapies received and the response to those therapies.

Over the past ten years, 15 new drugs and greater than 20 new indications have approved for the treatment of patients with MM (FDA Appendix Table 1). This has resulted in substantial improvements in the outcomes for patients in all settings. For example, the median OS for patients with NDMM has increased from approximately 3.5 years in the late 1990s to over 10 years recently (2). However, despite the availability of multiple therapies, MM remains an incurable disease with a 5-year relative survival rate of 59.8%. There remains a need for new safe and effective therapies.

2.2 Approval Pathways

Both accelerated and regular (or traditional) approval pathways have supported the approval of new therapies and combinations in MM. Regular approval is based on an effect on a clinical

benefit endpoint or a validated surrogate. Although, overall survival is the ultimate clinical benefit endpoint, in MM, PFS has supported regular approval with an assessment of OS.

Accelerated Approval (AA) is an approval pathway designed to expedite the approval of new drugs and biologics intended to treat serious or life-threating diseases. To meet the requirements for AA, the new application or supplement must treat a serious or life-threatening disease, demonstrate an advantage over available therapy and is based on a surrogate endpoint that is reasonably likely to predict clinical benefit or an intermediate clinical endpoint that is reasonably likely to predict clinical benefit.[3] When an application is granted AA, the Applicant may be required to conduct a post-marketing study to confirm clinical benefit.

Traditionally for MM, the accepted endpoint to support AA has been Overall Response Rate (ORR) which includes patients with partial response (PR) or better as defined by the International Myeloma Working Group criteria (IMWG) supported by duration of response.[4]

2.3 Minimal Residual Disease (MRD) in MM

Recent clinical trials have demonstrated substantially improved progression-free survival (PFS) and overall survival (OS) results.[5] Like the improvements in PFS and OS, recent clinical trials in MM have demonstrated high ORRs in the newly diagnosed setting. High ORRs have also been reported in the relapsed or refractory setting including in recent trials supporting accelerated approvals in patients who have received 4 or more prior lines including the major classes of drugs (FDA Appendix Table 1). There is an interest in developing endpoints other than ORR to potentially expediate drug development.

Recent improvements in technologies to detect the presence of malignant cells at orders of magnitude below the limit of conventional ORR, has allowed an assessment of Minimal Residual Disease (MRD) in MM. MRD is a measure of tumor burden assessed in the bone marrow sample. MRD as a biomarker has multiple regulatory uses including for response assessments and as a prognostic marker. In multiple myeloma, the International Myeloma Working Group (IMWG) has established uniform response criteria for MRD for use in MM. MRD has been included as an exploratory endpoint and secondary endpoint to assess response to therapies in MM clinical trials and when the data is robust, MRD data has been included in the prescribing information.[6] Several studies and metanalyses have reported the prognostic value of MRD in MM and the achievement of MRD negativity has been associated with depth of clinical response and prolongation of PFS and OS. These analyses have increased interest in evaluating MRD as an endpoint to support approval in MM clinical trials.

2.4 Regulatory Considerations for MRD Endpoint Development

The FDA Guidance for Industry, *Hematologic Malignancies: Regulatory Considerations for Use of Minimal Residual Disease in Development of Drug and Biological Products for Treatment*, outlines general considerations for developing novel endpoints as surrogates including endpoints based on MRD.[7]

The methodology for assessing surrogacy for validation typically involves conducting a meta-analysis that includes patient level data from multiple trials. The goal of the meta-analysis is

typically to assess the strength of two associations: association at the individual level and association at the trial level (Section 10.3.1).

- **Individual-level** association assesses the impact of the earlier endpoint on the long-term endpoint within an individual patient.
- **Trial-level** association assesses the correlation between the treatment effect on the earlier clinical endpoint and the treatment effect on the long-term clinical endpoint.

Depending on the strength of the evidence supporting the ability of a marker to predict clinical benefit, the marker may be a surrogate endpoint that is known to predict clinical benefit (a validated surrogate endpoint that can be used for traditional approval), or a surrogate endpoint that is reasonably likely to predict the drug's intended clinical benefit (and that could therefore be used as a basis for AA). A similar approach, using a patient-level meta-analysis, was previously used to support validation of CR 30 as a validated surrogate for follicular lymphoma and for path CR as an endpoint reasonably likely to predict clinical benefit for neoadjuvant trials in early breast cancer.[8, 9]

This briefing document includes results of a meta-analysis conducted by two independent Applicants and the FDA using patient level data from multiple trials to evaluate MRD as a potential regulatory endpoint to support accelerated approval.

3 Efficacy

3.1 Summary of Clinical Trials Supporting Efficacy

3.1.1 University of Miami Applicant Position

Evaluation of Minimal Residual Disease as a Prognostic Factor for Progression-Free Survival

Several meta-analyses have evaluated the prognostic value of MRD for PFS or OS in clinical studies of treatments for multiple myeloma, and these meta-analyses have indicated that MRD negativity has strong prognostic value for clinical benefit as measured by PFS or OS.[2,9,23,31,32] To further examine the potential role of MRD as an early clinical endpoint reasonably likely to predict clinical benefit in patients with multiple myeloma, the Applicant conducted a new meta-analysis including recent studies enrolling patients with NDMM, using patient-level data and methodology similar to that employed in the meta-analysis published in 2020 by Avet-Loiseau and colleagues.[2] The new analysis, which is reported in this briefing document, incorporated FDA guidance and industry feedback for considerations for a meta-analysis to be used for validation of MRD as a clinical endpoint and potential basis for accelerated approval.[44] The meta-analysis was designed to include studies that enrolled patients with NDMM as well as studies that enrolled patients with RRMM, but data from RRMM studies were not sufficient to support meaningful interpretation. Thus, this report focuses on NDMM studies.

Meta-Analysis to Evaluate MRD as a Clinical Endpoint for Progression-Free Survival

The Applicant designed and conducted the study entitled "Evaluating Minimal Residual Disease as Intermediate Clinical Endpoint for Multiple Myeloma," a meta-analysis of patient-level data from multiple randomized, controlled, Phase 2 and Phase 3 confirmatory clinical trials. Meta-analysis results are included in this summary for studies that met the following criteria:

- Phase 2 or 3 randomized, controlled clinical trials that enrolled patients with NDMM (transplant -eligible or transplant-ineligible).
- Performed validated MRD assays by either MFC and/or NGS in accordance with guidelines from the FDA, National Cancer Institute, and IMWG, as well as institutional standards of care for the treatment of patients with multiple myeloma.[20,24,38,39,40]
  - NGS analyses were conducted using the FDA-cleared Adaptive Biotechnologies clonoSEQ 2.0 diagnostic test, which has a sensitivity of 10^{-5} or better.
  - MFC analyses were conducted using a 10-color method.[38,39] Note that the validation of this method using antibody (CD117, PC5.5, CD19, BV421, CD138 APC, CD56, PC7, CD45, APC-H7, and CD38 BV510) and fluorescent stain (κ fluorescein isothiocynate phycoerythrin) cocktails has been reported by Royston et al[39]; standardized flow cytometric instrument settings were reported by Kalina et al.[18] Minimal residual disease monitoring using this method has a sensitivity of 10^{-5} or better.[15]
- MRD negativity was specified as a primary, secondary, or exploratory endpoint in the clinical trial protocol.

- Had a median follow-up of at least 6 months following the end of the time chosen to be the *a priori* defined time point of 12 months after randomization for the assessment of MRD negativity, determined by a Kaplan-Meier estimate of the censoring distribution.

Maintenance studies, as well as studies in which the primary endpoint was safety, toxicity, quality of life, or feasibility, were excluded.

Studies for inclusion were identified by searching PubMed, clinical trial registries (including ClinicalTrials.gov, the ISRCTN registry, European Union Clinical Trial Register, and Australian New Zealand Clinical Trials Registry), cooperative groups' websites, research organization meeting websites, review of citations in publications including meta-analysis papers, and other sources such as personal communications. Searches were restricted to randomized controlled trials with human subjects for which the publication or other relevant documentation is written in English. For studies identified from all these sources, title and abstract review or full text review were performed to further assess the studies' eligibility for inclusion. The bibliographies of retained articles were examined to identify additional studies. The final clinical study lists were reviewed and approved by the study principal investigator. All literature search procedures were performed in accordance with the PRISMA guidelines for meta-analysis.[36]

Clinical trial sponsors were contacted to provide secure transfer of information addressing patient and disease characteristics at diagnosis and randomization, number of lines of prior therapy, prior autologous stem cell transplant, MRD evaluation technique, and follow-up data on disease and outcomes.

Objectives

The objectives and endpoints of the reported analyses were included in a statistical analysis plan (SAP) jointly agreed upon by the principal investigator, the study statistician, collaborating agencies, and the FDA prior to the meta-analysis. The primary objectives were as follows:

- To evaluate whether MRD negativity while in a CR at an *a priori* defined time point (jointly agreed upon by the principal investigator, the study statistician, collaborating agencies, and the FDA before the meta-analysis to be 12 months with a window of ± 3 months for an MRD assessment to have taken place) is a reasonably likely endpoint for clinical benefit as measured by PFS in newly diagnosed, transplant-eligible patients with multiple myeloma.
- To evaluate whether MRD negativity while in a CR at an *a priori* defined time point (jointly agreed upon by the principal investigator, the study statistician, collaborating agencies, and the FDA before the meta-analysis to be 12 months with a window of ± 3 months for an MRD assessment to have taken place) is a reasonably likely endpoint for clinical benefit as measured by PFS in newly diagnosed, transplant-ineligible patients with multiple myeloma.

Key secondary objectives were as follows:

- To evaluate whether MRD negativity at an *a priori* defined time point (jointly agreed upon by the principal investigator, the study statistician, collaborating agencies, and FDA before the meta-analysis to be 12 months with a window of ± 3 months for an MRD assessment to have taken place) is a reasonably likely endpoint for clinical benefit as measured by PFS in patients with NDMM, regardless of transplant eligibility (i.e., in the combined population of transplant-eligible and transplant-ineligible NDMM).

- To evaluate whether attainment of MRD negativity at least once is reasonably likely to predict clinical benefit as measured by PFS.
- To evaluate whether sustained MRD (e.g., attainment of MRD negativity twice in succession, with ≥6 months between assessments) is reasonably likely to predict clinical benefit as measured by PFS.
- To evaluate whether MRD negativity is reasonably likely to predict clinical benefit as measured by OS.

Before any analysis, for each time point and corresponding window under consideration for the timing of the MRD assessment to be used for the primary definition of the endpoint of MRD negativity (3 ± 2 months, 6 ± 2 months, 9 ± 3 months, and 12 ± 3 months), the number of patients who achieved MRD negativity and the number of patients with a missing MRD status (e.g., because no MRD assessments occurred within the window) were reported for each clinical trial available for inclusion in the meta-analysis. In April of 2020, FDA and the Applicant, including the principal investigator and the study statistician, jointly selected one time point and one corresponding window of 12 ± 3 months to be the *a priori* defined time point and window to be used for the primary definition of the endpoint based on MRD for the meta-analysis in the setting of NDMM. This window was clinically relevant and provided the most complete dataset to analyze MRD negativity.

Statistical Methods

The complete analysis plan was detailed in the SAP (see Appendix), which was finalized prior to analysis. Briefly, a correlation approach was used to evaluate MRD negativity as reasonably likely to predict clinical benefit in clinical trials of treatments for multiple myeloma by assessing individual-level association and trial-level association of MRD negativity with PFS. This approach assessed the following 2 items:

- The endpoint (MRD negativity) is prognostic for clinical benefit (PFS).
- A treatment effect on the endpoint (MRD negativity) in a clinical trial is reasonably likely to predict the treatment effect on PFS.

The primary analysis used a correlation approach.[5,6,7,8] For individual-level associations, a bivariate Plackett copula model was used to estimate the association, on individual patient data, of the MRD negativity endpoint with PFS, while controlling for trial-specific and treatment-specific effects on MRD and on PFS.[8] The bivariate Plackett copula model quantifies the association of MRD negativity with PFS by providing an odds ratio of two odds: a) the odds of having a PFS event at or after a certain time (e.g., 4 years) for patients who achieve the MRD negativity endpoint, and b) the odds of having a PFS event at or after the same time (4 years) for patients who do not achieve the MRD negativity endpoint. For example, if the probability that a patient in a study who has MRD-negative CR at 12 months has a PFS of 4 years or longer is 75% (odds of 3:1), and if the probability that a patient in a study who does not have MRD-negative CR at 12 months has a PFS of 4 years or longer is 33% (odds of 0.5:1), then the odds ratio is 6 (=3/0.5). Because of the controlled nature of RCTs, an odds ratio of ≥ 3 that is statistically significant (i.e., its confidence interval excludes 1) should be interpreted as an analysis result that the MRD negativity endpoint is highly prognostic for the traditional clinical

endpoint (PFS). The same approach was used to evaluate the prognostic value of the MRD negativity endpoint for OS.

Two approaches were used to assess trial-level associations. The first approach used a weighted linear regression model, across clinical trials, where the explanatory variable is the estimated treatment effect on the MRD negativity endpoint in a clinical trial (log odds ratio comparing MRD negativity rate in the experimental arm to MRD negativity rate in the control arm in the trial) and the dependent variable is the treatment effect on PFS (log hazard ratio from Cox proportional hazards regression comparing the same experimental arm to the same control arm in the same trial). The regression was weighted by the inverse variances of the log odds ratio for MRD negativity. A sensitivity analysis was additionally weighted by the respective sample size of each trial. R^2 was used to describe the proportion of variance explained by regression (i.e., the extent to which the treatment effect on the MRD negativity endpoint in a trial can be used quantitatively to numerically predict what the treatment effect will be on PFS in a trial, using a straight-line relationship). This measure of R^2 for PFS was considered the primary outcome for this analysis. The second approach assessed the trial-level association based on the two-stage copula model.[8] Under this approach, the bivariate Placket copula model estimated the treatment effects on MRD negativity rate and PFS within each trial simultaneously. The treatment effect was based on marginal models equivalent to logistic regression and Weibull model for MRD rate and PFS, respectively. The same two approaches were repeated with OS replacing PFS.

An individual-level association can be demonstrated in a single trial, but an assessment of trial-level association is informed by a meta-analysis of more than one randomized trial. The assessment of individual-level association was quantified by an odds ratio comparing PFS of patients who achieve MRD negativity to those who did not. The evaluation of trial-level association was informed by an R^2 regarding using treatment effect of the early endpoint to predict the treatment effect on a long-term endpoint; the trial-level association was also evaluated by a planned concordance of significance analysis, which uses a statistically significant treatment effect on MRD negativity to predict whether the treatment effect on PFS will be statistically significant.

Studies Included in the Meta-Analysis

Studies providing data for the initial meta-analysis are listed in . These studies are described in greater detail in and . These Phase 3 studies were conducted internationally (apart from EudraCT 2010-019173-16), and they did not include data for practice setting type (e.g., urban or rural, academic or community cancer center).

University of Miami Table 1: Studies Providing Data for the Meta-Analysis

Registry #	Study/Sponsor
Transplant-Eligible Newly Diagnosed Multiple Myeloma	
EudraCT 2010-019173-16	Randomised Phase III Trial for Previously Untreated Multiple Myeloma to Evaluate Two Regimens of Bortezomib Based Induction Therapy and Lenalidomide Consolidation Followed by Lenalidomide Maintenance Treatment (MM5)/University Hospital Heidelberg[16]
NCT02874742	Phase 2, Randomized, Open-label Study Comparing Daratumumab, Lenalidomide, Bortezomib, and Dexamethasone (D-RVd) versus Lenalidomide, Bortezomib, and Dexamethasone (RVd) in Subjects with Newly Diagnosed Multiple Myeloma Eligible for High-dose Chemotherapy and Autologous Stem Cell Transplantation (GRIFFIN-MMY2004)/Janssen R&D[43]
NCT02541383	Study of Daratumumab in Combination with Bortezomib (VELCADE), Thalidomide, and Dexamethasone (VTD) in the First Line Treatment of Transplant Eligible Subjects with Newly Diagnosed Multiple Myeloma (CASSIOPEIA-MMY3006)/Intergroupe Francophone du Myelome (IFM)[29,30]
Transplant-Ineligible Newly Diagnosed Multiple Myeloma	
NCT01818752	A Randomized, Open-label Phase 3 Study of Carfilzomib, Melphalan, and Prednisone versus Bortezomib, Melphalan, and Prednisone in Transplant-ineligible Patients with Newly Diagnosed Multiple Myeloma (CLARION)/Amgen[11]
NCT01850524	A Phase 3, Randomized, Double-blind, Multicenter Study Comparing Oral MLN9708 Plus Lenalidomide and Dexamethasone (Rd) versus Placebo Plus Lenalidomide and Dexamethasone (Rd) in Adult Patients with Newly Diagnosed Multiple Myeloma (TOURMALINE-MM2)/Takeda[14]
NCT02195479	A Phase 3, Randomized, Controlled, Open-label Study of VELCADE (Bortezomib) Melphalan-Prednisone (VMP) Compared to Daratumumab in Combination with VMP (D-VMP) in Subjects with Previously Untreated Multiple Myeloma Who are Ineligible for High-dose Therapy (ALCYONE-MMY3007) Janssen R&D[27,28]
NCT02252172	A Phase 3 Study Comparing Daratumumab, Lenalidomide, and Dexamethasone (DRd) vs Lenalidomide and Dexamethasone (Rd) in Subjects with Previously Untreated Multiple Myeloma Who are Ineligible for High Dose Therapy (MAIA-MMY3008)/Janssen R&D[12,13]
NCT03217812	A Phase 3, Multicenter, Randomized, Controlled, Open-label Study of VELCADE (Bortezomib) Melphalan-Prednisone (VMP) Compared to Daratumumab in Combination with VMP (D-VMP), in Subjects With Previously Untreated Multiple Myeloma Who Are Ineligible for High-Dose Therapy (Asia Pacific Region-OCTANS-MMY3011)/Janssen R&D

University of Miami Table 2: PFS and OS Data for Studies Included in the Meta-Analysis

Registry #	Treatment: n (%)	N	Median Follow-up, months (95% CI)	Median PFS, months (95% CI)	Median OS, months (95% CI)
Transplant Eligible					
EudraCT 2010-019173-16	A1*: 149 (24.8%) A2*: 151 (25.1%) B1*: 150 (25%) B2*: 151 (25.1%)	601	57.6 (56.4-59.2)	40.5 (36.5-43.2)	Not reached
NCT02874742	D-RVd: 120 (53.8%) RVd: 103 (46.2%)	223	27.2 (26.7-28)	Not reached	Not reached
NCT02541383	D-VTd: 543 (50%) VTd: 542 (50%)	1085	18.8 (18.3-19.4)	Not reached	Not reached
Transplant Ineligible					
NCT01818752	KMP: 428 (50.2%) VMP: 425 (49.8%)	853	22.3 (22.1-22.6)	22.2 (21-24.2)	Not reached
NCT01850524	IRd: 351 (49.8%) Rd: 354 (50.2%)	705	58.1 (57.2-59.3)	27.9 (23.9-35.8)	Not reached
NCT02195479	D-VMP: 350 (49.6%) VMP: 356 (50.4%)	706	40.1 (39.6-40.5)	24 (21.6-27.4)	Not reached
NCT02252172	DRd: 368 (49.9%) Rd: 369 (50.1%)	737	64.5 (64-65.4)	44.8 (40.9-52.4)	73.7 (69.7-NR)
NCT03217812	D-VMP: 146 (66.4%) VMP: 74 (33.6%)	220	23.8 (22.8-24.6)	28.2 (24.4-NR)	41.6 (41.6-NR)

CI=confidence interval; DRd=daratumumab, lenalidomide, and dexamethasone; D-RVd=daratumumab, lenalidomide, bortezomib, and dexamethasone; D-VTd=daratumumab, bortezomib, thalidomide, and dexamethasone; D-VMP=daratumumab, bortezomib, melphalan, and prednisone; IRd=isatuximab, lenalidomide, and dexamethasone; KMP=carfilzomib, melphalan, and prednisone; NA=not available; OS=overall survival; PFS=progression-free survival; Rd=lenalidomide and dexamethasone; RVd=lenalidomide, bortezomib, and dexamethasone; VMP=bortezomib, melphalan, and prednisone; VTd=bortezomib, thalidomide, and dexamethasone

*A1=bortezomib, doxorubicin, and dexamethasone, high-dose melphalan, autologous blood stem cell transplantation and lenalidomide consolidation followed by lenalidomide maintenance therapy for 2 years

B1=bortezomib, doxorubicin, and dexamethasone, high-dose melphalan, autologous blood stem cell transplantation and lenalidomide consolidation followed by lenalidomide maintenance until achievement of complete response

A2=bortezomib, cyclophosphamide, and dexamethasone, high-dose melphalan, autologous blood stem cell transplantation and lenalidomide consolidation followed by lenalidomide maintenance therapy for 2 years

B2=bortezomib, cyclophosphamide, and dexamethasone, high-dose melphalan, autologous blood stem cell transplantation and lenalidomide consolidation followed by lenalidomide maintenance until achievement of complete response

University of Miami Table 3: MRD Data for Studies Included in the Meta-Analysis

Registry #	N	MRD Assay	MRD Detection	Best Response <CR n (%)	CR After Window n (%)	CR in Window			
						MRD Negative n (%)	MRD Positive n (%)	MRD Outside Window n (%)	No MRD Info n (%)
Transplant Eligible									
EudraCT 2010-019173-16	601	MFC	10^{-4}	392 (65)	55 (9)	37 (6)	15 (2)	57 (9)	45 (7)
NCT02874742	223	NGS	10^{-5}	70 (31)	13 (6)	26 (12)	16 (7)	76 (34)	22 (10)
NCT02541383	1085	NGS	10^{-5}	578 (54)	38 (4)	193 (18)	58 (5)	82 (8)	127 (12)
Transplant Ineligible									
NCT01818752	853	MFC	10^{-4}	660 (77)	1 (0)	45 (5)	38 (4)	30 (4)	79 (9)
NCT01850524	705	MFC	10^{-4}	560 (79)	44 (6)	21 (3)	16 (2)	38 (5)	26 (4)
NCT02195479	706	NGS	10^{-5}	465 (66)	18 (3)	80 (11)	82 (12)	28 (4)	33 (5)
NCT02252172	737	NGS	10^{-5}	444 (60)	99 (13)	63 (9)	86 (12)	27 (4)	18 (2)
NCT03217812	220	MFC	10^{-5}	142 (65)	2 (1)	48 (22)	18 (8)	9 (4)	1 (0)

CR=complete response; MFC=multiparameter-flow cytometry; MRD=minimal residual disease; NGS=next-generation sequencing

Analysis Sets and Demographic and Baseline Characteristics

A total of 8 studies contributed to the all-NDMM population (i.e., transplant eligible and transplant ineligible combined). One study provided 2 randomized comparisons.[16]

A total of 5130 subjects were randomized. The number of subjects randomized per study ranged from 220 to 1085 subjects with a median sample size of 306 subjects. Demographic and baseline disease characteristics of these subjects are provided in University of Miami Table 4. Of the nine 2-arm comparisons available for analysis for NDMM, 8 comparisons with 4907 subjects fulfilled the data criteria for the 12-month MRD negative endpoint and were included in the primary individual-level and trial-level surrogacy analysis for all NDMM.

University of Miami Table 4: Demographic Data for Studies Included in the Meta-Analysis

Registry #	Treatment: n (%)	Age (y) Median (range)	Sex (F/M) n (%))	Months from Diagnosis	ISS Stage
Transplant Eligible					
EudraCT 2010-019173-16	A1*: 149 (24.8%) A2*: 151 (25.1%) B1*: 150 (25%) B2*: 151 (25.1%)	59 (32-70)	249 (41.4%)/352 (58.6%)	NA	I: 230 (38.3%) II: 206 (34.3%) III: 165 (27.5%)
NCT02874742	D-RVd: 120 (53.8%) RVd: 103 (46.2%)	60 (29-70)	97 (43.5%)/126 (56.5%)	0.8 (0-60.8) NA: 2	NA
NCT02541383	D-VTd: 543 (50%) VTd: 542 (50%)	58 (22-65)	450 (41.5%)/635 (58.5%)	NA	NA
Transplant Ineligible					
NCT01818752	KMP: 428 (50.2%) VMP: 425 (49.8%)	72 (42-91)	NA	1 (0.1-46.2)	NA
NCT01850524	IRd: 351 (49.8%) Rd: 354 (50.2%)	74 (48-90)	351 (49.8%)/354 (50.2%)	1.1 (0.2-52.8) NA: 20	I or II: 582 (82.6%) III: 123 (17.4%)
NCT02195479	D-VMP: 350 (49.6%) VMP: 356 (50.4%)	69 (57-84)	379 (53.7%)/327 (46.3%)	0.8 (0.1-24.8)	NA
NCT02252172	DRd: 368 (49.9%) Rd: 369 (50.1%)	NA	353 (47.9%)/384 (52.1%)	0.9 (0-14.5)	NA
NCT03217812	D-VMP: 146 (66.4%) VMP: 74 (33.6%)	71 (40-93)	89 (40.5%)/131 (59.5%)	NA	NA

DRd=daratumumab, lenalidomide, and dexamethasone; D-RVd=daratumumab, lenalidomide, bortezomib, and dexamethasone; D-VMP daratumumab, bortezomib, melphalan, and prednisone; D-VTd=daratumumab, bortezomib, thalidomide, and dexamethasone; F=Female; IRd= isatuximab, lenalidomide, and dexamethasone; KMP=carfilzomib, melphalan, and prednisone; M=male; NA=not available; Rd=lenalidomide and dexamethasone; RVd=lenalidomide, bortezomib, and dexamethasone; VMP=bortezomib, melphalan, and prednisone; VTd=bortezomib, thalidomide, and dexamethasone; y=years

*See Table 3 footnotes for descriptions of A1, A2, B1, and B2.

MRD Methods and Calibration Data

The MRD methodology and the calibration information are provided by study in University of Miami Table 5. For earlier studies, these data were not available. For the studies that did have these data, the calibration success ranged from 91.2% to 93.5%.

University of Miami Table 5: MRD Assay and Calibration Information

Study	MRD Method	Calibration Success / Notes
EudraCT 2010-019173-16	MFC	Not applicable
NCT02874742	NGS	91.2%
NCT02541383	NGS	Data not available
NCT01818752	MFC	Not applicable
NCT01850524	MFC	Not applicable
NCT02195479	NGS	92.0%
NCT02252172	NGS	93.5%

MFC = multiparameter-flow cytometry; MRD = minimal residual disease; NDMM=newly diagnosed multiple myeloma; NGS = next-generation sequencing

Note: the calibration success rates for the Janssen studies are only for the subgroup of patients with a confirmed complete response.

3.1.2 The i2TEAMM Applicant's Position:

In March 2020, an exhaustive search was performed of the Medline database maintained by the US National Library of Medicine for publications on randomized studies conducted in MM, in order to identify clinical trials that would be eligible for the planned retrospective meta-analysis. A search using the strategy of the MeSH terms "multiple myeloma" AND "neoplasm, residual" AND the nonMeSH terms "MRD", "myeloma", AND "minimal residual disease" yielded 143 references. These publications were then individually examined using the following parameters to exclude studies for further considerations:

- Non-randomized studies (lack of statistical relevance and potential bias)
- Studies with a total sample size < 100 subjects (to ensure sufficient precision to estimate the treatment effect on both surrogate and the true endpoint within the study)
- Single-center studies (potential patient selection bias and lack of robustness)
- Studies published before 2006 (in 2006, the International Myeloma Working Group [IMWG] established the uniform response criteria for MM[22])
- Any evidence that MRD testing with 10^{-4} or higher sensitivity level was never performed

After initial screening, 29 trials were identified with potential data for MRD endpoint development and surrogacy evaluation. Hence, the owners of these trials were contacted with the aim of acquiring the trial individual patient-level data (IPD). The required IPD include clinical/pathological baseline factors, tumor response data, MRD data, progression and survival data, randomization and treatment data, and other relevant data fields. *A total of 20 trials, comprising a total of 12,926 patients, transferred IPD as of the date of finalizing the Statistical Analysis Plan (SAP v3.1 Jan 28, 2021; Finaly results report was produced in May 2021).* The details of these trials are included in the i2TEAMM Appendix Table 1, i2TEAMM Appendix Table 2, and i2TEAMM Appendix Table 3 per disease types: newly diagnosed transplant eligible (NDTE), newly diagnosed transplant in-eligible (NDTinE) and relapsed/refractory (RR) MM, respectively. Note, one study (Myeloma IX) includes both NDTE and NDTinE populations. The reasons that other trials' IPD were not available for data sharing at the time of analysis are: the original trial primary endpoints are maturing (5 studies), MRD testing is in progress (1 study), and no clear commitments from study owners (3 studies).

In order to estimate treatment effects on both the proposed surrogate endpoint and the true long-term endpoint, "two-arm comparisons" were defined as the comparison between an experimental arm and a control arm within a given clinical trial. Multiple two-arm comparisons were pre-defined, per SAP v3.1, when there are more than two experimental treatment groups and/or more than one randomization pre-planned in the original protocol of the study. In such cases, each experimental arm was compared to the common control arm in order to provide a treatment effect for each comparison and therefore maximize use of the available data to reduce bias of the analysis and improve precision. A total of 34 two-arm comparisons (21 in NDTE, 9 in NDTinE, and 4 in RR population) were formed (see details in i2TEAMM Table 1). i2TEAMM Table 2 shows the patients' demographics and baseline characteristics by experimental and control groups, pooling the two-arm comparisons per population. These

baseline characteristics were well balanced between control and experimental arms across all trials. Overall, the median follow-up for PFS across all studies per the inverse Kaplan-Meier method was 48.9 months providing mature long-term data to inform the surrogacy evaluation.

For a two-arm comparison to be included in the final trial-level surrogacy analyses, *it was pre-defined to require ≥ 80% of patients whose MRD surrogate endpoint status can be determined with either "success" or "failure" status, and with at least (≥) 50 patients. SAP v3.1 also defines, for each disease population, the trial-level surrogacy was only performed when there are at least 10 two-arm comparisons with sufficient MRD endpoint data.* The principal surrogate endpoint candidate to be evaluated was defined as the proportion of subjects *who achieved a complete response (CR) with at least one MRD negativity status at 9 months (±3 months)* after randomization (9m-MRDneg). This could be measured by any MFC or NGS technology as long as it was validated, and included data at thresholds of 10^{-4}, 10^{-5} or 10^{-6}. If a trial tested MRD at multiple sensitivity levels, the data at 10^{-5} was preferred for the meta-analysis if meeting the data sufficiency criteria. The secondary MRD endpoint was defined as the proportion of subjects who achieved a CR with at least one MRD negativity status at 12 months (±3 months) after randomization (12m-MRDneg). The determinations of MRD endpoint timepoint(s) and sensitivity level, as well as only incorporating CR (not including very good partial response [VGPR]), were based on clinical discussions within i2TEAMM consortium, interactions with FDA, and data availabilities. Detailed justifications were included in SAP v3.1 and briefing documentations supporting multiple Sponsor-FDA communications.

As shown in i2TEAMM Appendix Table 1, i2TEAMM Appendix Table 2, and i2TEAMM Appendix Table 3, the limitations of collected data include:

- The original purposes of MRD testing varies across trials. MRD endpoints were considered exploratory endpoints in about half of the studies. The timing and number of timepoints where MRD was measured varied across trials. In some trials, these were further varied across patients. Because of this, in some of the 2-arm comparisons, many patients did not have an MRD measurement within the 9 months +/- 3 months window when at least one CR was recorded to define MRD negativities. Per the SAP, these 2-arm comparisons that were lacking in sufficient MRD collection within 9 +/- 3 months window were not included in the analysis.
- Within the NDTE and NDTinE population, there were different MRD testing methods applied and, accordingly, variable sensitivity levels to determine MRD negativity. As agreed by the FDA in the process of SAP development and discussions, it is critical that the MRD surrogate endpoint is defined using consistent sensitivity level across patients and timepoints *within each two-arm comparison*. For the trial-level association analysis, the two-arm comparisons with MRD endpoint determined by different sensitivity levels may be pooled.
- In RR population, although all trials used Next Generation Sequencing to test MRD, there were only four trials with IPD available.

i2TEAMM Table 1: Two-arm Comparisons and % of Patients with 9 Months MRD Negativities that Can be Determined Based on IPD Collected

Population	Study Name	Treatment Comparison Description	% of Patients with 9 month MRDneg status that can be determined based on IPD provided			Selection when multiple sensitivity levels meet the inclusion criteria
			Sensitivity 10-4	Sensitivity 10-5	Sensitivity 10-6	
NDTE	GMMG MM5	PAd (Control) vs VCD (Experimental)	86.3%	0.0%	0.0%	MRD data at 10-4
NDTE	GMMG MM5	Lenalidomide for 2 years (Control) vs Lenalidomide until CR or 2 years (Experimental)	17.0%	0.0%	0.0%	No comparisons could be selected
NDTE	BMT CTN 0702	Auto-SCT+HDM (Control) vs Auto-SCT+HDM + Auto-SCT+HDM (Experimental)	53.4%	53.9%	73.3%	No comparisons could be selected
NDTE	BMT CTN 0702	Auto-SCT+HDM (Control) vs Auto-SCT+HDM -> RVD (Experimental)	50.8%	50.8%	72.4%	No comparisons could be selected
NDTE	BMT CTN 0702	Auto-SCT+HDM + Auto-SCT+HDM (Control) vs Auto-SCT+HDM -> RVD (Experimental)	51.1%	51.6%	81.3%	MRD data at 10-6
NDTE	EMN02/HO95 MM	VMP (Control) vs HDM (Experimental)	26.3%	19.1%	0.0%	No comparisons could be selected
NDTE	EMN02/HO95 MM	No Consolidation (Control) vs VRD (Experimental)	23.1%	17.8%	0.0%	No comparisons could be selected
NDTE	C16019	Placebo (Control) vs Ixazomib (Experimental)	0.0%	72.9%	0.0%	No comparisons could be selected
NDTE	GEM2012MENOS65	MEL-200 (Control) vs BUMEL (Experimental)	0.0%	98.2%	94.5%	MRD data at 10-5
NDTE	GEM2005MENOS65	TD (Control) vs VMBCP-VBAD/Velcade (Experimental)	96.1%	0.0%	0.0%	MRD data at 10-4
NDTE	GEM2005MENOS65	TD (Control) vs TD/Velcade (Experimental)	94.6%	0.0%	0.0%	MRD data at 10-4
NDTE	GEM2005MENOS65	TD/Velcade (Control) vs VMBCP-VBAD/Velcade (Experimental)	94.6%	0.0%	0.0%	MRD data at 10-4
NDTE	IFM DFCI2009	RVD (Control) vs High-Dose RVD -> Auto-SCT (Experimental)	76.4%	76.4%	76.4%	No comparisons could be selected
NDTE	Myeloma IX	Clodronic Acid + CVAD (Control) vs Zoledronic Acid + CVAD (Experimental)	83.8%	0.0%	0.0%	MRD data at 10-4
NDTE	Myeloma IX	Clodronic Acid + CTD (Control) vs Zoledronic Acid + CTD (Experimental)	78.9%	0.0%	0.0%	No comparisons could be selected
NDTE	Myeloma IX	No Maintenance (Control) vs Thalidomide (Experimental)	48.7%	0.0%	0.0%	No comparisons could be selected

NDTE	FORTE	CRd + Mobilization + CRd (Control) vs CCyd + Mobilization + ASCT + CCyd (Experimental)	0.0%	83.2%	64.6%	MRD data at 10-5
NDTE	FORTE	CRd + Mobilization + CRd (Control) vs CRd + Mobilization + ASCT + CRd (Experimental)	0.0%	81.9%	63.2%	MRD data at 10-5
NDTE	FORTE	CRd + Mobilization + ASCT + CRd (Control) vs CCyd + Mobilization + ASCT + CCyd (Experimental)	0.0%	86.8%	69.1%	MRD data at 10-5
NDTE	FORTE	Lenalidomide (Control) vs Lenalidomide/Carfilzomib (Experimental)	0.0%	70.1%	26.1%	No comparisons could be selected
NDTE	GRIFFIN	RVd + Mobilization + ASCT + RVd (Control) vs D-RVd + Mobilization + ASCT + D-RVd (Experimental)	85.3%	85.3%	85.3%	MRD data at 10-5
NDTinE	CLARION	VMP (Control) vs CMP (Experimental)	0.0%	80.6%	0.0%	MRD data at 10-5
NDTinE	C16014	LenDex (Control) vs Ixazomib/LenDex (Experimental)	94.0%	0.0%	0.0%	MRD data at 10-4
NDTinE	ALCYONE	VMP (Control) vs D-VMP (Experimental)	84.1%	84.1%	84.1%	MRD data at 10-5
NDTinE	GEM2005MAS65	VMP (Control) vs VTP (Experimental)	75.7%	0.0%	0.0%	
NDTinE	GEM2010MAS65	Sequential MPV -> Rd (Control) vs Alternating MPV + Rd (Experimental)	0.0%	87.6%	0.0%	MRD data at 10-5
NDTinE	MAIA	Rd (Control) vs DRd (Experimental)	93.7%	93.7%	93.7%	MRD data at 10-5
NDTinE	Myeloma IX	Clodronic Acid + MP (Control) vs Zoledronic Acid + MP (Experimental)	78.9%	0.0%	0.0%	No comparisons could be selected
NDTinE	Myeloma IX	Clodronic Acid + CTDa (Control) vs Zoledronic Acid + CTDa (Experimental)	84.5%	0.0%	0.0%	MRD data at 10-4
NDTinE	Myeloma IX	No Maintenance (Control) vs Thalidomide (Experimental)	74.3%	0.0%	0.0%	No comparisons could be selected
RR	POLLUX	Rd (Control) vs DRd (Experimental)	91.5%	91.5%	91.5%	MRD data at 10-5
RR	CASTOR	Vd (Control) vs DVd (Experimental)	91.9%	91.9%	91.9%	MRD data at 10-5
RR	ICARIA	Pd (Control) vs IPd (Experimental)	96.1%	96.1%	95.8%	MRD data at 10-5
RR	CANDOR	Carfilzomib/Dexamethasone (Control) vs Carfilzomib/Dexamethasone/Daratumumab (Experimental)	89.9%	89.5%	88.2%	MRD data at 10-5

Note: All two-arm comparisons had > 50 patients. The 9m MRDneg status at different sensitivity level was determined based on the actual data transferred.

i2TEAMM Table 2: Patient Demographics and Baseline Characteristics Pooling Two-arm Comparisons by Disease Populations

Population	NDTE		NDTinE		RR	
Treatment Type	Control (N=4501)	Experimental (N=4835)	Control (N=2364)	Experimental (N=2358)	Control (N=825)	Experimental (N=996)
Age						
N	4501	4835	2364	2358	822	996
Mean (SD)	56.8 (7.35)	56.6 (7.54)	72.8 (5.89)	72.8 (5.88)	64.4 (9.30)	63.9 (9.56)
Median	58.0	58.0	73.0	73.0	65.0	65.0
Range	28.0, 78.0	24.2, 74.0	43.0, 91.0	40.0, 93.0	33.0, 87.0	29.0, 89.0
Age Group, n (%)						
18-64	3985 (88.5%)	4284 (88.6%)	97 (4.1%)	102 (4.3%)	402 (48.9%)	479 (48.1%)
65+	516 (11.5%)	551 (11.4%)	2267 (95.9%)	2256 (95.7%)	420 (51.1%)	517 (51.9%)
Missing	0	0	0	0	3	0
Gender, n (%)						
Female	1927 (42.8%)	1960 (40.5%)	1141 (48.3%)	1158 (49.1%)	357 (43.3%)	423 (42.5%)
Male	2574 (57.2%)	2875 (59.5%)	1223 (51.7%)	1200 (50.9%)	468 (56.7%)	573 (57.5%)
Race, n (%)						
American Indian or Alaska Native	0 (0.0%)	0 (0.0%)	3 (0.1%)	3 (0.1%)	0 (0.0%)	0 (0.0%)
Asian/Native Hawaiian/Pacific Islander	60 (2.1%)	78 (2.6%)	222 (9.9%)	218 (9.9%)	35 (11.7%)	67 (14.8%)
Black	70 (2.5%)	69 (2.3%)	37 (1.7%)	37 (1.7%)	2 (0.7%)	7 (1.5%)
Multiracial	2 (0.1%)	2 (0.1%)	2 (0.1%)	3 (0.1%)	0 (0.0%)	0 (0.0%)
White	2678 (95.2%)	2832 (94.6%)	1953 (87.4%)	1937 (87.6%)	249 (83.3%)	361 (79.5%)
Other	4 (0.1%)	13 (0.4%)	17 (0.8%)	14 (0.6%)	13 (4.3%)	19 (4.2%)
Missing	1687	1841	130	146	526	542
ECOG Performance Score, n (%)						
0	1912 (42.6%)	2144 (44.5%)	634 (26.9%)	614 (26.1%)	412 (49.9%)	431 (43.4%)
1	1923 (42.8%)	2015 (41.9%)	1200 (50.9%)	1199 (51.0%)	357 (43.3%)	507 (51.1%)
2	514 (11.5%)	532 (11.1%)	456 (19.3%)	469 (19.9%)	56 (6.8%)	55 (5.5%)
3	130 (2.9%)	116 (2.4%)	63 (2.7%)	58 (2.5%)	0 (0.0%)	0 (0.0%)
4	9 (0.2%)	7 (0.1%)	5 (0.2%)	11 (0.5%)	0 (0.0%)	0 (0.0%)
Missing	13	21	6	7	0	3
Myeloma Type, n (%)						
IgA	839 (18.8%)	973 (20.2%)	550 (23.5%)	557 (23.8%)	180 (21.8%)	216 (21.7%)
IgD	66 (1.5%)	70 (1.5%)	21 (0.9%)	37 (1.6%)	11 (1.3%)	13 (1.3%)
IgE	29 (0.6%)	47 (1.0%)	2 (0.1%)	2 (0.1%)	0 (0.0%)	0 (0.0%)
IgG	2754 (61.6%)	2895 (60.2%)	1486 (63.5%)	1428 (60.9%)	495 (60.0%)	580 (58.2%)
IgM	11 (0.2%)	22 (0.5%)	6 (0.3%)	6 (0.3%)	1 (0.1%)	5 (0.5%)
Bence Jones	383 (8.6%)	413 (8.6%)	28 (1.2%)	27 (1.2%)	0 (0.0%)	0 (0.0%)
Non-secretory	45 (1.0%)	67 (1.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Bence Jones/Non-secretory	227 (5.1%)	214 (4.4%)	219 (9.4%)	255 (10.9%)	125 (15.2%)	171 (17.2%)
Biclonal	13 (0.3%)	16 (0.3%)	25 (1.1%)	28 (1.2%)	3 (0.4%)	3 (0.3%)

Population	NDTE		NDTinE		RR	
Treatment Type	Control (N=4501)	Experimental (N=4835)	Control (N=2364)	Experimental (N=2358)	Control (N=825)	Experimental (N=996)
Unknown	105 (2.3%)	94 (2.0%)	5 (0.2%)	4 (0.2%)	10 (1.2%)	8 (0.8%)
Missing	29	24	22	14	0	0
Light Chain Type, n (%)						
Kappa	2822 (63.9%)	3037 (64.0%)	1477 (63.7%)	1474 (63.6%)	520 (63.0%)	638 (64.1%)
Lambda	1485 (33.6%)	1609 (33.9%)	825 (35.6%)	826 (35.6%)	286 (34.7%)	335 (33.7%)
Kappa/Lambda	8 (0.2%)	10 (0.2%)	7 (0.3%)	6 (0.3%)	3 (0.4%)	2 (0.2%)
Unknown/ Not Detected	101 (2.3%)	90 (1.9%)	9 (0.4%)	12 (0.5%)	16 (1.9%)	20 (2.0%)
Missing	85	89	46	40	0	1
% of Bone Marrow Plasma Cell Involvement						
N	3968	4310	2276	2291	786	919
Mean (SD)	41.4 (27.85)	41.6 (28.83)	37.5 (25.48)	37.6 (25.19)	31.0 (25.62)	31.5 (25.92)
Median	38.0	40.0	31.5	32.0	23.7	24.0
Range	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0
Indicator for Plasma Cell Bone Marrow infiltration (calculated), n (%)						
No	519 (13.0%)	626 (14.5%)	229 (10.1%)	225 (9.8%)	160 (20.4%)	196 (21.3%)
Yes	3462 (87.0%)	3691 (85.5%)	2047 (89.9%)	2066 (90.2%)	626 (79.6%)	723 (78.7%)
Missing	520	518	88	67	39	77
ISS (calculated), n (%)						
Stage I	1204 (29.5%)	1262 (28.4%)	551 (23.8%)	526 (22.9%)	364 (44.3%)	450 (45.4%)
Stage II	1200 (29.4%)	1265 (28.5%)	1007 (43.6%)	992 (43.1%)	283 (34.4%)	336 (33.9%)
Stage III	1684 (41.2%)	1916 (43.1%)	754 (32.6%)	781 (34.0%)	175 (21.3%)	206 (20.8%)
Missing	413	392	52	59	3	4
Risk Group (Neben et. al.), n (%)						
Favorable Prognosis	591 (26.3%)	654 (26.1%)	318 (22.3%)	311 (22.0%)	208 (36.1%)	252 (36.3%)
Intermediate Prognosis	1288 (57.4%)	1449 (57.9%)	920 (64.6%)	913 (64.7%)	290 (50.3%)	359 (51.7%)
Poor Prognosis	365 (16.3%)	399 (15.9%)	186 (13.1%)	187 (13.3%)	78 (13.5%)	83 (12.0%)
Missing	2257	2333	940	947	249	302

3.1.3 The FDA's Position

FDA's advice in developing the SAP was to ensure the goals of the analyses were aligned with the potential use of such endpoints in a regulatory setting. FDA generally agrees with the meta-analyses conducted by both Applicants. The meta-analyses conducted by the two applicants used similar methodology. This methodology has been used previously to understand the level of surrogacy of earlier clinical endpoints with long-term clinical endpoints.[8, 9]

The odds ratio utilized at the individual-level is based on copula modeling. When the lower bound of a 95% CI for the odds ratio excludes 1, this is often interpreted as evidence of a statistically significant individual-association.

The i2TEAMM pre-specified a threshold for the trial-level associations. This criterion was: "If either of R^2_{WLS} or R^2_{Copula} is at least 0.8 with lower bound of the 95% confidence interval greater than 0.6, and neither estimate is lower than 0.7." The stated purpose of this criteria was to "qualify a Validated Surrogate Endpoint". Note that there is no formal FDA guidance which specifies such criteria.

For each Applicant, trials were included based on the criteria as specified in the respective SAP. Due to the nature of the methods implemented, trials were only included if there was >0% MRD negative CR rate in at least one of the arms. This is because the methods rely on estimation of odds, which are 0 or undefined when one of the response rates is 0%.

For all the analyses, the time points of 9 months and 12 months include a window of +/- 3 months and the primary analysis of MRD was assessed in patients who achieved a complete response (CR).

FDA Appendix Table 3 in Section 7.3.3 details the differences in the analyses submitted by the two Applicants. The major difference in approach was that the i2TEAMM SAP specified that meta-analyses for MRD- CR would only be conducted in clinical settings for which there were at least 10 two-arm comparisons. Previous research has suggested that when the number of two-arm comparisons is 6 or less, the estimates from such analyses may be poor.[11] In addition to such pre-specified analyses, the i2TEAMM reports additional exploratory analyses below.

An additional minor difference is the handling of missing MRD data. The i2TEAMM elected to remove patients with missing MRD data, while the University of Miami team retained these patients in the analyses and assigned their MRD status to be "MRD positive". For this reason, many of the i2TEAMM analyses utilized fewer patients from a given trial than the University of Miami analyses. As can be observed in the individual-level results presented in Section 3.2.3, this difference in approach impacts the estimation of individual-level associations. However, the trial-level association estimates are similar under either approach.

FDA performed additional meta-analysis based on all the data submitted by both Applicants, with duplicate trials removed and patients with missing MRD status were imputed as non-

responders. Refer to Section 7.3.4 for further details regarding derivation of this population. FDAs analyses used the same methodology as the two Applicants.

The purpose of these pooled analyses was to determine whether utilization of all available evidence would impact the results or conclusions. In these pooled analyses, the analysis population was all randomized patients available among the two Applicant analysis data sets. Patients with missing MRD status were assigned as "MRD positive". Additional details regarding the dataset construction are given in Section 7.3.4.

The baseline characteristics of the FDA pooled population (FDA Appendix Table 4) are similar to those reported in University of Miami Table 4 and i2TEAMM Table 2. Section 7.3.4 presents summary statistics for response and time-to-event endpoints in the FDA pooled population, respectively.

FDA also explored "MRD negative CR at any time" in the RR setting using data submitted to the FDA in support of an NDA/BLA application or in an IND submission. Multiple Myeloma trials typically did not mandate a landmark time for MRD assessment. Additionally, this assessment for MRD negative CR is consistent with the measurement for ORR that is estimated as best ORR rather than at a specific timepoint. The analyses for MRD negative CR at any time utilized all randomized patients as the analysis population and assigned patients with missing MRD status as "MRD positive".

If the surrogate endpoint is found to have significant association based on both trial-level and individual-level association, a subsequent analysis is typically conducted to quantify the predictive ability of the surrogate on the ultimate endpoint. This is typically accomplished via estimation of the surrogate threshold effect (STE), which is defined as, "the smallest treatment effect on the surrogate necessary to be observed to predict a treatment effect on the true endpoint that is statistically significantly different from zero". This measure has several limitations which are outlined in Section 7.3.2.

While the overall methodology proposed by the two Applicants is generally reasonable, there are some limitations which should be considered when interpreting the results.

- The treatment types represented are largely small molecules and monoclonal antibodies. None of the trials contained in these meta-analyses include evaluation of a chimeric antigen receptor (CAR) T cell therapy. It is uncertain whether any correlations observed would translate to new types of therapy. Recently published data suggests that MRD is correlated with progression-free survival after treatment with idecabtagene vicleucel.[10]
- The majority of the trials included did not assess 9-month or 12-month MRD rate as a key secondary endpoint. Thus, it is unclear what level of missingness one might expect in a future trial where such an endpoint is a primary or key secondary endpoint.
- Assays and assay sensitivities varied across trials.
- The trials vary in design and follow-up duration, resulting in many differences, including differences in assessment times, follow-up duration, and regions of enrollment.

- The patient populations vary considerably from trial to trial. However, only a few baseline characteristics were routinely collected across trials (e.g., age and sex; FDA Appendix Table 4. Given this and the small number of trials in any particular setting, there is limited ability to assess the impact of subgroups.
- The trials were included based on pre-specified inclusion criteria. While these inclusion criteria are reasonable, the results may not be fully representative of the relationship between MRD negative CR and PFS/OS if the trials not included in this analysis are qualitatively different than the trials analyzed.

Estimation approaches for the associations above are described in more detail in Section 7.3.2.

3.2 Efficacy Summary

3.2.1 University of Miami Applicant's Position

Individual-Level Associations

MRD and PFS

In the all NDMM population, the global odds ratio (4.72) demonstrated a strong individual-level association between 12-month MRD negativity and PFS; that is, being MRD negative at 12 months is prognostic of better long-term outcomes (University of Miami Table 6).

University of Miami Table 6: 12-Month MRD negative vs PFS (Individual-Level Results) – Primary Analysis

Population	Total Sample Size	Copula Global Odds Ratio (95% CI)
All NDMM	4907	4.72 (3.53-5.90)
Transplant-eligible NDMM	1686	2.45 (1.40-3.51)
Transplant-ineligible NDMM	3221	6.15 (4.27-8.03)

CI=confidence interval; NDMM=newly diagnosed multiple myeloma; PFS=progression-free survival

University of Miami Table 7 provides the results of sensitivity analyses of other groupings of clinical trial results. The individual-level association is also strong in these sensitivity analyses, further supporting the value of MRD as a prognostic marker for better long-term outcomes.

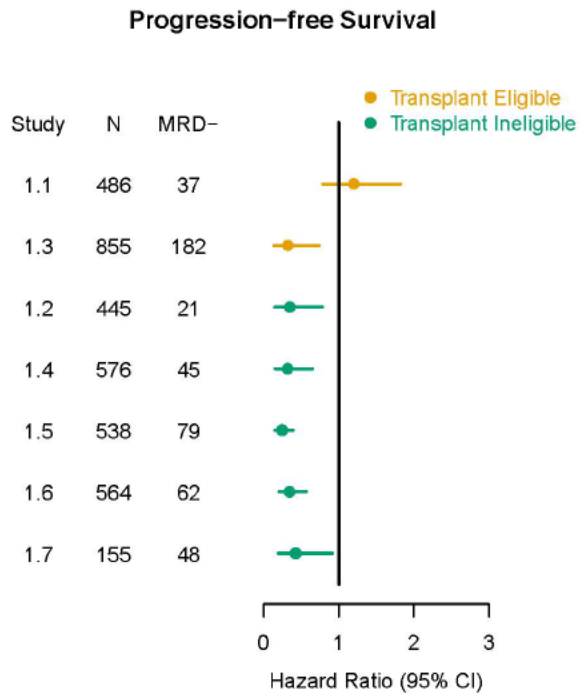
University of Miami Table 7: 12-Month MRD negative vs PFS (Individual-Level Results) – Sensitivity Analysis

Sensitivity Analysis	Total Sample Size	Copula Global Odds Ratio (95% CI)
Adding study 2.1 to all NDMM	5130	4.68 (3.51-5.85)
Adding study 2.1 to transplant-eligible NDMM	1909	2.46 (1.42-3.49)
All NDMM without study 1.1A and 1.1B	4306	6.08 (4.33-7.83)
All NDMM without study 1.2	4202	4.56 (3.37-5.76)
All NDMM without study 1.3	3822	4.64 (3.41-5.86)
All NDMM without study 1.4	4054	4.43 (3.25-5.61)
All NDMM without study 1.5	4201	4.10 (2.94-5.24)
All NDMM without study 1.6	3822	4.73 (3.41-6.06)
All NDMM without study 1.7	4687	4.71 (3.47-5.96)

CI=confidence interval; NDMM=newly diagnosed multiple myeloma; PFS=progression-free survival

The prognostic value of MRD is further illustrated in University of Miami Figure 1, which shows that individuals who achieved MRD negativity had a reduced risk of a PFS event compared in MRD-positive individuals in most studies. This analysis investigated the association between MRD-negativity at 12 months and PFS by creating a post-randomization landmark that included only patients alive, progression-free, and under follow-up at 12 months in the analysis.

University of Miami Figure 1: Forest Plot of Time to Progression-free Survival by MRD status for Patients Who Were Alive and Progression-free at 12 Months, by Study



MRD=minimal residual disease

MRD and OS

In the all-NDMM population, the global odds ratio (4.02) demonstrated a strong individual-level association between 12-month MRD negativity and OS; that is, being MRD negative at 12 months is prognostic of better long-term OS. Similar results were observed in the transplant-eligible and transplant-ineligible subgroups (University of Miami Table 8).

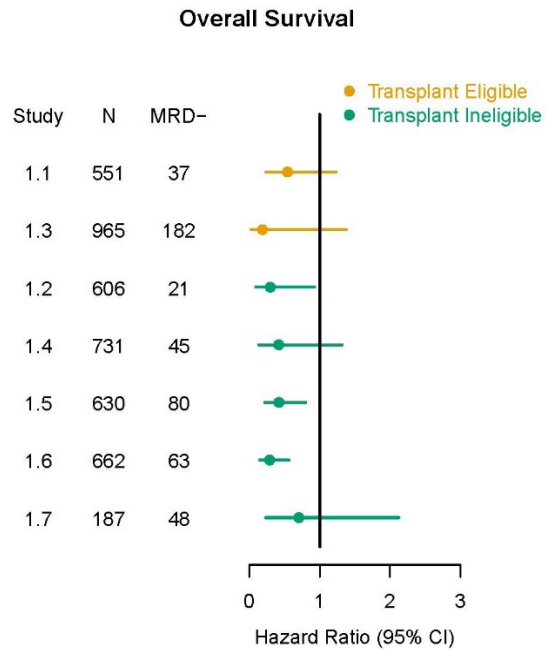
University of Miami Table 8: 12-Month MRD-Negative vs OS (Individual-Level Results) – Primary Analysis

Population	Total Sample Size	Copula Global Odds Ratio (95% CI)
All NDMM	4907	4.02 (2.57-5.46)
Transplant-eligible NDMM	1686	3.78 (0.78-6.78)
Transplant ineligible NDMM	3221	4.08 (2.44-5.72)

CI=confidence interval; NDMM=newly diagnosed multiple myeloma; OS=overall survival

The prognostic value of MRD is further illustrated in University of Miami Figure 2, which shows that individuals who achieved MRD negativity had a reduced risk of death compared to MRD-positive individuals.

University of Miami Figure 2: Forest Plot of Overall Survival by MRD Status for Patients Who Were Alive at 12 Months, by Study



MRD=minimal residual disease

Trial-Level Associations

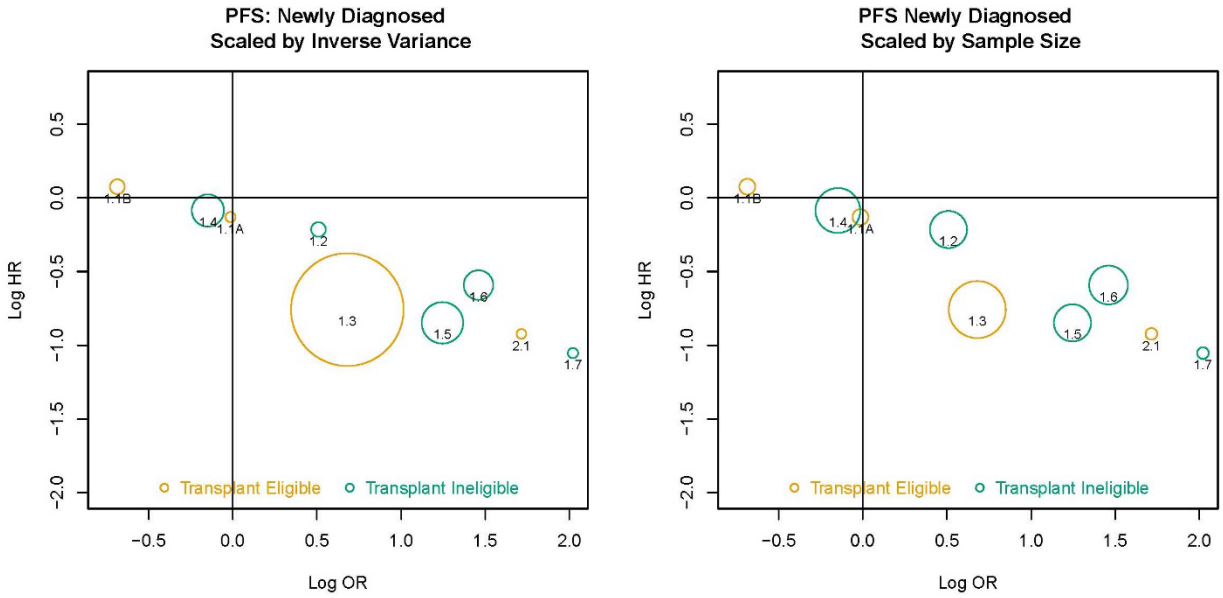
MRD and PFS

Of the nine 2-arm comparisons available for analysis, 8 comparisons with 4907 subjects fulfilled the data criteria for the 12-month MRD negativity endpoint and were included in the trial-level surrogacy analysis for all NDMM (i.e., transplant-eligible and transplant-ineligible combined):

- Overall, the median follow-up for PFS was 29 months (IQR: 18.9-57.7).
- Based on the weighted linear regression model, $R^2_{\text{weighted least squares (WLS; inverse variance)}}$=0.67 (95% CI: 0.43, 0.91) and $R^2_{\text{WLS(sample size)}}$=0.72 (95% CI: 0.51, 0.93) for PFS (University of Miami Figure 3).
- Based on the 2-stage copula model, R^2_{copula}=0.84 (95% CI: 0.64, >0.99) for PFS.
- The available data did not allow for calculation of R^2 using either model for the transplant-eligible subgroup of the NDMM population. Therefore, calculation of R^2 for transplant-eligible NDMM was not performed.
- For the transplant-ineligible subgroup of the NDMM population, based on the weighted linear regression model (University of Miami Figure 3), $R^2_{\text{WLS (inverse variance)}}$=0.83 (95% CI: 0.71,

0.96) and $R^2_{WLS\ (sample\ size)}$=0.84 (95% CI: 0.72, 0.97) and based on the 2-stage copula model R^2_{copula}=0.85 (95% CI: 0.62, >0.99) for PFS.

University of Miami Figure 3: Correlation Between Treatment Effect on 12-month MRD Negativity and Treatment Effect on PFS (All NDMM Population)



HR=hazard ratio; OR=odds ratio

Study 2.1 was not included in the primary analysis due to >20% of patients being assigned a value of missing for the primary endpoint definition based on MRD. This study was included in sensitivity analyses (University of Miami Table 9).

University of Miami Table 9: Trial-Level R^2 Estimates PFS – Sensitivity Analyses (All NDMM Population)

Sensitivity Analysis	Total Sample Size	Total Follow-up, months (IQR)	R^2_{copula} (95% CI)	$R^2_{WLS\ (inverse\ variance)}$ (95% CI)	$R^2_{WLS\ (sample\ size)}$ (95% CI)
Adding study 2.1* to all NDMM	5130	28.6 (19,56.4)	0.82 (0.62,1.03)	0.68 (0.45,0.91)	0.74 (0.54,0.94)
All NDMM without study 1.1A and 1.1B	4306	26.9 (18,53.4)	0.72 (0.35,>0.99)	0.53 (0.22,0.83)	0.63 (0.38,0.89)
All NDMM without study 1.2	4202	28.2 (18.8,55.6)	0.87 (0.69,>0.99)	0.69 (0.47,0.92)	0.76 (0.57,0.95)
All NDMM without study 1.3	3822	43.8 (24.3,61.1)	0.91 (0.79,>0.99)	0.88 (0.78,0.98)	0.87 (0.77,0.98)
All NDMM without study 1.4	4054	38 (18.9,59.5)	0.82 (0.58,>0.99)	0.56 (0.27,0.86)	0.63 (0.37,0.89)
All NDMM without study 1.5	4201	27.4 (18.1,58.9)	0.84 (0.63,>0.99)	0.62 (0.36,0.89)	0.68 (0.45,0.92)
All NDMM without study 1.6	4170	26.5 (18,44.4)	0.88 (0.72,>0.99)	0.82 (0.67,0.96)	0.82 (0.67,0.97)
All NDMM without study 1.7	4687	30.5 (18.9,58.3)	0.78 (0.47,>0.99)	0.64 (0.39,0.9)	0.68 (0.44,0.92)

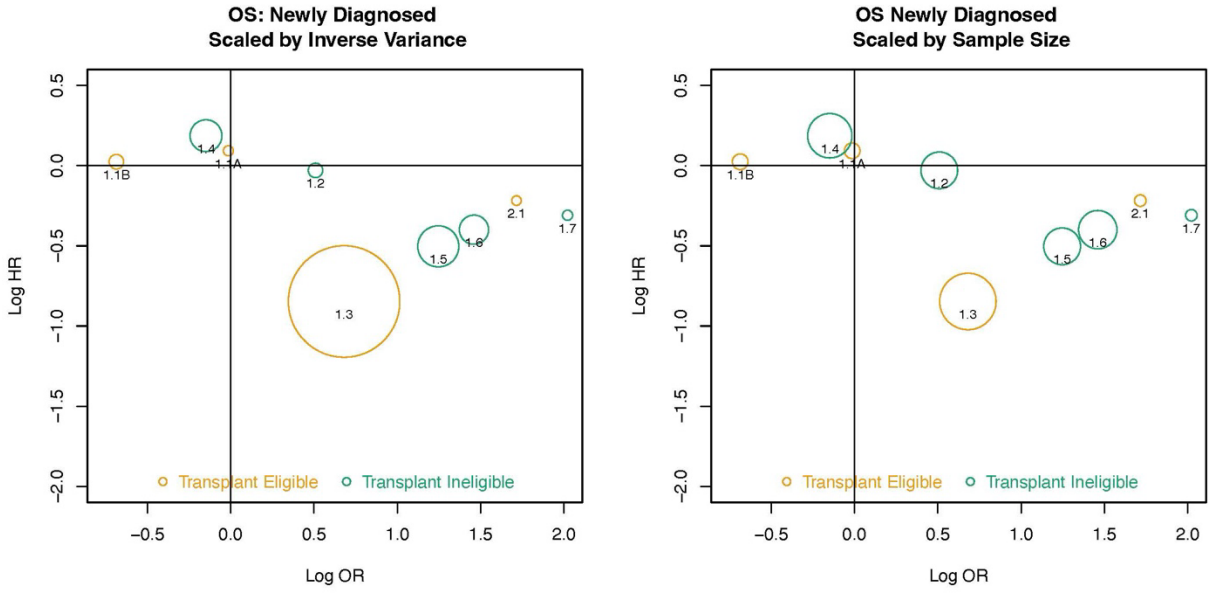
*The study 2.1 was not included in the primary analysis due to >20% of patients being assigned a value of missing for the primary endpoint definition based on MRD.

MRD and OS

Median OS was not reached for most of the studies (University of Miami Table 2). Additionally, non-randomized use of subsequent therapies likely confounded analyses of OS by attenuating the effect of randomized study treatment on OS. Of the nine 2-arm comparisons available for analysis, 8 comparisons with 4907 subjects fulfilled the data criteria for the 12-month MRD negative endpoint and were included in the primary trial-level surrogacy analysis for all NDMM (i.e., transplant-eligible and transplant-ineligible combined):

- Overall, the median follow-up for OS was 37.2 months (95% CI: 21.5,59.2).
- All studies had at least 1 arm that received FDA-approved drugs for NDMM.
- Based on the weighted linear regression model $R^2_{WLS\ (inverse\ variance)}$=0.21 (95% CI: <0.01-0.53) and $R^2_{WLS\ (sample\ size)}$=0.33 (95% CI: <0.01-0.67) for OS (University of Miami Figure 4).
- Based on the 2-stage copula model, R^2_{copula}=0.32 (95% CI: <0.01-0.86) for OS.
- The available data did not allow for calculation of R^2 using either model for the transplant-eligible NDMM population.
- For transplant-ineligible NDMM population, based on the weighted linear regression model (University of Miami Figure 4), $R^2_{WLS\ (inverse\ variance)}$=0.79 (95% CI: 0.63-0.95) and $R^2_{WLS\ (sample\ size)}$=0.83 (95% CI: 0.69-0.96) and based on the 2-stage copula model R^2_{copula}=0.63 (95% CI: 0.12-0.>0.99) for OS.

University of Miami Figure 4: Correlation Between Treatment Effect on 12-month MRD Negativity and Treatment Effect on OS (All NDMM Population)



HR=hazard ratio; OR=odds ratio; OS=overall survival

Concordance of Significance

As shown in University of Miami Table 10, among all NDMM studies (i.e., transplant eligible and transplant ineligible combined):

- The treatment effect on MRD negativity was statistically significant in 4 of the 8 treatment comparisons in the primary analysis. All 4 (100%) clinical trials with a statistically significant treatment effect on MRD negativity had a statistically significant treatment effect on PFS ($p < 0.05$).
- The treatment effect on MRD negativity was not statistically significant in 4 of the 8 treatment comparisons. Of the 4 treatment comparisons that did not have a statistically significant treatment effect on MRD negativity, 1 (25%) had a statistically significant treatment effect on PFS.

University of Miami Table 10: Concordance of Significance for MRD with PFS and OS for Trials Included in the Primary Analysis (All NDMM Population)

Study	Treatment Effect on MRD[a] (2-sided test) p-value	Treatment Effect on PFS[a] (2-sided test) p-value	Treatment Effect on OS[a] (2-sided test) p-value
Transplant-eligible NDMM			
1.1A	0.98	0.385	0.686
1.1B	0.131	0.605	0.901
1.3	<0.001	<0.001	0.008
Transplant-ineligible NDMM			
1.4	0.629	0.399	0.232
1.2	0.264	0.038	0.806
1.5	<0.001	<0.001	<0.001
1.6	<0.001	<0.001	<0.001
1.7	<0.001	<0.001	0.377

NDMM=newly diagnosed multiple myeloma; OS=overall survival; PFS=progression-free survival

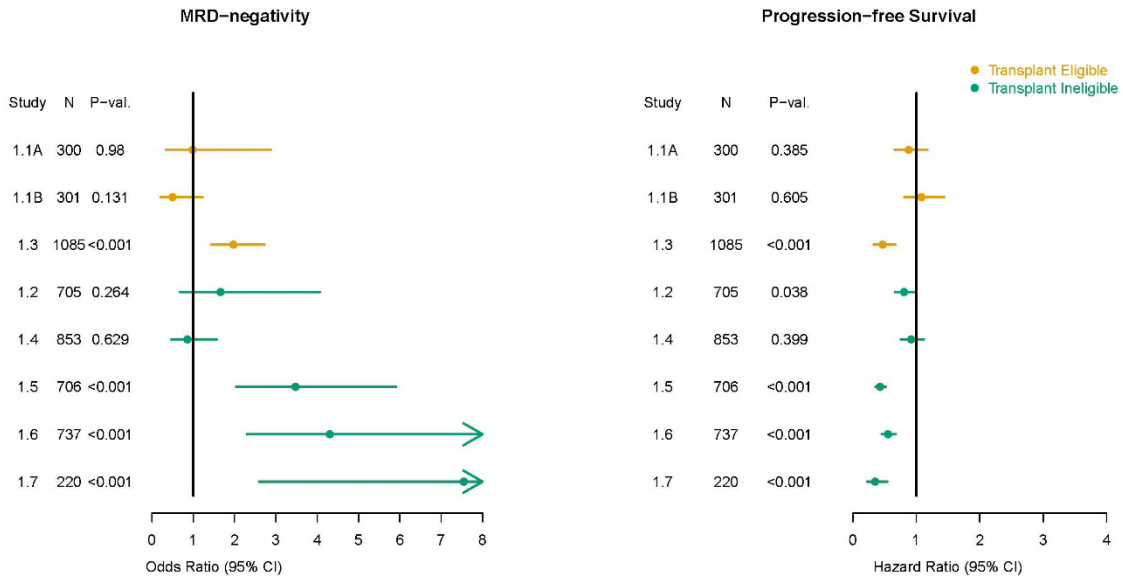
[a]Does not include stratification factors used in randomization

Forest Plots of Treatment Effect

MRD and PFS

Side-by-side forest plots of the treatment effect on MRD and the treatment effect on PFS are shown in University of Miami Figure 5. This comparison shows that studies with strong treatment effects on MRD also had strong treatment effects on PFS.

University of Miami Figure 5: Forest Plot of Treatment Effect on MRD and Time to Progression-free Survival



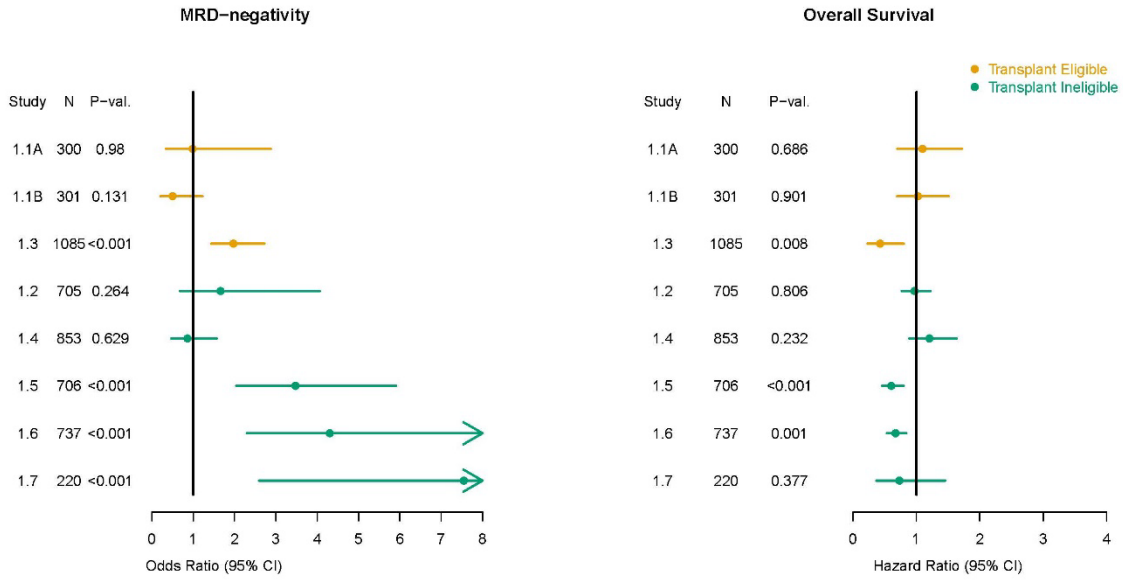
MRD and OS

As shown in University of Miami Table 10, among all NDMM (i.e., transplant eligible and transplant ineligible combined) studies:

- The treatment effect on MRD negativity was statistically significant in 4 of the 8 treatment comparisons. Of the 4 clinical trials with a statistically significant treatment effect on MRD negativity, 3 (75%) also had a statistically significant treatment effect on OS.
- The treatment effect on MRD negativity was not statistically significant in 4 of the 8 treatment comparisons. Of the 4 treatment comparisons that did not have a statistically significant treatment effect on MRD negativity, no trial had a statistically significant treatment effect on OS.

Side-by-side forest plots of the treatment effect on MRD and the treatment effect on OS are shown in University of Miami Figure 6. This comparison also shows that studies with a strong treatment effect on MRD tended to show a treatment effect on OS in the same direction.

University of Miami Figure 6: Forest Plot of Treatment Effect on MRD and Time to Overall Survival



Evaluation of Any MRD

The correlation of the attainment of MRD negativity at least once with clinical benefit (PFS or OS) was evaluated. The results of these analyses (described below) align with the primary analysis and support MRD as an endpoint reasonably likely to predict clinical benefit.

Attainment of MRD at Least Once During the Study

The clinical benefit (PFS and OS) achieved following attainment of MRD negativity at any time during the study (at least once) was evaluated. For this analysis data from 9 NDMM studies were included (1.1A, 1.1B, 1.2, 2.1, 1.3, 1.4, 1.5, 1.6, and 1.7). With attainment of MRD negativity (at least once), the R^2 values for PFS were as follows:

- Based on the weighted linear regression model, $R^2_{WLS\ (inverse\ variance)}$=0.54 (95% CI: 0.23,0.84)
- Based on the 2-stage copula model, R^2_{copula}=0.76 (95% CI: 0.49-0.99)
- With attainment of MRD negativity (at least once), the R^2 values for OS were as follows:
- Based on the weighted linear regression model, $R^2_{WLS\ (inverse\ variance)}$=0.07 (95% CI: <0.01-0.28)
- Based on the 2-stage copula model, R^2_{copula}=0.11 (95% CI: <0.01-0.49)

3.2.2 The i2Team Applicant's Position

Individual-level association:

Individual-level surrogacy is measuring the association between 9m/12m-MRDneg status and PFS at the individual patient level, i.e., the prognostic value of 9m/12m-MRDneg status. The analytic unit is each patient, instead of 2-arm comparison. Since the accuracy of this estimate is dependent on the number of patients and NOT the number of trial-level analytic units, the individual-level association analysis was conducted, per SAP, if there were *at least 200* patients with sufficient MRD endpoint data in each setting.

When there was sufficient data, the individual-patient level surrogacy was evaluated within each disease population separately and based on a uniform sensitivity level, by global odds ratio (OR) from the same Plackett copula model used to estimate trial-level surrogacy.[23] This method incorporates the entire PFS time from the randomization and treats binary MRD endpoint and time-to-event PFS endpoint as true bivariate endpoints, i.e., the association covers the entire timeline from randomization to end of follow-up. The global OR can be interpreted as the ORs for a patient being alive and progression-free beyond a time point when comparing patients with MRD negativity at 9 months versus those without MRD negativity, adjusting for treatments. For example, a global OR of 3.0 indicates that the odds of being alive and progression-free beyond a timepoint (giving that the patient is known to be alive and progression-free up to that timepoint) for a patient achieving CR up to 9 months and MRD negativity at 9 (+/-3) months is three times of the odds for a patient with positive status of 9m-MRDneg endpoint status. For global OR, a value higher than one indicates that patients who achieve MRD negativity (beyond CR) have a longer PFS outcome in general. It is considered as strong individual-level association if the global OR is high (e.g. ≥ 3.0) and the 95% confidence interval excludes 1.0.

The global OR and corresponding 95% CI estimated by a bivariate Plackett Copula model are reported in i2TEAMM Table 3 for 9m- and 12m-MRDneg rates, respectively. The scenarios where the global OR was not possible to be estimated were mainly due to insufficient data or zero MRDneg rate. The estimated global ORs were generally very large, i.e. > 5.0 (i.e. the odd of being alive and progress-free beyond a time point in patients with MRD negativity (beyond CR) is > 5 times higher than those without), with lower bound of 95% CI >> 1.0 (i.e., strongly statistically significant at type I error rate of 0.05). Therefore, at the individual patient level, the associations between both the 9m- or 12m-MRDneg status and PFS were very strong in all settings with sufficient data, indicating *consistently high patient-level surrogacy, for all three disease populations of NDTE, NDTinE, and RR MM at MRD testing sensitivity levels of 10^{-4} and 10^{-5}.* Empirically, there seems to be a trend of increased global OR values for the MRDneg endpoint measured at the 10^{-5} sensitivity level compared to the 10^{-4} sensitivity level.

i2TEAMM Table 3: Global Odds Ratio Estimates of Measuring Individual Level Associations Between 9m/12m MRDneg and PFS

	MRD Testing Sensitivity Level					
	10^{-4}		10^{-5}		10^{-6}	
Disease Population	N Subjects	Global OR (95% CI)	N Subjects	Global OR (95% CI)	N Subjects	Global OR (95% CI)
9m-MRDneg						
NDTE	3,700	5.68 (4.82 - 6.53)	3,061	8.27 (6.53 - 10.01)	2,297	5.88 (4.13 - 7.62)
NDTinE	2,516	8.32 (4.85 - 11.79)	2,235	9.80 (5.14 - 14.46)	Not sufficient data	
RR	1,380	7.13 (4.63 - 9.62)	1,378	8.24 (4.41 - 12.07)	Not sufficient data	
12m-MRDneg						
NDTE	3,410	6.95 (5.79 - 8.12)	3,009	9.15 (7.27 - 11.03)	2,064	5.46 (3.89 - 7.03)
NDTinE	2,310	9.08 (6.04 - 12.13)	2,281	11.95 (7.32 - 16.58)	Not sufficient data	
RR	1,375	8.82 (5.80 - 11.84)	863	16.24 (5.77 - 26.71)	Not sufficient data	

In addition, landmark analyses were performed to compare PFS between subjects who achieved MRD negativity (beyond CR) status at 9 months after randomization vs those who did not. These analyses were according to disease population and MRD sensitivity levels. The 9m-MRDneg Kaplan-Meier curves for the NDTE (i2TEAMM Appendix Figure 1and i2TEAMM Appendix Figure 2), NDTinE (i2TEAMM Appendix Figure 3) and RR (i2TEAMM Appendix Figure 4) populations are provided in Appendix. The HR estimates comparing PFS between 9m-MRDneg ranges from 0.19 to 0.69 (i.e., large effect size) with p-values of <0.0001 for all scenarios, except one (p = 0.0105). As can be seen from the plots in the appendix (i2TEAMM Appendix Figure 1, i2TEAMM Appendix Figure 2, i2TEAMM Appendix Figure 3, and i2TEAMM Appendix Figure 4), *the log-term outcomes for subjects achieving MRD negativity (beyond CR) were substantially improved compared to those who did not achieve MRD negativity: PFS was prolonged across all settings, regardless of disease settings or sensitivity levels.*

Trial-level association:

The primary 'true' endpoint for which MRD would be considered a surrogate was pre-specified and agreed with the Agency as **PFS**. Per the predefined statistical analysis plan agreed with the Agency (v3.1), trial-level surrogacy was measured by two methods; weighted least-squares regression (R^2_{WLS}) and a bivariate copula model (R^2_{Copula}). These two R^2 measures are very similar to the coefficient of determination in common linear regression. Hence, both R^2_{WLS} and R^2_{Copula} measure the proportion of variabilities in the treatment effect of PFS (i.e., log(HR)) that can be explained by the treatment effect of 9/12m-MRDneg (i.e., log(OR)). Values closer to 1.0 indicate stronger prediction of log(HR) on PFS based on the observed value of log(OR) on 9/12m-MRDneg. The formal qualification criteria to fully validate MRD negativity endpoint as a surrogate endpoint for traditional full regulatory approvals, such that it could be used to replace the clinical endpoint of PFS in future trials, was pre-specified in SAP v3.1:

If either of the R^2 values is at least 0.80 with a lower bound of the 95% confidence interval (CI) of >0.6 and neither estimate is <0.7, then the candidate MRD surrogate endpoint provides sufficient trial-level surrogacy to be used as a replacement for observation of treatment effect on the true clinical outcome.

Primary Analysis: NDTE Population

Among the 3 populations defined in the SAP, only the NDTE population has a sufficient number of two-arm comparisons to meet the data sufficiency criteria for performing trial-level surrogacy evaluation. Nevertheless, data allowed us to perform exploratory pooled analyses with NDTinE and RR populations as later described.

A total of 11 2-arm comparisons can be used to estimate trial-level R^2, with total of 3,298 patients. The MRD negativity was determined at 10^{-4}, 10^{-5}, and 10^{-6} sensitivity level for 5, 5 and 1 two-arm comparisons, respectively. All these 11 2-arm comparisons had at least one arm that included proteasome inhibitor (PI) and all are pre-maintenance treatment comparisons. Among these 11 2-arm comparisons, the treatment effect on MRD was based on the odds ratio (OR) of comparing 9m-MRDneg rate between experimental and control arms, and ranges from 0.56 to 4.36. Note, OR greater than 1.0 indicates higher (i.e., better) 9m-MRDneg rate in experimental arm than control arm. The hazard ratio (HR) of comparing PFS between experimental and control arms ranges from 0.36 to 1.90. Note, HR smaller than 1.0 indicates longer (i.e., better) PFS in experimental arm than control arm.

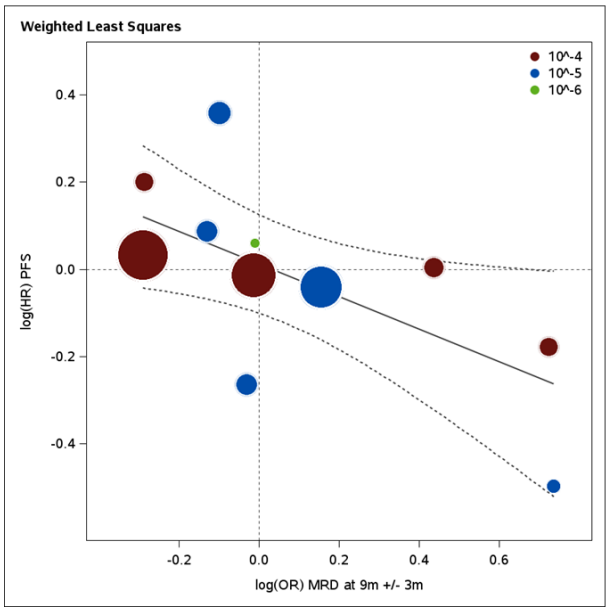
i2TEAMM Figure 1ashows the association between treatment effects (log(OR)) on 9m-MRDneg and treatment effects (log(HR)) on PFS at the trial level. There is a moderate association between treatment effects on the two endpoints (true and surrogate) at the trial level with wide confidence interval, demonstrated by R^2_{WLS} of 0.40 (95% CI, 0.01 to 0.79) and R^2_{Copula} of 0.44 (95% CI, 0.01 to 0.88). See i2TEAMM Table 4**.** Both primary trial-level surrogacy measures were lower than 0.8 and the lower bounds of the 95% CIs were lower than 0.5. Therefore, the principal surrogate endpoint candidate, 9m-MRDneg, does not meet the prospectively defined surrogacy criteria per SAP v3.1. However, as can be seen in i2TEAMM Figure 1b, the results indicate that the strength of relationship between treatment effects on MRD and PFS appears to differ depending on the sensitivity of the MRD level used for analysis. There appears to be a steeper slope if the regression line was estimated only based on two-arm comparisons using the 10^{-5} MRD sensitivity level (blue dots in i2TEAMM Figure 1b) compared to the regression line estimated based on those using the 10^{-4} sensitivity level (dark red dots in i2TEAMM Figure 1b). This may reflect the increased depth of sensitivity and the greater specificity to detect MRD, and therefore, *suggesting the 10^{-5} sensitivity level as a better predictor of treatment effect towards long-term PFS*. Pooling the data from different sensitivity levels may hamper interpretation of the overall R^2 values being estimated from these data.

In a post hoc analysis (i.e., not pre-specified in the SAP) based on five 2-arm comparisons that classified MRD negativities at 10^{-5} MRD sensitivity level (blue dots in Figure 1b) the association increased; R^2_{WLS} was 0.54 and R^2_{Copula} was 0.52. However, both estimates had very wide 95% CIs, due to the limited number of 2-arm comparisons. *This supports the*

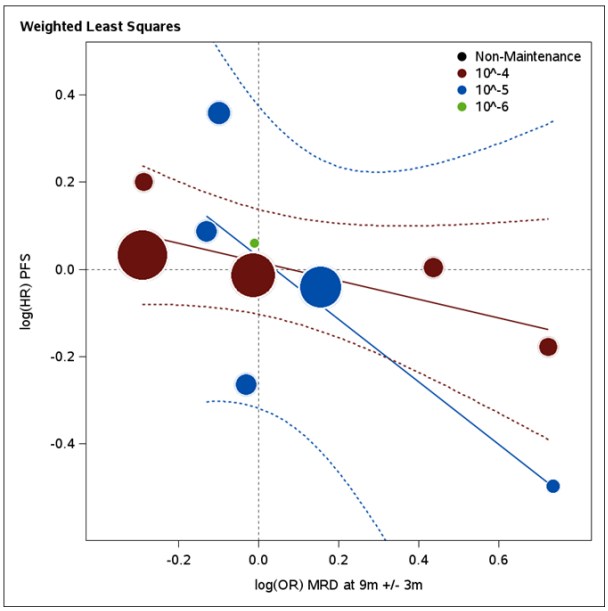
hypothesis that pooling data at different sensitivity levels may have underestimated the overall predictive ability of MRD.

i2TEAMM Figure 1: Association Between Treatment Effect on 9m-MRDneg Endpoint, Pooling Sensitivity Levels, and Treatment Effect on PFS in NDTE Population

1a



1b

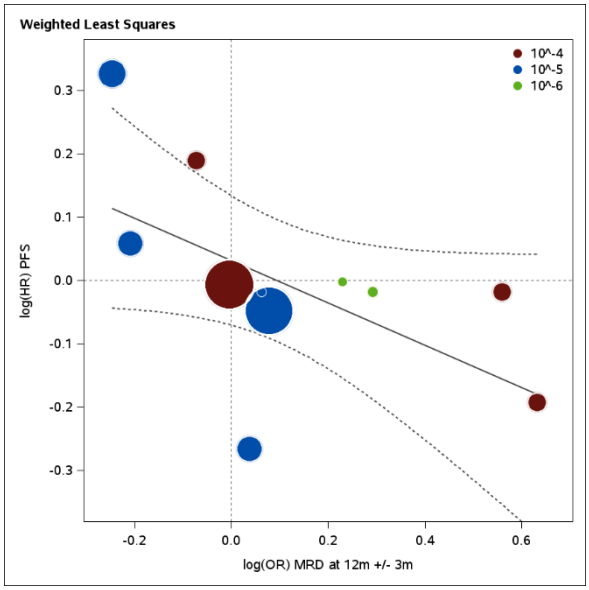


HR=hazard ratio; MRD=minimal residual disease; OR=odds ratio; PFS=progression-free survival; Dark red indicates two-arm comparisons with MRD tested at 10^{-4} sensitivity level; Blue indicates 10^{-5} sensitivity level; Green indicates 10^{-6} sensitivity level. The size of the dots is proportional to the sample size. The solid line indicates the fitted weighted least squares regression line. The dashed lines indicate 95% prediction limits. The horizontal dashed line corresponds to the $\log(HR_{PFS})$ of 0 (ie, HR of 1). The vertical dashed line corresponds to the $\log(OR_{9mMRDneg})$ of 0 (ie, OR of 1).

The results of further sensitivity and subgroup analyses as predefined in the SAP are provided in i2TEAMM Appendix Table 4and i2TEAMM Appendix Table 5, respectively. The trial-level surrogacy estimates were consistent with primary analyses when imputing missing MRD endpoint status by failure status, or when lowering the threshold in the non-missing MRD endpoint status. Similar results were seen in subgroup analyses.

An additional surrogacy candidate was pre-specified in the SAP as MRD tested at 12 (rather than 9) months, +/-3 months (12m-MRDneg). When considering this candidate, 11 two-arm comparisons (2,917 patients) met the data requirements. Note that these are not the same eleven comparisons as for the MRDneg9m analysis, hence the total number of patients differs. All 11 two-arm comparisons were pre-maintenance treatment comparisons, and the MRD tests were by MFC or NGF. At the trial level, the estimated R^2_{WLS} was 0.32 (95% CI: 0.00, 0.74), and the R^2_{Copula} was 0.30 (95% CI: 0.00, 0.75), showing slightly lower evidence for surrogacy than the earlier MRD timepoint. See i2TEAMM Table 4 and i2TEAMM Figure 2.

i2TEAMM Figure 2: Association Between Treatment Effect on 12m-MRDneg Endpoint, Pooling Sensitivity Levels, and Treatment Effect on PFS in NDTE Population



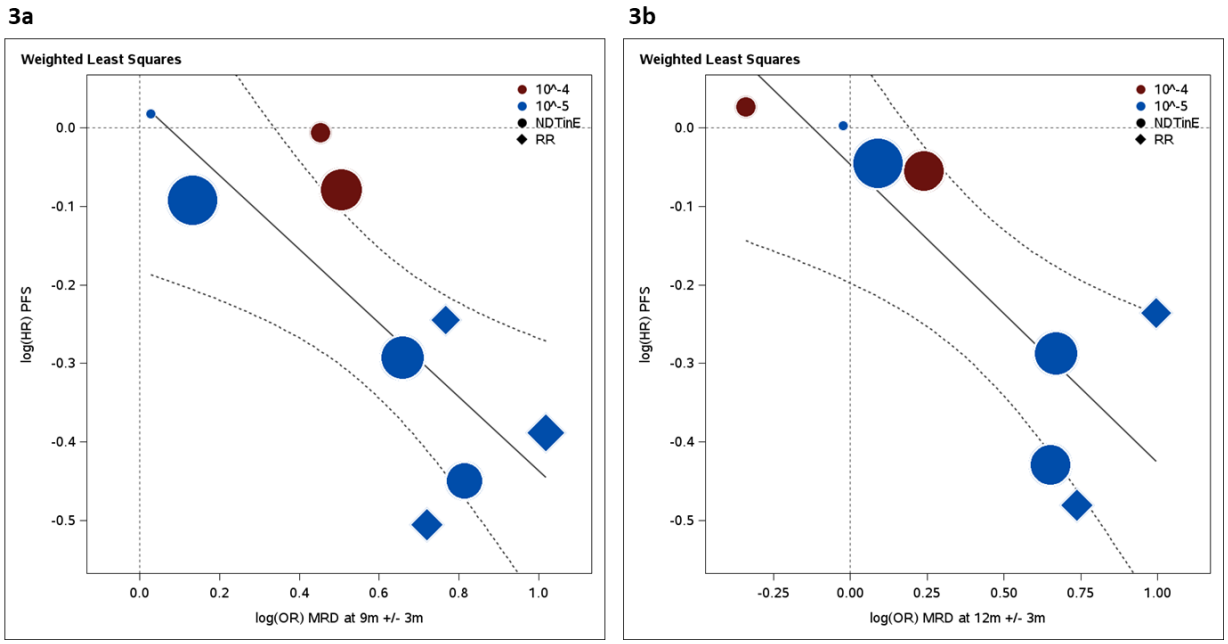
HR=hazard ratio; MRD=minimal residual disease; OR=odds ratio; PFS=progression-free survival; Dark red indicates two-arm comparisons with MRD tested at 10^{-4} sensitivity level; Blue indicates 10^{-5} sensitivity level; Green indicates 10^{-6} sensitivity level. The size of the dots is proportional to the sample size. The solid line indicates the fitted weighted least squares regression line. The dashed lines indicate 95% prediction limits. The horizontal dashed line corresponds to the log(HR_{PFS}) of 0 (ie, HR of 1). The vertical dashed line corresponds to the log($OR_{9mMRDneg}$) of 0 (ie, OR of 1).

<u>Exploratory Analysis per SAP: Pooling NDTinE and RR Populations</u>

The SAP also pre-specified an exploratory analysis in which the NDTinE and RR populations would be pooled together. In this analysis, there are 10 2-arm comparisons that met the trial-level surrogacy analysis inclusion criteria with ≥ 50 patients and ≥ 80% of patients that had 9m-MRDneg status. However, in one of these 10 2-arm comparisons (ICARIA), the 9m-MRDneg rate was 0% in control arm. The OR of comparing 9m-MRDneg endpoint in experimental arm to control arm is not estimable, i.e. OR = ∞. Therefore, only 9 2-arm comparisons with 4630 patients were included in the exploratory trial-level surrogate evaluation.

i2TEAMM Figure 3a shows the association between treatment effects (log(OR)) on 9m-MRDneg and treatment effects (log(HR)) on PFS at the trial level in this pooled population. There is an encouraging association between treatment effects on the two endpoints (true and surrogate) at the trial level, demonstrated by R^2_{WLS} of 0.62 (95% CI, 0.38 to 0.85) and R^2_{Copula} of 0.55 (95% CI, 0.11 to 0.98). See i2TEAMM Table 4. i2TEAMM Figure 3b shows the association between treatment effects on 12m-MRDneg and that on PFS at the trial level. The R^2_{WLS} was 0.67 (95% CI, 0.33 to 1.00) and R^2_{Copula} was 0.58 (95% CI, 0.14 to 1.00). See i2TEAMM Table 4.

i2TEAMM Figure 3: Association Between Treatment Effect on 9m-MRDneg (3a) and 12m-MRDneg (3b) Endpoint, Pooling Sensitivity Levels, and Treatment Effect on PFS in Pooled NDTinE and RR Populations



HR=hazard ratio; MRD=minimal residual disease; OR=odds ratio; PFS=progression-free survival; NDTinE=newly diagnosed transplant in-eligible population; RR=relapsed/refractory population; Dark red indicates two-arm comparisons with MRD tested at 10^{-4} sensitivity level; Blue indicates 10^{-5} sensitivity level. Square shape indicates RR population. The size of the dots is proportional to the sample size. The solid line indicates the fitted weighted least squares regression line. The dashed lines indicate 95% prediction limits. The horizontal dashed line corresponds to the log(HR_{PFS}) of 0 (ie, HR of 1). The vertical dashed line corresponds to the log($OR_{9mMRDneg}$) of 0 (ie, OR of 1).

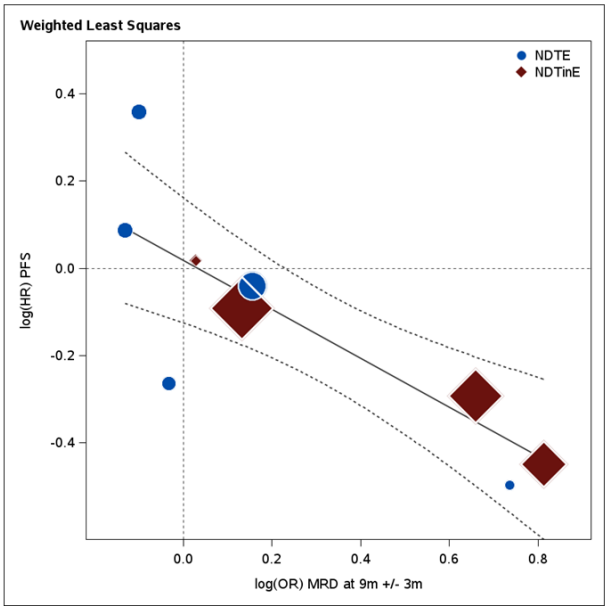
<u>Post Hoc Analysis Which Was Not Pre-Specified in SAP:</u>

Note, in the pooled analysis of NDTinE and RR populations, majority of the 2-arm comparisons had MRD negativity status classified at the sensitivity level of 10^{-5}. <u>*This further supports the importance of homogeneous sensitivity level in trial-level surrogacy estimation in MM.*</u> The rationale for pooling NDTinE and RR population was that transplantation is not an indication for both populations. Notwithstanding, we believe that the encouraging R^2_{WLS} and R^2_{Copula} results are not due to the absence of transplantation, but rather the fact that most of the NDTinE and RR trials used MRD methodologies (NGS and NGF) achieving a sensitivity level of 10^{-5}. This reflects the momentum of these trials: when newer and more effective therapies became available for NDTinE and RR patients, the sensitive NGF and NGS MRD methodologies were already available and were adopted. By contrast, when transplantation was incorporated several decades ago in the treatment algorithm of NDTE patients, a higher rate of CR was observed and this triggered the investigation of MRD in this population. Obviously, many decades ago the sensitive NGF and NGS methodologies were not available and MRD was assessed using less sensitive techniques (with a limit of detection of 10^{-4}). To have a better understanding of the less encouraging results in NDTE, we conducted two post hoc analysis (i.e., not pre-specified in SAP) by pooling NDTE and NDTinE populations using MRDneg status

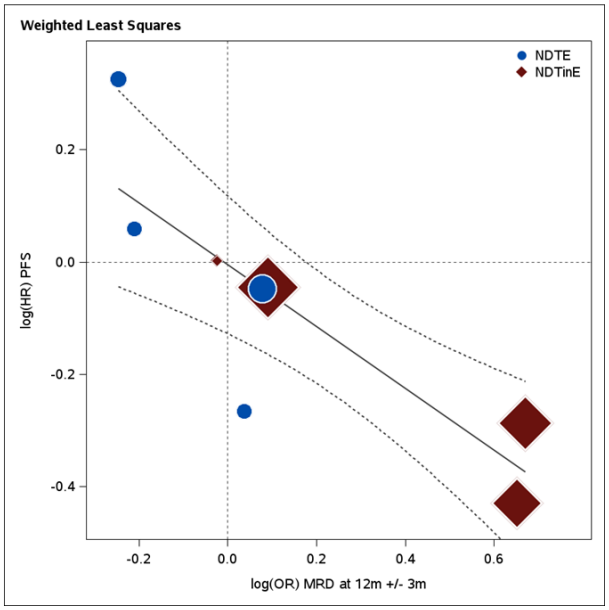
classified at sensitivity level of 10^{-5} only (see i2TEAMM Figure 4 and i2TEAMM Table 4), and by pooling all patients (NDTE, NDTinE and RR) using MRDneg status classified at sensitivity level of 10^{-5} only (see i2TEAMM Figure 5 and i2TEAMM Table 4) *A substantial increase in both R^2 values is observed, which further supports the importance of homogeneous sensitivity level in trial-level surrogacy estimation in MM. However, these results should be interpreted with caution due to the post-hoc nature.*

i2TEAMM Figure 4: Association Between Treatment Effect on 9m-MRDneg (4a) and 12m-MRDneg (4b) Endpoint, at 10-5 Sensitivity Level, and Treatment Effect on PFS in Pooled NDTE and NDTinE Populations

4a

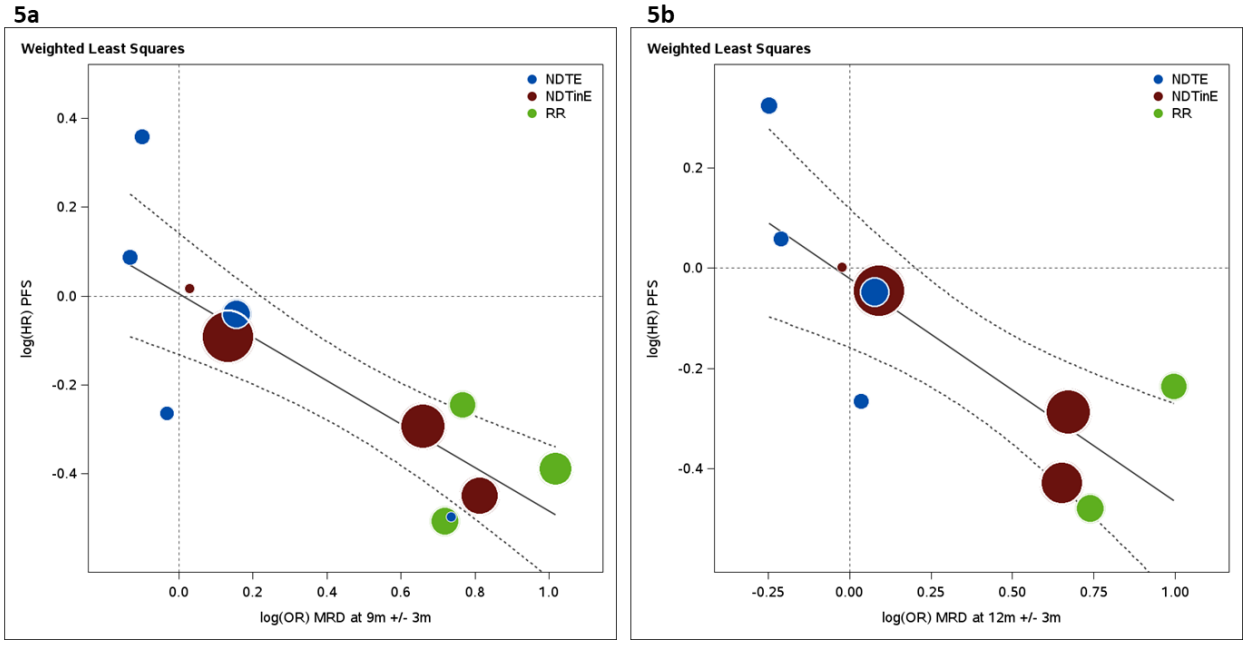


4b



HR=hazard ratio; MRD=minimal residual disease; OR=odds ratio; PFS=progression-free survival; NDTE=newly diagnosed transplant eligible population; NDTinE=newly diagnosed transplant in-eligible population; Dark red and square shape indicate NDTinE population; Blue and circle shape indicate NDTE population. The size of the dots is proportional to the sample size. The solid line indicates the fitted weighted least squares regression line. The dashed lines indicate 95% prediction limits. The horizontal dashed line corresponds to the $\log(HR_{PFS})$ of 0 (ie, HR of 1). The vertical dashed line corresponds to the $\log(OR_{9mMRDneg})$ of 0 (ie, OR of 1).

i2TEAMM Figure 5: Association Between Treatment Effect on 9m-MRDneg (5a) and 12m-MRDneg (5b) Endpoint, at 10-5 Sensitivity Level, and Treatment Effect on PFS in Pooled NDTE, NDTinE, and RR Populations



HR=hazard ratio; MRD=minimal residual disease; OR=odds ratio; PFS=progression-free survival; NDTE=newly diagnosed transplant eligible population; NDTinE=newly diagnosed transplant in-eligible population; RR=relapsed/refractory population; Dark red indicates NDTinE population; Blue indicates NDTE population; Green indicates RR population. The size of the dots is proportional to the sample size. The solid line indicates the fitted weighted least squares regression line. The dashed lines indicate 95% prediction limits. The horizontal dashed line corresponds to the log(HR_{PFS}) of 0 (ie, HR of 1). The vertical dashed line corresponds to the log($OR_{9mMRDneg}$) of 0 (ie, OR of 1).

i2TEAMM Table 4: Trial-Level Surrogacy Estimations

		9 months MRD Negativity Rate			12 months MRD Negativity Rate		
Analysis Population (Analysis type in SAP)	**Sensitivity Level**	**N Comparisons (N pts)**	**R^2 using Weighted Least Squares**	**R^2 using Plackett**	**N Comparisons (N pts)**	**R^2 using Weighted Least Squares**	**R^2 using Plackett**
NDTE (Primary analysis in SAP)	10^{-4}, 10^{-5}, and 10^{-6} levels [52% and 40% at 10^{-4} for 9 and 12m MRDneg, respectively]	11 (3,298)	0.40 (0.01, 0.79)	0.44 (0.01, 0.88)	11 (2,917)	0.32 (0.00, 0.74)	0.30 (0.00, 0.75)
NDTinE and RR (Exploratory analysis in SAP)	10^{-4} and 10^{-5} levels [21% and 24% at 10^{-4} for 9 an 12m MRDneg, respectively]	9 (4,630)	0.62 (0.38, 0.85)	0.55 (0.11, 0.98)	8 (4,134)	0.67 (0.33 to 1.00)	0.58 (0.14, 1.00)
NDTE and NDTinE (Post hoc analysis, not in SAP)	10^{-5} level [0% at 10^{-4}]	9 (3,665)	0.73 (0.38, 1.00)	0.67 (0.31, 1.00)	8 (3,566)	0.78 (0.47, 1.00)	0.71 (0.36, 1.00)
NDTE, NDTinE and RR (Post hoc analysis, not in SAP)	10^{-5} level [0% at 10^{-4}]	12 (5,043)	0.70 (0.48, 0.92)	0.66 (0.36, 0.97)	10 (4,429)	0.66 (0.34, 0.98)	0.61 (0.23, 0.99)

Although confidence intervals remain wide, the overall evidence supports that treatment effects on MRD measured at 10^{-5} sensitivity level explains a reasonable amount of the variability in treatment effects on PFS observed across these trials.

Combining primary analysis, pre-planned exploratory and post-hoc analyses, based on the data used so far, the results of the analyses demonstrate that while there is not sufficient evidence to support use of MRD as an validated surrogate endpoint in place of longer-term endpoints such as PFS, there is sufficient evidence to consider use of MRD as an endpoint 'reasonably likely' to predict clinical benefit and potentially enabling earlier access to novel therapies, in parallel with collection of confirmatory efficacy data on PFS in the same trial.

Conclusions:

This initial trial-level meta-analysis contained 20 robust, randomized, controlled Phase 3 clinical trials with mature PFS data and large sample sizes. *The collected trials enrolled patients from US and global countries including Europe, the Middle East, Africa, and Asia.* The trials were varied in their design, including different lines of therapy, different treatment strategies, different MRD testing methods, different timings and/or number of MRD assessments, and different MRD sensitivity levels. Therefore, the results of the initial meta-analysis are largely representative of a wide spectrum of the treatment options and clinical practice. *Comparing to previous published individual studies and literature-based meta-analyses, an important and critical feature of the current analysis is that, in the definition of MRD negativity, a uniform timepoint and method incorporating conventional CR were applied to all studies and settings.*

At the individual patient level, both the bivariate association analysis via the Copula model and the landmark analysis showed *very strong associations between MRD negativity (measured at both 9 and 12 months) after achieving conventional CR and PFS. This is consistently seen across different disease settings and MRD sensitivity levels.* Furthermore, empirically, the strength of the association seems to increase at higher sensitivity levels (i.e., 10^{-5} compared with the 10^{-4} sensitivity level); this is consistent with previous research[24,25]. With a uniformly defined MRD negativity endpoint evaluated based on IPD from a large collection of randomized clinical trials, the results provide much stronger evidence than previous studies: that patients achieving MRD negativity beyond CR have a much better long-term prognosis than patients who do not. The results of these analyses increase the confidence in the use of MRD negativity as an endpoint to predict patient-level long-term outcomes, since newer therapies that can increase the proportion of patients who are able to achieve MRD negativity are, therefore, potentially more likely to also increase the longer-term treatment effect on PFS.

At the trial-level, the number of 2-arm comparisons (as opposed to number of individual patients) needs to be sufficiently large to provide robust trial-level surrogacy estimates. Despite collection of a large number of trials, this number is still limited within each of the disease populations, especially when considering a homogeneous MRD sensitivity level. In the NDTE population, the trial-level surrogacy estimates are encouraging with R^2_{WLS} and R^2_{Copula} of 0.40 and 0.44, respectively. The post-hoc analysis in NDTE population, the R^2 estimates increased to 0.54 and 0.52 if only sensitivity level of 10^{-5} was considered. In pre-planned exploratory analysis: pooled NDTinE and RR populations, most of the trials used NGS or NGF for MRD

testing with increased sensitivity and standardization. The results again demonstrated encouraging R^2 values of approximately 0.6 where majority of two-arm comparisons were based on the 10^{-5} sensitivity level. Further post hoc analysis by pooling NDTE and NDTinE populations and restricting to 10^{-5} level showed highest R^2 values of > 0.7. Pooling all three populations at 10^{-5} sensitivity level, the higher R^2 values remained.

The totality of evidence presented herein therefore supports the following conclusions:

(1) The strong prognostic value of MRD negativity beyond CR for PFS in patients with MM is confirmed, supporting the existing literature and further adding robust evidence due to the standardized approach to assessing and defining the MRD endpoint, using patient-level data and pre-specified statistical analysis of data, and
(2) Use of an MRD-based endpoint as a full replacement of PFS in Phase 3 trials is not demonstrated by the analysis conducted by i2TEAMM, however the data support that treatment effects on MRD (tested at a sensitivity of 10^{-5} or better) are reasonably likely to predict treatment effects on PFS. Based on the results observed in this meta-analysis of multiple large randomized studies, i2TEAMM believe there is sufficient evidence to support use of MRD as an endpoint for accelerated approval, with PFS maintained as a long-term endpoint for confirmation of clinical benefit.

3.2.3 The FDA's Position

In general, FDA agrees with the results presented by the Applicants.

FDA's results are presented below. In the text that follows, the terms "weak", "moderate", and "strong" are used to describe various association measures. There is no universally accepted definition for these magnitudes. In practice, correlations above 0.8 have been referred to as "strong", other publications have utilized lower values such as 0.7 when categorizing a correlation as "strong" or "high". In addition, correlations below 0.50 have been referred to as "weak", although similarly, this threshold ranges from 0.25 to 0.60 among various publications. Note that these terms typically refer to the point estimate of R^2 rather than the confidence interval. These terms are used loosely below in the interpretation of the results. The results are interpreted with reference to the i2TEAMM thresholds for establishing a validated surrogate endpoint. FDA reiterates that these thresholds are not established anywhere in FDA guidance.

In the NDTE population, 8 trials were included in the analysis. Among the 8 trials, 3 used MRD threshold at 10^{-4}, 4 used MRD threshold at 10^{-5}, and 1 used MRD threshold at 10^{-6}. In the NDTinE population, 7 trials were included in the analysis. Among the 7 trials, 2 used MRD threshold at 10^{-4} and 5 used MRD threshold at 10^{-5}. In the RR population, 4 trials were included with MRD threshold at 10^{-5}.

Individual-Level Association

Individual-level associations for PFS and OS are shown in FDA Table 1 and FDA Table 2.

For MRD negative CR at 9 months vs PFS and 12-months MRD negative CR vs PFS, the copula global odds ratio ranged from 2.85 to 7.40 and 3.39 to 7.67, respectively. For 9-months MRD negative CR vs OS and 12-months MRD negative CR vs OS, the copula global odds ratio ranged from 2.77 to 6.46 and 3.83 to 6.03, respectively. All confidence intervals excluded 1. These results suggest that there is a strong positive individual-level association between MRD negative CR and both PFS and OS for all populations considered.

FDA Table 1: Individual-level association – MRD Negative CR and PFS

	9 months		12 months	
	N comparison (N Patients)	**Copula Global Odds Ratio (95% CI)**	**N comparison (N Patients)**	**Copula Global Odds Ratio (95% CI)**
NDTE	12 (4820)	2.85 (2.37, 3.34)	13 (4993)	3.39 (2.87, 3.92)
NDTinE	7 (3974)	6.55 (4.48, 8.63)	7 (3974)	7.30 (5.21, 9.38)
RR	4 (1835)	7.40 (4.17, 10.62)	4 (1835)	7.67 (4.24, 11.1)

Source: FDA analysis
MRD negative CR = MRD negativity with complete response; PFS = progression-free survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Table 2: Individual-level association – MRD Negative CR and OS

	9 months		12 months	
	N comparison (N Patients)	Copula Global Odds Ratio (95% CI)	N comparison (N Patients)	Copula Global Odds Ratio (95% CI)
NDTE	12 (4820)	2.77 (2.15, 3.38)	13 (4993)	3.83 (3.00, 4.67)
NDTinE	7 (3974)	5.02 (2.82, 7.21)	7 (3974)	4.75 (2.91, 6.58)
RR	4 (1835)	6.46 (2.54, 10.38)	4 (1835)	6.03 (2.48, 9.59)

Source: FDA analysis
MRD negative CR = MRD negativity with complete response; OS=overall survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

These results are generally similar to the results reported by the two Applicants with the following exception. The i2TEAMM reported higher global odds ratios for the three populations: 9.15 for NDTE, 11.95 for NDTinE, and 16.24 for RR. This difference appears to be due to the removal of patients with missing MRD data, and difference in the trials included in the two analyses.

Overall, the available data suggest that the individual-level association between MRD negative CR at 9 or 12 months and PFS or OS is high. That is, there is strong evidence that patients who achieve MRD negative CR at 9 or 12 months tend to have long PFS and OS compared to patients who do not achieve MRD negative CR at these timepoints.

Trial-Level Associations

Trial-level associations for PFS and OS are presented in FDA Table 3 and FDA Table 4, respectively.

MRD Negative CR at 9 months

In NDTE population, 8 trials were included and total of 12 two-arm comparisons were formed. In NDTinE population, 7 trials were included and total of 7 two-arm comparisons were formed. In RR population, 4 trials were included and total of 4 two-arm comparisons were formed.

There was weak correlation between the treatment effects of MRD negative CR at 9 months and PFS in the NDTE and RR population (R^2_{copula} range 0.01 to 0.45, R^2_{wls} range 0.01 to 0.47), because these values were low and did not meet the prespecified thresholds. Moderate correlation was observed in NDTinE population (R^2_{copula}= 0.72 (95% CI= 0.36, >0.99), R^2_{wls}= 0.72 (95% CI: 0.51, 0.93)), close to but below the prespecified thresholds of 0.8 for R^2 and 0.6 for the lower bound.

For 9-months MRD negativity vs OS, there was weak correlation found between the treatment effects on the two endpoints in the NDTE and RR population (R^2_{copula} range 0.26 to 0.34, R^2_{wls} range 0.29 to 0.30), because these values were low and did not meet the prespecified

thresholds. Moderate correlation was observed in NDTinE population based on the R^2_{wls} value (R^2_{copula}= 0.50 (95% CI <0.01, >0.99), R^2_{wls}= 0.76 (95% CI: 0.58, 0.95)), close to but below the prespecified thresholds.

MRD Negative CR at 12 months

For NDTE population, 7 trials were included and total of 13 two-arm comparisons were formed. For NDTinE population, 7 trials were included and total of 7 two-arm comparisons were formed. For RR population, 4 trials were included and total of 4 two-arm comparisons were formed.

For 12-months MRD negativity vs PFS, there was weak correlation found between the treatment effects on the two endpoints in the NDTE and RR population (R^2_{copula} range 0.00 to 0.35, R^2_{wls} range <0.01 to 0.45), because these values did not meet the prespecified thresholds. Strong correlation was observed in NDTinE population (R^2_{copula}= 0.83 (95% CI= 0.61, >0.99), R^2_{wls}= 0.82 (95% CI: 0.67, 0.97)), and the R^2 (95% CI) values met the thresholds prespecified by the i2TEAMM. Note that the lower bound of the confidence interval would be consistent with a moderate association.

For 12-months MRD negativity vs OS, weak or moderate association was found for all three populations (R^2_{copula} range 0.12 to 0.36, R^2_{wls} range 0.13 to 0.52), because these values did not meet the prespecified thresholds.

FDA Table 3: Trial-level association – MRD Negative CR and PFS

	9 months		12 months	
	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)
NDTE	12 (4820)	0.45 (0.04, 0.87) 0.47 (0.16, 0.78)	13 (4993)	0.35 (<0.01, 0.77) 0.45 (0.14, 0.76)
NDTinE	7 (3974)	0.72 (0.36, >0.99) 0.72 (0.51, 0.93)	7 (3974)	0.83 (0.61, >0.99) 0.82 (0.67, 0.97)
RR	4 (1835)	0.01 (<0.01, 0.21) 0.01 (<0.01, 0.08)	4 (1835)	0.00 (<0.01, 0.10) <0.01 (<0.01, 0.03)

Source: FDA analysis
MRD negative CR = MRD negativity with complete response; PFS = progression-free survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Table 4: Trial-level association – MRD Negative CR and OS

	9 months		12 months	
	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)
NDTE	12 (4820)	0.26 (<0.01, 0.68) 0.30 (<0.01, 0.62)	13 (4993)	0.36 (<0.01, 0.78) 0.44 (0.13, 0.75)
NDTinE	7 (3974)	0.50 (<0.01, >0.99) 0.76 (0.58, 0.95)	7 (3974)	0.34 (<0.01, 0.91) 0.52 (0.21, 0.83)
RR	4 (1835)	0.34 (<0.01, >0.99) 0.29 (<0.01, 0.58)	4 (1835)	0.12 (<0.01, 0.70) 0.13 (<0.01, 0.37)

Source: FDA analysis
MRD negative CR = MRD negativity with complete response; OS=overall survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

Taken together, FDA Table 3 and FDA Table 4 show that either endpoint (MRD negative CR at 9 months or 12 months) have generally weak-to-moderate associations with PFS and OS. The correlations for MRD negative CR at 12 months meet the pre-specified thresholds for PFS in the NDTinE population, although these results are not replicated in other populations. This population only included 7 2-arm comparisons. Other reasons such as sensitivity of the MRD at this time may have also contributed to these discrepant results between the different disease settings. Similarly, there are limitations to the data for the MRD data for the RRMM patient population. Despite pooling data from both the Applicants, only 4 two-arm comparisons were available for analysis.

Given that moderate trial-level correlations with PFS were observed for some populations, STE was calculated to provide additional context for the correlations observed. Estimates of the STE for PFS and OS are presented in FDA Table 5 and FDA Table 6, respectively. Values of "NA" reflect the fact that the trial-level correlation was low and that no such threshold can be calculated.

The STE estimates presented in FDA Table 5 suggest that, to observe a statistically significant treatment effect on PFS, the treatment effect on MRD negative CR needed would be an odds ratio between 2.12 and 4.95, depending on exact endpoint and setting. As noted in Section 6.3.2, such treatment effects assume that a future trial has 100% power for PFS. In practice, the treatment effect on MRD negative CR may need to be larger to predict a treatment effect on PFS.

FDA Table 5: Surrogate Threshold Effect (STE) for PFS

MRD Negative CR vs PFS	9 months		12 months	
	N comparison (N Patients)	STE odds ratio	N comparison (N Patients)	STE odds ratio
NDTE	12 (4820)	4.71	13 (4993)	4.95
NDTinE	7 (3974)	2.72	7 (3974)	2.12
RR	4 (1835)	NA	4 (1835)	NA

Source: FDA analysis
MRD negative CR = MRD negativity with complete response; PFS = progression-free survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Table 6: Surrogate Threshold Effect (STE) for OS

MRD Negative CR vs OS	9 months		12 months	
	N comparison (N Patients)	STE odds ratio	N comparison (N Patients)	STE odds ratio
NDTE	12 (4820)	NA	13 (4993)	5.81
NDTinE	7 (3974)	6.49	7 (3974)	12.3
RR	4 (1835)	NA	4 (1835)	NA

Source: FDA analysis
MRD negative CR = MRD negativity with complete response; OS=overall survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

To understand how the STE values above would theoretically translate to a future clinical trial, FDA Table 7 gives example MRD- CR rates based on assumed rates in the control arm of a future trial and selected STE estimates. These rates are for illustrative purposes only.

FDA Table 7: MRD- CR Rates Needed in Treatment Arms Based on Estimated STEs (Example)

STE Odds Ratio	Assumed MRD Negative CR Rate in Control Arm	MRD Negative CR Rate Needed in Treatment Arm (Based on STE)
2.12	10%	19%
	20%	35%
	30%	48%
4.95	10%	25%
	20%	55%
	30%	68%

Source: FDA analysis
MRD negative CR = MRD negativity with complete response
STE= surrogate threshold effect; MRD Negative CR = MRD negativity with complete response

MRD Negative CR at Any Time: RR Setting

FDA performed additional analyses in the RR setting to understand the potential use of MRD negative CR at Any Time as a potential endpoint.

Results are presented in 7.3.5.3. The results are similar to results presented above for the RR population: the individual-level association yielded an odds ratio of 8.70 (4.84, 12.55) and the trial-level association was weak with an R^2_{Copula} of 0.11 (<0.01, 0.62).

There are some limitations to this MRD data in the RR setting.

The MRD negative CR at Any Time rate was low (<5%) in the control arms of multiple trials, calling into question whether the odds ratio is a representative quantification of the treatment effect. Note that this limitation applies to MRD negative CR at 9 or 12 months also, as the rates of MRD negative CR at any time are similar to those for the MRD negative CR at 9 months or 12 months.

FDA conducted additional sensitivity analysis by pooling all three patient populations. These analyses assume that MRD negative CR at 9 months or 12 months predict PFS or OS similarly across the various clinical settings. If this assumption is tenable, then results from the pooled population NDTE+NDTinE+RR may provide further insight to support the results in the RRMM setting (Section 7.3.5.2). At the trial-level, these analyses yield R^2 values for PFS ranging from 0.51 to 0.61. The corresponding STEs for PFS range from 3.25 to 3.82. These results provide data to support the use of MRD as an endpoint across all disease settings in MM (Section 7.3.5.2). We note that a similar analysis was conducted by both Applicants, pooling different patient populations with results supportive of the association of MRD with long term outcomes across different disease settings.

Conclusions

- MRD-negative CR at 9 or 12 months have strong individual-level correlations with PFS and OS across multiple MM disease settings. This indicates that MRD-negative CR and CR are strong prognostic factors for OS and PFS.
- For trial-level association, a weak-to-moderate association between 9-month/12-month MRD-negative CR and PFS was observed across the various populations studied. No association between 9-months MRD-negative CR or 12-months MRD-negative CR and OS was demonstrated in the trial-level analysis for any of the 3 populations.

4 Points for the Advisory Committee to Consider

4.1 University of Miami Applicant's Position

Over the past decade, MRD has been increasingly used as a clinical endpoint in studies evaluating new treatments for multiple myeloma in patients with NDMM. Many of these studies have been included in the meta-analysis described in this report. In examining the results of these studies, a longer period is required to reach mature PFS and OS data with newer treatments, and MRD is an objective measure of anti-myeloma clinical activity that can reliably be used to predict a clinically meaningful treatment effect on PFS. While waiting for PFS or OS endpoints to mature, patients may be denied access – for several years – to an effective therapy or alternatively may be kept on a study that may reveal a lack of benefit of a new therapy. Furthermore, with the rapidly evolving treatment of multiple myeloma, during a long study the comparator arm could become obsolete.

FDA guidance recommends that before initiating a clinical trial, sponsors should consider and discuss with FDA whether based on the available preliminary clinical data, the expected effect on response rate or other early endpoint is of a sufficient magnitude to be reasonably likely to predict clinical benefit.[45] Results of this meta-analysis demonstrate a strong correlation between MRD and PFS, supporting MRD as an intermediate endpoint reasonably likely to predict clinical benefit. Reasonably likely intermediate endpoints are supported by strong mechanistic and/or epidemiologic rationale. MRD negativity has been found to be associated with longer PFS and OS in several epidemiological studies. As MRD has been rigorously analyzed, the Applicant believes that MRD is an intermediate clinical endpoint reasonably likely to predict clinical benefit and that benefit could be confirmed within the same study by following patients in a fully powered registration trial, as per their assigned treatment assignment, for long-term outcomes such as PFS. It is particularly relevant to identify an early biomarker among patients with NDMM, who are likely to have long times to disease progression or death as new treatments and combinations are studied.

Modern 3- and 4-drug combination therapies for NDMM have been found to deliver high rates of MRD negativity in the absence of autologous stem cell transplant.[10,12,26,34,37,43] The use of a modern 4-drug combination has been associated with 71% MRD negativity in the absence of transplant.[26] In the MANHATTAN trial, after completion of 4-drug combination therapy, patients who had collected stem cells during combination therapy had the freedom either to receive a transplant followed by maintenance, or to keep their stem cells for storage and proceed with maintenance (i.e., delayed transplant).[26] The majority of the 71% of patients who achieved MRD negativity chose to keep their stem cells for storage and move forward with maintenance. Similarly, in the standard of care setting, an increasing proportion of patients are choosing to delay transplant in the US. Based on the results of recent studies with 3- and 4-drug combination therapies, it is reasonable to conjecture that it may no longer be appropriate to subdivide patients with NDMM into transplant-eligible and transplant-ineligible categories. This concept has been extensively discussed at recent meetings and is supported by key opinion

leaders and highly experienced physicians treating multiple myeloma patients.[19] Thus, it was appropriate to combine transplant-eligible and transplant-ineligible patients in this analysis.

The Applicant believes that achieving a deep response as measured by MRD-negative CR is clinically meaningful and would benefit patients as an early endpoint for regulatory approval. This meta-analysis, in combination with the substantial scientific, mechanistic, and clinical knowledge regarding MRD negativity, has substantially increased the level of confidence in the role of MRD negativity as an early clinical endpoint. The rationale for this conclusion includes the very strong prognostic value of MRD negativity shown in multiple independent analyses of large meta-analytic datasets[22,31] and the strong trial-level association between treatment effects on MRD negativity and PFS demonstrated with patient level data in this meta-analysis. Based on these findings, a therapy that demonstrates a strong, statistically significant treatment effect on MRD-negative CR in a randomized clinical trial is highly likely ultimately to demonstrate a statistically significant treatment effect on PFS.

4.2 The i2TEAMM Applicant's Position

Statement 1:
The rate of MRD negativity beyond CR classified at 10^{-5} or higher sensitivity level can be considered as an early endpoint reasonably likely to predict clinical benefit to be used to support accelerated approval (AA) of new agents in NDTE, NDTinE and RR MM settings.

Sponsor's Position

The MM treatment landscape is in a unique scenario. Successive improvements in therapies have significantly prolonged survival outcomes for many years, but unfortunately most patients are not cured. Fortunately, there are many potential new drugs and combinations that could continue improving outcomes, but the time needed for the readout of new clinical trials precludes patients having access to improved treatment options in a reasonable amount of time. This is the motivation behind our proposal of using MRD negativity rate as an early endpoint reasonably likely to predict clinical benefit to be used to support AA in MM. It stems from the strong and consistent patient-level association between MRD negativity and PFS demonstrated based on a uniformly defined MRD endpoint and the largest IPD collected from high quality randomized clinical trials worldwide, in combination with the substantial scientific, mechanistic, and clinical knowledge regarding MRD negativity. Altogether these have substantially increased the level of confidence in the role of MRD negativity as an early clinical endpoint. Given the substantial improvements in PFS for individual patients who achieved MRD negativity at 9 or 12 months, a novel regimen that produces a marked absolute increase in MRD negativity rate compared with control regimen in the intent-to-treat population (i.e., all randomized patients) may be reasonably likely to result in long-term improvement in PFS. On the other hand, lack of differences in MRD negativity rates may identify years in advance a regimen that will not produce a benefit in PFS. The encouraging trial-level surrogacy estimates further support this notion. Of note, we believe that even more

encouraging trial-level surrogacy is impractical given the nature of our IPD data: heterogeneous MRD techniques with highly variable sensitivity; heterogeneous time points of MRD assessment, heterogeneous treatment settings and even trials including two randomization steps. Despite the heterogeneity, we found encouraging trial-level surrogacy estimates that are aligned with the strong and consistent patient-level association between MRD negativity and PFS, and that together, provide confidence in the role of MRD negativity as an early endpoint reasonably likely to predict clinical benefit to be used to support AA in MM.

Statement 2:

i2TEAMM consortium proposed few recommendations in designing future trials in MM using MRD negativity beyond CR classified at 10^{-5} or higher sensitivity level as the primary endpoint to seek accelerated approval (AA).

Sponsor's Position:

MM is currently an incurable disease, and an environment conducive to timely clinical development of new therapies for patients should continue to address the high unmet medical need in this population, and to bring new therapies that are developed with signs of marked increases in efficacy to patients who urgently need them. MRD negativity can be measured much earlier than PFS differences in MM regardless of subpopulations. Biologically, MRD negativity represents much deeper responses to therapies than conventional CR measurements. AAs based on a therapy with significant absolute improvement in MRD negativity rate in the intent-to-treat population with acceptable safety will ensure timely the drug development in MM. Such AAs can benefit all stakeholders including patients, physicians, sponsors of innovative treatments, reimbursement agencies and regulators.

Approvals based on intermediate or early endpoints which do not have a formal qualification of surrogacy carry a level of risk, in that the clinical benefit has yet to be confirmed at the time of approval. To mitigate such risk when considering an endpoint of MRD negativity rate as a primary endpoint in future trials, we propose the following design recommendations:

- The long-term outcome data need to be continuously collected even after the trial results based on MRD negativity rate are released. A pre-specified PFS/OS hypothesis testing(s) or a set of hypotheses (with clear testing strategy) needs to be defined in the original trial protocol and statistical analysis plan. It is recommended to design such trial with a sample size providing sufficient power to test treatment effect on PFS/OS at a desired significance level.
- In addition to the toxicity profile of the new drug, on-treatment deaths and early deaths need to be closely monitored and compared between patients receiving new drug and standard of care treatment.
- The definition of MRD negativity rate endpoint is critical for supporting AA and provide comparability and reproducibility across trials.
    - Testing method: Should be clearly described in terms of the technology being used, samples (collection, shipment, storage), sites of testing, etc Data supports the use of the next-generation methods endorsed by the IMWG response criteria

published in 2016.

- Sensitivity level: The estimated sensitivity level should be prespecified and so the cutoff that will be used to define MRD negativity. Sensitivity level of 10^{-5} or higher is recommended, which is also consistent with the regulatory guidance. In addition, the LOD achieved in each sample should be recorded and the median (range) LOD achieved in the ITT population should be analyzed.
- Measurement timepoints and frequency: We recommend that all patients in the intent-to-treat population to be tested for MRD approximately 9-12 months after initiation of the treatment. From that timepoint, MRD should be reassessed at least every 12 months. Other timepoints can be considered are at the end of induction or consolidation.
- Follow standard operating procedures for collecting and handling specimens needed for MRD testing, as well as interpretation of MRD testing findings backed up by key data elements on each patient enrolled on the trial.
- Most importantly, to minimize missing data, all patients in intent-to-treat population achieving a certain depth of response (e.g., CR) need to be tested for MRD at the timepoint(s) which is (are) critical for primary hypothesis testing.
- It is also important to uniformly use the same test method and procedures throughout the trial by all participant sites.

4.3 The FDA's Position

Given the recent therapeutic advances in MM, with observation of high response rates and long overall survival times in recent clinical trials, MRD has the potential to expedite drug development if it is used as an endpoint to support accelerated approval in MM.
FDA's accelerated approval program is intended to facilitate expedited approval of novel therapies for serious and life-threatening conditions based on a surrogate endpoint reasonably likely to predict clinical benefit or an intermediate clinical endpoint that can be measured earlier than irreversible morbidity or mortality. Determining whether an endpoint is reasonably likely to predict clinical benefit will depend on the biological plausibility of the relationship between the disease, the endpoint, and the desired effect, and the empirical evidence to support that relationship. Clinical data should be provided to support a conclusion that a relationship of an effect on the intermediate clinical endpoint to an effect on the clinical outcome is reasonably likely. To make this determination, the FDA considers all relevant evidence.

The use of MRD as an intermediate clinical endpoint reasonably likely to predict clinical benefit is supported by biologic plausibility. ORR has been used an intermediate clinical endpoint to support accelerated approval, as reduction in myeloma tumor burden is considered reasonably likely to predict clinical benefit. MRD is also a measure of tumor burden, and, like ORR, it is biologically plausible that achieving a deeper level of response with MRD will be associated with improvement in long term outcomes.

- The data from the meta-analysis conducted by the Applicants and the FDA show strong patient level association for MRD negative CR with PFS and OS. This indicates that MRD negative CR is a strong prognostic biomarker for OS and PFS.
- The individual-level association has been demonstrated in multiple disease settings including the NDTE, NDTInE, and the RRMM settings.
- Another strength of these analyses is the consistent assessment of MRD based on pre-specified timepoints of 9-month and 12-months. The results of the individual-level association were consistent across both time points for PFS and OS.

There is a risk that improvement in MRD may not predict clinical benefit with long term follow up. However, this is a risk with the use of any early endpoint. Notably, in the BELLINI trial, a phase 3 trial that evaluated the addition of venetoclax to standard of care bortezomib and dexamethasone in patients with relapsed or refractory MM, a detriment in OS was noted in the venetoclax arm despite improvement in PFS, ORR and MRD. This trial underscores the importance of assessment of both early and late endpoints.

If MRD is used as an endpoint to support AA in MM, subsequent verification of clinical benefit will be required. This can be accomplished by a single trial model, where the accelerated approval can be based on assessment of MRD in a well-controlled randomized trial with subsequent follow up of the same trial for long term clinical outcomes to confirm clinical benefit. Alternatively, two trials can be conducted, one to support the initial accelerated approval and a separate trial to confirm the clinical benefit. In the recent Consolidated Appropriations Act, 2023, Congress provided FDA with the authority to require a confirmatory trial to be "underway" prior to accelerated approval. Additionally, the Consolidated Appropriations Act created a formal expedited withdrawal procedure for drugs approved through accelerated approval if, among other reasons, "a study required to verify and describe the predicted effect on irreversible morbidity or mortality, or other clinical benefit of the product fails to verify and describe such effect or benefit". These authorities minimize the risk of granting accelerated approval based on an intermediate clinical endpoint such as MRD.

Considerations for Future Trials

The available data appears to support the use of MRD as an intermediate clinical endpoint in MM clinical trials to support accelerated approval across different disease settings, NDTE, NDTInE and the RRMM settings. The analyses conducted by the Applicants and the FDA provide robust data to characterize the relationship between MRD and long-term outcomes of interest. MRD negative CR is a strong prognostic factor for OS and PFS. This has been demonstrated in multiple disease settings including the NDTE, NDTInE, and the RRMM settings and at multiple timepoints. This analysis may allow the use of MRD as an endpoint to support accelerated approval in MM. At this time, there is no data to support the use of MRD as an endpoint to support accelerated approval in other disease settings such as smoldering MM or patients with MM precursor conditions such as monoclonal gammopathy of undetermined significance.

If MRD is used as an endpoint, both 9-month and 12-month time points may be appropriate for assessment of the MRD endpoint across the different disease settings. Although these time points were not pre-specified in the original trials, the MRD assessments at 9-months and 12-months were prespecified for the current meta-analysis. As noted previously, the results of the patient level association were consistent across both the 9-month and 12-month time points for PFS and OS. The use of 9-month versus 12-month may depend on a particular disease setting. For example, in the newly diagnosed transplant eligible setting, MRD negativity at 12 months may be more appropriate as it allows for capture of the multiple treatment components that impact long term outcomes, including induction and transplant. In a treatment setting after multiple relapses or in a refractory patient population, a 9-month timepoint may be more appropriate. The assessment of durability is built into the timepoint of assessment of MRD at 9-months and 12-months and additional assessment of durability may not be needed to support the robustness of the results. However, this may depend on other considerations including the number of patients assessed for MRD at a specific time point. When ORR is used as an endpoint in MM clinical trials, a minimum of 9-12 months of follow-up has been required to ensure the durability of the response. Longer follow up for assessment of durability may be required if MRD is assessed earlier than 9 months. The 9-month and 12-month timepoints included a window of +/-3 months. Durability of MRD may be important to assess if the majority of the patients are assessed earlier than 9 months (i.e., at 6 months (-3-month window of a 9-month timepoint)).

An MRD endpoint assessing MRD negative CR at any time, similar to best ORR, may also be appropriate. However, as detailed above, durability of MRD negativity may also be needed to support the robustness of the MRD endpoint if patients are assessed and achieve MRD negative CR earlier than 9-12 months. This may include follow up of patients with confirmation of MRD after a specified duration of time or duration of response. The available data did not allow for an assessment of MRD durability (sustained MRD) on long-term outcomes.

The magnitude of MRD negative CR that would be considered to provide a meaningful advantage over available therapy is unclear. In the NDTE+NDTinE+RR population, the STE for MRD- CR vs. PFS is 3.82, suggesting that in a randomized trial in which a 25% MRD negative CR rate is observed in the control arm, a 56% MRD negative CR rate in the treatment arm would be needed to predict a positive treatment effect on PFS. Of course, in a single-arm setting, comparison to an historical rate is typically confounded by difference in population, follow-up, trial design, etc.

If MRD negative CR is used to support an accelerated approval based on a randomized trial in MM, a key component of the approval will be early information on PFS and OS relative to control. The information available on either of these endpoints will depend on accrual rate, sample size, and outcomes. Note that fewer patients are typically required to power a response endpoint compared to a time-to-event endpoint. Given this and the improved outcomes in PFS and OS in NDMM, it is likely that a well-powered analysis of MRD negative CR at 9 or 12 months will be accompanied by limited PFS and OS information. However, the same trial can be followed for long term outcomes and confirmation of clinical benefit.

While FDA recommends randomized trials be used to support accelerated approvals where feasible, many accelerated approvals in RRMM have been based on single-arm trials. If MRD negative CR is to be used in single-arm trials, a minimum follow-up time should be specified to ensure most MRD negative CR responses could be observed regardless of the timepoint specified.

There are other regulatory considerations related to the development of MRD as a potential endpoint to support AA in MM including assay considerations. Currently, there are two general technologies used for bone marrow MRD assessment in MM: multiparametric flow cytometry (MPFC) and next generation sequencing (NGS). The FDA is agnostic to which technology platform is used in clinical trials assessing MRD. However, the assay should be analytically validated for its context of use and should be sensitive to detect a prespecified MRD negativity threshold. Additionally, clinical studies should prespecify the measurement of MRD threshold. Although different thresholds were used in the trials included in the meta-analyses, the majority of the trials included assessed MRD negativity at a threshold of 10^{-5}. Additionally, the IMWG uniform response criteria for MRD includes interpretation of MRD at the threshold of 10^{-5}.[4] While there is emerging data that lower thresholds 10^{-6} may have better correlation, based on the data available at this time, a threshold of 10^{-5} would be appropriate. MRD should be conducted at prespecified times and missing data should be minimized.

Conclusion

In general, the development of novel endpoints is challenging. The Applicants have worked with the broader MM community to develop a novel endpoint of MRD that has the potential to expedite drug development in MM. While there are still outstanding questions on how to best use MRD, the meta-analyses conducted represent robust assessments of MRD that support its prognostic value, provide information regarding the appropriate timing of MRD assessment, and suggest that MRD may be appropriate to use as an intermediate clinical endpoint to support AA. While there are risks with use of any early endpoint, the accelerated approval paradigm addresses these risks by requiring confirmation of the anticipated clinical benefit and providing FDA with authority to seek prompt withdrawal of a product if clinical benefit is not verified. We look forward to continued engagement with the community and further development of MRD to help expedite the availability of effective therapies to patients.

5 Draft Topics for Discussion by the Advisory Committee

Discuss the adequacy of the data to support the use of MRD as an endpoint to support accelerated approval in MM clinical trials.

6 References

6.1 University of Miami References

1. Anderson KC, Auclair D, Adam SJ, et al. Minimal residual disease in myeloma: application for clinical care and new drug registration. Clin Cancer Res. 2021;27:5195-5212.
2. Avet-Loiseau H, Ludwig H, Landgren O, et al. Minimal residual disease status as a surrogate endpoint for progression-free survival in newly diagnosed multiple myeloma studies: a meta-analysis. Clin Lymphoma, Myeloma, Leukemia. 2020;20(1):e30-7.
3. Avet-Loiseau H, San-Miguel J, Casneuf T, 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. J Clin Oncol. 2021;39(10):1139-1149.
4. Bahlis NJ, Dimopoulos MA, White DJ, et al. Daratumumab plus lenalidomide and dexamethasone in relapsed/refractory multiple myeloma: extended follow-up of POLLUX, a randomized, open-label, phase 3 study. Leukemia. 2020;34:1875-1884.
5. Buyse M, Molenberghs G, Burzykowski T, et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1(1):49-67.
6. Buyse M, Sargent DJ, Grothey A, et al. Biomarkers and surrogate end points--the challenge of statistical validation. Nat Rev Clin Oncol. 2010; 7(6): 309-317.
7. Burzykowski T, Molenberghs G, Buyse M, et al: Validation of surrogate end points in multiple randomized clinical trials with failure time endpoints. Appl Stat. 2001; 50:405-422.
8. Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. J R Statist Soc A. 2004;167:103-124.
9. Cavo M, San-Miguel JFF, Usmani SZ, et al. Prognostic value of minimal residual disease negativity in myeloma: combined analysis of POLLUX, CASTOR, ALCYONE, MAIA. Blood. 2021;139(6):835-844.
10. Costa L, Chhabra, S, Medvedova E, et al. Daratumumab, carfilzomib, lenalidomide, and dexamethasone with minimal residual disease response-adapted therapy in newly diagnosed multiple myeloma. J Clin Oncol. 2022;49(25):2901-2912.
11. Facon T, Lee JG, Moreau P. et al. Carfilzomib or bortezomib with melphalan-prednisone for transplant-ineligible patients with newly diagnosed multiple myeloma. Blood. 2019;133 (28):1953-1963.
12. Facon T, Kumar S, Plesner T, et al. Daratumumab plus lenalidomide and dexamethasone for untreated myeloma. N Engl J Med. 2019;380(22):2104-2115.

13. Facon T, Kumar S, Plesner T, et al. Daratumumab, lenalidomide, and dexamethasone versus lenalidomide and dexamethasone alone in newly diagnosed multiple myeloma (MAIA): overall survival results from a randomized, open-label, phase 3 trial. Lancet Oncology. 2021:22(11):1582-1596.
14. Facon T, Venner CP, Bahlis NJ, et al. Oral ixazomib, lenalidomide, and dexamethasone for transplant-ineligible patients with newly diagnosed multiple myeloma. Blood. 2021;137(26):3616-3628.
15. Flores-Montero J, Sanoja-Flores L, Paiva B, et al. Next generation flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia. 2017;31:2094-2103.
16. Goldschmidt H, Mai EK, Düring J, et al. Response-adapted lenalidomide maintenance in newly diagnosed myeloma: results from the phase III GMMG-MM5 trial. Leukemia. 2020;34:1853-1865.
17. Gormley N. Fireside Chat with Nicole Gormley and Ola Landgren. https://vjhemonc.com/event/miami-myeloma-mrd-2021/. Accessed: 13 October 2021.
18. Kalina T, Flores-Montero J, van der Velden VHJ, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia. 2012;26:1986-2010.
19. Kazandjian D, Mo CC, Landgren O, et al. The role of high-dose melphalan with autologous stem-cell transplant in multiple myeloma: is it time for a paradigm shift? Br J Haematol. 2020;191(5):692-703.
20. Kumar S, Paiva B, Anderson KC, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol. 2016;17:e328-346.
21. Kumar SK, Callander NS, Adekola K, et al for the National Comprehensive Cancer Network. Multiple myeloma, version 3.2021, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Network. 2020;18(12):1685-1717.
22. Lahuerta JJ, Paiva B, Vidriales MB, et al. Depth of response in multiple myeloma: A pooled analysis of three PETHEMA/GEM clinical trials. J Clin Oncol 2017;35:2900-2918.
23. Landgren O, Devlin S, Boulad M, et al. Role of MRD status in relation to clinical outcomes in newly diagnosed multiple myeloma patients: a meta-analysis. Bone Marrow Transplant. 2016;51:1565-1568.
24. Landgren O, Gormley N, Turley D, et al. Flow cytometry detection of minimal residual disease in multiple myeloma: lessons learned at FDA-NCI roundtable symposium. Am J Hematol. 2014;89:1159-1160.
25. Landgren O, Iskander K. Modern multiple myeloma therapy: deep, sustained treatment response and good clinical outcomes. J Intern Med. 2017;281:365-382.

26. Landgren O, Hultcrantz M, Diamond B, et al. Safety and effectiveness of weekly carfilzomib, lenalidomide, dexamethasone, and daratumumab combination therapy for patients with newly diagnosed multiple myeloma: The MANHATTAN nonrandomized clinical trial. JAMA Oncol. 2021;7(6):862–868.
27. Mateos MV, Cavo M, Blade J, et al. Overall survival with daratumumab, bortezomib, melphalan, and prednisone in newly diagnosed multiple myeloma (ALCYONE): a randomised, open-label, phase 3 trial. Lancet. 2020;395(10218):132-141.
28. Mateos MV, Dimopoulos MA, Cavo M, et al. Daratumumab plus bortezomib, melphalan, and prednisone for untreated myeloma. N Engl J Med. 2018;378:518-28.
29. Moreau P, Attal M, Hulin C, et al. Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (CASSIOPEIA): a randomized, open-label, phase 3 study. Lancet. 2019;394:29-38.
30. Moreau P, Hulin C, Perrot A, et al. Maintenance with daratumumab or observation following treatment with bortezomib, thalidomide, and dexamethasone with or without daratumumab and autologous stem-cell transplantation in patients with newly diagnosed multiple myeloma (CASSIOPEIA): an open-label, randomized, phase 3 study. Lancet Oncol. 2021;22:1378-1390.
31. Munshi NC, Loiseau HA, Rawstron AC, et al. Association of minimal residual disease with superior survival outcomes in patients with multiple myeloma: a meta-analysis. JAMA Oncol. 2017;3(1):28-35.
32. Munshi NC, Avet-Loiseau H, Anderson KC, et al. A large meta-analysis establishes the role of MRD negativity in long-term survival outcomes in patients with multiple myeloma. Blood Adv. 2020;4:5988-5999.
33. National Comprehensive Cancer Network. Clinical practice guidelines in multiple myeloma. 2020: Version 4.2020.
34. Perrot A, Lauwers-Cances V, Cazaubiel T, et al. Early versus late autologous stem cell transplant in newly diagnosed multiple myeloma: long-term follow-up analysis of the IFM 2009 Trial. Blood. 2020;136(Supp 1): 39.
35. Polsinelli B, Tsigkos S, Naumann-Winter F, et al. Evolving prevalence of haematological malignancies in orpha designation procedures in the European Union. Orphanet J Rare Dis. 2017;12(1):17. doi: 10.1186/s13023-017-0567-7.
36. PRISMA Transparent Reporting of Systematic Reviews and Meta-analyses. Available at http://www.prisma-statement.org. Accessed 01 December 2021.
37. Richardson P, Jacobus M, Weller E, et al. Triplet therapy, transplantation, and maintenance until progression in myeloma. N Engl J Med. 2022;387:132-147.
38. Roshal M, Flores-Montero JA, Gao Q, et al. MRD detection in multiple myeloma: comparison between MSKCC 10-color single-tube and EuroFlow 8-color 2-tube methods. Blood Adv. 2017;1:728-732.

39. Royston DJ, Gao Q, Nguyen N, et al. Single-tube 10-fluorochrome analysis for efficient flow cytometric evaluation of minimal residual disease in plasma cell myeloma. Am J Clin Pathol. 2016;146:41-49.

40. San-Miguel JF, Avet-Loiseau H, Paiva B, et al. Sustained minimal residual disease negativity with daratumumab in newly diagnosed multiple myeloma: MAIA and ALCYONE. Blood. 2021;Jul 16.

41. SEER 2023. Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: Myeloma. Available at: https://seer.cancer.gov/statfacts/html/mulmy.html. Accessed 29 January 2024.

42. Turesson I, Bjorkholm M, Blimark CH, et al. Rapidly changing myeloma epidemiology in the general population: increased incidence, older patients, and longer survival. Eur J Haematol. 2018;101(2):237-244.

43. Voorhees P, Kaufman J, Laubach J, et al. Daratumumab, lenalidomide, bortezomib and dexamethasone for transplant-eligible newly diagnosed multiple myeloma: the GRIFFIN trial. Blood. 2020;136(8):936-945.

44. US Department of Health and Human Services. Food and Drug Administration. Guidance for Industry. Hematologic malignancies: Regulatory considerations for use of minimal residual disease in development of drug and biological products for treatment. January 2020. https://www.fda.gov/media/134605/download. Accessed 19 November 2021.

45. US Department of Health and Human Services. Food and Drug Administration. Guidance for Industry. Clinical trial considerations to support accelerated approval of oncology therapeutics. March 2023. https://www.fda.gov/media/166431/download. Accessed 19 January 2024.

6.2 i2TEAMM References

1. 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 7:e456-e468, 2020
2. Rosinol L, Oriol A, Teruel AI, et al: Bortezomib and thalidomide maintenance after stem cell transplantation for multiple myeloma: a PETHEMA/GEM trial. Leukemia 31:1922-1927, 2017
3. Attal M, Lauwers-Cances V, Hulin C, et al: Lenalidomide, Bortezomib, and Dexamethasone with Transplantation for Myeloma. N Engl J Med 376:1311-1320, 2017
4. Voorhees PM, Sborov DW, Laubach J, 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. Lancet Haematol 10:e825-e837, 2023
5. Sonneveld P, Dimopoulos MA, Boccadoro M, et al: Daratumumab, Bortezomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. N Engl J Med 390:301-313, 2024
6. Facon T, Kumar S, Plesner T, et al: Daratumumab plus Lenalidomide and Dexamethasone for Untreated Myeloma. N Engl J Med 380:2104-2115, 2019
7. San-Miguel J, Dhakal B, Yong K, et al: Cilta-cel or Standard Care in Lenalidomide-Refractory Multiple Myeloma. N Engl J Med 389:335-347, 2023
8. Anderson KC, Auclair D, Kelloff GJ, et al: The Role of Minimal Residual Disease Testing in Myeloma Treatment Selection and Drug Development: Current Value and Future Applications. Clin Cancer Res 23:3980-3993, 2017
9. Kane RC, Bross PF, Farrell AT, et al: Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. Oncologist 8:508-13, 2003
10. Dimopoulos MA, Oriol A, Nahi H, et al: Daratumumab, Lenalidomide, and Dexamethasone for Multiple Myeloma. N Engl J Med 375:1319-1331, 2016
11. Administration USFD: Guidance for Industry: Hematologic Malignancies: Regulatory Considerations for Use of Minimal Residual Disease in Development of Drug and Biological Products for Treatment, 2020
12. Munshi NC, Avet-Loiseau H, Anderson KC, et al: A large meta-analysis establishes the role of MRD negativity in long-term survival outcomes in patients with multiple myeloma. Blood Adv 4:5988-5999, 2020
13. Perrot A, Lauwers-Cances V, Corre J, et al: Minimal residual disease negativity using deep sequencing is a major prognostic factor in multiple myeloma. Blood 132:2456-2464, 2018
14. Avet-Loiseau H, San-Miguel J, Casneuf T, 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. J Clin Oncol 39:1139-1149, 2021
15. Bahlis NJ, Dimopoulos MA, White DJ, et al: Daratumumab plus lenalidomide and dexamethasone in relapsed/refractory multiple myeloma: extended follow-up of POLLUX, a randomized, open-label, phase 3 study. Leukemia 34:1875-1884, 2020
16. Cavo M, San-Miguel J, Usmani SZ, et al: Prognostic value of minimal residual disease negativity in myeloma: combined analysis of POLLUX, CASTOR, ALCYONE, and MAIA. Blood 139:835-844, 2022
17. Landgren O, Iskander K: Modern multiple myeloma therapy: deep, sustained treatment response and good clinical outcomes. J Intern Med 281:365-382, 2017

18. San-Miguel J, Avet-Loiseau H, Paiva B, et al: Sustained minimal residual disease negativity in newly diagnosed multiple myeloma and the impact of daratumumab in MAIA and ALCYONE. Blood 139:492-501, 2022

19. Paiva B, Zherniakova A, Nunez-Cordoba JM, et al: Impact of treatment effect on MRD and PFS: an aggregate data analysis from randomized clinical trials in multiple myeloma. Blood Adv 8:219-223, 2024

20. Shi Q, Sargent DJ: Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. Int J Clin Oncol 14:102-11, 2009

21. Shi Q, Renfro LA, Bot BM, et al: Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials. Computational Statistics and Data Analysis 55, 2011

22. Durie BG, Harousseau JL, Miguel JS, et al: International uniform response criteria for multiple myeloma. Leukemia 20:1467-73, 2006

23. Burzykowski T, Molenberghs G, Buyse M: The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. Journal of the Royal Statistical Society Series A: Statistics in Society 167, 2004

24. Lahuerta JJ, Paiva B, Vidriales MB, et al: Depth of Response in Multiple Myeloma: A Pooled Analysis of Three PETHEMA/GEM Clinical Trials. J Clin Oncol 35:2900-2910, 2017

25. Munshi N, Loiseau HA, Anderson KC, et al: Expanded meta-analysis confirms the association between MRD and long-term survival outcomes in multiple myeloma. Amer. Soc. Hematol., 2019

26. Merz M, Salwender H, Haenel M, et al: Subcutaneous versus intravenous bortezomib in two different induction therapies for newly diagnosed multiple myeloma: an interim analysis from the prospective GMMG-MM5 trial. Haematologica 100:964-9, 2015

27. Stadtmauer EA, Pasquini MC, Blackwell B, et al: Autologous Transplantation, Consolidation, and Maintenance Therapy in Multiple Myeloma: Results of the BMT CTN 0702 Trial. J Clin Oncol 37:589-597, 2019

28. Dimopoulos MA, Gay F, Schjesvold F, et al: Oral ixazomib maintenance following autologous stem cell transplantation (TOURMALINE-MM3): a double-blind, randomised, placebo-controlled phase 3 trial. Lancet 393:253-264, 2019

29. Rosinol L, Oriol A, Rios R, et al: Bortezomib, lenalidomide, and dexamethasone as induction therapy prior to autologous transplant in multiple myeloma. Blood 134:1337-1345, 2019

30. Rosinol L, Oriol A, Teruel AI, et al: Superiority of bortezomib, thalidomide, and dexamethasone (VTD) as induction pretransplantation therapy in multiple myeloma: a randomized phase 3 PETHEMA/GEM study. Blood 120:1589-96, 2012

31. Morgan GJ, Davies FE, Gregory WM, et al: First-line treatment with zoledronic acid as compared with clodronic acid in multiple myeloma (MRC Myeloma IX): a randomised controlled trial. Lancet 376:1989-99, 2010

32. Mina R, Musto P, Rota-Scalabrini D, et al: Carfilzomib induction, consolidation, and maintenance with or without autologous stem-cell transplantation in patients with newly diagnosed multiple myeloma: pre-planned cytogenetic subgroup analysis of the randomised, phase 2 FORTE trial. Lancet Oncol 24:64-76, 2023

33. Facon T, Lee JH, Moreau P, et al: Carfilzomib or bortezomib with melphalan-prednisone for transplant-ineligible patients with newly diagnosed multiple myeloma. Blood 133:1953-1963, 2019

34. Facon T, Venner CP, Bahlis NJ, et al: Oral ixazomib, lenalidomide, and dexamethasone for transplant-ineligible patients with newly diagnosed multiple myeloma. Blood 137:3616-3628, 2021

35. Mateos MV, Dimopoulos MA, Cavo M, et al: Daratumumab plus Bortezomib, Melphalan, and Prednisone for Untreated Myeloma. N Engl J Med 378:518-528, 2018

36. Mateos MV, Oriol A, Martinez-Lopez J, et al: Bortezomib, melphalan, and prednisone versus bortezomib, thalidomide, and prednisone as induction therapy followed by maintenance treatment with bortezomib and thalidomide versus bortezomib and prednisone in elderly patients with untreated multiple myeloma: a randomised trial. Lancet Oncol 11:934-41, 2010

37. Mateos MV, Martinez-Lopez J, Hernandez MT, et al: Sequential vs alternating administration of VMP and Rd in elderly patients with newly diagnosed MM. Blood 127:420-5, 2016

38. Palumbo A, Chanan-Khan A, Weisel K, et al: Daratumumab, Bortezomib, and Dexamethasone for Multiple Myeloma. N Engl J Med 375:754-66, 2016

39. Attal M, Richardson PG, Rajkumar SV, 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. Lancet 394:2096-2107, 2019

40. Dimopoulos M, Quach H, Mateos MV, et al: Carfilzomib, dexamethasone, and daratumumab versus carfilzomib and dexamethasone for patients with relapsed or refractory multiple myeloma (CANDOR): results from a randomised, multicentre, open-label, phase 3 study. Lancet 396:186-197, 2020

6.3 FDA References

1. American Cancer Society, 2024, Key Statistics About Multiple Myeloma https://www.cancer.org/cancer/types/multiple-myeloma/about/what-is-multiple-myeloma.html

2. National Cancer Institute, 2024, Cancer Stat Facts: Multiple Myeloma, 2024, https://seer.cancer.gov/statfacts/html/mulmy.html.

3. Code of Federal Regulations, 21 CFR Part 314 Subpart H: https://www.ecfr.gov/current/title-21/chapter-I/subchapter-D/part-314/subpart-H

4. Kumar S, Paiva B, Anderson KC, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol. 2016;17:e328-346.

5. Citation: Ghandili, Susanne, Katja C. Weisel, Carsten Bokemeyer, and Lisa Beatrice Leypoldt. "Current treatment approaches to newly diagnosed multiple myeloma." Oncology Research and Treatment 44, no. 12 (2021): 690-699

6. Baines AC, Yazdy MS, Kasamon YL, Ershler R, Jen EY, Kanapuru B, Richardson NC, Lane A, Carioti T, Theoret MR, Pazdur R, Gormley NJ. Minimal Residual Disease Data in Hematologic Malignancy Drug Applications and Labeling: An FDA Perspective. Clin Cancer Res. 2023 Aug 1;29(15):2748-2752.

7. Administration USFDA: Guidance for Industry: Hematologic Malignancies: Regulatory Considerations for Use of Minimal Residual Disease in Development of Drug and Biological Products for Treatment, 2020

8. Shi Q, Flowers CR, Hiddemann W, Marcus R, Herold M, Hagenbeek A, Kimby E, Hochster H, Vitolo U, Peterson BA, Gyan E, Ghielmini M, Nielsen T, De Bedout S, Fu T, Valente N, Fowler NH, Hoster E, Ladetto M, Morschhauser F, Zucca E, Salles G, Sargent DJ. Thirty-Month Complete Response as a Surrogate End Point in First-Line Follicular Lymphoma Therapy: An Individual Patient-Level Analysis of Multiple Randomized Trials. J Clin Oncol. 2017 Feb 10;35(5):552-560.

9. Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval, July 2020: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pathological-complete-response-neoadjuvant-treatment-high-risk-early-stage-breast-cancer-use

10. Paiva B, Manrique I, Rytlewski J, et al. Time-Dependent Prognostic Value of Serological and Measurable Residual Disease Assessments after Idecabtagene Vicleucel. *Blood Cancer Discov*. 2023;4(5):365-373

11. Shi, Qian, Lindsay A. Renfro, Brian M. Bot, Tomasz Burzykowski, Marc Buyse, and Daniel J. Sargent. "Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials." *Computational Statistics & Data Analysis* 55, no. 9 (2011): 2748-2757.

7 Appendix

7.1 University of Miami Appendix

Statistical Analysis Plan (SAP)

Study Title: **Evaluating Minimal Residual Disease as Intermediate Clinical Endpoint for Multiple Myeloma**

Principal Investigator (Lead): **C. Ola Landgren, MD PhD**

Co- Principal Investigator (Statistician): **Sean M Devlin, PhD**

SAP Version **1.3**

SAP Date **December 1, 2021**

This statistical analysis plan (SAP) describes the planned analysis and reporting for the study titled “Evaluating Minimal Residual Disease as Intermediate Clinmissical Endpoint for Multiple Myeloma.” The planned analyses outlined in this SAP will be reviewed by Dr. Landgren, Dr. Devlin, each of the collaborating agencies, and the United States Food and Drug Administration (FDA).

1 OBJECTIVES

Broadly, the objective of the study is to evaluate the extent to which an endpoint based on measurement of minimal residual disease (MRD) is an early clinical endpoint reasonably likely to predict clinical benefit in the context of the clinical development of treatments for multiple myeloma (reasonably likely endpoint). In support of this objective, this SAP describes plans for a retrospective meta-analysis of multiple clinical trials in which MRD was measured and clinical benefit was also measured by progression-free survival (PFS) and overall survival (OS).

Secondary objectives include the evaluation of the extent to which other endpoints based on MRD are also reasonably likely endpoints.

1.1 Primary Objectives

To evaluate whether MRD negativity at an *a priori* defined time point (to be jointly agreed upon by the principal investigator, the study statistician, collaborating agencies, and the FDA before the meta-analysis takes place) is a reasonably likely endpoint for clinical benefit as measured by PFS in newly diagnosed transplant eligible patients (NDTE) with multiple myeloma.

To evaluate whether MRD negativity at an *a priori* defined time point (to be jointly agreed upon by the principal investigator, the study statistician, collaborating agencies, and the FDA before the meta-analysis takes place and possibly different from the *a priori* defined time point described above) is a reasonably likely endpoint for clinical benefit as measured by PFS in newly diagnosed transplant ineligible (NDTIE) patients with multiple myeloma.

To evaluate whether MRD negativity at an *a priori* defined time point (to be jointly agreed upon by the principal investigator, the study statistician, collaborating agencies, and the FDA before the meta-analysis takes place and possibly different from the *a priori* defined time points described above) is a reasonably likely endpoint for clinical benefit as measured by PFS in patients with relapsed/refractory (RR) multiple myeloma.

1.2 Secondary Objectives

To evaluate whether MRD negativity at the *a priori* defined time point(s) described above is a reasonably likely endpoint for clinical benefit as measured by PFS in patients with newly diagnosed (ND) multiple myeloma, regardless of transplant eligibility (i.e., in the combined population of NDTE and NDTIE).

To evaluate whether MRD negativity at the *a priori* defined time point(s) described above is a reasonably likely endpoint for clinical benefit as measured by PFS in patients with multiple myeloma, regardless of whether the multiple myeloma is ND or RR (i.e., in the combined population of NDTE, NDTIE, and RR).

To evaluate whether assessments of MRD negativity at pre-specified time points other than the *a priori* defined time point(s) described above are reasonably likely endpoints for clinical benefit as measured by PFS.

To evaluate whether attainment of MRD negativity at least once is a reasonably likely endpoint for clinical benefit as measured by PFS.

To evaluate whether sustained MRD (e.g., attainment of MRD negativity twice, in succession) is a reasonably likely endpoint for clinical benefit as measured by PFS.

To evaluate whether MRD negativity is a reasonably likely endpoint for clinical benefit as measured by OS.

2 STUDY DESIGN AND DATA COLLECTION PROCEDURES

This study is a retrospective meta-analysis of multiple randomized, controlled, Phase 2 and Phase 3 confirmatory clinical trials. The reporting of this study will be guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (http://www.prisma-statement.org).

2.1 Trial Eligibility

- The clinical trial is a Phase 2 or Phase 3 randomized, controlled clinical trial that enrolled patients with NDTE, NDTIE, or RR multiple myeloma.
- MRD assays were performed in the clinical trial by multiparameter flow cytometry (MFC) and/or next-generation sequencing (NGS) in accordance with guidelines from the FDA and National Cancer Institute (NCI) as well as institutional standards of care for the treatment of patients with multiple myeloma (Landgren 2014, Roshal 2017, Royston 2016).
    - NGS analyses were conducted using the FDA-approved Adaptive Biotechnologies 2.0 diagnostic test (Harris 2018), which has a sensitivity of 10^{-5} or better.
    - MFC analyses were conducted using a 10-color method (Roshal 2017, Royston 2016). Note that the validation of this method using antibody (CD117, PC5.5, CD19, BV421, CD138 APC, CD56, PC7, CD45, APC-H7, and CD38 BV510) and fluorescent stain (κ fluorescein isothiocynate phycoerythrin) cocktails has been reported by Royston et al (2016), and standardized flow cytometric instrument settings were reported by Kalina et al (2012). MRD monitoring using this method has a sensitivity of 10^{-5} or better (Flores-Montero 2017).
- MRD negativity was specified as a primary, secondary, or exploratory endpoint in the clinical trial protocol.
- The clinical trial has a median follow-up of at least six months following the end of the time chosen to be the *a priori* defined time point for the assessment of MRD negativity.

Note that median follow-up will be determined by a Kaplan-Meier estimate of the censoring distribution.

- The clinical trial is not a study of a maintenance therapy. Note that such studies are excluded from this analysis as patients might have achieved MRD negativity prior to randomization.
- The clinical trial does not have a primary endpoint such as safety, toxicity, quality of life, or feasibility.

2.2 Identification of Clinical Trials

The following sources are to be searched to identify clinical trials:

- PubMed: (https://pubmed.ncbi.nlm.nih.gov)
- Clinical trial registries:
  - ClinicalTrials.gov: (https://www.clinicaltrials.gov)
  - The ISRCTN registry: (https://www.isrctn.com)
  - EU Clinical Trials Register: (https://www.clinicaltrialsregister.eu)
  - Australian New Zealand Clinical Trials Registry: (https://www.anzctr.org.au)
- Cooperative Groups' websites
- Research organization meeting websites
- Review of citations in publications including meta-analysis papers
- Other sources (e.g., personal communications)

Searches are to be restricted to randomized, controlled trials with human subjects for which the publication or other relevant documentation is written in English. For studies identified from all these sources, title and abstract review or full text review are to be performed to further assess the studies' eligibility for inclusion. The bibliographies of retained articles are to be examined to identify additional studies. A four-phase (identification, screening, eligibility, and included) flow diagram will be created and reported to show how to the final list of clinical trials was identified. The final trial lists are to be reviewed and approved by the principal investigator. All literature search procedures will be performed in accordance with the PRISMA guidelines for meta-analysis.

2.3 Data Collection Procedure

For clinical trials with sponsors that agree to participate by contributing the relevant clinical trial data to the meta-analysis, further details, including the full clinical trial protocol, data

transfer agreement, data dictionary and timelines will be provided. The data dictionary provides general information on (1) data formats; (2) set of variables being requested, variable definition and coding; and (3) data transfer procedure. Relevant clinical trial data items being requested for the meta-analysis include patient and disease characteristics at diagnosis and randomization, number of lines of prior therapy, information regarding occurrence of prior autologous stem cell transplant (ASCT), MRD evaluation technique, and follow up data on disease and outcomes.

Electronic data files and related documents from contributors of clinical trial data will be delivered via encrypted secure email to the Memorial Sloan Kettering Cancer Center (MSKCC). The contributors of clinical trial data may choose to use their own encrypted secure email system or use a secure file transfer service. Delivery of data files through a regular email under password-protection is acceptable but not encouraged.

2.4 Database Maintenance and Safety

All data files received, as well as the central database of Statistical Analysis Software (SAS) and Microsoft Excel datasets that will be created to combine data files from all trials, will be stored at MSKCC. The data repository will only be accessible to members of the study team and only internally from within MSKCC's computer network.

To protect patient confidentiality, data files will use only research identifiers; there are no paper or computer records with personally identifiable information such as subject name, date of birth, contact information, social security numbers, or medical record numbers. The only data items for subjects in the databases are to be dates of diagnoses, laboratory measurements, procedures and treatment administration, and disease progression and death.

2.5 Data Quality Control

The following aspects of the clinical trial datasets received will be assessed and compared with publications of the results of the clinical trial as a mechanism to control the quality of the data and analyses:

- Number of subjects randomized, number of subjects included in the primary analysis of the trial, number of subjects by treatment arm
- Data completion on key variables (date of randomization, date of treatment initiation, disease recurrence/progression, initiation date of first post-randomization anti-myeloma therapy, and survival data)
- Distribution of baseline characteristics by treatment assignment
- Determination of MRD at select time points, distribution of missing MRD values across study time points, and MRD assay and sensitivity
- Number of failures for progression-free survival and overall survival

Data queries may be sent to the sponsor of a clinical trial if there is:

- A large discrepancy in baseline characteristics between received data and published data
- Any discrepancy in treatment assignment between received data and published data
- Any discrepancy in the comparison of PFS by treatment arm between the received data and published data
- Errors or issues raised from assessment of the disease outcome data which need clarification

2.6 Primary Definition of Clinical Endpoints

The primary analyses will be performed based on the below definitions of OS and PFS. In the primary analyses, all time-to-event (TTE) variables are to be calculated from date of randomization, and the primary analyses will be intention-to-treat (ITT) analyses.

2.6.1 OS

OS should be measured from the date of randomization to the date of death due to any cause. Subjects who are lost to follow-up will should be censored at the last known alive date prior to the time of lost to follow-up. Subjects who are still alive at the study clinical cutoff date for the analysis should be censored at the last known alive date. The definition of OS and the corresponding censoring status definition that was used in the original SAP of each respective clinical trial included in the meta-analysis will be used for the primary analysis.

2.6.2 PFS

PFS should be defined as the duration from the date of randomization to either progressive disease, according to the International Myeloma Working Group (IMWG) response criteria, or death, whichever occurs first. Subjects who are lost to follow-up should be censored at the date of the last disease assessment prior to the time of lost to follow-up. The definition of PFS and the corresponding censoring status definition that was used in the original SAP of each respective clinical trial included in the meta-analysis will be used for the primary analysis. Sensitivity analyses will investigate the analysis results when different definitions of censoring for PFS are used.

2.7 Primary Definition of Endpoint Based on MRD

It is anticipated that in most clinical trials included in the meta-analysis, MRD will have been assessed at multiple time points following randomization for many subjects. For the purpose of the construction of a primary definition of an endpoint based on MRD, MRD negativity status will be assigned to each subject based on results of that subject's MRD assessments, if any, with reference to an *a priori* defined time point following randomization along with a corresponding window of time (e.g., a time point of 9 months after randomization +/- 3 months) (see details of the algorithm for the primary definition of the endpoint below).

Prior to any analysis, for each time point and corresponding window under consideration (3 months +/- 2 months, 6 months +/- 2 months, 9 months +/- 3 months, and 12 months +/- 3 months), the number of patients who achieve MRD negativity as per the endpoint described

above and the number of patients with a missing MRD status for that endpoint (e.g., because no MRD assessments occurred within the window; see details of the algorithm for the primary definition of the endpoint below) will be reported for each clinical trial available for inclusion in the meta-analysis. The principal investigator, the study statistician, the collaborating agencies, and the FDA will jointly select one time point and one corresponding window for each of the NDTE, NDTIE, and RR settings to be the *a priori* defined time point to be used for the primary definition of the endpoint based on MRD for that respective setting; these jointly selected *a priori* defined time points may differ. The time points and windows that are not selected will be considered for secondary analyses. This selection of the *a priori* defined time point(s) will be made before investigating any of the primary or secondary objectives.

The details of the algorithm for the primary definition of the endpoint based on MRD for a given patient is below; the algorithm assigns one of three values (MRD-negative, MRD-positive, or missing) for that endpoint.

1. If the patient does not achieve an IMWG response of complete response (CR) or better by the end of the time window under consideration, the patient is assigned a value of **MRD-positive** for the endpoint.
2. If the patient achieves an IMWG response of complete response (CR) or better by the end of the time window under consideration:
    a. If the patient does not have any MRD assessments in the time window under consideration:
        i. If the patient has an IMWG response of progressive disease or dies by the end of the time window under consideration, the patient is assigned a value of **MRD-positive** for the endpoint.
        ii. If the patient does not have an IMWG response of progressive disease or die by the end of the time window under consideration, the patient is assigned a value of **missing** for the endpoint.
    b. If the patient does have at least one MRD assessment in the time window under consideration, the patient is assigned a value for the endpoint according to the result of the MRD assessment closest in time to the time point that was used to define the middle of the time window under consideration (e.g., for the time point and window of 9 months +/- 3 months, the MRD assessment closest to the time point of 9 months is used to assign a value for the endpoint), as follows: if the MRD assessment closest in time to that time point indicated that the patient achieved MRD negativity according to the protocol for the relevant clinical trial in which the patient participated, then the patient is assigned a value of **MRD-negative** for the endpoint; otherwise, the patient is assigned a value of **MRD-positive** for the endpoint.

Notes regarding planned analyses of this endpoint are below:

1. The analysis population for the primary analysis is ITT.

2. Generally, in the analysis, patients with a value for the endpoint of missing will be considered as MRD-positive unless otherwise noted.
3. A sensitivity analysis will be conducted that will consider an alternative version of the endpoint based on MRD; in that sensitivity analysis, the algorithm above will be repeated except with the response of CR or better being replaced with a response of VGPR or better.
4. A sensitivity analysis will be conducted that will consider an alternative version of the endpoint based on MRD; in that sensitivity analysis, the algorithm above will be repeated except that in order to have assigned a value of MRD-negative for the endpoint, an MRD assessment indicating that the patient achieved MRD negativity must occur after an IMWG response of CR or better.
5. A sensitivity analysis will be conducted that will consider an alternative version of the endpoint based on MRD; in that sensitivity analysis, the algorithm above will be repeated except that, instead of using MRD negativity according to the protocol for the relevant clinical trial in which the patient participated, specific thresholds (e.g., 10^{-4} or 10^{-5}) will be applied to the results of the MRD assessment.
6. If, according to the above algorithm, more than 20% of the patients in a clinical trial are assigned a value of missing for the endpoint, that clinical trial will be excluded from the primary analysis. However, such clinical trials will be included in a sensitivity analysis.

2.8 Secondary Definitions of Endpoint Based on MRD

To evaluate whether attainment of MRD negativity at least once is a reasonably likely endpoint for clinical benefit as measured by PFS, a secondary definition of an endpoint based on MRD will be used; this endpoint will be defined similarly as to the primary definition, with the exception that MRD negativity that occurs at any time (i.e., not necessarily within a pre-specified 6-month window), if the patient is also understood to be experiencing an IMWG response of CR or better at that time, results in the assignment of a value of MRD-negative for the endpoint for that patient.

To evaluate whether sustained MRD (e.g., attainment of MRD negativity twice, in succession, at least six months apart) is a reasonably likely endpoint for clinical benefit as measured by PFS, another secondary definition of an endpoint based on MRD will be used; this endpoint will be defined similarly as to the above, except that two assessments indicating MRD negativity, in succession, if the patient is also understood to be experiencing an IMWG response of CR or better at both of those times, results in the assignment of a value of MRD-negative for the endpoint for that patient.

3 STATISTICAL METHODS

3.1 Analysis Populations/Units

At patient level

Individual patient data of the trials will be pooled to perform an analysis of individual-level association as part of the evaluation of MRD as a reasonably likely endpoint for clinical benefit as measured by PFS or OS. Unless specifically noted otherwise, patients will be analyzed

according to the treatment group to which they were assigned by randomization, regardless of the treatment actually received or any treatment error (i.e., the analyses will be ITT).

At trial level

Study or trial is the analysis unit for the analysis of trial-level association as part of the evaluation of MRD as a reasonably likely endpoint.

3.2 Summary of Trials and Populations

The following data will be summarized:

- Number of subjects randomized by trial
- Number of subjects analyzed by trial
- Patient demographics
- Patient characteristics at diagnosis (age, fluorescence in situ hybridization [FISH]/cytogenetic subtype)
- Patient characteristics at randomization (age, prior lines of therapy, prior ASCT, time from diagnosis to randomization)
- Treatment assignment by randomization
- Accrual period
- Median follow-up using a Kaplan-Meier estimate of the censoring distribution
- MRD ascertainment at select time points, along with the MRD missing rate
- MRD assay, along with the sensitivity of the assay
- Number of events for PFS and OS
- Median estimates or other descriptive statistics for PFS and OS

These data will be summarized by type of therapy (classes of therapy where at least one agent in a particular multiagent combination belongs to class) and by individual trial. Qualitative data will be presented as frequencies and percentages. Quantitative data will be summarized as mean, standard deviation, median, interquartile range and range. The distribution of TTE endpoints will be estimated using the Kaplan-Meier method.

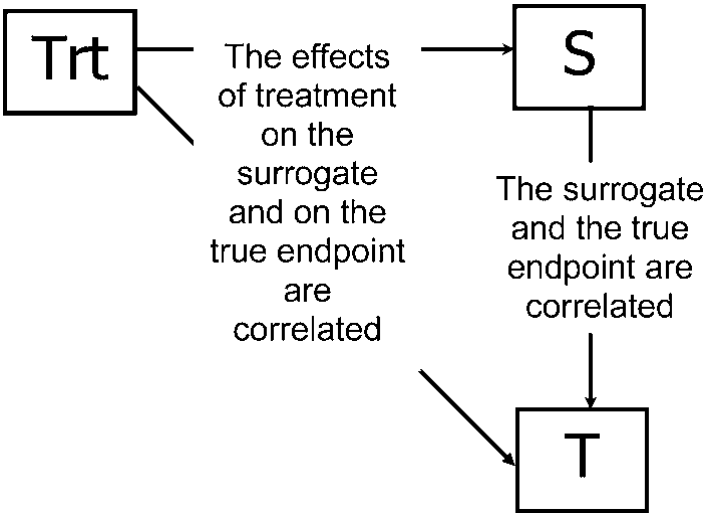
3.3 Correlation approach (Buyse et al. 2000)

A correlation approach has been described to evaluate an endpoint by assessing individual-level association and trial-level association as described in Figure 2 (Buyse et al. 2000, Buyse et al., 2008). Essentially, this approach requires two conditions be fulfilled: (1) the endpoint is prognostic for clinical benefit; and (2) a treatment effect on the endpoint in a clinical trial is predictive of a treatment effect on an endpoint traditionally used to measure clinical benefit in same clinical trial.

Standard correlation coefficients are the most commonly used approach to quantify statistical associations. An individual-level association can be demonstrated in a single trial, but an assessment of trial-level association is based on a meta-analysis of several randomized trials. The assessment of individual-level association can be quantified by a correlation coefficient, and the assessment of trial-level association can also be quantified by a (separate) correlation

coefficient; if both of these correlation coefficients are high, they are supportive of an endpoint's being reasonably likely to predict clinical benefit in the context of the clinical development of new treatments.

Figure 2. Correlation approach (Buyse et al., 2008)



3.4 Analyses

The primary analysis will use a correlation approach as described by Burzykowski and Buyse (Buyse et al. 2000, Burzykowski and Buyse 2001, Buyse et al. 2010).

3.4.1 Individual-level association: Correlation between 1) MRD and PFS, and 2) MRD and OS

A bivariate copula model will be used to estimate the correlation on individual patient data while allowing and accounting for trial-specific treatment effects on MRD and on PFS (Burzykowski et al, 2004). For each trial, the effect of treatment on MRD and PFS will be modeled using a logistic regression and Cox proportional hazards regression, respectively, where treatment effects are quantified as log odds ratios and log hazard ratios estimated with treatment as the only factor. With these models, a bivariate Plackett copula model will be constructed to estimate R^2, which takes into account patient-level correlation between the two endpoints. A similar approach will be used for OS.

3.4.2 Trial-level association: Correlation between treatment effects on 1) MRD and PFS, and on 2) MRD and OS

Two different approaches will be used to assess the trial-level association between treatment effects on MRD and PFS.

The first approach will use a weighted linear regression model on the estimated treatment effects on MRD (log odds ratios from logistic regression) and on PFS (log hazard ratio from Cox proportional hazards regression). The regression will be weighted by the inverse variances of the log odds ratio for MRD negativity. A sensitivity analysis will additionally weight by the respective sample size of each trial. R^2 will be used to describe the proportion of variance explained by the regression. In addition to the point estimate of R^2, a confidence interval will be reported.

The second approach will assess the trial-level association based on the two-stage copula model of Burzykowski et al, 2004. Under this approach, the bivariate Placket copula model estimates the treatment effects on MRD negativity rate and PFS within each of the trials simultaneously. The treatment effect is based on marginal models which are equivalent to logistic regression and Weibull proportional hazard model for MRD rate and PFS, respectively. Similarly as to the first approach, a point estimate and confidence interval for a correlation coefficient will be reported.

The same two approaches described in the preceding two paragraphs will be repeated, with OS replacing PFS, to assess the trial-level association between MRD and OS.

3.4.3 Surrogate Threshold Effect

Burzykowski and Buyse (2006) introduced the concept of surrogate threshold effect (STE), which is related to the minimum treatment effect on one endpoint required to predict a non-zero treatment effect for another endpoint. An STE is computed based on a linear regression model between treatment effects; the STE is given by the intersection of the 95% prediction limits obtained from the model and the x-axis (corresponding to no treatment effect for the latter endpoint). Respective STE analyses will be performed for MRD negativity for PFS and OS.

3.4.4 Concordance of significance

The proportion of trials for which the same conclusion is reached for both MRD negativity and PFS in terms of the estimated treatment effect (i.e., significant for both endpoints, or non-significant for both endpoints) will be reported. If the original trial included stratification factors, these factors will be included in this comparison. A similar approach will be used for OS

3.4.5 Studies with three arms

For randomized trials with three treatment arms (i.e., two experimental arms, Arm A and Arm B, and a standard-of-care arm, Arm C), the analysis of trial-level association will consider the two pairwise comparisons (Arm A vs Arm C; Arm B vs Arm C) in the primary analysis. Sensitivity analyses will be conducted that only include one of the two pairwise comparisons.

3.4.6 Sensitivity Analysis

The planned sensitivity analyses include:

- A sensitivity analysis will use a weight defined by the respective sample size of each trial when estimating the trial-level correlation for the weighted linear regression approach.

- A sensitivity analysis will be conducted that will consider an alternative version of the endpoint based on MRD; in that sensitivity analysis, the algorithm from Section 2.7 will be repeated except with the response of CR or better being replaced with a response of VGPR or better.
- A sensitivity analysis will be conducted that will consider an alternative version of the endpoint based on MRD; in that sensitivity analysis, the algorithm from Section 2.7 will be repeated except that in order to have assigned a value of MRD-negative for the endpoint, an MRD assessment indicating that the patient achieved MRD negativity must occur after an IMWG response of CR or better.
- In the subset of trials with available data, the primary analysis will be repeated considering MRD assessments that occur at the time of suspected IMWG response of CR or better instead of at a fixed time point following randomization.
- Single imputation for the MRD endpoint will be evaluated based on the following rules.

Table 1. Rules to determine MRD status for patients in CR/sCR who have missing MRD samples/results during the pre-defined time window

MRD-status before the pre-defined time window	MRD-status following the pre-defined time window	Patient status during or immediately following the pre-defined time window	Imputed MRD status
No sample obtained	No sample obtained	N/A	MRD positive
MRD negative[a]	MRD negative[b]	N/A	MRD negative
MRD negative[a]	No sample obtained	Censored due to clinical data cut AND no evidence of PD AND no evidence of subsequent systemic therapy	MRD negative
MRD negative[a]	MRD positive[b]	N/A	MRD positive
MRD negative[a]	No sample obtained	Disease progression OR start of subsequent systemic therapy	MRD positive
[a]last MRD results before pre-defined time window [b]first MRD results after the pre-defined time window			

- Multiple imputation for patients who have a missing MRD status. Imputation will be based on the patient's age, FISH/cytogenetic subtype, and prior lines of therapy (when available). All imputation will be conducted intra-trial and within the randomized arm.

- Leave-one-out cross-validation, which compares the predicted with the observed log hazard ratios on PFS on the basis of the estimated trial-level model that leaves one trial out at a time, will used to assess the prediction performance of the regression model.
- Leave-one-out estimation, which re-estimates the R^2 when one trial is excluded at a time, will be used to identify potential influential trials.
- Exclusion of trials exhibiting extreme treatment benefits on either MRD or PFS.
- Adjustment for prognostic factors (e.g., the stratification factors at randomization) in the estimation of the treatment effects on MRD and PFS.
- A sensitivity analysis will include the trials for which more than 20% of patients were assigned a value of missing for the primary endpoint definition based on MRD and were therefore excluded from the primary analysis.
- A sensitivity analysis will include the trials for which the median follow-up is not at least six months longer than the primary MRD assessment time point and were therefore excluded from the primary analysis.
- Sensitivity analyses will evaluate various PFS censoring definitions. These definitions will be evaluated in the subset of trials with sufficient information on the proposed modified censoring rule. These include:
    - Patients who start subsequent antimyeloma therapies without disease progression will be censored at the last disease assessment before the start of subsequent therapies.
    - Patients who have 2 or more consecutive missing visits will be censored at the date of last disease assessment prior to the first missed visit.

3.4.7 Subgroup Analysis

Multiple subgroup analyses by patient and disease characteristics are planned; the planned subgroups were chosen based on review of the literature and direct input from this study's principal investigator, collaborating agencies, and the FDA.

3.4.7.1 By Patient Population

Subgroup analysis will be conducted since MRD may be different for different types of patient populations.

Analysis are planned for the following subgroups:

- ND
- NDTE versus NDTIE
- RR
- RR with 1-3 prior lines of therapy versus RR with 4 or more prior lines of therapy
- Subgroup analyses based on cytogenetic risk groups (high-risk versus non-high-risk)
- Subgroup analyses by age, with separate analyses for those <65 years vs ≥65 years
- Subgroup analyses in patients with and without extramedullary disease, in the subset of trials with available data

3.4.7.2 By MRD assessment technique

Subgroup analyses will be performed based on the MRD assessment technique in each study. The pre-defined analyses include:

- Separate analyses based on MRD assessment technique (NGS or MFC)
- Separate analyses based on different assay cutoffs ($<10^{-4}$, $<10^{-5}$, $<10^{-6}$)

3.4.8 Statistical Methods in Accounting for Heterogeneity

Multi-level analysis and subgroup analyses are planned to account for possible heterogeneity.

(1) Study or trial is the analysis unit for the analysis of trial-level association as part of the evaluation of MRD as a reasonably likely endpoint; the study or trial will also be used as a stratification factor for the analysis of the individual-level patient data.
(2) For the assessment of trial-level association between treatment effects, both fixed effects and random effects models will be applied. The fixed effects model assumes the relationship between the two treatment effects does not vary by type of therapy or patient risk stratum, whereas the random effects model allows and accounts for such differences across type of therapy or patient risk strata.
(3) Subgroup analyses by type of therapy and patient population will be conducted, as described in Section 3.4.7.

3.5 Tables, Listings, and Figures

The following will be provided for each meta-analysis:

- A list of the clinical trials included in the meta-analysis, including the title, treatment arms, accrual period, number of patients, and the number of events for primary TTE endpoints
- Tables regarding patient demographics and relevant clinical characteristics
- The resulting correlation coefficients from the assessment of trial-level association and individual-level association (e.g., between MRD negativity and PFS or between MRD negativity and OS)

The following figures will be provided:

- Kaplan–Meier estimates of OS and PFS by treatment arm separately for each included clinical trial
- A scatter plot of trial-specific proportion of patients who are MRD-negative at t-months (x-axis) by treatment arm and trial-specific Kaplan-Meier estimates of PFS at t-years (y-axis) by treatment arm along with the estimated regression line and correlation coefficient
- A scatter plot of treatment effect (odds ratio) on MRD negativity and treatment effect (hazard ratio) on PFS along with the estimated regression line and correlation coefficient (and a 95% confidence interval)
- A scatter plot of treatment effect (odds ratio) on MRD negativity and treatment effect (hazard ratio) on OS along with the estimated regression line and correlation coefficient (and a 95% confidence interval).

- Assessment of the prediction of the treatment effect (hazard ratio) on PFS on the basis of the estimated regression model at the trial level by leave-one-out cross validation.
- Forest plots of treatment effects on MRD and PFS for the clinical trials included in the meta-analysis

3.6 Analysis Software

All analysis will be performed using SAS Software version 9.4 or later (SAS Institute Inc, Carey, NC), or the R software (*www.r-project.org*).

3.7 Additional Follow up

Analyses may be repeated in the future if additional follow-up becomes available for one or more of the trials included in the meta-analysis.

Methods for analysis of individual-level and trial-level associations

1. Akristas MG, Siebert J: A test for partial correlation with censored astronomical data. Mon Not R Astron Soc. 1995. 278:919-919-924.

2. Baker, Stuart G, Sargent, Daniel J, Buyse, Marc, Burzykowski, Tomasz Predicting treatment effect from surrogate endpoints and historical trials: an extrapolation involving probabilities of a binary outcome or survival to a specific time. Biometrics 2012; 68: 1. 248-257

3. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. Nature reviews Clinical oncology. 2010; 7(6): 309-17.

4. Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D. & Geys, H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000; 1:49–67.

5. Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D. & Geys, H. Statistical validation of surrogate endpoints: problems and proposals. Drug Inf J 2000; 34: 447-454.

6. Burzykowski T, Molenberghs G, Buyse M, et al: Validation of surrogate end points in multiple randomized clinical trials with failure time endpoints. Appl Stat 2001; 50:405-422.

7. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharmaceutical Stat 2006. 5: 173-186.

8. Fleming, T.R, Prentice, R. L., Pepe, M. S. and Gledden, D. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. Statistics in Medicine 1994.13:955–968

9. Freedman, L.S., Graubard, B.I. and Schatzkin, A. Statistical validation of intermediate endpoints for chronic disease. Statistics in Medicine 1992. 11:167-78.

10. Ghosh D. On assessing surrogacy in a single trial setting using a semicompeting risks paradigm. Biometrics. 2009; 65:521-529.

11. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints.Stat Methods Med Res. 2008. 17(3):303-40.

12. Lin D.Y., Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. Stat Med. 1997. 16(13):1515-27.

13. Prentice, R.L. Surrogate endpoints in clinical trials: definitions and operational criteria. Statistics in Medicine 1989. 8:431-40.

14. Renfro LA, Shi Q, Sargent DJ, Carlin BP. Bayesian adjusted R2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. Stat Med 2012; 31:743--61.

15. Schemper, M. The relative importance of prognostic factors in studies of survival. Statistics in Medicine 1993. 12: 2377-82.

16. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review Stat Med. 2006 Jan 30;25(2):183-203.

Endpoints investigated for other oncologic diseases

1. Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Luck HJ, et al. 2008. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. J Clin Oncol. 26(12): 1987-92.

2. Burzykowski T, Buyse M, Yothers G, Sakamoto J, Sargent D. 2008. Exploring and validating surrogate endpoints in colorectal cancer. Lifetime data analysis. 2008; 14(1): 54-64.

3. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer. Lancet. 2000.356(9227):373-8.

4. Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent DJ, Miller LL, Elfring GL, Pignon JP, Piedbois P. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol. 2007.25(33):5218-24.

5. Buyse M, Michiels S, Squifflet P, Lucchesi KJ, Hellstrand K, Brune ML,Castaigne S, Rowe JM. Leukemia-free survival as a surrogate end point for overall survival in the evaluation of maintenance therapy for patients with acute myeloid leukemia in complete remission. Haematologica. 2011.96(8):1106-12.

6. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P, Swain SM, Prowell T, Loibl S,Wickerham DL, Bogaerts J, Baselga J, Perou C, Blumenthal G, Blohmer J, Mamounas EP, Bergh J, Semiglazov V, Justice R, Eidtmann H,Paik S, Piccart M, Sridhara R, Fasching PA, Slaets L, Tang S, Gerber B, Geyer CE Jr, Pazdur R, Ditsch N, Rastogi P, Eiermann W, von Minckwitz G. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014.384(9938):164-72.

7. de Gramont A, Hubbard J, Shi Q, O'Connell MJ, Buyse M, Benedetti J, et al. Association between disease-free survival and overall survival when survival is prolonged after recurrence in patients receiving cytotoxic adjuvant therapy for colon cancer: simulations based on the 20,800 patient ACCENT data set. J Clin Oncol. 2010; 28(3): 460-5.

8. Flaherty KT, Hennig M, Lee SJ, Ascierto PA, Dummer R, Eggermont AM, Hauschild A, Kefford R, Kirkwood JM, Long GV, Lorigan P, Mackensen A, McArthur G, O'Day S, Patel PM, Robert C, Schadendorf D. Surrogate endpoints for overall survival in metastatic melanoma: a meta-analysis of randomized controlled trials. Lancet Oncol. 2014 Mar;15(3):297-304.

9. Giessen C, Laubender RP, Ankerst DP, Stintzing S, Modest DP, Mansmann U, Heinemann V. Progression-free survival as a surrogate endpoint for median overall survival in metastatic colorectal cancer:literature-based analysis from 50 randomized first-line trials. Clin Cancer Res. 2013;19(1):225-35.

10. Halabi S, Rini B, Escudier B, Stadler WM, Small EJ. Progression-free survival as a surrogate endpoint of overall survival in patients withmetastatic renal cell carcinoma. Cancer. 2014 ,120(1):52-60.

11. Laporte S, Squifflet P, Baroux N, Fossella F, Georgoulias V, Pujol JL, Douillard JY, Kudoh S, Pignon JP, Quinaux E, Buyse M. Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: evidence from a meta-analysis of 2334 patients from 5 randomised trials. BMJ 2013; 3: 3. 03

12. Mauguen A, Pignon JP, Burdett S, Domerg C, Fisher D, Paulus R, Mandrekar SJ, Belani CP, Shepherd FA, Eisen T, Pang H, Collette L, Sause WT, Dahlberg SE, Crawford J, O'Brien M, Schild SE, Parmar M, Tierney JF, Le Pechoux C, Michiels S; Surrogate Lung Project Collaborative Group. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data. Lancet Oncol. 2013;14(7):619-26.

13. Michiels S, Le Maitre A, Buyse M, Burzykowski T, Maillard E, Bogaerts J, et al. Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data. The lancet oncology. 2009; 10(4): 341-50.

14. Oba K, Paoletti X, Alberts S, Bang YJ, Benedetti J, Bleiberg H, Catalano P, Lordick F, Michiels S, Morita S, Ohashi Y, Pignon JP, Rougier P, Sasako M, Sakamoto J, Sargent D, Shitara K, Cutsem EV, Buyse M, Burzykowski T; GASTRIC group. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. J Natl Cancer Inst. 2013;105(21):1600-7.

15. Özer-Stillman I, Strand L, Chang J, Mohamed AF, Tranbarger-Freier KE. Meta-analysis for the Association between Overall Survival and Progression-Free Survival in Gastrointestinal Stromal Tumor. Clin Cancer Res. 2015;21(2):295-302.

16. Paoletti X, Oba K, Bang YJ, Bleiberg H, Boku N, Bouché O, Catalano P, Fuse N, Michiels S, Moehler M, Morita S, Ohashi Y, Ohtsu A, Roth A, Rougier P, Sakamoto J, Sargent D, Sasako M, Shitara K, Thuss-Patience P, Van Cutsem E, Burzykowski T,Buyse M; GASTRIC group. Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: a meta-analysis. J Natl Cancer Inst. 2013;105(21):1667-70.

17. Petrelli F, Barni S. Surrogate endpoints in metastatic breast cancer treated with targeted therapies: an analysis of the first-line phase III trials. Med Oncol. 2014;31(1):776.

18. Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M, Labianca R, Seitz JF, O'Callaghan CJ, Francini G, Grothey A, O'Connell M, Catalano PJ, Blanke CD, Kerr D, Green E, Wolmark N, Andre T, Goldberg RM, De Gramont A. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol. 2005 Dec 1;23(34):8664-70.

7.2 i2TEAMM Appendix

i2TEAMM Appendix Table 1: Randomized Studies in Newly Diagnosed and Transplant Eligible MM Transferred IPD i2TEAMM

Publication Author (Year) Study Name Identifier Group or Owner	Sample size	No. of centers and/or countries	Study design / Treatment schema (Duration) [PFS/OS median follow-up (months)*]	Primary endpoint of original trial	MRD endpoint in original trial (2nd vs. exploratory)	MRD Method*	MRD Sensitivity level$	MRD testing timepoints *	Comments
Merz, et al (2015) [26] **GMMG MM5** 2010-019173-16[A] GMMG	604	31 centers and 75 sites / 1 country	Induction (up to 3 months): VCD vs PAd SC mobilization + Leukapheresis Single or tandem ASCT (if no nCR/CR) Maintenance (up to 2 years): Lenalidomide for 2 years vs Lenalidomide until CR or 2 years 1st Randomization prior to Induction (randomization included induction + maintenance combination) [58m / 58m]	i) RR (non-inferiority of VCD to Pad) ii) PFS (best treatment strategy)	Exploratory	Flow Cytometry	10-4	At screening and to confirm CR	
Stadtmauer et al. (2019) [27] **BMT CTN 0702** NCT01109004 BMT CTN	758	54 centers / 1 country	Transplant: 1) AHCT + AHCT vs 2) AHCT + RVD vs 3) AHCT Consolidation (1 arm: up to 12 weeks): RVD Maintenance (all arms: until toxicity, PD, withdrawal): Lenalidomide 1st Randomization prior to first transplant [38m / 38m]	PFS at 38 months	Exploratory	Flow Cytometry	10-4, 10-5, 10-6	Prior to maintenance and at end of maintenance.	Although in the study's publication, only 10-4 data were reported, study team transferred additional data where the specimens were re-tested for MRD at higher sensitivity level (10-5, 10-6) which were not included in the previous publication. Sensitivity level varies across patients and timepoints
Cavo et al. (2020)[1] **EMN02 HO95 MM** NCT01208766	1197	172 centers / 13 countries	Induction (up to 12 weeks): VCD Intensification: VMP (up 24 weeks) vs HDM vs HDM + HDM Consolidation: VRD (up to 8 weeks) vs No Consolidation	PFS	Exploratory	Flow Cytometry	10-4, 10-5	-Screening, after 4th VCD, after 4th VMP, after each	Sensitivity level varies across patients and timepoints

EMN/Hovon			Maintenance (until PD): Lenalidomide 1st Randomization prior to Intensification 2nd Randomization prior to Consolidation [61m / 60m]					course of HDM, after 2nd VRD, and every 6 months during maintenance when there is immunofixation negativity in serum and urine.	
Dimopoulos et al. (2019) [28] **C16019 TOURMALINE-MM3** NCT02181413 Takeda	656	167 sites / 30 countries	[Prior to enrollment: Induction: Standard-of-care Intensification: HDM At least PR to be enrolled] Maintenance (up to 2 years): Ixazomib vs Placebo 1st Randomization prior to maintenance [31m / NA]	PFS	Secondary	Flow Cytometry	10-5	Pts in CR/VGPR get additional collection at cycle 13 and end of tx.	Only data of MRD negativity status based on 10-5 cutoff were transferred.
Rosinol et al. (2019) [29] **GEM2012MENOS 65** NCT01916252 GEM/Pethema	458	69 sites / 1 country	Induction (up to 24 weeks): VRD Conditioning Regimen: BUMEL vs MEL-200 Maintenance (up 8 weeks): VRD 1st Randomization prior to Induction [30m / 30m]	PFS	Secondary	Next Generation Flow	10-5, 10-6	Start and end of induction, to confirm CR Before and after consolidation	Only data of MRD negativity status based on 10-5 and 10-6 cutoff were transferred. Sensitivity level varies across patients and timepoints
Rosinol et al. (2012) [30] **GEM2005MENOS 65** NCT00461747 GEM/Pethema	386	66 sites / 1 country	Induction (up to 24 weeks): 1) VBMCP/VBAD/B vs 2) TD vs 3) VTD ASCT + HDM Maintenance (up to 3 years): Interferon alfa-2b vs Thalidomide 1st Randomization prior to Induction 2nd Randomization prior to Maintenance	PFS	Exploratory	Flow Cytometry	10-4	Before HDT/ASCT, 100 days after HDT/ASCT	

			[114m / 114m]						
Attal et al. (2017) [3] **IFM DFCI2009** NCT01191060 IFM	700	69 centers / 3 countries	Induction (up to 9 weeks): RVD SC mobilization + Cyclophosphamide + G-CSF Consolidation: 1) RVD (up to 15 weeks) vs 2) Melphalan + AutoSCT + RVD (up to 6 weeks) Maintenance (up to 1 year): Lenalidomide 1st Randomization during 1st cycle of Induction [57m / 57m]	PFS	Exploratory	Next Generation Sequencing	10-4, 10-5, 10-6	If pt at least VGPR: -Arm A: at pre-maintenance and post-maintenance. -Arm B at pre-RVD cycle 4, pre-maintenance and post-maintenance	Sensitivity level varies across patients and timepoints
Morgan et al. (2010) [31] **Myeloma IX (intensive pathway)** ISRCTN68454111 MRC	1111	120 centers / 1 country	Induction (CVAD up to 12 weeks; CTD up to 18 weeks): 1) Clodronic acid (with CVAD or CTD) vs 2) Zoledronic acid (with CVAD or CTD) SC mobilization and harvest + HDM + AutoSCT Maintenance (until PD): Thalidomide vs No Thalidomide 1st Randomization prior to Induction 2nd Randomization prior to Maintenance [72m / 71m]	OS, PFS, ORR	Exploratory	Flow Cytometry	10-4	Baseline, post-treatment, 3months post-HDT, at relapse	
Mina et al. (2023)[32] **FORTE** NCT02203643 EMN	477	42 centers / 1 country	Induction: CCyd (up to 16 weeks) vs CRd (up to 16 weeks) vs CRd long treatment (up to 48 weeks) ASCT Consolidation: Cyd Maintenance: Lenalidomide vs Lenalidomide + Carfilzomib [43m / Not reached]	≥VGPR rate	Secondary	Next Generation Flow Flow Cytometry	10-5, 10-6	Diagnosis, at CR confirmation and after around 6 months, every six months in maintenan	Only data of MRD negativity status based on 10-5 and 10-6 cutoff were transferred. Sensitivity level varies across patients and timepoints, dependent on method

								ce, and at clinical relapse.	
Voorhees et al. (2023)[4] **GRIFFIN** NCT02874742 Janssen	207	35 sites / 1 country	Induction + Maintenance : DVRd vs VRd [Not reached / Not reached]	sCR rate	Secondary	Next Generation Sequencing	10-4, 10-5, 10-6	Screening, at CR/sCR/VGPR, after induction, post-ASCT consolidation, at 12 and 24 months during maintenance	

* per IPD transferred or supporting documents; $ Sensitivity level refers to cutoff used for the MRD negativity status determination by study owner in the transferred data (across all patients, all time points)

i2TEAMM Appendix Table 2: Randomized Studies in Newly Diagnosed and Transplant Ineligible MM Transferred IPD

Publication Author (Year) Study Name Identifier Group or Owner	**Sample size***	**No. of centers and/or countries**	**Study Design / Treatment Schema (Duration)**	**Primary endpoint of original trial**	**MRD endpoint in original trial (2nd vs. exploratory)**	**MRD Method***	**MRD Sensitivity level$**	**MRD testing timepoints ***	**Comments**
Facon et al. (2019) [33] **CLARION** NCT01818752 AMGEN	955	183 sites	Induction (up to 54 weeks): KMP vs VMP 1ST Randomization prior to Induction [23 m / 22m]	PFS	Exploratory	Next Generation Flow	10-5	Screening, at first CR/sCR, and end of treatment	Only data of MRD negativity status based on 10-5 cutoff were transferred.
Facon et al. (2021)[34] **C16014 TOURMALINE-MM2** NCT01850524 Takeda	705	157 sites / 8 countries	(up to 72 weeks): Ixazomib +Lenalidomide+Dexamthasone vs Placebo + Lenalidomide + Dexamethasone [54m / Not reached]	PFS	Secondary	Flow Cytometry	10-4	At CR, cycle 18 (if in CR)	

Mateos et al. (2018) [35] **ALCYONE** NCT02195479 Janssen	706	162 sites / 25 countries	Induction (up to 54 weeks): D-VMP vs VMP 1ST Randomization prior to Induction [40m / 40m]	PFS	Secondary	Next Generation Sequencing	10-4, 10-5, 10-6	Screening, confirmation of CR/sCR and at PD. For those with CR/sCR: additional collections at 12,18,24, & 30 months after first dose.	
Mateos et al. (2010) [36] **GEM2005MAS65** NCT00443235 GEM/Pethema	260	63 centers / 1 country	Induction (up to 31 weeks): VMP vs VTP Maintenance (up to 3 years): VP vs VT 1st Randomization prior to Induction 2nd Randomization prior to Maintenance [76m / 76m]	PFS	Secondary	Flow Cytometry	10-4	After induction	
Mateos et al. (2016) [37] **GEM2010MAS65** NCT00443235 GEM/Pethema	233	44 centers / 1 country	Induction: 1) Sequential VMP->Rd (up to 38 weeks) vs 2) Alternating VMP/Rd (up to 36 weeks) 1st Randomization prior to Induction [60m / 60m]	18m PFS, Safety	Exploratory	Flow Cytometry	10-5	Screening, to confirm CR, end of cycle 9 and end of treatment.	Only data of MRD negativity status based on 10-5 cutoff were transferred.
Facon et al. (2019) [6] **MAIA** NCT02252172 Janssen	737	176 sites / 14 countries	Induction (until PD): DRd vs Rd 1st Randomization prior to Induction [27m / 28m]	PFS	Secondary	Next Generation Sequencing	10-4, 10-5, 10-6	At suspected CR/sCR, 12, 18, 24, 30 months after first dose	
Morgan et al. (2010) [31] **Myeloma IX (non-intensive pathway)**	849	120 centers / 1 country	Induction (MP up to 24 weeks; CTDa up to 36 weeks): 1) Clodronic acid (with MP or CTDa) vs 2) Zoledronic acid (with MP or CTDa) SC mobilization and harvest + HDM + AutoSCT	OS, PFS, ORR	Exploratory	Flow Cytometry	10-4	Baseline, post-treatment, at relapse	

ISRCTN68454111 MRC			Maintenance (until PD): Thalidomide vs No Thalidomide 1st Randomization prior to Induction 2nd Randomization prior to Maintenance [73m / 69m]						

* per IPD transferred or supporting documents; $ Sensitivity level refers to cutoff used for the MRD negativity status determination by study owner in the transferred data (across all patients, all time points)

i2TEAMM Appendix Table 3: Randomized Studies in Relapsed and Refractory MM Transferred IPD

Publication Author (Year) Study Name Identifier Group or Owner	**Sample size***	**No. of centers and/or countries**	**Study Design / Treatment Schema (Duration)**	**Primary endpoint of original trial**	**MRD endpoint in original trial (2nd vs. exploratory)**	**MRD Method***	**MRD Sensitivity level$**	**MRD testing timepoints ***	**Comments**
Dimopoulos et al. (2016) [10] **POLLUX** NCT02076009 Janssen	569	NA	Induction (until PD): DRd vs Rd 1st Randomization prior to Induction [44m / 44m]	PFS	Secondary	Next Generation Sequencing	10-4, 10-5, 10-6	Baseline, to confirm CR, at PD. For CR pts, at 3 and 6 months post-CR	
Palumbo et al. (2016) [38] **CASTOR** NCT02136134 Janssen	498	115 centers / 16 countries	Induction (up to 24 weeks): DVd vs Vd 1st Randomization prior to Induction [39m / 40m]	PFS	Exploratory	Next Generation Sequencing	10-4, 10-5, 10-6	-Screening, to confirm CR and at PD. Also done for pts with VGPR and suspected daratumumab interference. For those with CR/VGPR done at	

								cycle 9 and cycle 15	
Attal et al. (2019)[39] **ICARIA** NCT02990338 Sanofi	307	102 sites / 24 countries	Induction (until PD): Pd vs IPd [12m / 12m]	PFS	Exploratory	Next Generation Sequencing	10-4, 10-5, 10-6	Screening, at time of CR, If the first MRD is positive, BMA collection for MRD is to be repeated 3 months later for late negativity (one additional sample can be collected if patient remains MRD positive). No more than 3 post treatment samples are to be obtained	
Dimopoulos et al. (2020)[40] **CANDOR** NCT03158688 Amgen	466	102 sites	Induction (until PD): KdD vs Kd [28m / Not reached]	PFS	Secondary	Next Generation Sequencing	10-4, 10-5, 10-6	Baseline, at CR, 12 months, 24 months (if CR)	

[A]Eudra CT number; * per IPD transferred or supporting documents; [$] Sensitivity level refers to cutoff used for the MRD negativity status determination by study owner in the transferred data (across all patients, all time points)

i2TEAMM Appendix Table 4: Trial-Level R2 Estimates of 9m MRDneg Rate - Sensitivity Analyses (Newly Diagnosed Transplant Eligible Population)

Sensitivity analyses	N of comparisons (N of subjects)	R^2_{WLS} (95% CI)	R^2_{Copula} (95% CI)
Per strict ITT population	11 (3734)	0.37 (0.00, 0.75)	0.44 (0.01, 0.88)
Alternative censoring rules for PFS	6 (1578)	NA*	NA*
Each 2-arm comparison with ≥75% of subjects with MRD endpoint status can be determined	13 (4271)	0.27 (0.00, 0.64)	0.35 (0.00, 0.77)

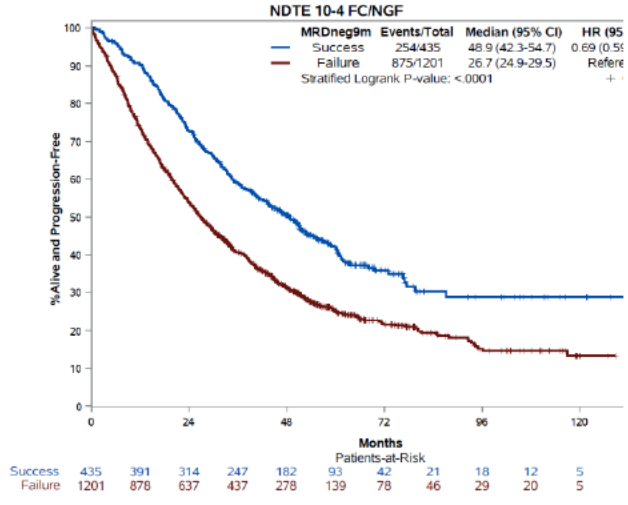
CI=confidence interval; ITT=intent-to-treat; MRD=minimal residual disease; NA=not applicable; PFS=progression-free survival;
*There are only 6 two-arm comparisons with sufficient MRD endpoint data and where the PFS could be derived with alternative censoring rules. The two-arm comparisons that dropped out were due to lack of data on alternative treatment.

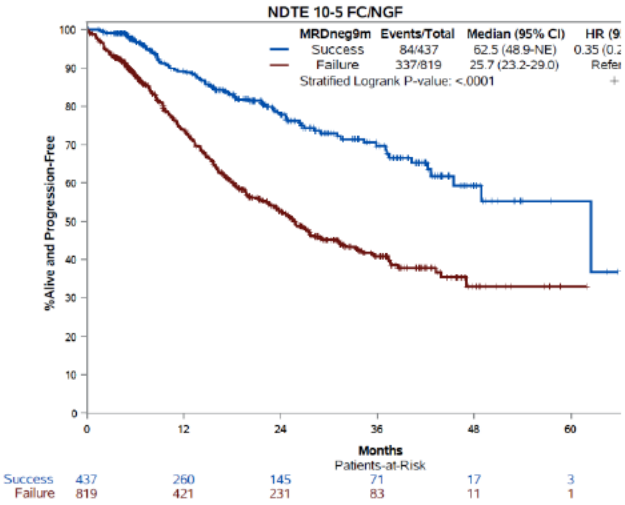
i2TEAMM Appendix Table 5: Trial-Level R2 Estimates of 9m MRDneg Rate - Subgroup Analyses (Newly Diagnosed Transplant Eligible Population)

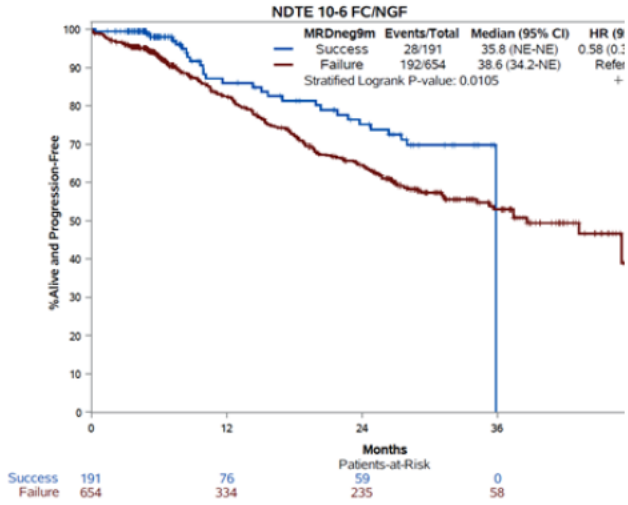
Subgroups	N of comparisons (N of subjects)	R^2_{WLS} (95% CI)	R^2_{Copula} (95% CI)
Treatment with IMiD	10 (2849)	0.40 (0.00, 0.80)	0.44 (0.00, 0.90)
Treatment with PI and IMiD	10 (2849)	0.40 (0.00, 0.81)	0.44 (0.00, 0.90)
MRD tested by MFC or NGF method	10 (3113)	0.22 (0.00, 0.61)	0.24 (0.00, 0.69)
ISS stage II disease	11 (1315)	0.19 (0.00, 0.56)	0.23 (0.00, 0.67)
Age <65 years	10 (2760)	0.26 (0.00, 0.64)	0.32 (0.00, 0.80)

CI=confidence interval; IMiD=immunomodulatory drug; ISS=International Staging System; MFC=multiparameter flow cytometry; MRD=minimal residual disease; NGF=next-generation flow; PI=proteasome inhibitor
Note: Trial-level R^2 estimates per subgroups with ≥10 two-arm comparisons with sufficient data (ie, ≥80% of subjects with MRD surrogate endpoint status determined as either “success” or “failure” status and with at least ≥50 subjects.

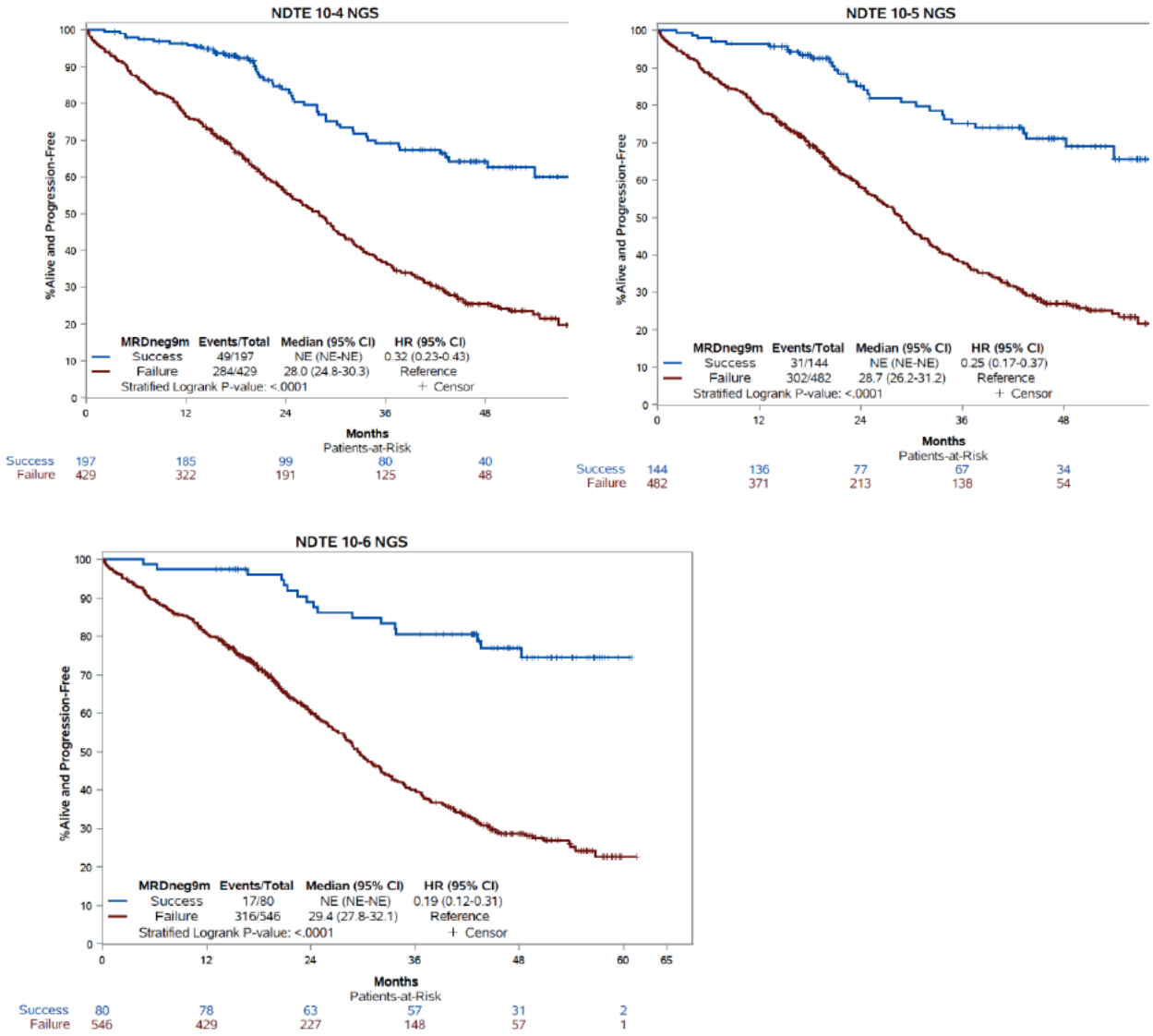
i2TEAMM Appendix Figure 1: Landmark Analysis Kaplan Meier Curves – Newly Diagnosed Transplant-Eligible Multiple Myeloma (Analysis by Multiparameter Flow Cytometry/Next-Generation Flow and MRD Sensitivity Level)



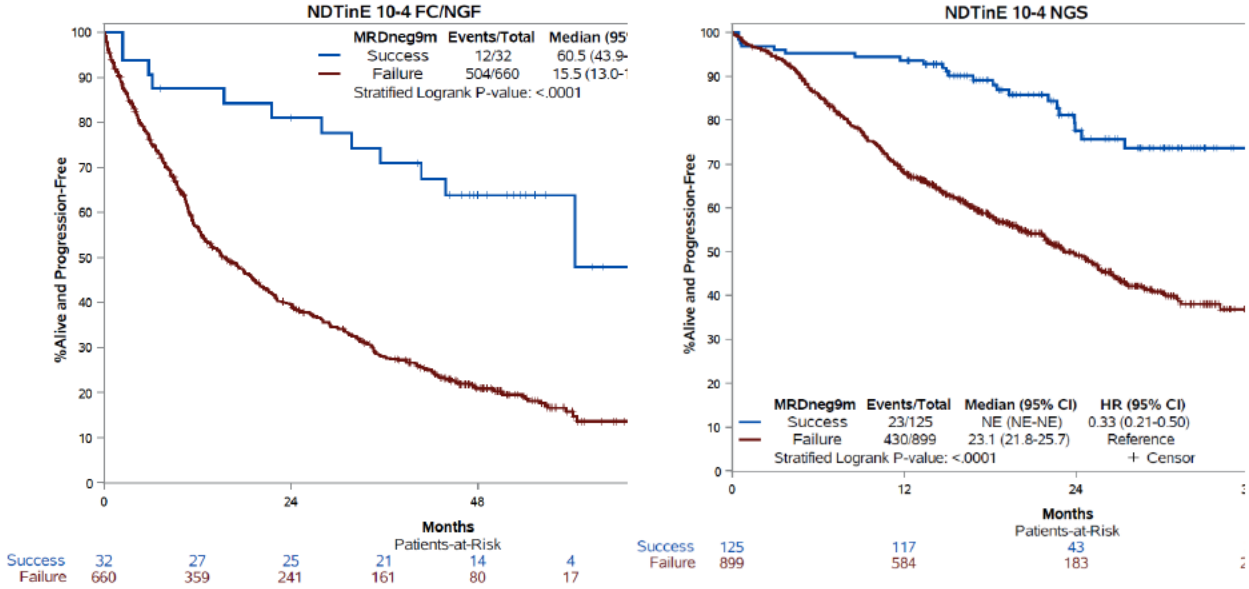




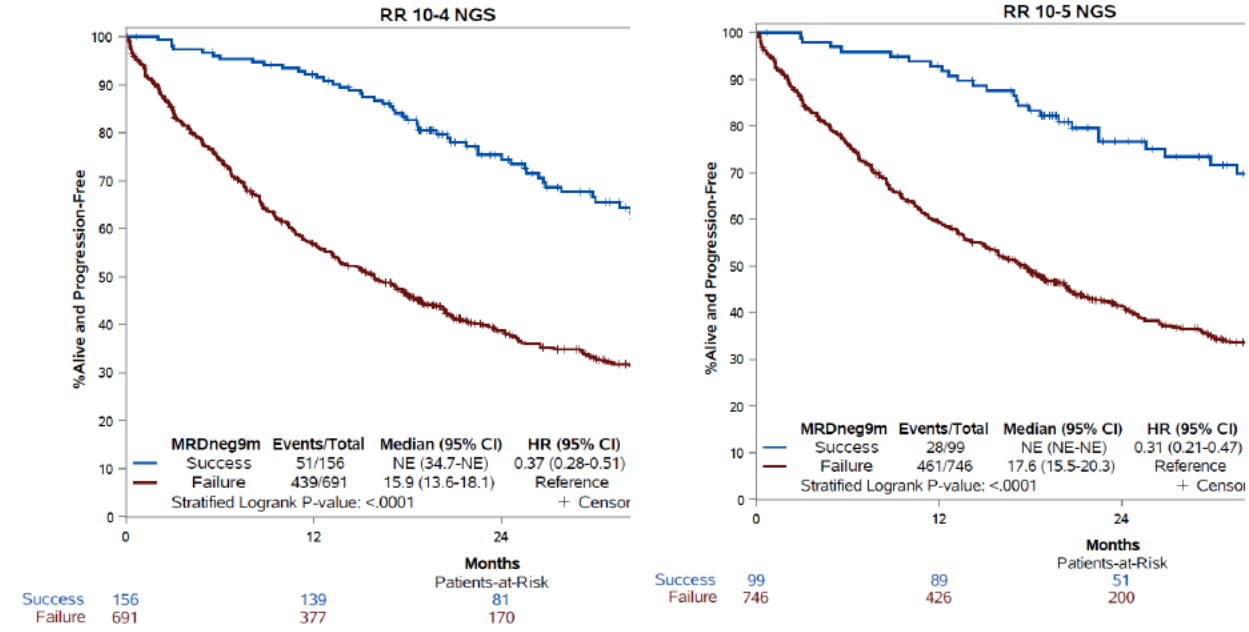
i2TEAMM Appendix Figure 2: Landmark Analysis Kaplan Meier Curves – Newly Diagnosed Transplant-Eligible Multiple Myeloma (Analysis by Next-Generation Sequencing and MRD Sensitivity Level)



i2TEAMM Appendix Figure 3: Landmark Analysis Kaplan Meier Curves – Newly Diagnosed Transplant-Ineligible Multiple Myeloma (Multiparameter Flow Cytometry/Next-Generation Flow vs Next Generation Sequencing and MRD Sensitivity Level)



i2TEAMM Appendix Figure 4: Landmark Analysis Kaplan Meier Curves – Relapsed Refractory Multiple Myeloma (Next Generation Sequencing and MRD Sensitivity Level)Appendix Figure 4: Landmark Analysis Kaplan Meier Curves – Relapsed Refractory Multiple Myeloma (Next Generation Sequencing and MRD Sensitivity Level)



7.3 FDA Appendix

7.3.1 Approved Therapies for MM.

FDA Appendix Table 1: New Therapies and Combinations Approved for Patients with Multiple Myeloma (MM)

Drug	Approval	Indication	Endpoint	Trial Design/Results [95%CI]
Velcade (bortezomib)	Treatment of patients with multiple myeloma			
	Accelerated (2003)	RRMM after at least 2 prior therapies and progressed on the last therapy	ORR	Single-arm trial ORR 27.7% [21, 35] mDOR: 365 days [224, NE]
	Regular (2005)	RRMM after at least 1 prior therapy	TTP/OS	RCT: V vs. dex TTP: HR=0.55 [0.40, 0.81] mTTP: 6.2 vs. 3.5 months OS: HR=0.57 [0.40, 0.81]
	Regular (2008)	Newly diagnosed MM	TTP/PFS	RCT: VMP vs. MP TTP: HR=0.54 [0.42, 0.70] mTTP: 20.7 vs. 15.0 months PFS: HR=0.61 [0.49, 0.76], mPFS: 18.3 vs. 14.0 months OS: HR=0.65 [0.51, 0.84] mOS: NR vs. 43.1 months
Doxil Liposomal (doxorubicin HCl)	Regular (2007)	In combination with bortezomib in patients who have not previously received bortezomib and have received at least one prior therapy.	TTP	RCT: Doxil + V vs. V TTP: HR=0.55 [0.43, 0.71] mTTP: 9.3 vs. 6.5 months
Revlimid (lenalidomide)	Treatment of adult patients with multiple myeloma in combination with dexamethasone			
With dex	Regular (2006)	RRMM, at least 1 prior line	TTP	RCT: Rd vs. dex Study 1: TTP: HR=0.285 [0.210, 0.386] mTTP: 13.9 vs. 4.7 months Study 2: TTP: HR=0.32 [0.240, 0.438] mTTP: 12.1 vs. 4.7 months
With dex	Regular (2015)	Newly diagnosed MM	PFS	RCT: Rd (continuous) vs. MPT PFS: HR=0.72 [0.61, 0.85] mPFS: 21.2 vs. 20.7 months
Single agent	Regular (2017)	Maintenance following auto-HSCT	PFS and OS	Study 1 (RCT): PFS: HR=0.38 [0.27, 0.54] mPFS: 33.9 vs. 19.0 months OS: HR=0.59 [0.44, 0.78] mOS: 111 vs. 84.2 months

				Study 2 (RCT): PFS: HR=0.50 [0.39, 0.64] mPFS: 41.2 vs. 23.0 months OS: HR=0.90 [0.72, 1.13] mOS: 105.9 vs. 88.1 months
Kyprolis (carfilzomib)	Accelerated (2012)	RRMM, one or more lines of therapy	ORR	Single-arm trial ORR 23% [18, 28] mDOR: 7.8 months [5.6, 9.2]
Kyprolis with Rd	Regular (2015)	RRMM, 1-3 prior lines	PFS /OS	RCT: KRd vs. Rd PFS: HR=0.69 [0.57, 0.83] mPFS: 26.3 vs. 17.6 months OS: HR: 0.79 [0.67, 0.95] mOS: 48.3 vs. 40.4 months
Kyprolis with dex	Regular (2016)	RRMM, 1-3 prior lines	PFS	RCT: Kd vs. Vd PFS: HR=0.533 [0.437, 0.651] mPFS: 18.7 vs. 9.4 months OS: HR=0.79 [0.65, 0.96] mOS: 47.6 vs. 40.0 months
Pomalyst (pomalidomide)	Accelerated (2013)	RRMM, at least 2 prior lines, including len and bortezomib	ORR	RCT: P vs Pd ORR: 7.4% [3.3, 14.1] vs. 29.2% [21.0, 38.5]
Pomalyst with dex	Regular (2015)	RRMM, at least 2 prior lines, including len and PI	PFS/OS	RCT: Pd vs. dex PFS: HR=0.45 [0.35, 0.59] mPFS: 3.6 vs. 1.8 months OS: HR=0.70 [0.54, 0.92] mOS: 12.4 vs. 8.0 months
Farydak (panobinostat) with Vd (Withdrawn 2021)	Accelerated (2015)	RRMM, at least 2 prior lines, including bortez and IMiD	PFS	RCT: PVd vs. Vd PFS: HR=0.52 [0.36, 0.76] mDOR: 10.6 vs. 5.8 months
Ninlaro (ixazomib) with Rd	Regular (2015)	RRMM, at least 1 prior line	PFS	RCT: Ixaz + Rd vs. placebo + Rd PFS: HR=0.74 [0.59, 0.94] mPFS: 20.6 vs. 14.7 months
Darzalex (daratumumab)	Accelerated (2015)	RRMM, at least 3 prior lines, including PI and IMiD	ORR	Single-arm trial ORR: 29% [20.8, 38.9] mDOR: NE [2.2, 13.1+ months]
Darzalex with Rd	Regular (2016)	RRMM, at least 1 prior line	PFS	RCT: DRd vs. Rd PFS: HR=0.37 [0.27, 0.52] mPFS: 45 vs. 17.5 months ORR: 91.3% vs. 74.6%
Darzalex with Vd*	Regular (2016)	RRMM, at least 1 prior line	PFS	RCT: DVd vs. Vd PFS: HR=0.39 [0.28, 0.53] mPFS: NE vs. 7.2 months ORR: 79.3% vs. 59.9%
Darzalex with Pd	Regular (2017)	RRMM, at least 2 prior lines, including len and PI	ORR	Single-arm trial ORR: 59.2% mDOR: 13.6 months [0.9+, 14.6]
Empliciti (elotuzumab) with Rd	Regular (2015)	RRMM, 1-3 prior lines	PFS	RCT: ERd vs. Rd PFS: HR=0.70 [0.57, 0.85] mPFS: 19.4 vs. 14.9 months OS: HR=0.82 [0.68, 1.00]

				mPFS: 48.3 vs 39.6 months
Empliciti (elotuzumab) with Pd	Regular (2018)	RRMM, at least 2 prior lines, including len and PI	PFS	RCT: EPd vs. Pd PFS: HR=0.54 [0.34, 0.86] mPFS: 10.3 vs. 4.7 months OS: HR=0.59 [0.37, 0.93], mOS: 29.80 vs. 17.41 months
Xpovio (selinexor) with dex	Accelerated (2019)	RRMM, at least 4 prior lines, refractory to 2 PIs, 2 IMiDs, and anti-CD38 mAb	ORR	Single-arm trial ORR: 25.4% [16.4, 36] mDOR 3.8 months [2.3, NE]
Xpovio with Vd	Regular (2020)	RRMM, at least 1 prior line	PFS	RCT: SVd vs. Vd PFS: HR=0.70 [0.53, 0.93] mPFS: 13.9 vs. 9.5 months
Darzalex with VMP	Regular (2018)	MM, newly diagnosed, transplant-ineligible	PFS	RCT: D-VMP vs. VMP PFS: HR=0.50 [0.38, 0.65] mPFS: 36.4 vs. 19.3 months OS: HR=0.60 [0.46, 0.80] mOS: NE vs. NE
Darzalex with Rd	Regular (2019)	MM, newly diagnosed, transplant-ineligible	PFS	RCT: DRd vs. Rd PFS: HR=0.56 [0.43, 0.73] mPFS NR vs. 31.9 months OS: HR=0.66 [0.53, 0.86] mOS: NE vs. NE
Darzalex with Kd Also includes cross labelled indication	Regular (2020)	RRMM, 1-3 prior lines	PFS	RCT: DKd vs. Kd PFS: HR=0.63 [0.46, 0.85] mPFS: NR vs. 15.8 months
Darzalex with VTd*	Regular (2019)	MM, newly diagnosed, transplant-eligible	PFS, sCR/CR (Day +100)	RCT: D-VTd vs. VTd sCR: 28.9% vs. 20.3% CR: 9.9% vs. 5.7% PFS: HR=0.47 [0.33, 0.67] mPFS: NR vs. NR
Darzalex Faspro (daratumumab and hyaluronidase)	Regular (2020)	RRMM, at least 3 prior lines, including PI and IMiD or PI/IMiD double-refractory	ORR, Max C_{trough}	RCT (Non-Inferiority): Dara IV vs. Dara SC ORR: 41% [35, 47] vs. 37% [31, 43]
Darzalex Faspro with VMP	Regular (2020)	MM, newly diagnosed, transplant-ineligible	ORR	Single-arm trial ORR: 88% [78, 95]
Darzalex Faspro with Rd	Regular (2020)	RRMM, at least 1 prior line	ORR	Single-arm trial ORR: 91% [81, 97]
Blenrep (belantamab mafodotin) Withdrawn (2023)	Accelerated (2020)	RRMM, 4 prior lines, including anti-CD38 mAb, PI, IMiD	ORR	Single-arm trial ORR: 31% mDOR: NR [NR, NR] 73% of responders had DOR ≥6 months
Sarclisa (isatuximab) with Pd	Regular (2020)	RRMM, at least 2 prior therapies, including len and PI	PFS	RCT: Isa-Pd vs. Pd PFS: HR=0.59 [0.44, 0.81] mPFS: 11.5 vs. 6.5 months

Sarclisa with Kd Also includes cross labelled indication	Regular (2021)	RRMM, 1-3 prior lines	PFS	RCT: Isa-Kd vs. Kd PFS: HR=0.548 [0.366, 0.822] mPFS: NR vs. 20.3 months
Pepaxto (melphalan flufenamide) Withdrawn (2024)	Accelerated (2021)	RRMM, at least 4 prior lines, refractory to PI, IMiD, anti-CD38 mAb	ORR	Single-arm trial ORR: 23.7% [15.7, 33.4] mDOR: 4.2 months [3.2, 7.6]
Abecma (idecabtagene vicleucel)	Regular (2021)	RRMM, at least 4 prior lines, including anti-CD38 mAb, PI, and IMiD	ORR	Single-arm trial ORR: 72% [62, 81] mDOR: 11 months [10.3, 11.4]
Darzalex Faspro with Pd	Regular (2021)	RRMM, at least 1 prior line including len and PI	PFS	RCT: Dara SC-Pd vs. Pd PFS: HR=0.63 [0.47, 0.85] mPFS: 12.4 vs. 6.9 months
Darzalex Faspro with Kd Also includes cross labelled indication	Regular (2021)	RRMM, 1-3 prior lines of therapy	ORR	Single-arm trial ORR: 84.8% [73.9, 92.5] mDOR: NR; 85.2% maintained response for ≥6 months and 82.5% for ≥9 months
Carvykti (Ciltacabtagene autoleucel)	Regular (2022)	RRMM, at least 4 prior lines including PI, IMiD, and anti-CD38	ORR	Single-arm trial ORR: 97.9% [92.7, 99.7] mDOR: 21.8 months [21.8, NE]
Teclistamab	Accelerated (2022)	RRMM, at least 4 prior lines including PI, IMiD, and anti-CD38	ORR	Single arm trial ORR: 61.8% (95% CI: 52.1, 70.9) DOR rate was 90.6% (95% CI: 80.3%, 95.7%) at 6 months and 66.5% (95% CI: 38.8%, 83.9%) at 9 months.
Elranatamab	Accelerated (2023)	RRMM, at least 4 prior lines including PI, IMiD, and anti-CD38	ORR	Single arm trial ORR: 57.7% (95% CI: 47.3%, 67.7%) mDOR: NR DOR rate at 6 months was 90.4% (95% CI: 78.4%, 95.9%) and at 9 months was 82.3% (95% CI: 67.1%, 90.9%).
Talquetamab	Accelerated (2023)	RRMM, at least 4 prior lines including PI, IMiD, and anti-CD38	ORR	Single arm trial <u>0.4mg/kg weekly</u> ORR: 73% (63.2%, 81.4%) mDOR: 9.5 months 0.8 mg/kg biweekly ORR: 73.6% (63.0%, 82.4%) mDOR: Not estimable

7.3.2 Details Regarding Methodology Used in Evaluating Surrogacy

Individual-level and trial-level associations are summarized in FDA Appendix Table 2.

FDA Appendix Table 2: Associations Estimated in the Surrogacy Methodology

Level	Measures Associated	Measure Quantifying Association	Question Addressed
Individual	MRD- CR and PFS/OS	Odds Ratio	Do patients who achieve MRD- CR tend to have longer PFS/OS than those who do not?
Trial	Odds ratio of MRD-CR and hazard ratio of PFS/OS	R^2	Do trials which observe larger treatment effects on MRD- CR tend to observe larger treatment effects on PFS/OS?

Associations between two variables are often reported in terms of the correlation coefficient *r*, where an *r* value of 0 reflects no association between two endpoints and an *r* of 1/-1 reflects complete dependence (positive or negative). The correlation coefficient *r* is most appropriate when describing the relationship between two normally-distributed endpoints. Because MRD-CR and PFS/OS are binary and time-to-event endpoints, a more generalized approach is needed to capture this dependence. One such approach is copula modeling, which allows modeling the dependence between two endpoints that may not be normally distributed. The odds ratio utilized at the individual-level is based on copula modeling.

Trial-level associations assume that larger treatment effects on the surrogate endpoint should translate to larger effects on the final endpoint. These associations are typically quantified by regressing the treatment effect for the final endpoint (log HR for PFS/OS) on the treatment effect for the surrogate endpoint (log odds ratio for MRD- CR). The resulting associations are quantified via the coefficient of determination R^2, which describes the proportion of variance in the dependent variable explained by the independent variable. In the analysis of surrogate endpoints, there are two commonly-used approaches for estimating R^2. One approach is to compute R^2 based on results from the copula model described above (R^2_{Copula}). The trial-level association can also be estimated without use of copula modeling. The typical approach for estimating this association is via weighted regression, where the weighting is inversely proportional to the sample size or variance of each trial. These estimates are referred to as R^2_{WLS}.

One applicant pre-specified thresholds for the trial-level associations. In addition, the i2TEAMM proposed criteria to "qualify a reasonably likely surrogate endpoint". These criteria were:

- *Criteria to Qualify a Validate Surrogate Endpoint (i2TEAMM only):* If either of R^2_{WLS} or R^2_{Copula} is at least 0.8 with lower bound of the 95% confidence interval greater than 0.6, and neither estimate is lower than 0.7.

- *Criteria to Qualify a Reasonably Likely Surrogate Endpoint (i2TEAMM only):* If either of R^2_{WLS} or R^2_{Copula} is at least 0.8 with lower bound of the 95% confidence interval greater than 0.5.

These approaches are generally appropriate for supporting development of a new validated surrogate or intermediate endpoint. Note that the goal for such analyses is often to show that both the patient level and trial-level association is high. This approach is aligned with the goal of "replacing" the true endpoint with the proposed surrogate, as high correlations at both levels would reflect that measuring the surrogate is nearly as informative as measuring the true endpoint. Such "replacement endpoints" would be considered validated surrogates. No formal criteria exist for development of intermediate endpoints for the purpose of accelerated approval. The evidentiary standard may be different for such endpoints, as the intermediate would not replace the true endpoint but simply serve to support accelerated approval earlier than the true endpoint can be measured.

Limitations of the STE

A number of caveats apply to estimation of STE. For instance, all calculations of the STE assume the future trial has an infinite sample size, resulting in 100% power for the true endpoint. Some calculations also assume that there is no estimation error in the meta-analysis, which can only be achieved when there are an infinite number of trials with infinite sample size in the meta-analysis. In addition, typically simplifying assumptions are made the preceding meta-analyses, and such assumptions typically underestimate the total variability in the estimates. Thus, while a theoretical threshold for what type of treatment effect on the surrogate endpoint would translate to the true endpoint, such thresholds could be inherently optimistic when applied to future trials.

Additional Limitations of Analyses of Surrogate Endpoints

In general, the approaches implemented in this document depend heavily on the validity of the data utilized. While this is true for any statistical analysis, the nature of meta-analysis is retrospective, such that methodology for data collection cannot be pre-specified. This is in contrast to a prospective clinical trial, for which the analysis is specified to correspond to carefully planned data collection.

In addition to the limitations outlined in Section 2.1.3, the following limitations apply to the results presented herein. These are more technical in nature.

- Strong individual-level association does not imply strong trial-level association. This is because individual-level association and baseline risk may vary by treatment.
    - For instance, it is possible that for two treatments, responders survive equally well but non-responders in the treatment group have shorter survival times than non-responders in the control group. In this example, even if the treatment group contains more responders, the overall treatment effect may not be positive.
- The models implemented make simplifying assumptions which may not accurately

reflect the true relationships modeled. This is due to computational constraints and the limited number of trials contained in the various meta-analyses.

  - The copula models utilize a "reduced" model approach, which does not include trial-level intercepts for either endpoint.

- A Weibull proportional hazards model is assumed for the time-to-event endpoints, which enforces assumptions on the baseline hazard as well as treatment effects.
- A linear relationship is assumed between the treatment effects (log HR and log OR). The validity of this assumption impacts the calculation of any R^2 value.
- As for any regression model, inference and prediction based on the model of log HR vs. log OR is limited by the range of the log OR values observed. Inferences or predictions for log OR values outside of this range are considered extrapolation, with no data to inform them. In some analyses, the spread of log OR values is quite limited.

7.3.3 Methodological Differences Between the Two Applicants' Approaches

FDA Appendix Table 3: Methodological Differences Between the Two Applicants' Analyses

	I2TEAMM	Univ of Miami
Analysis Population	• NDTE population • NDTinE population • RR population • Pooled NDTinE and RR population (Exploratory)	• All newly diagnosed multiple myeloma population. • NDTE population • NDTinE population • RR population[1]
Definition of "missing" MRD	• Definition: MRD data not available within 3-month window of 9/12 months and achieved CR/sCR prior to 9/12+3 months or no response data available. • Patients with missing endpoints data are excluded from the primary analysis.	• Definition: patient achieves CR or better by the end of the time window, does not have any MRD assessments in the time window, and does not have progressive disease or death by the end of the time window. • Patients with a value for the endpoint of missing will be considered as MRD-positive.
Trials Included in the meta-analyses	GMMG MM5 BMT CTN 0702 GEM2012MENOS65 GEM2005MENOS65 Myeloma IX FORTE GRIFFIN- MMY2004 CLARION MM2 (C16014) ALCYONE-MMY3007 GEM2010MAS65 MAIA-MMY3008 POLLUX-MMY3003 CASTOR-MMY3004 CANDOR	GMMG MM5 GRIFFIN- MMY2004 MM2 (C16014) mmy3006 CLARION ALCYONE-MMY3007 MAIA-MMY3008 OCTANS-MMY3011 IKEMA[1] CANDOR[1] POLLUX-MMY3003[1] CASTOR-MMY3004[1]
Primary Analyses Specified in SAP	• Evaluation of trial-level association between 9-month (+/- 3 months) MRD vs	• Evaluation of trial-level association between 12-month (+/- 3 months) MRD vs

	PFS in NDTE and NDTinE populations, separately[2] • Evaluation of individual-level association between 9-month (+/- 3 months) MRD vs PFS within each disease population, MRD sensitivity level, and MRD testing methods, separately[2]. • Surrogate threshold effect for PFS.	PFS in NDTE, NDTinE, and RR populations. • Evaluation of individual-level association between 12-month (+/- 3 months) MRD vs PFS in NDTE, NDTinE and RR populations. • Surrogate threshold effect for PFS.

[1] The RR population data were included in the Univ. of Miami's study report submitted previously, but not included in the briefing document.
[2] For I2TEAMM's meta-analysis, overall survival and 12-month (+/- 3 months) MRD are secondary clinical outcome and key secondary candidate MRD surrogate endpoint, respectively.
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

7.3.4 Derivation and Description of FDA's Pooled Population

FDA utilized data from both Applicants to derive a pooled analysis population. The pooled population utilized the following inclusion/exclusion criteria:

1. <20% missingness for the primary endpoint definition based on MRD
2. >0% MRD- CR rate in all treatment arms
3. All randomized patients (whose data were available)

To the extent possible, Applicant derivations of endpoints (MRD- CR, PFS, and OS) were utilized. If a trial was submitted to either Applicant, the dataset which contained all randomized patients was prioritized for inclusion.
The endpoints of MRD- CR at 9 and 12 months was explored in FDA's analyses. The same +/- 3 months window was utilized. Patients who had missing data at either time point were considered non-responders. All analyses were conducted using the treatment arm assigned at randomization.

FDA Appendix Table 4 presents selected baseline characteristics in FDA's Analysis Population. Note that data collection for such characteristics varied by trial, resulting in missing data for trials that did not collect or report a particular characteristic. The characteristics reported below are those which were collected by most trials and which are relevant to understanding the population.

FDA Appendix Table 4: Baseline Demographic and Disease Characteristics (FDA Analysis Population)

	NDTE* (N=4063)	NDTinE (N=3974)	RR (N=1835)
Age			
N	4063	3237	1067
Mean (SD)	56.79 (7.50)	72.28 (6.02)	63.88 (9.32)
Median	58	72	65
Min, Max	22, 74	40, 93	30, 89
Age Group, n(%)			
18-64	3562 (87.7%)	152 (3.8%)	530 (28.9%)
65+	501 (12.3%)	3085 (77.6%)	537 (29.3%)
Missing	0	737 (18.6%)	768 (41.9%)
Gender, n(%)			
Female	1720 (42.3%)	1934 (48.7%)	578 (31.5%)
Male	2343 (57.7%)	2040 (51.3%)	791 (43.1%)
Missing	0	0	466 (25.4%)
Race, n (%)			
American Indian or Alaska Native	0	3 (0.1%)	0
Asian/Native Hawaiian/Pacific Islander	14 (0.3%)	246 (6.2%)	0
Black	56 (1.4%)	8 (0.2%)	0
Multiracial	2 (0.1%)	2 (0.1%)	0
White	1598 (39.3%)	1081 (27.2%)	0
Other	2 (0.1%)	10 (0.3%)	0
Missing	2379 (48.0%)	2624 (66%)	1835 (100%)
ISS, n (%)			
Stage I	937 (23.1%)	289 (7.3%)	395 (21.5%)
Stage II	966 (23.8%)	671 (16.9%)	236 (12.9%)
Stage III	363 (15.7%)	596 (15%)	134 (7.3%)
Missing	1524 (37.5%)	2418 (60.9%)	1070 (58.3%)

Source: FDA analysis

*Duplicated subjects were removed from NDTE population.

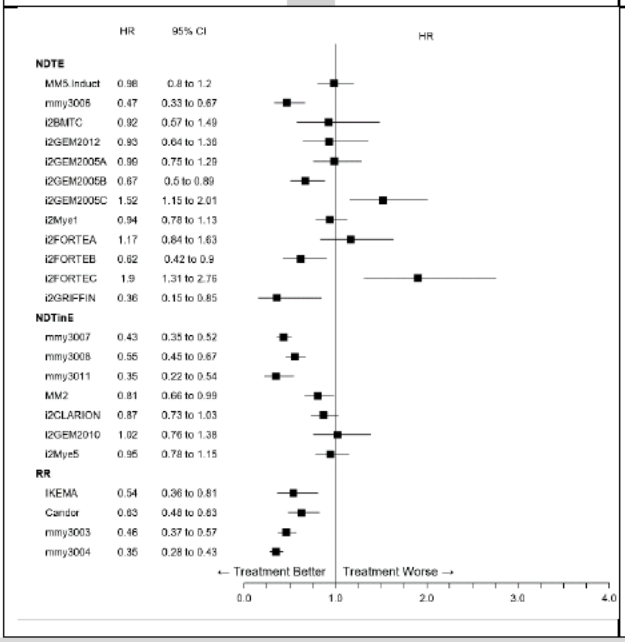
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

7.3.5 FDA's Results

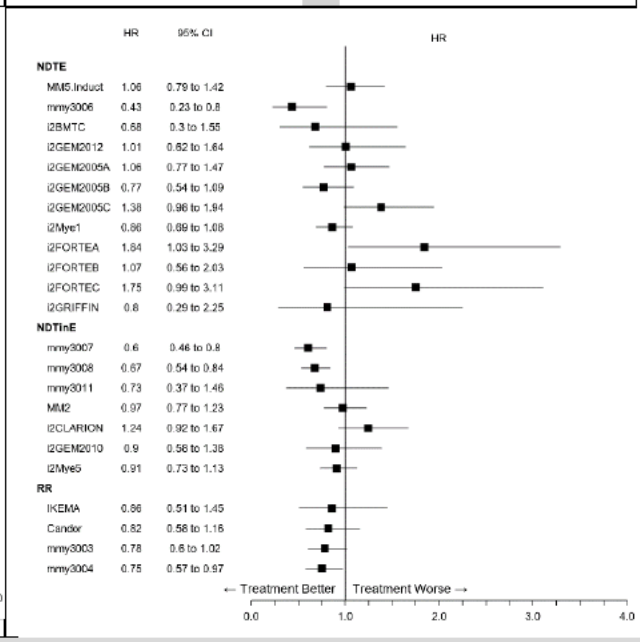
7.3.5.1 Results for 9/12-months MRD-CR vs PFS and OS.

FDA Appendix Figure 1: Forest Plot of Treatment Effects on PFS and OS (FDA Analysis Population, Trials Contributing to Analyses for MRD- CR at 9 Months)

PFS



OS



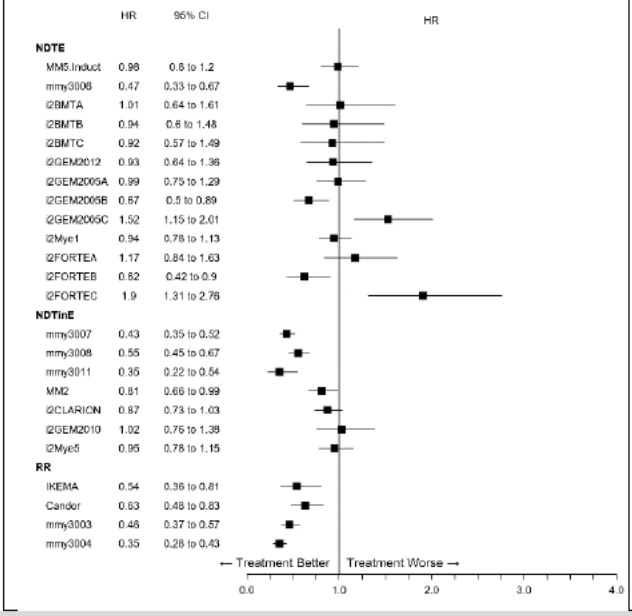
Source: FDA analysis

PFS = progression-free survival; OS = overall survival; HR = hazard ratio; CI = confidence interval

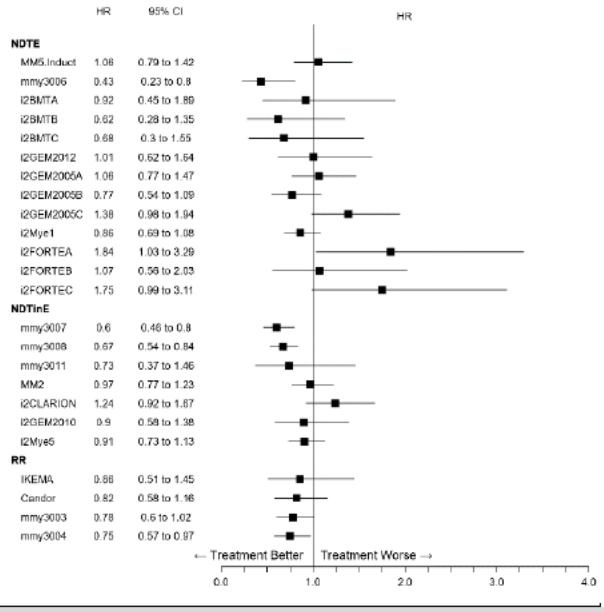
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Appendix Figure 2: Forest Plot of Treatment Effects on PFS and OS (FDA Analysis Population, Trials Contributing to Analyses for MRD- CR at 12 Months)

PFS



OS



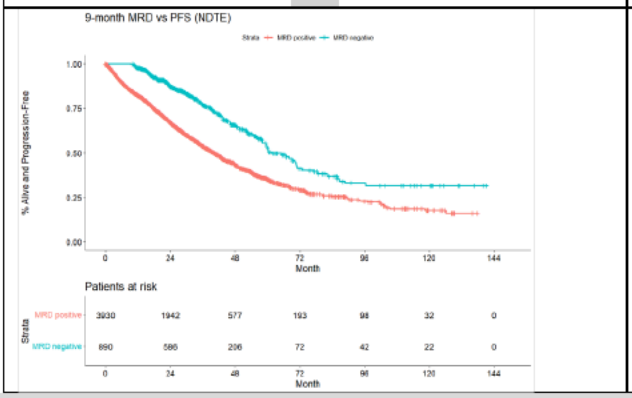
Source: FDA analysis

PFS = progression-free survival; OS = overall survival; HR = hazard ratio; CI = confidence interval

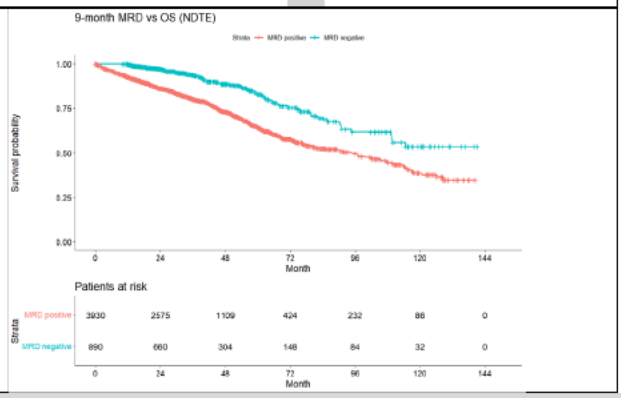
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Appendix Figure 3: Kaplan-Meier Plots of PFS and OS by 9-Month MRD- CR Status (NDTE, FDA Analysis Population)

PFS



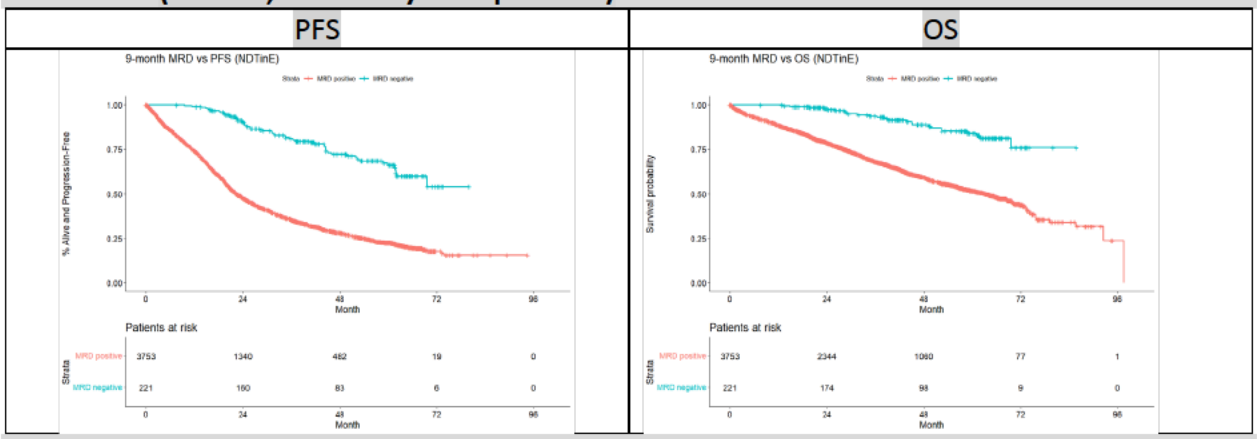
OS



Source: FDA analysis

PFS = progression-free survival; OS = overall survival; NDTE = newly diagnosed transplant eligible; HR=hazard ratio

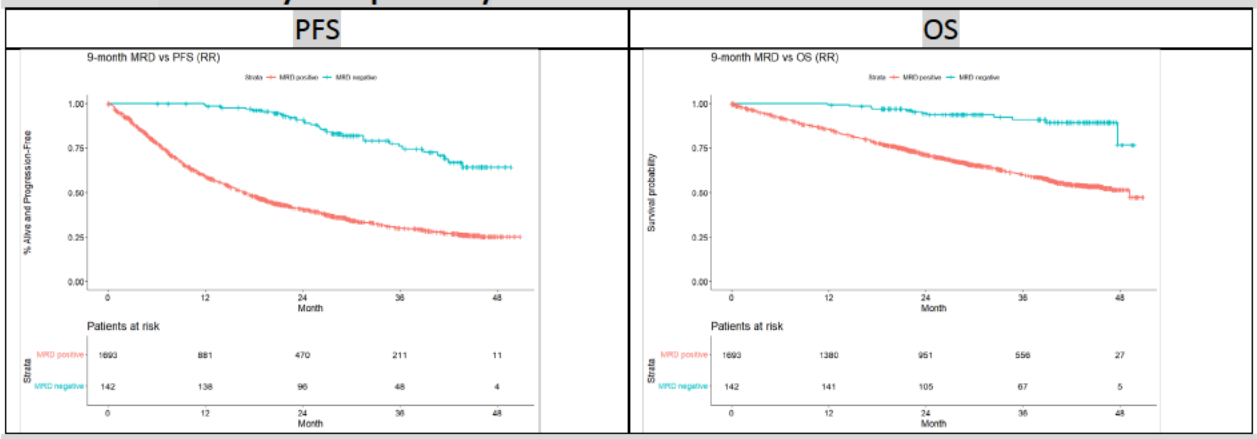
FDA Appendix Figure 4: Kaplan-Meier Plots of PFS and OS by 9-Month MRD- CR Status (NDTinE, FDA Analysis Population)



Source: FDA analysis

PFS = progression-free survival; OS = overall survival; NDTinE = newly diagnosed transplant ineligible; HR=hazard ratio

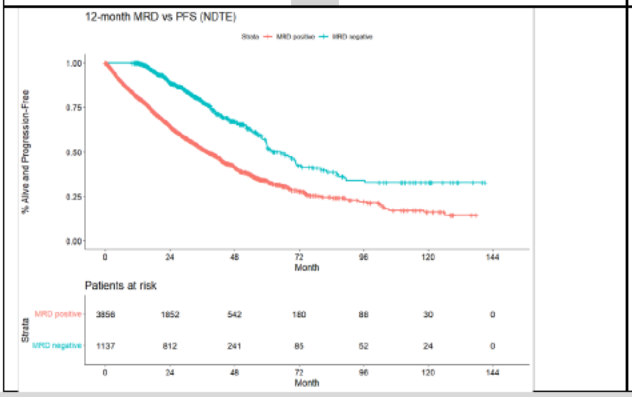
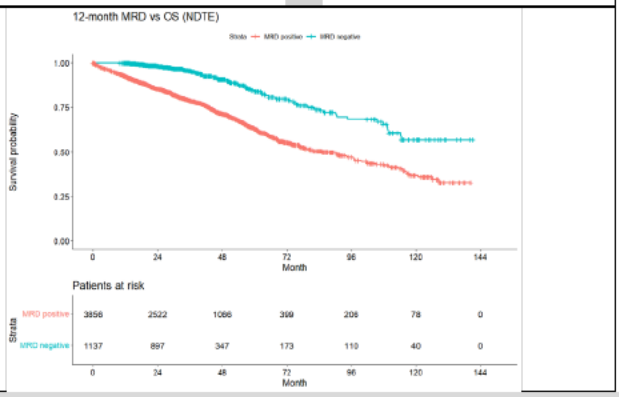
FDA Appendix Figure 5: Kaplan-Meier Plots of PFS and OS by 9-Month MRD- CR Status (RR, FDA Analysis Population)



Source: FDA analysis

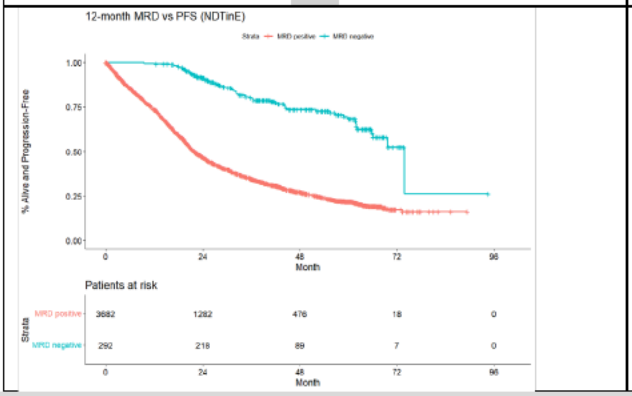
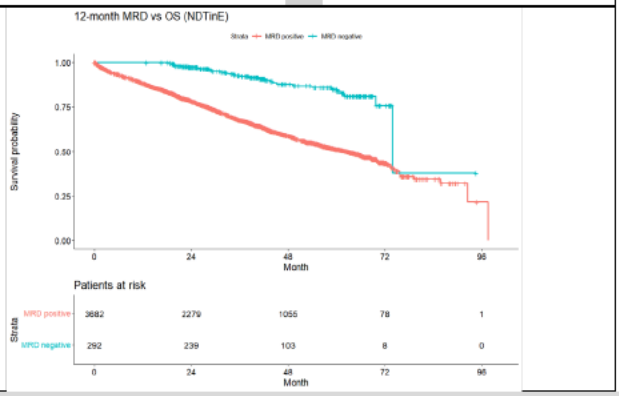
PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory; HR=hazard ratio

FDA Appendix Figure 6: Kaplan-Meier Plots of PFS and OS by 12-Month MRD- CR Status (NDTE, FDA Analysis Population)

PFS	OS
	

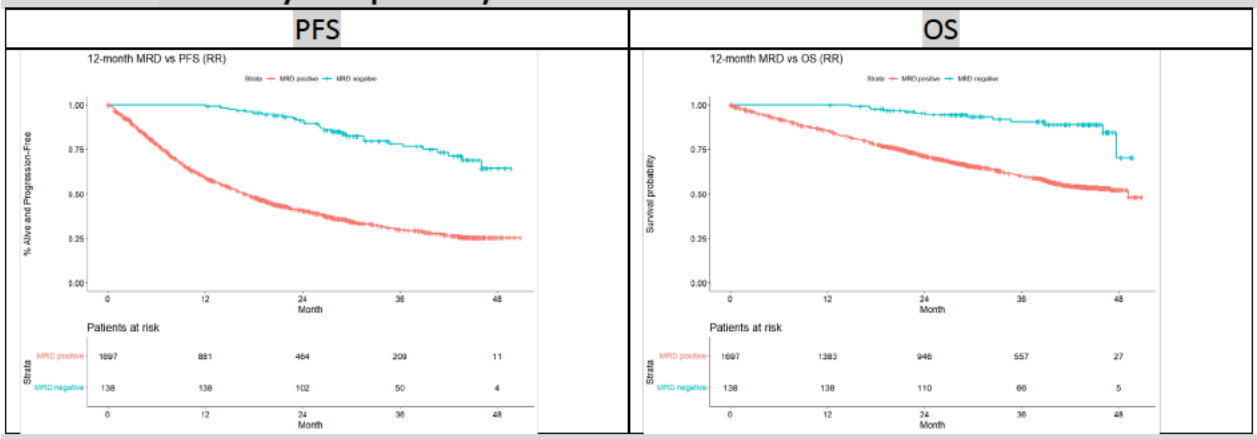
Source: FDA analysis
PFS = progression-free survival; OS = overall survival; NDTE = newly diagnosed transplant eligible; HR=hazard ratio

FDA Appendix Figure 7: Kaplan-Meier Plots of PFS and OS by 12-Month MRD- CR Status (NDTE, FDA Analysis Population)

PFS	OS
	

Source: FDA analysis
PFS = progression-free survival; OS = overall survival; NDTinE = newly diagnosed transplant ineligible; HR=hazard ratio

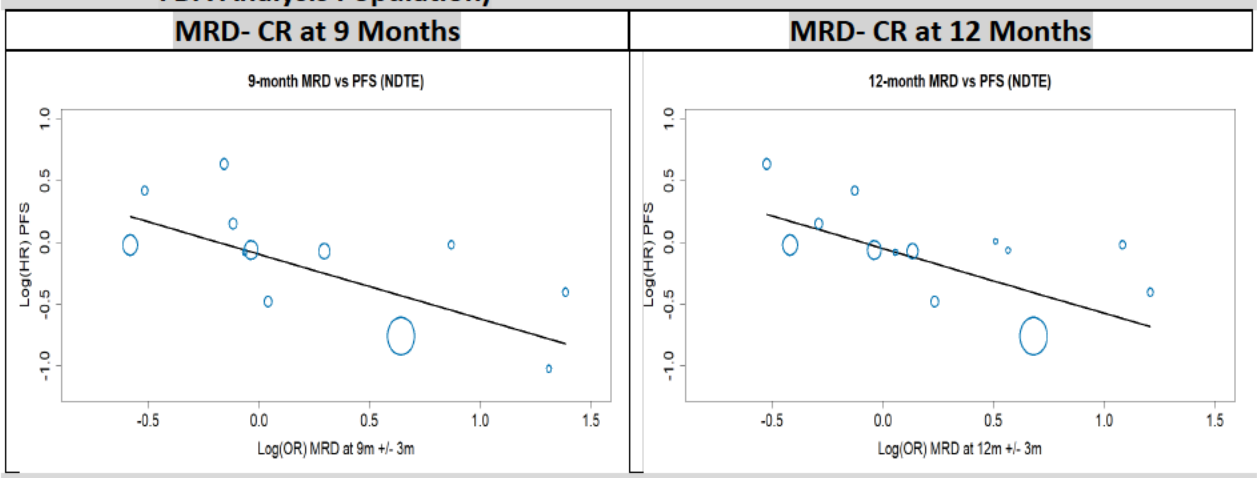
FDA Appendix Figure 8: Kaplan-Meier Plots of PFS and OS by 12-Month MRD- CR Status (RR, FDA Analysis Population)



Source: FDA analysis

PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory; HR=hazard ratio

FDA Appendix Figure 9: Trial-Level Association of MRD- CR at 9/12 Months with PFS (NDTE, FDA Analysis Population)

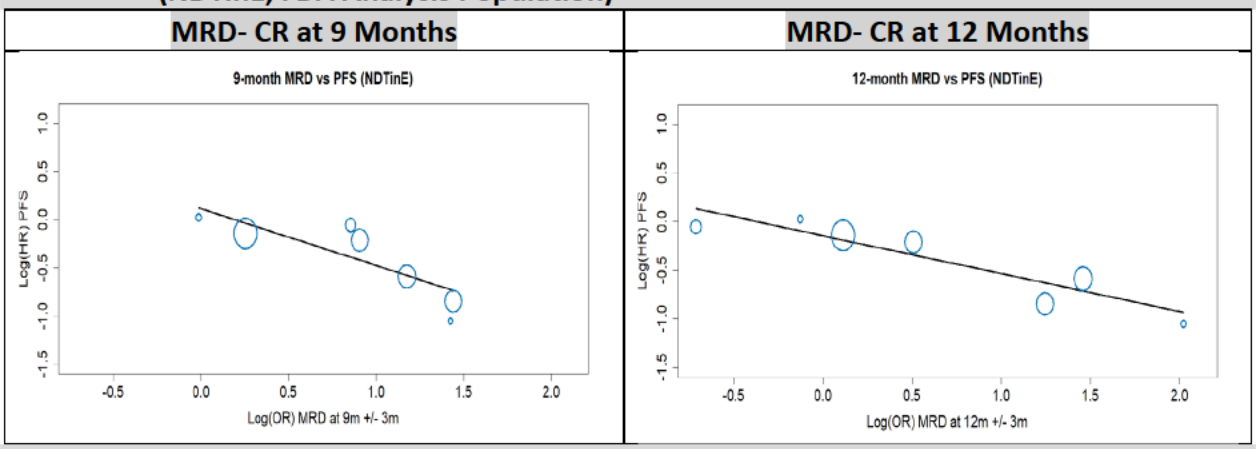


Source: FDA analysis

PFS = progression-free survival; HR=hazard ratio; OR=odds ratio; NDTE = newly diagnosed transplant eligible.

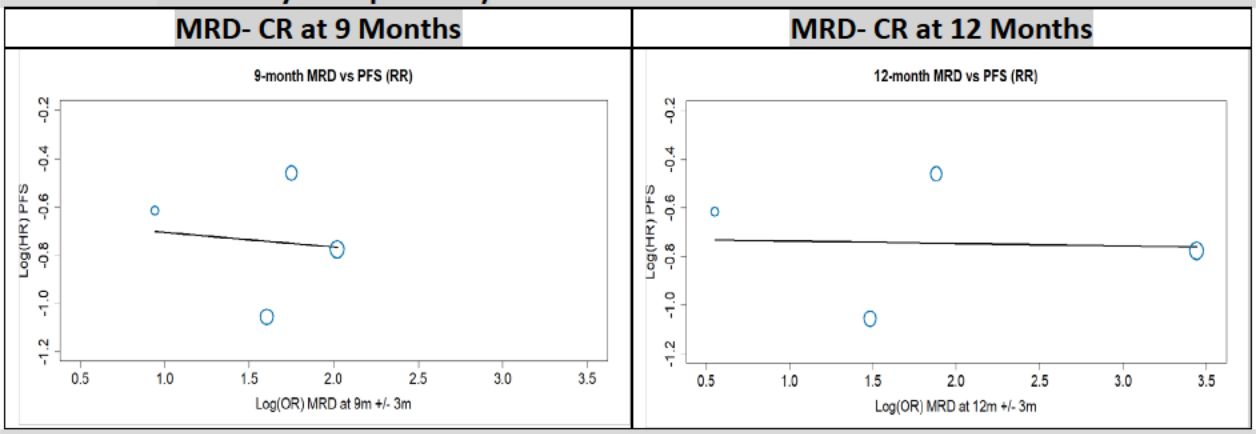
Circles are scaled based on sample size of each trial

FDA Appendix Figure 10: Trial-Level Association of MRD- CR at 9/12 Months with PFS (NDTinE, FDA Analysis Population)



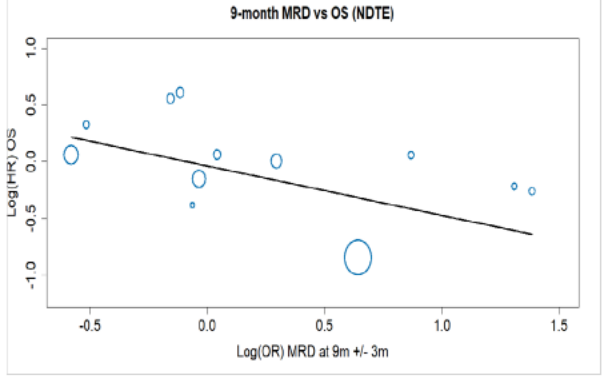
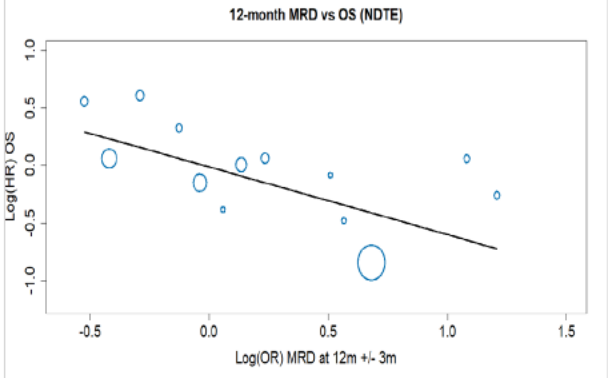
Source: FDA analysis
PFS = progression-free survival; HR=hazard ratio; OR=odds ratio; NDTinE = newly diagnosed transplant ineligible
Circles are scaled based on sample size of each trial.

FDA Appendix Figure 11: Trial-Level Association of MRD- CR at 9/12 Months with PFS (RR, FDA Analysis Population)



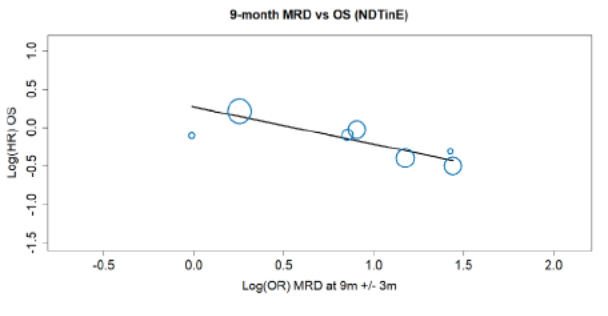
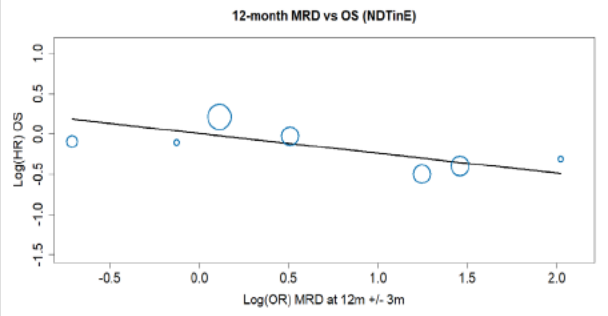
Source: FDA analysis
PFS = progression-free survival; HR=hazard ratio; OR=odds ratio; RR = relapsed and refractory
Circles are scaled based on sample size of each trial.

FDA Appendix Figure 12: Trial-Level Association of MRD- CR at 9/12 Months with OS (NDTE, FDA Analysis Population)

MRD- CR at 9 Months	MRD- CR at 12 Months
	

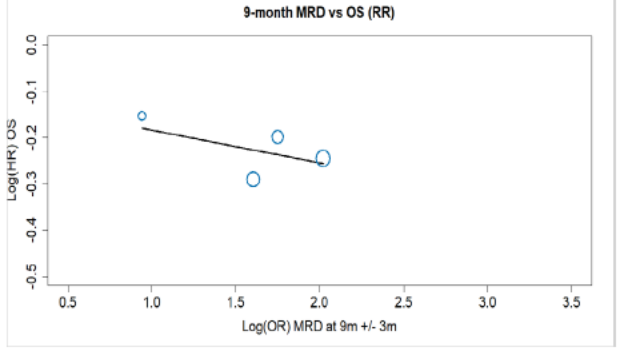
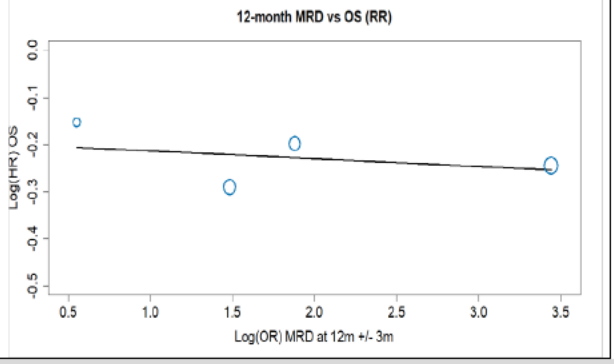
Source: FDA analysis
OS = overall survival; HR=hazard ratio; OR=odds ratio; NDTE = newly diagnosed transplant eligible
Circles are scaled based on sample size of each trial.

FDA Appendix Figure 13: Trial-Level Association of MRD- CR at 9/12 Months with OS (NDTinE, FDA Analysis Population)

MRD- CR at 9 Months	MRD- CR at 12 Months
	

Source: FDA analysis
OS = overall survival; HR=hazard ratio; OR=odds ratio; NDTinE = newly diagnosed transplant ineligible
Circles are scaled based on sample size of each trial.

FDA Appendix Figure 14: Trial-Level Association of MRD- CR at 9/12 Months with OS (RR, FDA Analysis Population)

MRD- CR at 9 Months	MRD- CR at 12 Months
	

Source: FDA analysis
OS = overall survival; HR=hazard ratio; OR=odds ratio; RR = relapsed and refractory
Circles are scaled based on sample size of each trial.

7.3.5.2 Analyses for MRD- CR at 9 or 12 Months in Pooled Populations

FDA Appendix Table 5: Trial-level association results for PFS (Pooled Populations, FDA Analysis Population)

	9 months		12 months	
	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)
NDTE+NDTinE	19 (8794)	0.58 (0.29, 0.87) 0.54 (0.28, 0.79)	20 (8967)	0.58 (0.29, 0.86) 0.60 (0.38, 0.83)
NDTE+NDTinE+RR	23 (10629)	0.61 (0.36, 0.86) 0.56 (0.33, 0.79)	24 (10802)	0.53 (0.25, 0.80) 0.51 (0.27, 0.75)

Source: FDA analysis
PFS = progression-free survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Appendix Table 6: Trial-level association results for OS (Pooled Populations, FDA Analysis Population)

	9 months		12 months	
	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)	N comparison (N Patients)	R2-Copula (95% CI) R2-WLS (95% CI)
NDTE+NDTinE	19 (8794)	0.33 (<0.01, 0.67) 0.30 (0.02, 0.59)	20 (8967)	0.29 (<0.01, 0.63) 0.34 (0.06, 0.63)
NDTE+NDTinE+RR	23 (10629)	0.31 (<0.01, 0.63) 0.24 (<0.01, 0.50)	24 (10802)	0.22 (<0.01, 0.51) 0.19 (<0.01, 0.43)

Source: FDA analysis
OS = overall survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Appendix Table 7: Surrogate Threshold Effect (STE) for PFS (Pooled Populations, FDA Analysis Population)

MRD vs PFS	9 months		12 months	
	N comparison (N Patients)	STE odds ratio	N comparison (N Patients)	STE odds ratio
NDTE+NDTinE	19 (8794)	3.35	20 (8967)	3.25
NDTE+NDTinE+RR	23 (10629)	3.46	21 (10802)	3.82

Source: FDA analysis
PFS = progression-free survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

FDA Appendix Table 8: Surrogate Threshold Effect (STE) for OS (Pooled Populations, FDA Analysis Population)

MRD vs PFS	9 months		12 months	
	N comparison (N Patients)	STE odds ratio	N comparison (N Patients)	STE odds ratio
NDTE+NDTinE	19 (8794)	18.92	20 (8967)	6.82
NDTE+NDTinE+RR	23 (10629)	38.09	21 (10802)	12.06

Source: FDA analysis
PFS = overall survival
NDTE = newly diagnosed transplant eligible; NDTinE = newly diagnosed transplant ineligible; RR = relapsed and refractory

7.3.5.3 MRD- CR at Any Time in the RR Population

FDA conducted additional analyses to explore use of MRD- CR at Any Time in the RR population. Any trial submitted to FDA was included in the analysis. This resulted in 7 trials (Table X). No criteria for missingness were utilized, as MRD was routinely captured following CR in all trials. The resulting analysis set will be referred to as "FDA RR Sensitivity Analysis Population".
The following endpoints were explored:

- **MRD- at any time**: defined as achievement of MRD-negativity at 10^-5 level at any time following randomization, regardless of whether or not CR was achieved.
- **MRD- and CR at any time**: defined as achievement of MRD-negativity at 10^-5 level at any time following randomization and achievement of CR at any time following randomization. MRD-negativity and CR could occur at any time relative to each other.
- **MRD- CR at any time**: defined as achievement of MRD-negativity at 10^-5 following CR at any time following randomization. MRD negativity could be achieved up to 3 months prior to achievement of CR.

The individual-level and trial-level measures used by the Applicants were derived for each of these endpoints. For these analyses, trials which had 0% response rate for an endpoint in any arm were excluded from the analyses.

Because MRD-based response rates are generally low in the RR population, it is unclear whether the odds ratio is the most appropriate measure for quantifying the treatment effect. Additional analyses were conducted to explore the potential trial-level association of such endpoints with PFS and OS based on the risk difference (e.g., correlation of the risk difference for MRD- CR at any time with the log HR for PFS). All seven trials were included in these exploratory analyses, as the risk difference is estimable even when the response rate is 0% in a treatment arm.

FDA Appendix Table 9: Patient-level association for MRD- Endpoints vs PFS (FDA RR Sensitivity Analysis Population)

	N comparison (N Patients)	Patient level Global odds ratio (95% CI)
MRD- at any time	6 (2430)	8.92 (5.58, 12.26)
MRD- and CR at any time[1]	5 (2139)	9.40 (5.21, 13.59)
MRD- CR at any time[2]	5 (2139)	8.70 (4.84, 12.55)

Source: FDA analysis
[1]Defined as achievement of MRD any time prior to progression and CR any time prior to progression
[2]Defined as achievement of CR any time prior to progression with MRD-negativity observed after CR or at most 3 months prior to achievement of CR.
CR=complete response; PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

FDA Appendix Table 10: Patient-level association for MRD- Endpoints vs OS (FDA RR Sensitivity Analysis Population)

	N comparison (N Patients)	Patient level Global odds ratio (95% CI)
MRD- at any time	6 (2430)	6.92 (3.47, 10.37)
MRD- and CR at anytime[1]	5 (2139)	6.89 (2.81, 10.97)
MRD- CR at any time[2]	5 (2139)	6.47 (2.65, 10.30)

Source: FDA analysis
[1]Defined as achievement of MRD any time prior to progression and CR any time prior to progression
[2]Defined as achievement of CR any time prior to progression with MRD-negativity observed after CR or at most 3 months prior to achievement of CR.
CR=complete response; PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

FDA Appendix Table 11: Trial-level association for MRD- Endpoints vs PFS (FDA RR Sensitivity Analysis Population)

	N comparison (N Patients)	Wls-R2 (95% CI)	Copula-R2(95% CI)
MRD- at any time	6 (2430)	<0.01 (<0.01, 0.04)	<0.01 (<0.01, 0.05)
MRD- and CR at any time[1]	5 (2139)	0.09 (<0.01, 0.31)	0.09 (<0.01, 0.58)
MRD- CR at any time[2]	5 (2139)	0.1 (<0.01, 0.35)	0.11 (<0.01, 0.62)

Source: FDA analysis
[1]Defined as achievement of MRD any time prior to progression and CR any time prior to progression
[2]Defined as achievement of CR any time prior to progression with MRD-negativity observed after CR or at most 3 months prior to achievement of CR.
CR=complete response; PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

FDA Appendix Table 12: Trial-level association for MRD- Endpoints vs OS (FDA RR Sensitivity Analysis Population)

	N comparison (N Patients)	Wls-R2 (95% CI)	Copula-R2(95% CI)
MRD- at any time	6 (2430)	0.55 (0.26, 0.85)	0.62 (0.15, >0.99)
MRD- and CR at any time[1]	5 (2139)	0.28 (<0.01, 0.61)	0.27 (<0.01, 0.93)
MRD- CR at any time[2]	5 (2139)	0.29 (<0.01, 0.62)	0.26 (<0.01, 0.92)

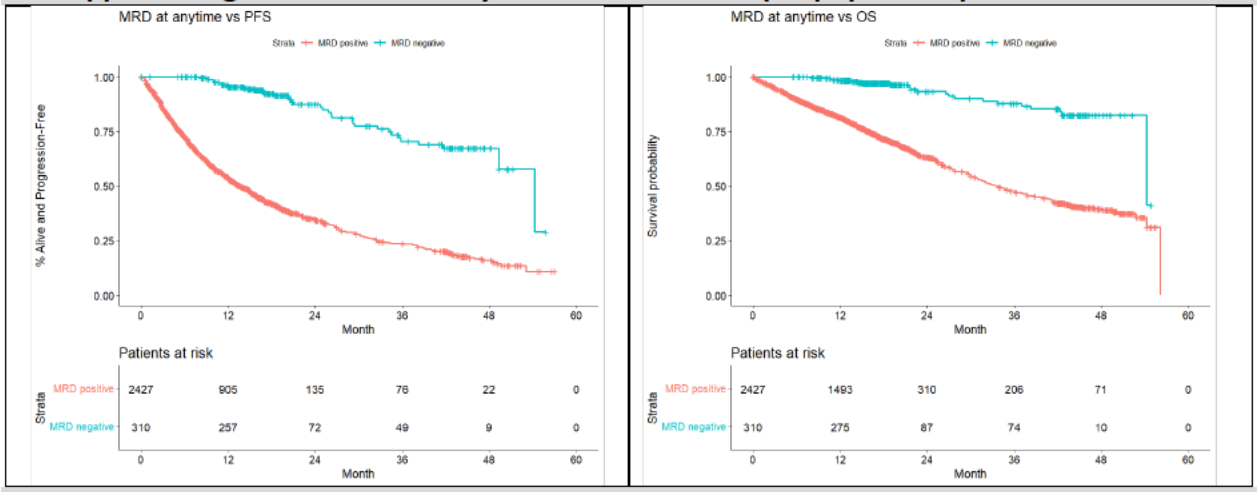
Source: FDA analysis
[1]Defined as achievement of MRD any time prior to progression and CR any time prior to progression
[2]Defined as achievement of CR any time prior to progression with MRD-negativity observed after CR or at most 3 months prior to achievement of CR.
CR=complete response; PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

FDA Appendix Table 13: Surrogate Threshold Effect for MRD- Endpoints vs. PFS and OS (FDA RR Sensitivity Analysis Population)

	N comparison (N Patients)	STE for PFS	STE for OS
MRD- at any time	6 (2430)	NA	NA
MRD- and CR at any time[1]	5 (2139)	NA	NA
MRD- CR at any time[2]	5 (2139)	NA	NA

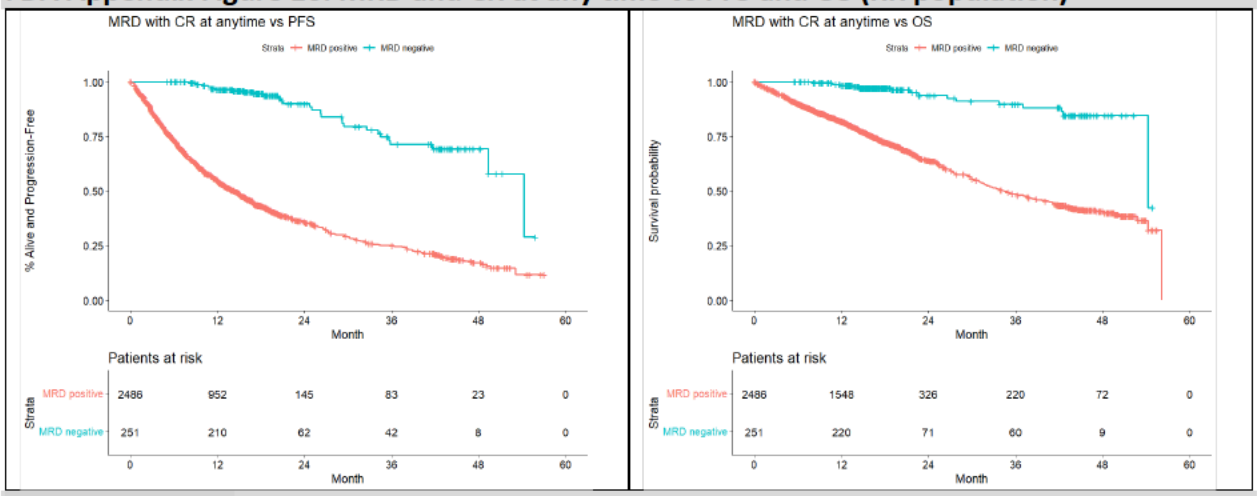
Source: FDA analysis
[1]Defined as achievement of MRD any time prior to progression and CR any time prior to progression
[2]Defined as achievement of CR any time prior to progression with MRD-negativity observed after CR or at most 3 months prior to achievement of CR.
CR=complete response; PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

FDA Appendix Figure 15: MRD at any time vs PFS and OS (RR population)



Source: FDA analysis
PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

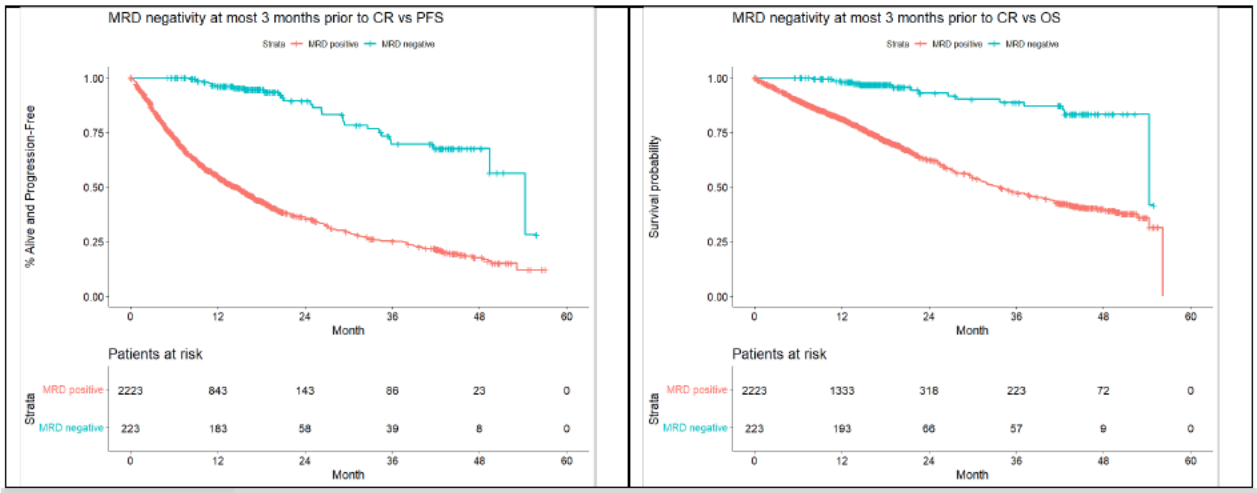
FDA Appendix Figure 16: MRD and CR at any time vs PFS and OS (RR population)



Source: FDA analysis

CR=complete response; PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

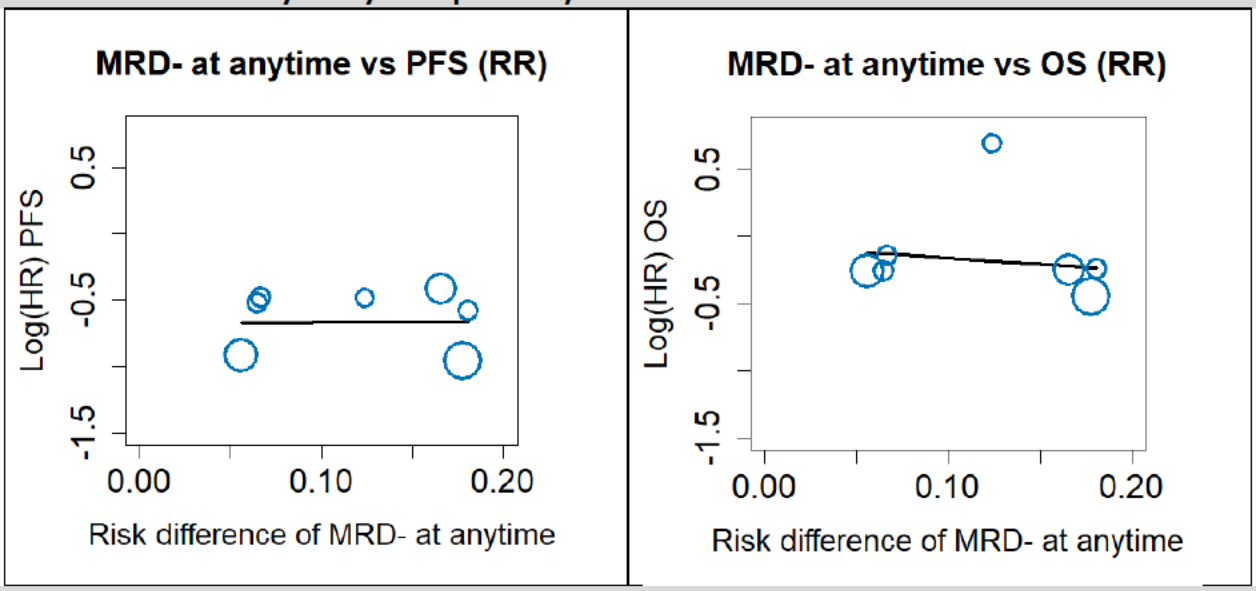
FDA Appendix Figure 17: MRD negativity at most 3 months prior to CR vs PFS and OS (RR population)



Source: FDA analysis

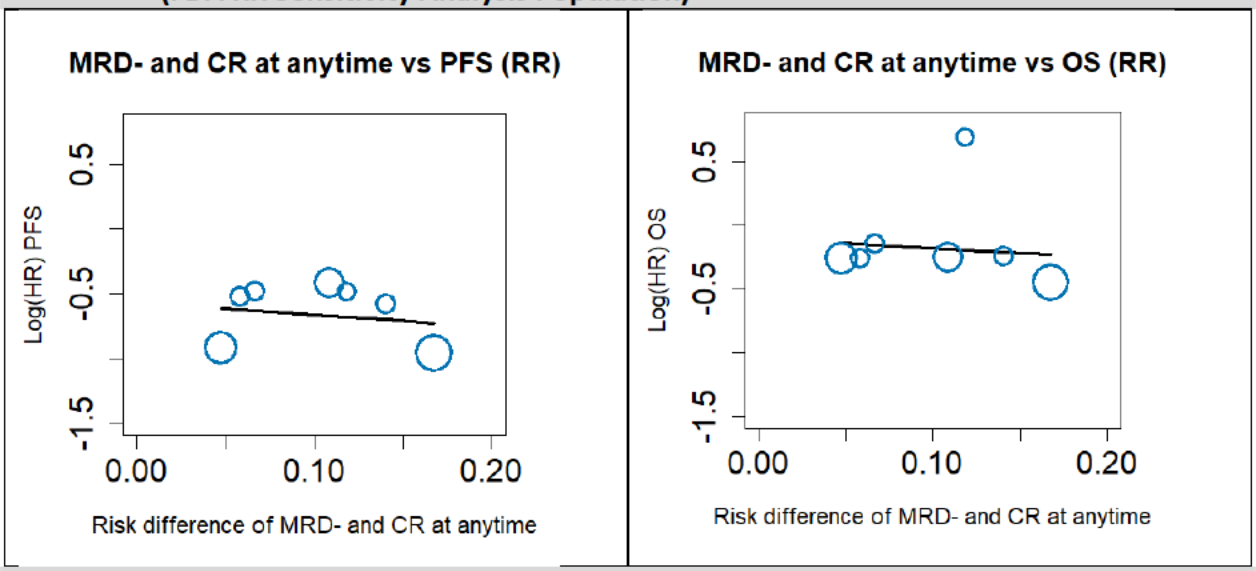
CR=complete response; PFS = progression-free survival; OS = overall survival; RR = relapsed and refractory

FDA Appendix Figure 18: Log HR of PFS/OS by Risk Difference of MRD- at Any Time (FDA RR Sensitivity Analysis Population)



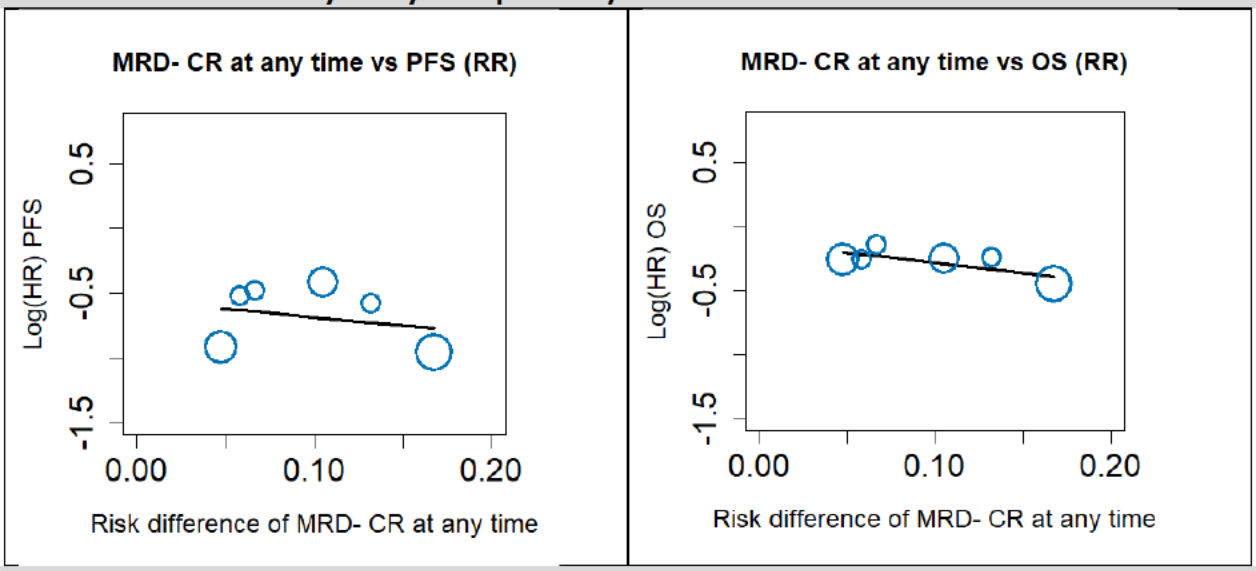
Source: FDA analysis
PFS = progression-free survival; OS = overall survival; HR=hazard ratio; RR = relapsed and refractory

FDA Appendix Figure 19: Log HR of PFS/OS by Risk Difference of MRD- and CR at Any Time (FDA RR Sensitivity Analysis Population)



Source: FDA analysis
CR=complete response; PFS = progression-free survival; OS = overall survival; HR=hazard ratio; RR = relapsed and refractory

FDA Appendix Figure 20: Log HR of PFS/OS by Risk Difference of MRD- CR at Any Time (FDA RR Sensitivity Analysis Population)



Source: FDA analysis
CR=complete response; PFS = progression-free survival; OS = overall survival; HR=hazard ratio; RR = relapsed and refractory