1	TITLE: Effect modification in network meta-analyses of treatments for relapsing refractory									
2	multiple myeloma (RRMM): systematic review, meta-analysis, and simulation									
3										
4	RUNNING HEAD: Effect modification in NMAs for RRMM									
5										
6	Christopher James Rose ^{1,2,*} ; Ingrid Kristine Ohm ¹ ; Liv Giske ¹ ; Gunn Eva Næss ¹ ;									
7	Atle Fretheim ²									
8										
9	1. Reviews and Health Technology Assessments, Division for Health Services, Norwegian									
10	Institute of Public Health, Oslo, Norway									
11	2. Center for Epidemic Interventions Research, Norwegian Institute of Public Health, Oslo,									
12	Norway									
13	* Corresponding author (cjro@fhi.no)									
14										
15	Counts:									
16	Abstract: 250 words (limit = 250 words)									
17	Main text: 4306 words (limit = 4000 words)									
18	Number of figures/tables: 6 (limit = 7 figures/tables)									
19	Number of references: 71 (limit = 100 references)									
20										

Working Paper — This document has not yet been submitted for peer review and may be revised.

21 ABSTRACT

Aims Network meta-analysis (NMA) has been used in several systematic reviews on relapsing refractory multiple myeloma (RRMM). NMAs have been questioned on the basis that effect modification may invalidate the underpinning assumptions. We aimed to systematically review and meta-analyze the evidence for effect modification of hazard ratios (HRs) for overall survival (OS) and progression-free survival (PFS) with respect to refractory status and number of treatment lines.

Methods We extracted stratified HR estimates from 42 phase 2 and 3 randomized controlled trials (RCTs). We tested for within-study effect modification and used meta-analyses to estimate ratios of hazard ratios (RHRs) across trial under assumptions that strongly favor the modification hypothesis. RHR estimates were used in simulations to estimate how many NMA results would be expected to differ in the presence versus absence of effect modification.

Results Most (95%) publications could have reported stratified estimates but only 14% (OS) and 43% (PFS) did. Within-study evidence for effect modification is very weak (p > 0.05 for 47 of 49 sets of stratified estimates). The largest RHR estimated was 1.31 (95% CI 1.16–1.47), for the modifying effect of refractory status on HR for PFS. Simulations suggest that, in the worst case, effect modification would result in 4.48% (95% CI 4.42%–4.53%) of NMA estimates differing statistically significantly in the presence versus absence of effect modification.

Conclusions Effect modification is essentially undetectable in phase 2 and 3 trials. In the worst
case, it is unlikely to affect more than about 5% of random-effects NMA estimates.

Working Paper — This document has not yet been submitted for peer review and may be revised.

42 KEYWORDS

- 43 relapsing refractory multiple myeloma; RRMM; effect modification; stratification; systematic
- 44 review; meta-analysis; network meta-analysis; simulation study

Working Paper — This document has not yet been submitted for peer review and may be revised.

45 **INTRODUCTION**

A defining characteristic of relapsing refractory multiple myeloma (RRMM) is that patients either do not respond, or stop responding — i.e., are refractory — to specific treatments.¹ Refractory patients must switch to alternative treatments, if available. Multiple treatment options now exist, and the treatment regimens often comprise multiple drugs in combination. This naturally leads to questions about treatment superiority. These have been addressed in several systematic reviews that have used network meta-analysis (NMA).^{2,3,4,5,6,7}

52

If the assumptions underpinning an NMA model are sufficiently satisfied, NMA facilitates metaanalytical estimation of all pairs of treatment effects, including between treatments that have not been compared directly in a trial. One of these assumptions is called the transitivity assumption^{8,9,10} which, informally, means that a treatment effect for one comparison can be calculated by adding or subtracting treatment effects for other comparisons in the network. This allows treatment effects to be estimated for pairs of treatments that have not been directly compared by the trials included in the network (i.e., indirect comparisons).

60

NMAs should assess and report on the likely validity of the transitivity assumption. This requires comparing distributions of effect modifiers across trials.⁸ An effect modifier is a variable that causes a difference in treatment effect but is not itself a treatment or an outcome.^{11,12} In plain English: effect modification is about stratification — when effect modification occurs, treatment effect is different for different subgroups of patients. It is important to distinguish between a variable that is associated with *treatment effect* (a comparison between treatments) and a

Working Paper — This document has not yet been submitted for peer review and may be revised.

67	variable that is only associated with outcome (e.g., overall survival for a particular patient). The
68	former is an effect modifier, but the latter is a risk factor. Number of lines of treatment (LOT) is
69	presumably a risk factor for overall survival, if for no other reason than patients who have
70	received many LOT will be older. However, that does not mean it is also an effect modifier.
71	
72	Unfortunately, non-statistical articles on NMA often conflate risk factors and effect modifiers
73	when considering the transitivity assumption. Risk factors are not a concern for NMAs of RCTs
74	because, in expectation, randomization excludes the possibility that they account for observed
75	treatment effects. In large part, this is why RCTs are so useful. However, if a <i>fixed-effects</i> NMA is
76	applied to estimates from trials with different distributions of effect modifiers, the transitivity
77	assumption will be threatened because the estimates have different interpretations, and with it,
78	the validity of the NMA. That said, the nature and extent to which an NMA may be invalidated by
79	effect modification depends on the magnitudes and directions of the modifications. If
80	modification is small compared to the precisions of the trial estimates, NMA estimates may still
81	be consistent with the true treatment effects (e.g., confidence intervals may contain the target
82	parameter values). Random-effects NMAs are designed specifically to address heterogeneity in
83	trial-level treatment effects.

84

The use of NMA in RRMM has been criticized¹³ on the basis that variables such as refractory status and LOT are effect modifiers, with the implication that NMAs that do not account for effect modification may be untrustworthy. The present article was motivated by an ongoing health technology assessment (HTA) we are conducting on treatments for RRMM that was

Working Paper — This document has not yet been submitted for peer review and may be revised.

89	commissioned via Norway's National System for Managed Introduction of New Health
90	Technologies within the Specialist Health Service ("Nye Metoder"). ¹⁴ One of our clinical advisors
91	highlighted concerns about effect modification with respect to refractory status and LOT. While
92	these concerns have been raised in previous work, ¹³ we could not find definitive quantitative
93	research on effect modification in RRMM that could inform our HTA. We therefore performed a
94	systematic review and meta-analysis of stratified estimates reported by the trials included in our
95	HTA. We then used the meta-analysis results in a simulation study to assess the degree to which
96	NMA estimates are likely to be affected by effect modification.

97

98 METHODS

99 This meta-analysis was not prespecified or registered because it was performed in response to 100 comments on an ongoing HTA. Further methodological details and completed PRISMA 101 checklists¹⁵ are available in Supplementary Materials.

102

103 Literature search strategy

The search was first performed in February 2020 and was regularly updated until January 2022 (ongoing trials until June 2021). We limited the search to RCTs, used the search term Multiple Myeloma, and used MeSH-terms and text words. Halfway through we limited the search to include the terms Relapse or Refractory. The full strategy is presented in Supplementary Materials. We also contacted project stakeholders, including industry, to solicit suggestions for potentially relevant publications. We did not systematically search beyond this work to support our HTA because we are primarily interested in effect modification within the trials included in

Working Paper — This document has not yet been submitted for peer review and may be revised.

- 111 our HTA. Via manual searching we found nine articles reporting stratified estimates for the
- included trials^{16,17,18,19,20,21,22,23,24} but used stratified estimates from the main trial publications,
- 113 because they are more likely to have been prespecified.
- 114
- 115 Inclusion and exclusion criteria

116 From the identified publications, we included those that provide estimates of hazard ratios (HRs)

117 for OS or PFS that could be included in NMAs (i.e., those that report point estimates and a

118 statement of precision such as a confidence interval or *p*-value). We excluded trials comparing

- 119 doses or schedules of the same treatment.
- 120

We excluded publications from meta-analysis if they did not report stratified estimates of HR for all strata for at least one of two potential effect-modifiers (e.g., we would have excluded a study if it did report an estimate for lenalidomide-refractory patients but did not report an estimate for patients not refractory to lenalidomide). We excluded publications that did not report numerical statements of uncertainty on stratified estimates (e.g., we excluded one study that reported point estimates numerically but only provided a graphical presentation of the confidence intervals).

128

129 Data extraction

CJR extracted data from trial reports and supplementary materials. Extracted data were checked
 independently by IKO. Disagreements were resolved via discussion. We extracted estimates of
 HR for OS and PFS, stratified by LOT and refractory status or previous use of immunomodulatory

Working Paper — This document has not yet been submitted for peer review and may be revised.

133	drugs (IMiDs; see below), where available. For LOT, we extracted data using the categorizations
134	used by the publications (e.g., Attal 2019 used 2–3 vs >3 LOT, while Dimopoulos 2016 used 1 vs
135	2 vs 3 vs >3 LOT). For refractory status we preferentially extracted estimates stratified by
136	refractory status with respect to lenalidomide (Revlimid), which was identified as being of
137	particular concern by our clinical advisor (i.e., if stratified estimates were available for
138	lenalidomide-refractory patients and patients not refractory to lenalidomide, we extracted these
139	estimates). If this information was not reported, we extracted estimates for refractory status with
140	respect to other named IMiDs such as thalidomide (of which lenalidomide is an analog). If this
141	information was not reported, we extracted estimates for refractory status with respect to IMiDs
142	in general. If this information was not available, we extracted estimates stratified by previous use
143	of lenalidomide or other IMiDs, on the assumption that previously IMiD use is a reasonable proxy
144	for being refractory to an IMiD. For the same reason, we did not extract estimates stratified by
145	LOT and refractory status simultaneously (e.g., comparing OS in lenalidomide-refractory patients
146	with one LOT). In summary, we used a pragmatic and inclusive definition of refractory status
147	rather than a strict definition that would have yielded very little evidence on effect modification.
148	For brevity we call this concept "refractory status" in the remainder of the article. However,
149	readers are reminded that the concept is broader than this name implies (the statistical analysis
150	section describes how we address the issue statistically).

151

152 Statistical analyses

We first performed pairwise random-effects meta-analyses of stratified HRs, grouped by trial, for
 refractory status and LOT. This facilitates testing for evidence of effect modification within each

Working Paper — This document has not yet been submitted for peer review and may be revised.

155	trial. Because these analyses yielded very weak evidence for effect modification, but there
156	nevertheless seems to be strong opinions that effect modification does occur and is a problem
157	for NMAs for RRMM, we then performed pairwise random-effects meta-analyses of ratios of
158	hazard ratios (RHRs; described below) for refractory status and LOT. This facilitates estimation of
159	relative magnitudes of effect modification and allows us to test for effect modification by pooling
160	all evidence of effect modification across trial and treatment comparison.

161

162 RHRs were computed for each trial as follows (a formal definition is provided in Supplementary 163 Materials). First, the trial's strata were sorted to ensure that the order of strata have similar 164 interpretations across trials and are therefore amenable to meta-analysis. For example, LOT 165 strata were sorted from fewest to most LOT, and previous lenalidomide use was nominated as 166 the first (i.e., reference) level of the refractory status factor variable. Then, we computed the 167 ratio between the HR for each stratum and the HR for its preceding stratum (except for the first stratum, which is the reference). Finally, we "inverted" any of these ratios with a point estimate 168 169 less than one to ensure that point estimates for all RHRs are greater than or equal to one. This 170 inversion step is necessary to prevent ratios less than one from cancelling ratios greater than one in the meta-analyses and thereby obscuring any evidence of effect modification (see below). 171 172 Standard errors on RHRs were computed as described in Supplementary Materials. We excluded 173 reference strata from meta-analysis because, as references, they are not defined with respect to 174 another stratum.

Working Paper — This document has not yet been submitted for peer review and may be revised.

176	The RHR scale removes heterogeneity in direction of treatment effect within and between trials
177	and facilitates meta-analysis across all trials such as to make evidence of effect modification
178	statistically detectable; it therefore strongly favors the effect modification hypothesis. A RHR tells
179	us how many times larger a stratified estimate is compared to the estimate for its preceding
180	stratum (or vice versa). If the meta-analytical estimate of mean RHR differs statistically from
181	RHR = 1, then we can reject the null hypothesis of no effect modification.

182

All meta-analyses were performed on the logarithmic scale. We used random-effects models 183 184 throughout because there are important differences in the definitions of refractory status and 185 LOT used across the trials which would be expected to manifest as heterogeneity, and which must be accounted for statistically. We present results using forest plots, sub-grouped by 186 187 publication, to report estimates of mean HRs or mean RHRs, 95% confidence intervals, and l² and 188 *p*-values throughout. We used the conventional p < 0.05 criterion for statistical significance. Statistical analyses were performed using Stata 16 (StataCorp LLC, College Station, Texas, USA). 189 Data and code are freely available (see the Data and Software Availability section). 190

191

192 *Simulation studies*

To help understand the degree to which effect modification may affect NMA results, we performed two simulation studies (plus various sensitivity analyses; see Discussion). The purpose of the simulations was to estimate the percentage of NMA estimates that would be expected to be statistically significantly different under effect modification compared to no effect

Working Paper — This document has not yet been submitted for peer review and may be revised.

197 modification, due to refractory status and LOT. Figure 1 shows a cartoon that illustrates the

- 198 design of these studies.
- 199

200

Insert Figure 1 about here

201 Each simulation used 1000 pairs of synthetic networks of evidence, generated to be similar in 202 distribution to the real network for PFS (the outcome for which RHRs were estimated to be 203 largest; see Results). Networks within a pair were identical except that one network was 204 subjected to simulated effect modification and the other was not, such that any differences in 205 NMA estimates between the two networks could only be attributed to the impact of effect modification. All networks had the same topology as the network for PFS. Simulated effect sizes 206 (log HRs) and their standard errors were drawn from distributions that matched those for the PFS 207 208 data.

209

210 We used estimates of RHR for PFS because they were larger than for OS (i.e., we assumed worst-211 case scenarios), simulating effect modification by sampling from normal distributions 212 parameterized by mean RHRs and their standard errors to account for uncertainty on the estimates of RHR. We fitted random-effects component-NMA models²⁵ to each pair of simulated 213 214 networks and tested null hypotheses of no differences between corresponding estimates. Testing 215 was performed using two-sided Z-tests using the estimated log HRs and their standard errors. 216 Corresponding estimates were deemed to differ if p < 0.05. We summarized the results of each 217 simulation as the percentage of estimates expected to be statistically significantly different under 218 effect modification compared to no effect modification. Simulations were performed using R

Working Paper — This document has not yet been submitted for peer review and may be revised.

- version 3.5.2²⁶ with component NMAs performed using the netmeta²⁷ package (version 1.3-0).
- 220 Further details are available in Supplementary Methods.

221 **RESULTS**

222 Systematic literature searching identified 810 references, of which 40 publications contributed 223 stratified estimates (see Supplementary Materials). Table 1 summarizes the included 224 publications. It shows which trials could have reported stratified estimates (because they 225 included patients that differ with respect to refractory status or LOT) and did so; trials and 226 publications that could have reported stratified estimates but chose not to; and trials that could 227 not report stratified estimates.

- 228
- 229

Insert Table 1 about here

230

231 Effect modification of HR for progression-free survival

Almost all trials could have reported stratified estimates, but only 17 (40%) publications, representing 8364 patients, did report estimates stratified by refractory status (Table 1). Similarly, 18 (43%) publications, representing 7,503 patients, did report estimates stratified by LOT (Table 1). Within-trial evidence for effect modification of HR for PFS by refractory status and LOT is weak (Figure 2 and Figure 3). Only one test for equality of stratified HRs was statistically significant with respect to refractory status (p < 0.01 for the comparison Kd vs Vd²⁸) and another with respect to LOT (p = 0.01 for the comparison DVd vs Vd²⁹).

Working Paper — This document has not yet been submitted for peer review and may be revised.

Insert Figure 2 about here
Insert Figure 3 about here
Mean RHR was estimated to be 1.31 (95% CI 1.16 to 1.47; $p < 0.005$; $l^2 = 0\%$) for refractory status
and 1.19 (95% CI 1.09 to 1.29; $p < 0.01$; $l^2 = 0\%$) for LOT (Figure 4). No statistical heterogeneity in
RHR was observed.
Insert Figure 4 about here
Effect modification of HR for overall survival
Almost all trials could have reported stratified estimates for OS. Only 6 publications (14%),
representing 3,471 patients, did report estimates stratified by refractory status (Table 1).
Similarly, only 7 (17%) publications, representing 4,063 patients, did report estimates stratified
by LOT (Table 1). Within-trial evidence for effect modification of HR for OS by refractory status
and LOT is very weak, with no tests for equality of stratified HRs demonstrating statistical
significance (Supplementary Materials).
Mean RHR was estimated to be 1.16 (95% CI 1.01 to 1.32; $p = 0.03$; $l^2 = 0\%$) for refractory status
and 1.09 (95% CI 0.98 to 1.20; $p = 0.12$; $l^2 = 0\%$) for LOT (Supplementary Materials). No statistical

257 heterogeneity in RHR was observed, suggesting that effect modification may be relatively

consistent across trial and comparison, and that our broad definitions of refractory status and

259 LOT did not introduce undue heterogeneity.

Working Paper — This document has not yet been submitted for peer review and may be revised.

260

261 Simulation study

262	Table 2 summarizes the results for the simulations. We would expect no more than 0.4% of NMA
263	estimates for PFS to differ if trial estimates are, versus are not, subject to effect modification by
264	refractory status. I.e., among the 595 possible comparisons of the 35 treatments in the included
265	trials on PFS, we would expect no more than about 2–3 comparisons to differ statistically
266	significantly due to effect modification. We would expect no more than 5% of NMA estimates for
267	PFS to differ if trial estimates are, versus are not, subject to effect modification by LOT. I.e.,
268	among the 595 possible comparisons, no more than about 30 comparisons would be expected
269	to differ statistically significantly due to effect modification.

270

271

Insert Table 2 about here

272 **DISCUSSION**

We have systematically reviewed and meta-analyzed the evidence for effect modification of HR for OS and PFS stratified by refractory status and LOT. Within-trial evidence for effect modification of HR for OS and PFS by refractory status and LOT is weak, with only 2 of 49 tests of heterogeneity demonstrating statistical significance (i.e., almost exactly the number of type I errors expected at the 95% significance level under the null hypothesis of no effect modification). Of the meta-analyses of cross-trial effect modification, the largest (i.e., worst-case) mean RHR estimated was 1.31 (95% CI 1.16 to 1.47) for HR for PFS with respect to refractory status.

Working Paper — This document has not yet been submitted for peer review and may be revised.

281 We then used worst-case estimates of RHR in simulations to estimate percentages of NMA 282 estimates that may be affected by effect modification. For refractory status, the simulations 283 suggest that even if effect modification is as large as the worst-case estimate, substantially fewer than 1% of NMA estimates are likely to be statistically different than they would be if effect 284 285 modification does not occur. For LOT, the simulations suggest that fewer than 5% of NMA 286 estimates are likely to be statistically different in the presence of effect modification and 287 heterogeneity. This is substantially higher than for refractory status, but putting this result in 288 perspective, 5% is the same as our typical tolerance for type I errors given by the conventional 289 significance level of 95%.

290

Absence of evidence is not evidence of absence, and we may simply not have sufficient data to detect the impact of effect modification. Still, if effect modification does occur, we would expect to see consistent patterns which support effect modification, which we do not. In some cases, estimates increase with refractory status or LOT, in others it is opposite, but in most cases the estimates are practically the same.

296

297 Cope et al. qualitatively assessed 12 NMAs or unanchored indirect comparisons and, based on 298 expert opinions on variables that may be effect modifiers, concluded that the NMA estimates 299 may have been compromised by differences in distributions of effect modifiers.¹³ We are aware 300 of one other attempt at quantifying effect modification through meta-analysis of results from 301 subgroups, e.g., patients with one previous LOT vs. 2 or more previous LOT.⁶ In this NMA, which 302 was limited to comparing different immunomodulatory-containing regimens for RRMM,

Working Paper — This document has not yet been submitted for peer review and may be revised.

- 303 Dimopoulos et al. reported that their subgroup analyses yielded results consistent with their 304 main findings (i.e., no apparent effect modification).
- 305

306 Strengths and limitations

307 To our knowledge, this is the first systematic review, meta-analysis, and simulation study of effect 308 modification in RRMM. However, it was not prespecified. While we were able to obtain stratified 309 estimates from up to 18 (43%) of the included publications, most publications did not report 310 stratified estimates. Because we did not perform a separate literature search, it is possible that 311 we do not have all the available data on effect modification. However, it would be unreasonable 312 to expect the smaller trials to report stratified estimates, as they would likely be so imprecise as 313 to be uninformative. Further, stratified estimates were published for about half as many analyses 314 of OS compared to PFS, despite there being about the same number of publications providing 315 estimates of HR for the two outcomes. Because stratified estimates are not reported in the main 316 trial reports for so many comparisons, it is possible that effect modification is larger than we 317 estimate, particularly for OS.

318

While we systematically reviewed evidence of effect modification with respect to refractory status and number of lines of treatment, we did not systematically review other variables that also may be effect modifiers. However, we did look at all stratified estimates when extracting data and did not notice any variables that appeared to consistently demonstrate convincing evidence of effect modification.

Working Paper — This document has not yet been submitted for peer review and may be revised.

325	Because there was heterogeneity in trial reporting, we were not able to use definitions of
326	refractory status and number of lines of treatment that measured exactly what we were
327	interested in, because doing so would have resulted in almost no synthesizable evidence. We
328	therefore used pragmatic and inclusive definitions, particularly for refractory status (see
329	Methods). We would have expected this to introduce heterogeneity, but this was not statistically
330	observable in the meta-analyses of RHR ($I^2 = 0\%$ in all analyses).

331

Because the within-trial evidence of effect modification is so weak, but there are nevertheless concerns in the RRMM research community about effect modification and NMA, we constructed RHR and designed the simulations to strongly favor the effect modification hypothesis. This likely resulted in exaggerated conclusions about whether effect modification occurs and the extent to which it is problematic.

337

Quantities such as ratios of hazard ratios, as used in meta-research,³⁰ are likely challenging to 338 interpret, and we suspect that few will have an intuitive understanding of what constitutes a 339 "large" or "important" RHR with respect to effect modification in RRMM. A major strength of this 340 341 work is that having estimated RHRs, we then used simulations to investigate how many NMA 342 estimates would be expected to be statistically significantly different under the estimated degree 343 of effect modification. We hope this helps readers understand the likely impact of any effect 344 modification on NMA estimates. However, we remind readers that we used random-effects NMAs,³¹ which are designed to account for heterogeneity in trial estimates. Our results do not 345 346 necessarily translate to fixed-effects NMAs, as used in some systematic reviews on treatments

Working Paper — This document has not yet been submitted for peer review and may be revised.

for RRMM.^{4,2} It is important to note that fixed- and random-effects NMAs make fundamentally different assumptions about transitivity. Fixed-effects NMAs assume that "trial-level" treatment effects can be added and subtracted to make indirect estimates. Random-effects NMAs assume that *average* treatment effects can be added and subtracted. Use of average treatment effects explicitly accounts for differences between trials, including different distributions of effect modifiers.

353

Finally, we also performed sensitivity analyses to investigate the implications of the assumptions we made in the simulations. For example, because RHR discards direction of modification, we assumed that direction of modification is consistent within treatment comparison but may vary between comparisons. This may not be true, so we performed a sensitivity analysis in which direction is assumed consistent within and between comparisons. The result of this analysis suggests that about half as many estimates would differ statistically compared to the main analysis (i.e., that the main result reflects the worst-case).

361

362 Implications for research

Understandably, the transitivity assumption underpinning NMA is often highly simplified in articles aimed at non-statisticians. For example, articles tend to use arguments about "similarity" of patients.⁸ Given this oversimplification, it is unsurprising there are concerns about using NMA in RRMM and other areas. The transitivity assumption does not concern patient similarity, nor whether treatment effect *estimates* can be added or subtracted, it concerns whether target parameters can be linearly combined. Patient similarity is a good place to start thinking about

Working Paper — This document has not yet been submitted for peer review and may be revised.

369 NMAs, but a terrible place to stop. Modern statistical methods should be communicated more370 carefully and received more studiously.

371

372 Understanding effect modification is important for making decisions based on individual trials, 373 and for assessing the assumptions and validity of NMAs. We therefore suggest that RRMM 374 trialists develop and adopt standardized definitions of potential effect modifiers that, where 375 possible, should be used to report stratified analyses in future trials. In addition to improving 376 transparency and improving consistency of reporting of effect estimates for patient subgroups, 377 standardization would facilitate more specific meta-analytical study of effect modification by 378 reducing methodological heterogeneity. Further, we suggest that stratified analyses be reported 379 for all patient-important outcomes, particularly OS, which has been dramatically underreported 380 compared to PFS.

381

382 The strength of concerns that effect modification, as it may occur in RRMM, may invalidate NMAs 383 appears to be inconsistent with the available evidence. This suggests that NMA can probably be 384 relied upon to estimate direct and indirect treatment effects, subject to some important caveats. 385 First, evidence on effect modification is limited to at most ~40% of comparisons, so it is possible 386 that modification is more severe in the remaining ~60% of comparisons. That said, it would be 387 concerning if large modification occurs but has been systematically unreported in the majority of 388 phase 2 and 3 trials. Second, more evidence on modification is available for PFS than OS, so it is possible that HR for OS is subject to greater modification than the available evidence suggests. 389 390 This may be because the PFS endpoint is typically reached earlier than that for OS. However,

Working Paper — This document has not yet been submitted for peer review and may be revised.

391 again, it would be concerning if large modification was not being reported for what is arguably 392 the most important outcome of cancer trials. Third, we are not suggesting that a particular NMA 393 estimate can be applied to patients who are refractory to one or both treatments involved in a 394 given comparison: such estimates would be subject to an obvious, if somewhat absurd, form of 395 effect modification. A method for ranking treatments for patients who are refractory to specific 396 treatments or components is presented in the Supplementary Appendix. Fourth, and crucially, 397 our simulations estimated random-effects (cf. fixed-effects). Random-effects models account for 398 heterogeneity in effect estimates due not just to sampling error, but also other factors, including 399 effect modification. Our findings are unlikely to generalize to fixed-effects NMAs. Finally, NMA 400 should be able to be used to make indirect estimates if effect modification is negligible for all 401 direct comparisons and there is no good reason to believe that non-negligible modification would 402 occur for treatment comparisons that have not been made directly. However, it would be 403 preferable to have direct evidence.

404

405 Summary and conclusions

There is very weak within-trial evidence for effect modification with respect to refractory status and number of previous lines of treatment. It is plausible that effect modification does not occur with respect to these variables or is so small as to be statistically undetectable, even in phase 3 trials. If this is true, then differences in the distributions of these variables across trials are unlikely to be a problem in NMAs. We were only able to detect effect modification by estimating ratios of hazard ratios (RHRs) across trials under assumptions that strongly favor the modification hypothesis. These assumptions may not hold, so our estimates of the magnitude of effect

Working Paper — This document has not yet been submitted for peer review and may be revised.

- 413 modification may be exaggerated, as may our estimates of the percentages of NMA estimates
- 414 that would be expected to be affected.

415

- 416 Adequately performed random-effects NMAs can probably be relied upon to provide direct and
- 417 indirect estimates of mean HRs for OS and PFS, subject to the caveats discussed above.

418

Working Paper — This document has not yet been submitted for peer review and may be revised.

420 AUTHOR CONTRIBUTIONS

- 421 Study conception and design: CJR
- 422 Acquisition, analysis, or interpretation of data: CJR, IKO, LG, GEN, AF
- 423 Drafting the manuscript: CJR, IKO, GEN, AF
- 424 Critical revision of the manuscript for important intellectual content: IKO, LG, GEN, AF
- 425 Statistical analysis: CJR
- 426 Supervision: AF
- 427 All authors approved the published version and agreed to be accountable for all aspects of the
- 428 work.

429

430 DATA AND SOFTWARE AVAILABILITY

- 431 Data and software are publicly available at: https://github.com/multinormal/fhi.rrmm-em.2022
- 432 The specific version used to generate the results presented herein is archived at Zenodo:
- 433 https://doi.org/10.5281/zenodo.6545584

434 DISCLOSURE OF CONFLICTS OF INTEREST AND FUNDING

- 435 None of the authors have any competing interests with respect to this work.
- 436 The authors conducted this research under the employ of the Norwegian Institute of Public
- 437 Health (Folkehelseinstituttet). The work was funded via Norway's National System for Managed
- 438 Introduction of New Health Technologies within the Specialist Health Service (Nye Metoder). The
- 439 funder had no role in the conduct of this work.

Working Paper — This document has not yet been submitted for peer review and may be revised.

440 SUPPLEMENTARY MATERIAL

- 441 The following supplementary material accompanies this manuscript:
- 442 1. Supplementary Materials.docx

Working Paper — This document has not yet been submitted for peer review and may be revised.

443 **REFERENCES**

- 444 1. van de Donk NWCJ, Pawlyn C, Yong KL. Multiple myeloma. *The Lancet*.
 445 2021;397(10272):410-427. doi:10.1016/S0140-6736(21)00135-5
- 446 2. van Beurden-Tan CHY, Franken MG, Blommestein HM, Uyl-de Groot CA, Sonneveld P.
- 447 Systematic literature review and network meta-analysis of treatment outcomes in 448 relapsed and/or refractory multiple myeloma. *Journal of Clinical Oncology*. 449 2017;35(12):1312-1319.
- Botta C, Ciliberto D, Rossi M, et al. Network meta-analysis of randomized trials in multiple
 myeloma: efficacy and safety in relapsed/refractory patients. *Blood Advances*.
 2017;1(7):455-466.
- 4. Luo XW, Du XQ, Li JL, Liu XP, Meng XY. Treatment options for refractory/relapsed multiple
 myeloma: an updated evidence synthesis by network meta-analysis. *Cancer Management and Research*. 2018;10:2817.
- Maiese EM, Ainsworth C, le Moine JG, Ahdesmäki O, Bell J, Hawe E. Comparative efficacy
 of treatments for previously treated multiple myeloma: a systematic literature review and
 network meta-analysis. *Clinical Therapeutics*. 2018;40(3):480-494.
- 459 6. Dimopoulos MA, Kaufman JL, White D, et al. A comparison of the efficacy of 460 immunomodulatory-containing regimens in relapsed/refractory multiple myeloma: a 461 network meta-analysis. *Clinical Lymphoma Myeloma and Leukemia*. 2018;18(3):163-173.
- Weisel K, Sonneveld P, Spencer A, et al. A comparison of the efficacy of
 immunomodulatory-free regimens in relapsed or refractory multiple myeloma: a network
 meta-analysis. *Leukemia & Lymphoma*. 2019;60(1):151-162.

- 465 8. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of
- 466 systematic reviews incorporating network meta-analyses of health care interventions:
- 467 checklist and explanations. *Ann Intern Med*. 2015;162(11):777-784.
- 468 9. Bagg MK, Salanti G, McAuley JH. Comparing interventions with network meta-analysis.
- 469 *Journal of Physiotherapy*. 2018;64(2):128-132.
- 470 10. Dias S, Caldwell DM. Network meta-analysis explained. Archives of Disease in Childhood -
- 471 *Fetal and Neonatal Edition*. 2019;104:F8-F12.
- 472 11. van der Weele TJ. On the distinction between interaction and effect modification.
- 473 *Epidemiology*. 2009;20(6):863-871.
- 474 12. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect
 475 modification and interaction. *Int J Epidemiol*. 2012;41(2):514-520.
- 476 13. Cope S, Clemens A, Hammes F, Noack H, Jansen JP. Critical appraisal of network meta-
- 477 analyses evaluating the efficacy and safety of new oral anticoagulants in atrial fibrillation
- 478 stroke prevention trials. *Value in Health*. 2015;18(2):234-249.
- 479 14. Desser A, Fretheim A, Ohm IK, et al. *Treatments for Relapsing, Refractory Multiple*480 *Myeloma: A Health Technology Assessment Project Description.*; 2021.
- 481 15. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated
 482 guideline for reporting systematic reviews. *International Journal of Surgery*.
 483 2021;88:105906.
- 484 16. San Miguel JF, Weisel KC, Song KW, et al. Impact of prior treatment and depth of response
- 485 on survival in MM-003, a randomized phase 3 study comparing pomalidomide plus low-

- dose dexamethasone versus high-dose dexamethasone in relapsed/refractory multiple
 myeloma. *Haematologica*. 2015;100(10):1334.
- 488 17. Richardson PG, Hungria VTM, Yoon SS, et al. Panobinostat plus bortezomib and
- 489 dexamethasone in previously treated multiple myeloma: outcomes by prior treatment.
- 490 *Blood*. 2016;127(6):713-721.
- 491 18. Dimopoulos MA, Stewart AK, Masszi T, et al. Carfilzomib–lenalidomide–dexamethasone vs
- 492 lenalidomide–dexamethasone in relapsed multiple myeloma by previous treatment. *Blood*
- 493 *Cancer Journal*. 2017;7(4):e554–e554.
- 494 19. Mateos MV, Masszi T, Grzasko N, et al. Impact of prior therapy on the efficacy and safety
- 495 of oral ixazomib-lenalidomide-dexamethasone vs. placebo-lenalidomide-dexamethasone
- 496 in patients with relapsed/refractory multiple myeloma in TOURMALINE-MM1.
 497 *Haematologica*. 2017;102(10):1767.
- Moreau P, Joshua D, Chng WJ, et al. Impact of prior treatment on patients with relapsed
 multiple myeloma treated with carfilzomib and dexamethasone vs bortezomib and
 dexamethasone in the phase 3 ENDEAVOR study. *Leukemia*. 2017;31(1):115-122.
- 501 21. Bringhen S, Pour L, Vorobyev V, et al. Isatuximab plus pomalidomide and dexamethasone
 502 in patients with relapsed/refractory multiple myeloma according to prior lines of
 503 treatment and refractory status: ICARIA-MM subgroup analysis. *Leukemia Research*.
 504 2021;104:106576.
- 505 22. Dimopoulos M, Weisel K, Moreau P, et al. Pomalidomide, bortezomib, and dexamethasone 506 for multiple myeloma previously treated with lenalidomide (OPTIMISMM): outcomes by 507 prior treatment at first relapse. *Leukemia*. 2021;35(6):1722-1731.

- 508 23. Mateos M v, Gavriatopoulou M, Facon T, et al. Effect of prior treatments on selinexor,
- bortezomib, and dexamethasone in previously treated multiple myeloma. *Journal of Hematology & Oncology*. 2021;14(1):1-5.
- 511 24. Quach H, Nooka A, Samoylova O, et al. Carfilzomib, dexamethasone and daratumumab in
- 512 relapsed or refractory multiple myeloma: results of the phase III study CANDOR by prior
- 513 lines of therapy. *British Journal of Haematology*. 2021;194(4):784-788.
- 514 25. Rücker G, Petropoulou M, Schwarzer G. Network meta-analysis of multicomponent 515 interventions. *Biometrical Journal*. 2020;62(3):808-821. doi:10.1002/bimj.201800167
- 516 26. R Core Team. *R: A Language and Environment for Statistical Computing.*; 2018. 517 https://www.R-project.org/
- 518 27. Rücker G, Krahn U, König J, Efthimiou O, Schwarzer G. *Netmeta: Network Meta-Analysis*519 Using Frequentist Methods.; 2021. https://CRAN.R-project.org/package=netmeta
- 520 28. Dimopoulos MA, Moreau P, Palumbo A, et al. Carfilzomib and dexamethasone versus
- 521 bortezomib and dexamethasone for patients with relapsed or refractory multiple
- 522 myeloma (ENDEAVOR): a randomised, phase 3, open-label, multicentre study. *The Lancet*
- 523 *Oncology*. 2016;17(1):27-38.
- 524 29. Mateos MV, Sonneveld P, Hungria V, et al. Daratumumab, bortezomib, and 525 dexamethasone versus bortezomib and dexamethasone in patients with previously 526 treated multiple myeloma: three-year follow-up of CASTOR. *Clinical Lymphoma Myeloma* 527 *and Leukemia*. 2020;20(8):509-518.

- 52830.Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of529methodological quality associated with estimates of treatment effects in controlled trials.
- 530 *JAMA*. 1995;273(5):408-412.
- 531 31. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment 532 comparisons. *Statistics in Medicine*. 2004;23(20):3105-3124.
- 352 compansons. statistics in medicine. 2004,25(20).5105 5124.
- 53332.Attal M, Richardson PG, Rajkumar SV, et al. Isatuximab plus pomalidomide and low-dose534dexamethasone versus pomalidomide and low-dose dexamethasone in patients with535relapsed and refractory multiple myeloma (ICARIA-MM): a randomised, multicentre,
- 536 open-label, phase 3 study. *The Lancet*. 2019;394(10214):2096-2107.
- 33. 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(7):1875-1884.
- 540 34. Dimopoulos M, Spencer A, Attal M, et al. Lenalidomide plus dexamethasone for relapsed
 541 or refractory multiple myeloma. *New England Journal of Medicine*. 2007;357(21):2123542 2132.
- 543 35. Dimopoulos M, Siegel DS, Lonial S, et al. Vorinostat or placebo in combination with 544 bortezomib in patients with multiple myeloma (VANTAGE 088): a multicentre, 545 randomised, double-blind study. *The Lancet Oncology*. 2013;14(11):1129-1140.
- 546 36. Dimopoulos MA, Dytfeld D, Grosicki S, et al. Elotuzumab plus pomalidomide and 547 dexamethasone for multiple myeloma. *New England Journal of Medicine*. 548 2018;379(19):1811-1822.

- 549 37. Dimopoulos MA, Lonial S, White D, et al. Elotuzumab, lenalidomide, and dexamethasone
 550 in RRMM: Final overall survival results from the phase 3 randomized ELOQUENT-2 study.
 551 *Blood Cancer Journal*. 2020;10(9):1-10.
 552 38. Dimopoulos MA, Terpos E, Boccadoro M, et al. Daratumumab plus pomalidomide and
- 553 dexamethasone versus pomalidomide and dexamethasone alone in previously treated 554 multiple myeloma (APOLLO): an open-label, randomised, phase 3 trial. *The Lancet*
- 555 *Oncology*. 2021;22(6):801-812.
- 556 39. Dimopoulos M, Bringhen S, Anttila P, et al. Isatuximab as monotherapy and combined with 557 dexamethasone in patients with relapsed/refractory multiple myeloma. *Blood*. 558 2021;137(9):1154-1165.
- Garderet L, Iacobelli S, Moreau P, et al. Superiority of the triple combination of
 bortezomib-thalidomide-dexamethasone over the dual combination of thalidomidedexamethasone in patients with multiple myeloma progressing or relapsing after
 autologous transplantation: the MMVAR/IFM 2005-04 Randomized Phase III Trial from the
 Chronic Leukemia Working Party of the European Group for Blood and Marrow
 Transplantation. *Journal of Clinical Oncology*. 2012;30(20):2475-2482.
- 565 41. Grosicki S, Simonova M, Spicka I, et al. Once-per-week selinexor, bortezomib, and 566 dexamethasone versus twice-per-week bortezomib and dexamethasone in patients with 567 multiple myeloma (BOSTON): a randomised, open-label, phase 3 trial. *The Lancet*. 568 2020;396(10262):1563-1573.

- 569 42. Hájek R, Masszi T, Petrucci MT, et al. A randomized phase III study of carfilzomib vs low-
- 570 dose corticosteroids with optional cyclophosphamide in relapsed and refractory multiple 571 myeloma (FOCUS). *Leukemia*. 2017;31(1):107-114.
- 572 43. Hou J, Jin J, Xu Y, et al. Randomized, double-blind, placebo-controlled phase III study of
- 573 ixazomib plus lenalidomide-dexamethasone in patients with relapsed/refractory multiple
- 574 myeloma: China Continuation study. *Journal of Hematology & Oncology*. 2017;10(1):1-13.
- 575 44. Iida S, Wakabayashi M, Tsukasaki K, et al. Bortezomib plus dexamethasone vs thalidomide
- 576 plus dexamethasone for relapsed or refractory multiple myeloma. *Cancer Science*.
- 577 2018;109(5):1552-1561.
- 578 45. Jakubowiak A, Offidani M, Pégourie B, et al. Randomized phase 2 study: elotuzumab plus
 579 bortezomib/dexamethasone vs bortezomib/dexamethasone for relapsed/refractory MM.
- 580 *Blood*. 2016;127(23):2833-2840.
- Kropff M, Vogel M, Bisping G, et al. Bortezomib and low-dose dexamethasone with or
 without continuous low-dose oral cyclophosphamide for primary refractory or relapsed
 multiple myeloma: a randomized phase III study. *Annals of Hematology*.
 2017;96(11):1857-1866.
- 585 47. Kumar SK, Harrison SJ, Cavo M, et al. Venetoclax or placebo in combination with 586 bortezomib and dexamethasone in patients with relapsed or refractory multiple myeloma 587 (BELLINI): a randomised, double-blind, multicentre, phase 3 trial. *The Lancet Oncology*. 588 2020;21(12):1630-1642.
- 589 48. Lonial S, Dimopoulos M, Palumbo A, et al. Elotuzumab therapy for relapsed or refractory
 590 multiple myeloma. *New England Journal of Medicine*. 2015;373(7):621-631.

Working Paper — This document has not yet been submitted for peer review and may be revised.

591	49.	Lu J, Fu W, Li W, et al. Daratumumab, bortezomib, and dexamethasone versus bortezomib
592		and dexamethasone in chinese patients with relapsed or refractory multiple myeloma:
593		Phase 3 LEPUS (MMY3009) study. Clinical Lymphoma Myeloma and Leukemia.
594		2021;21(9):e699–e709.

- 595 50. Mateos MV, Blacklock H, Schjesvold F, et al. Pembrolizumab plus pomalidomide and

dexamethasone for patients with relapsed or refractory multiple myeloma (KEYNOTE-

- 597 183): a randomised, open-label, phase 3 trial. *The Lancet Haematology*. 2019;6(9):e459–
- 598 e469.

- 599 51. Montefusco V, Corso A, Galli M, et al. Bortezomib, cyclophosphamide, dexamethasone 600 versus lenalidomide, cyclophosphamide, dexamethasone in multiple myeloma patients at 601 first relapse. *British Journal of Haematology*. 2020;188(6):907-917.
- Moreau P, Masszi T, Grzasko N, et al. Oral ixazomib, lenalidomide, and dexamethasone for
 multiple myeloma. *New England Journal of Medicine*. 2016;374(17):1621-1634.
- 53. Moreau P, Dimopoulos MA, Mikhael J, et al. Isatuximab, carfilzomib, and dexamethasone
- in relapsed multiple myeloma (IKEMA): a multicentre, open-label, randomised phase 3
 trial. *The Lancet*. 2021;397(10292):2361-2371.
- 607 54. Orlowski RZ, Nagler A, Sonneveld P, et al. Randomized phase III study of pegylated
 608 liposomal doxorubicin plus bortezomib compared with bortezomib alone in relapsed or
 609 refractory multiple myeloma: combination therapy improves time to progression. *Journal*610 of Clinical Oncology. 2007;25(25):3892-3901.
- 611 55. Orlowski RZ, Gercheva L, Williams C, et al. A phase 2, randomized, double-blind, placebo612 controlled study of siltuximab (anti-IL-6 mAb) and bortezomib versus bortezomib alone in

- 613 patients with relapsed or refractory multiple myeloma. *American Journal of Hematology*.
- 614 2015;90(1):42-49.
- 615 56. Orlowski RZ, Nagler A, Sonneveld P, et al. Final overall survival results of a randomized trial
- 616 comparing bortezomib plus pegylated liposomal doxorubicin with bortezomib alone in
- 617 patients with relapsed or refractory multiple myeloma. *Cancer*. 2016;122(13):2050-2056.
- 618 57. Orlowski RZ, Moreau P, Niesvizky R, et al. Carfilzomib-dexamethasone versus bortezomib-
- 619 dexamethasone in relapsed or refractory multiple myeloma: updated overall survival,
- 620 safety, and subgroups. *Clinical Lymphoma Myeloma and Leukemia*. 2019;19(8):522-530.
- 58. Palumbo A, Chanan-Khan A, Weisel K, et al. Daratumumab, bortezomib, and
 dexamethasone for multiple myeloma. *New England Journal of Medicine*.
 2016;375(8):754-766.
- 59. Raje NS, Moreau P, Terpos E, et al. Phase 2 study of tabalumab, a human anti-B-cell activating factor antibody, with bortezomib and dexamethasone in patients with previously treated multiple myeloma. *British Journal of Haematology*. 2017;176(5):783-795.
- 628 60. Richardson PG, Sonneveld P, Schuster M, et al. Extended follow-up of a phase 3 trial in
 629 relapsed multiple myeloma: final time-to-event results of the APEX trial. *Blood*.
 630 2007;110(10):3557-3560.
- 631 61. Richardson PG, Siegel DS, Vij R, et al. Pomalidomide alone or in combination with low-dose
 632 dexamethasone in relapsed and refractory multiple myeloma: a randomized phase 2
 633 study. *Blood*. 2014;123(12):1826-1832.

Working Paper — This document has not yet been submitted for peer review and may be revised.

634	62.	Richardson PG, Oriol A, Beksac M, et al. Pomalidomide, bortezomib, and dexamethasone
635		for patients with relapsed or refractory multiple myeloma previously treated with
636		lenalidomide (OPTIMISMM): a randomised, open-label, phase 3 trial. The Lancet Oncology.
637		2019;20(6):781-794.
638	63.	Richardson PG, Kumar SK, Masszi T, et al. Final overall survival analysis of the
639		TOURMALINE-MM1 phase III trial of ixazomib, lenalidomide, and dexamethasone in
640		patients with relapsed or refractory multiple myeloma. Journal of Clinical Oncology.
641		2021;39(22):2430-2442.

642 64. San Miguel J, Weisel K, Moreau P, et al. Pomalidomide plus low-dose dexamethasone
643 versus high-dose dexamethasone alone for patients with relapsed and refractory multiple
644 myeloma (MM-003): a randomised, open-label, phase 3 trial. *The lancet oncology*.
645 2013;14(11):1055-1066.

646 65. San-Miguel JF, Hungria VTM, Yoon SS, et al. Panobinostat plus bortezomib and 647 dexamethasone versus placebo plus bortezomib and dexamethasone in patients with 648 relapsed or relapsed and refractory multiple myeloma: a multicentre, randomised, double-649 blind phase 3 trial. *The Lancet Oncology*. 2014;15(11):1195-1206.

650 66. San-Miguel JF, Hungria VTM, Yoon SS, et al. Overall survival of patients with relapsed 651 multiple myeloma treated with panobinostat or placebo plus bortezomib and 652 dexamethasone (the PANORAMA 1 trial): a randomised, placebo-controlled, phase 3 trial.

653 *The Lancet Haematology*. 2016;3(11):e506–e515.

Working Paper — This document has not yet been submitted for peer review and may be revised.

- 654 67. Siegel DS, Dimopoulos MA, Ludwig H, et al. Improvement in overall survival with 655 carfilzomib, lenalidomide, and dexamethasone in patients with relapsed or refractory 656 multiple myeloma. *Journal of Clinical Oncology*. 2018;36(8):728-734.
- 657 68. Stewart AK, Rajkumar SV, Dimopoulos MA, et al. Carfilzomib, lenalidomide, and 658 dexamethasone for relapsed multiple myeloma. *New England Journal of Medicine*. 659 2015;372(2):142-152.
- 660 69. Usmani SZ, Quach H, Mateos MV, et al. Carfilzomib, dexamethasone, and daratumumab

661 versus carfilzomib and dexamethasone for patients with relapsed or refractory multiple

662 myeloma (CANDOR): updated outcomes from a randomised, multicentre, open-label,

663 phase 3 study. *The Lancet Oncology*. 2022;23(1):65-76.

- Weber DM, Chen C, Niesvizky R, et al. Lenalidomide plus dexamethasone for relapsed
 multiple myeloma in North America. *New England Journal of Medicine*.
 2007;357(21):2133-2142.
- 667 71. White D, Kassim A, Bhaskar B, Yi J, Wamstad K, Paton VE. Results from AMBER, a
 668 randomized phase 2 study of bevacizumab and bortezomib versus bortezomib in relapsed
 669 or refractory multiple myeloma. *Cancer*. 2013;119(2):339-347.

670

TABLES

Table 1: Overview of included publications

		Could report HRs stratified by [*]		Does report HRs stratified by**			
Publication (Trial)	Comparison [†]	Refractory status ^{††}	Lines of treatment —	Refractory status ^{††}		Lines of treatment	
				OS	PFS	OS	PFS
Attal 2019 (ICARIA-MM) ³²	IsPd vs Pd	\checkmark	\checkmark		\checkmark		\checkmark
Bahlis 2020 (POLLUX) ³³	DRd vs Rd	\checkmark	\checkmark		\checkmark		\checkmark
Dimopoulos 2007 (MM-010) ³⁴	Rd vs d	\checkmark	\checkmark				
Dimopoulos 2013 (VANTAGE 088) ³⁵	VorV vs V	\checkmark	\checkmark		\checkmark		\checkmark
Dimopoulos 2016 (ENDEAVOR) ²⁸	Kd vs Vd	\checkmark	\checkmark		\checkmark		\checkmark
Dimopoulos 2018 (ELOQUENT-3) ³⁶	EPd vs Pd	\checkmark	\checkmark				
Dimopoulos 2020 (ELOQUENT-2) ³⁷	ERd vs Rd	\checkmark	\checkmark		\checkmark		\checkmark
Dimopoulos 2021 (APOLLO) ³⁸	DPd vs Pd	\checkmark	\checkmark				
Dimopoulos 2021 ³⁹	ls vs Isd		\checkmark				
Garderet 2012 (MMVAR/IFM 2005-04) ⁴⁰	VTd vs Td	\checkmark					
Grosicki 2020 (BOSTON) ⁴¹	SeVd vs Vd	\checkmark	\checkmark		\checkmark		\checkmark
Hájek 2017 (FOCUS (PX-171-011)) ⁴²	K vs d	\checkmark	\checkmark				
Hou 2017 (China Cont. Study) ⁴³	IRd vs Rd	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
lida 2018 ⁴⁴	Vd vs Td		\checkmark				
Jakubowiak 2016 ⁴⁵	EVd vs Vd	\checkmark	\checkmark		\checkmark		\checkmark
Kropff 2017 ⁴⁶	Vd vs CyVd	\checkmark	\checkmark				
Kumar 2020 (BELLINI) ⁴⁷	VeVd vs Vd	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Lonial 2015 (ELOQUENT-2) ⁴⁸	Eld vs Ld	\checkmark	\checkmark		\checkmark		
Lu 2021 (LEPUS) ⁴⁹	DVd vs Vd	\checkmark	\checkmark		\checkmark		\checkmark
Mateos 2019 (KEYNOTE-183) ⁵⁰ PemPd v		\checkmark	\checkmark				
Mateos 2020 (CASTOR) ²⁹	DVd vs Vd	\checkmark	\checkmark		\checkmark		\checkmark
Montefusco 2020 ⁵¹	CyVd vs CyRd	\checkmark					

Moreau 2016 (TOURMALINE-MM1) ⁵²	IRd vs Rd	\checkmark	\checkmark					
Moreau 2021 (IKEMA) ⁵³	lsKd vs Kd	\checkmark	\checkmark					
Orlowski 2007 (DOXIL-MMY-3021) ⁵⁴	V vs DoxV	\checkmark	\checkmark		\checkmark			
Orlowski 2015 ⁵⁵	SV vs V	\checkmark	\checkmark					
Orlowski 2016 (DOXIL-MMY-3021) ⁵⁶	V vs DoxV	\checkmark	\checkmark	\checkmark				
Orlowski 2019 (ENDEAVOR) ⁵⁷	Kd vs Vd	\checkmark	\checkmark	\checkmark		\checkmark		
Palumbo 2016 (CASTOR) ⁵⁸	DVd vs Vd	\checkmark	\checkmark		\checkmark		\checkmark	
Raje 2017 ⁵⁹	TabVd vs Vd	\checkmark	\checkmark					
Richardson 2007 (APEX) ⁶⁰	V vs d	\checkmark	\checkmark					
Richardson 2014 (MM-02) ⁶¹	Pd vs P	\checkmark	\checkmark					
Richardson 2019 (OPTIMISMM) ⁶²	PVd vs Vd	\checkmark	\checkmark		\checkmark		\checkmark	
Richardson 2021 (TOURMALINE-MM1) ⁶³	IRd vs Rd	\checkmark	\checkmark	\checkmark		\checkmark		
San-Miguel 2013 (MM-03) ⁶⁴	Pd vs d	\checkmark	\checkmark			\checkmark	\checkmark	
San-Miguel 2014 (PANORAMA-1) ⁶⁵	FVd vs Vd	\checkmark	\checkmark				\checkmark	
San-Miguel 2016 (PANORAMA-1) ⁶⁶	FVd vs Vd	\checkmark	\checkmark	\checkmark		\checkmark		
Siegel 2018 (ASPIRE) ⁶⁷	KRd vs Rd	\checkmark	\checkmark			\checkmark		
Stewart 2015 (ASPIRE) ⁶⁸	KRd vs Rd	\checkmark	\checkmark		\checkmark			
Usmani 2022 (CANDOR) ⁶⁹	DKd vs Kd	\checkmark	\checkmark		\checkmark		\checkmark	
Weber 2007 (MM-09) ⁷⁰	Rd vs d	\checkmark	\checkmark					
White 2013 (AMBER) ⁷¹	BevV vs V	\checkmark	\checkmark					
Publications (%)	42	40 (95%)	40 (95%)	6 (14%)	17 (40%)	7 (17%)	18 (43%)	

*Trials that include only refractory patients, or patients who received a specific number of previous lines of treatment, cannot report stratified estimates. Determinations were made using tables of baseline characteristics and inclusion criteria.

**Publications that report stratified estimates graphically rather than numerically are classified as not reporting stratified estimates. Publications that report a stratified estimate for only one stratum (e.g., refractory but not also non-refractory) are classified as not reporting stratified estimates as they provide incomplete information. It would not be practical or possible for some trials to report stratified estimates due to very small sample sizes or no or very few events.
*Treatment names are listed in full in Supplementary Materials.

⁺⁺A broad definition of refractory status or previous treatment use was used (see Methods).

Working Paper — This document has not yet been submitted for peer review and may be revised.

675 **Table 2: Simulation study results: Percentages of NMA estimates for progression free survival**

676 *expected to differ in the presence of effect modification and heterogeneity*

Potential effect modifier	NMA estimates expected to differ (95% CI)
Refractory status*	0.35% (0.33%–0.36%)
Lines of Treatment ⁺	4.48% (4.42%–4.53%)
*A broad definition of refractory status or previous	treatment use was used (see Methods).
⁺ Results for lines of treatment assume four categor	ries, that effect modification consistently increases or
decreases with LOT, and that effect modification of	ompounds over categories (see Methods).

677

Working Paper — This document has not yet been submitted for peer review and may be revised.

679 **FIGURE LEGENDS**

680 Figure 1: Cartoon of the simulation study

Each panel shows one of the 1000 pairs of simulated networks. Within a pair, the estimates of 681 682 one network (the topmost in the cartoon) were only subject to simulated heterogeneity, while 683 the other was subject to heterogeneity and effect modification. Each network in the cartoon has 684 5 treatments (A, B, ..., E), but the simulations used 35 treatments. The magnitudes of direct 685 estimates of effect are indicated by the lengths of the links between treatments (heterogeneity and effect modification affect the magnitudes of the estimates, and in extreme cases, their 686 687 directions). Direct estimates that are particularly modified are shown as red links. NMA results 688 are indicated by the matrices. Diagonal elements are not considered further (shaded) because 689 there is no treatment effect between one treatment and itself. Lower triangles are not 690 considered further (shaded) because they are identical to the upper triangles except for direction (sign). Corresponding estimates within a pair are tested for equality, and those that differ 691 692 significantly are counted. Effect modification is quite severe in the first and final simulations 693 illustrated by the cartoon, with 7 and 8 estimates differing. No estimates are statistically different 694 in the final simulation. These numbers are merely illustrative. By performing many simulations it 695 is possible to estimate the proportion of NMA estimates that would expected to be affected by 696 the degree of effect modification observed in the literature.

697

698 Figure 2: Hazard ratios for PFS stratified by refractory status

699 Statistically significant stratified estimates of HR indicate likely treatment effect in specific patient

subgroups. Effect modification would be demonstrated by unequal stratified HRs within trial.

Working Paper — This document has not yet been submitted for peer review and may be revised.

- 701 Only one of the 17 within-trial tests for equality of stratified HRs give statistically significant
- results at the 95% significance level. Note the lack of a consistent pattern in the estimates across
- trial that would lend face-validity to the effect modification hypothesis.
- 704
- 705 Figure 3: Hazard ratios for PFS stratified by number of lines of
- 706 Statistically significant stratified estimates of HR indicate likely treatment effect in specific patient
- subgroups. Effect modification would be demonstrated by unequal stratified HRs within trial.
- 708 Only one of the 18 within-trial tests for equality of stratified HRs give statistically significant
- results at the 95% significance level. Note the lack of a consistent pattern in the estimates across
- trial that would lend face-validity to the effect modification hypothesis.
- 711

712 Figure 4: Estimates of ratios of hazard ratios for

The panels show estimates of ratios of hazard ratios (RHRs) constructed under conditions that favor the effect modification hypothesis. The top panel shows RHRs for refractory status and the bottom panel shows RHRs for number of lines of treatment. RHR = 1 corresponds to no effect modification.

Working Paper — This document has not yet been submitted for peer review and may be revised.

- 718 FIGURES
- 719

Working Paper — This document has not yet been submitted for peer review and may be revised.

721 Figure 1: Cartoon of the simulation study



Working Paper — This document has not yet been submitted for peer review and may be revised.

Figure 2: Hazard ratios for PFS stratified by refractory status 723

Study		Hazard Ratio with 95% Cl
Attal 2019 (ICARIA-MM): IsPd vs Pd		
Refractory to R		0.59 [0.43, 0.81]
Not refractory to R	<	0.18 [0.02, 1.55]
Heterogeneity: τ^{2} = 0.09, I^{2} = 12.27%, H^{2} = 1.14		0.54 [0.28, 1.01]
Test of $\theta_i = \theta_j$: Q(1) = 1.14, p = 0.29		
Bahlis 2020 (POLLUX): DRd vs Rd		
Previous R use		0.38 [0.21, 0.67]
No previous R use	-	0.45 [0.35, 0.58]
Heterogeneity: $\tau^{_2}$ = 0.00, $I^{_2}$ = 0.00%, $H^{_2}$ = 1.00		0.44 [0.35, 0.55]
Test of $\theta_i = \theta_j$: Q(1) = 0.28, p = 0.60		
Dimopoulos 2013 (VANTAGE 088): VorV vs V		
Previous IMiD use	-	0.76 [0.60, 0.96]
No IMiD use	-	0.93 [0.67, 1.28]
Heterogeneity: τ^{2} = 0.00, I^{2} = 0.00%, H^{2} = 1.00		0.81 [0.67, 0.99]
Test of $\theta_i = \theta_j$: Q(1) = 0.96, p = 0.33		
	1/32 1/8 1/2	2

Study		Hazard Ratio with 95% Cl
Dimopoulos 2016 (ENDEAVOR): Kd vs Vd		
Not refractory to R		0.44 [0.34, 0.56]
Refractory to R	-8-	0.80 [0.57, 1.12]
Heterogeneity: $\tau^2 = 0.16$, $I^2 = 87.38\%$, $H^2 = 7.92$	2	0.59 [0.33, 1.05]
Test of $\theta_i = \theta_j$: Q(1) = 7.92, p = 0.00		
Dimopoulos 2018 (ELOQUENT-3): EPd vs Pd		
Ref. to R and a PI		0.56 [0.33, 0.96]
Not ref. to R and a PI		0.51 [0.21, 1.21]
Heterogeneity: $\tau^{_2}$ = 0.00, $I^{_2}$ = 0.00%, $H^{_2}$ = 1.00		0.55 [0.35, 0.86]
Test of $\boldsymbol{\theta}_i = \boldsymbol{\theta}_j \text{:} \ Q(1) = 0.03, \ p = 0.86$		
Dimopoulos 2021 (APOLLO): DPd vs Pd		
Refractory to R		0.36 [0.15, 0.85]
Not refractory to R		0.66 [0.49, 0.89]
Heterogeneity: $\tau^2 = 0.08$, $I^2 = 41.61\%$, $H^2 = 1.71$		0.56 [0.33, 0.95]
Test of $\theta_i = \theta_j$: Q(1) = 1.71, p = 0.19		
	1/22 1/9 1/2	

Study					Hazard Ratio with 95% CI
Hou 2017 (China Cont.): IRd vs Rd					
Previous IMiD use					0.55 [0.33, 0.93]
No IMiD use					0.77 [0.15, 3.96]
Heterogeneity: $\tau^2 = 0.00, \ I^2 = 0.00\%, \ H^2 = 1.00$					0.57 [0.35, 0.94]
Test of $\theta_i = \theta_j$: Q(1) = 0.14, p = 0.71					
Jakubowiak 2016 (NCT01478048): EVd vs Vd					
Previous IMiD use					0.87 [0.58, 1.30]
No IMiD use		-			0.39 [0.19, 0.80]
Heterogeneity: $\tau^{_2}$ = 0.23, $I^{_2}$ = 72.65%, $H^{_2}$ = 3.66					0.62 [0.28, 1.34]
Test of $\theta_i = \theta_j$: Q(1) = 3.66, p = 0.06					
Kumar 2020 (BELLINI): VeVd vs Vd					
Naive					0.61 [0.31, 1.19]
Refractory to IMiD					0.75 [0.41, 1.37]
Heterogeneity: $\tau^{_2}$ = 0.00, $I^{_2}$ = 0.00%, $H^{_2}$ = 1.00					0.68 [0.44, 1.07]
Test of $\theta_i = \theta_j$: Q(1) = 0.20, p = 0.65					
	1/32	1/8	1/2	2	

Study					Hazard Ratio with 95% CI
Lonial 2015 (ELOQUENT-2): ELd vs Ld					
Previous R use				+	0.59 [0.25, 1.40]
No previous R use			-	F	0.70 [0.57, 0.86]
Heterogeneity: $\tau^2 = 0.00$, $I^2 = 0.00\%$, $H^2 = 1.00$					0.69 [0.56, 0.85]
Test of $\theta_i = \theta_j$: Q(1) = 0.14, p = 0.71					
Lu 2021 (LEPUS): DVd vs Vd					
Refractory to IMiD					0.39 [0.22, 0.69]
Not refractory to IMiD		_	—		0.16 [0.07, 0.38]
Heterogeneity: $\tau^2 = 0.26$, $I^2 = 64.38\%$, $H^2 = 2.81$					0.27 [0.11, 0.63]
Test of $\theta_i = \theta_j$: Q(1) = 2.81, p = 0.09					
Mateos 2020 (CASTOR): DVd vs Vd					
Refractory to IMiD					0.46 [0.32, 0.67]
Not refractory to IMiD			_		0.30 [0.20, 0.45]
Heterogeneity: $\tau^2 = 0.05$, $I^2 = 56.24\%$, $H^2 = 2.29$					0.37 [0.25, 0.57]
Test of $\theta_i = \theta_j$: Q(1) = 2.29, p = 0.13					
	1/32	1/8	1/2	2	_

Study					Hazard Ratio with 95% CI
Moreau 2021 (IKEMA): IsKd vs Kd					
Refractory to R					0.50 [0.28, 0.88]
Not refractory to R					0.48 [0.30, 0.77]
Heterogeneity: $\tau^{_2}$ = 0.00, $I^{_2}$ = 0.00%, $H^{_2}$ = 1.00					0.49 [0.34, 0.70]
Test of $\theta_i = \theta_j$: Q(1) = 0.01, p = 0.91					
Orlowski 2007 (NCT00103506): V vs DoxV					
Previous R or T use					1.62 [1.08, 2.42]
No previous R or T use					2.01 [1.42, 2.84]
Heterogeneity: τ^{2} = 0.00, I^{2} = 0.00%, H^{2} = 1.00					1.83 [1.41, 2.38]
Test of $\theta_i = \theta_j$: Q(1) = 0.64, p = 0.43					
Richardson 2019 (OPTIMISMM): PVd vs Vd					
Refractory to R			-		0.65 [0.50, 0.84]
Not refractory to R			_		0.48 [0.30, 0.76]
Heterogeneity: $\tau^{\scriptscriptstyle 2}$ = 0.01, $I^{\scriptscriptstyle 2}$ = 21.51%, $H^{\scriptscriptstyle 2}$ = 1.27					0.59 [0.45, 0.78]
Test of $\theta_i = \theta_j$: Q(1) = 1.27, p = 0.26					
	1/32	1/8	1/2	2	-

Study					Hazard Ratio with 95% CI
Stewart 2015 (ASPIRE): KRd vs Rd					
Refractory to IMiD					0.64 [0.45, 0.92]
Not refractory to IMiD			-		0.72 [0.58, 0.90]
Heterogeneity: $\tau^{\scriptscriptstyle 2}$ = 0.00, $I^{\scriptscriptstyle 2}$ = 0.00%, $H^{\scriptscriptstyle 2}$ = 1.00					0.70 [0.58, 0.84]
Test of $\theta_i = \theta_j$: Q(1) = 0.30, p = 0.59					
Usmani 2022 (CANDOR): DKd vs Kd					
Refractory to R					0.46 [0.28, 0.74]
Not refractory to R					0.63 [0.44, 0.90]
Heterogeneity: $\tau^2 = 0.00$, $I^2 = 5.89\%$, $H^2 = 1.06$					0.56 [0.42, 0.76]
Test of $\theta_i = \theta_j$: Q(1) = 1.06, p = 0.30					
	1/32	1/8	1/2	2	_

Working Paper — This document has not yet been submitted for peer review and may be revised.

725

726 Figure 3: Hazard ratios for PFS stratified by number of lines of treatment





Study		Hazard Ratio with 95% CI
Dimopoulos 2021 (APOLLO): DPd vs Pd		
1		0.70 [0.30, 1.65]
2-3	-=-	0.66 [0.48, 0.91]
>=4		0.40 [0.18, 0.89]
Heterogeneity: $\tau^2 = 0.00$, $I^2 = 0.00\%$, $H^2 = 1.00$		0.62 [0.47, 0.83]
Test of $\theta_i = \theta_j$: Q(2) = 1.36, p = 0.51		
Grosicki 2020 (BOSTON): SeVd vs Vd		
1		0.63 [0.41, 0.96]
2		0.65 [0.40, 1.06]
3		0.82 [0.45, 1.49]
Heterogeneity: $\tau^2 = 0.00$, $I^2 = 0.00\%$, $H^2 = 1.00$		0.68 [0.51, 0.90]
Test of $\theta_{_i}=\theta_{_j};$ Q(2) = 0.53, p = 0.77		
Hou 2017 (China Cont.): IRd vs Rd		
1		0.85 [0.40, 1.79]
2		0.37 [0.16, 0.83]
3		0.70 [0.21, 2.32]
Heterogeneity: $\tau^2 = 0.06$, $I^2 = 21.86\%$, $H^2 = 1.28$		0.60 [0.34, 1.06]
Test of $\theta_i = \theta_j$: Q(2) = 2.24, p = 0.33		
		_







Working Paper — This document has not yet been submitted for peer review and may be revised.

728 Figure 4: Estimates of ratios of hazard ratios for PFS

Ratio of hazard ratios for PFS (refractory statu
--

Study	BHBs	Mean RHR with 95% Cl c-value
Attal 2019 (ICARIA-MM): ISPd VS Pd	1	3.28 [0.55, 19.42] 0.191
Bahlis 2020 (POLLUX): DRd vs Rd	1 -	1.18 [0.78, 1.81] 0.434
Dimopoulos 2013 (VANTAGE 088): VorV vs V	1 + 🖬 -	1.22 [0.89, 1.68] 0.217
Dimopoulos 2016 (ENDEAVOR): Kd vs Vd	1	1.82 [1.20, 2.75] 0.005
Dimopoulos 2018 (ELOQUENT-3): EPd vs Pd	1	1.10 [0.59, 2.05] 0.769
Dimopoulos 2021 (APOLLO): DPd vs Pd	1	1.83 [0.83, 4.04] 0.133
Hou 2017 (China Cont.): IRd vs Rd	1	1.39 [0.46, 4.19] 0.564
Jakubowiak 2016 (NCT01478048): EVd vs Vd	1	2.23 [1.01, 4.90] 0.046
Kumar 2020 (BELLINI): VeVd vs Vd	1	1.23 [0.68, 2.22] 0.492
Lonial 2015 (ELOQUENT-2): ELd vs Ld	1 —	1.19 [0.67, 2.10] 0.557
Lu 2021 (LEPUS): DVd vs Vd	1	2.44 [0.92, 6.44] 0.072
Mateos 2020 (CASTOR): DVd vs Vd	1 -	1.53 [0.93, 2.54] 0.096
Moreau 2021 (IKEMA): IsKd vs Kd	1 —	1.04 [0.67, 1.63] 0.858
Orlowski 2007 (NCT00103506): V vs DoxV	1 -	1.24 [0.84, 1.84] 0.285
Richardson 2019 (OPTIMISMM): PVd vs Vd	1 +	1.35 [0.87, 2.10] 0.174
Stewart 2015 (ASPIRE): KRd vs Rd	1 -	1.13 [0.84, 1.50] 0.424
Usmani 2022 (CANDOR): DKd vs Kd	1 +	1.37 [0.85, 2.22] 0.202
Overall	•	1.31 [1.16, 1.47]
Heterogeneity: $\tau^2 = 0.00$, $I^2 = 0.00\%$, $H^2 = 1.00$		
Test of $\theta_i = \theta_i$: Q(16) = 10.89, p = 0.82		
Test of $\theta = 0$: $z = 4.44$, $p = 0.00$		
	1/2 1 2 4	8 16
Random-effects REML model		

Ratio of hazard r	atios	for PFS (lines of treatme	ent)	
			Mean RHR	
Study	RHRs	3	with 95% CI	<i>p</i> -value
Attal 2019 (ICARIA-MM): IsPd vs Pd	1		1.00 [0.68, 1.47]	1.000
Bahlis 2020 (POLLUX): DRd vs Rd	3		1.15 [0.88, 1.49]	0.307
Dimopoulos 2013 (VANTAGE 088): VorV vs V	1		1.33 [0.94, 1.90]	0.109
Dimopoulos 2016 (ENDEAVOR): Kd vs Vd	1		1.33 [0.93, 1.90]	0.112
Dimopoulos 2018 (ELOQUENT-2): ELd vs Ld	1		1.13 [0.87, 1.47]	0.349
Dimopoulos 2018 (ELOQUENT-3): EPd vs Pd	1		1.08 [0.61, 1.92]	0.798
Dimopoulos 2021 (APOLLO): DPd vs Pd	2		1.25 [0.80, 1.95]	0.320
Grosicki 2020 (BOSTON): SeVd vs Vd	2		1.11 [0.81, 1.52]	0.528
Hou 2017 (China Cont.): IRd vs Rd	2		- 2.10 [1.00, 4.41]	0.049
Jakubowiak 2016 (NCT01478048): EVd vs Vd	1		1.03 [0.67, 1.58]	0.904
Kumar 2020 (BELLINI): VeVd vs Vd	1		1.39 [0.80, 2.41]	0.243
Lu 2021 (LEPUS): DVd vs Vd	3		2.00 [0.99, 4.02]	0.053
Mateos 2020 (CASTOR): DVd vs Vd	3		1.62 [1.15, 2.29]	0.006
Moreau 2021 (IKEMA): IsKd vs Kd	1		1.23 [0.71, 2.11]	0.456
Richardson 2019 (OPTIMISMM): PVd vs Vd	2		1.17 [0.90, 1.53]	0.247
San-Miguel 2013 (MM-003): Pd vs GC	1 -		1.02 [0.56, 1.86]	0.945
San-Miguel 2014 (PANORAMA-1): FVd vs Vd	1		1.03 [0.82, 1.29]	0.790
Usamani 2022 (CANDOR): DKd vs Kd	1		1.20 [0.80, 1.79]	0.371
Overall		•	1.19 [1.09, 1.29]	
Heterogeneity: τ^2 = 0.00, I^2 = 0.00%, H^2 = 1.00				
Test of $\theta_i = \theta_j$: Q(17) = 12.15, p = 0.79				
Test of θ = 0: z = 3.94, p = 0.00				
	-	1 2	4	
Random-effects REML model				