TITLE: Effect modification in network meta-analyses of treatments for relapsing refractory multiple myeloma (RRMM): systematic review, meta-analysis, and simulation

RUNNING HEAD: Effect modification in NMAs for RRMM

Christopher James Rose[1,2,*]; Ingrid Kristine Ohm[1]; Liv Giske[1]; Gunn Eva Næss[1]; Atle Fretheim[2]

1. Reviews and Health Technology Assessments, Division for Health Services, Norwegian Institute of Public Health, Oslo, Norway
2. Center for Epidemic Interventions Research, Norwegian Institute of Public Health, Oslo, Norway

* Corresponding author (cjro@fhi.no)

Counts:

Abstract:	250 words	(limit = 250 words)
Main text:	4306 words	(limit = 4000 words)
Number of figures/tables:	6	(limit = 7 figures/tables)
Number of references:	71	(limit = 100 references)

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.

KEYWORDS

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

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]

If the assumptions underpinning an NMA model are sufficiently satisfied, NMA facilitates meta-analytical 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).

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

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 *fixed-effects* 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
85 The use of NMA in RRMM has been criticized[13] on the basis that variables such as refractory
86 status and LOT are effect modifiers, with the implication that NMAs that do not account for effect
87 modification may be untrustworthy. The present article was motivated by an ongoing health
88 technology assessment (HTA) we are conducting on treatments for RRMM that was

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 checklists[15] are available in Supplementary Materials.
102

103 ***Literature search strategy***

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

111 our HTA. Via manual searching we found nine articles reporting stratified estimates for the
112 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
121 We excluded publications from meta-analysis if they did not report stratified estimates of HR for
122 all strata for at least one of two potential effect-modifiers (e.g., we would have excluded a study
123 if it did report an estimate for lenalidomide-refractory patients but did not report an estimate for
124 patients not refractory to lenalidomide). We excluded publications that did not report numerical
125 statements of uncertainty on stratified estimates (e.g., we excluded one study that reported
126 point estimates numerically but only provided a graphical presentation of the confidence
127 intervals).
128
129 ***Data extraction***
130 CJR extracted data from trial reports and supplementary materials. Extracted data were checked
131 independently by IKO. Disagreements were resolved via discussion. We extracted estimates of
132 HR for OS and PFS, stratified by LOT and refractory status or previous use of immunomodulatory

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***

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

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

RHRs were computed for each trial as follows (a formal definition is provided in Supplementary Materials). First, the trial's strata were sorted to ensure that the order of strata have similar interpretations across trials and are therefore amenable to meta-analysis. For example, LOT strata were sorted from fewest to most LOT, and previous lenalidomide use was nominated as the first (i.e., reference) level of the refractory status factor variable. Then, we computed the 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 less than one to ensure that point estimates for all RHRs are greater than or equal to one. This 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). Standard errors on RHRs were computed as described in Supplementary Materials. We excluded reference strata from meta-analysis because, as references, they are not defined with respect to another stratum.

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
183 All meta-analyses were performed on the logarithmic scale. We used random-effects models
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
186 must be accounted for statistically. We present results using forest plots, sub-grouped by
187 publication, to report estimates of mean HRs or mean RHRs, 95% confidence intervals, and I^2 and
188 p-values throughout. We used the conventional $p < 0.05$ criterion for statistical significance.
189 Statistical analyses were performed using Stata 16 (StataCorp LLC, College Station, Texas, USA).
190 Data and code are freely available (see the Data and Software Availability section).
191
192 ***Simulation studies***
193 To help understand the degree to which effect modification may affect NMA results, we
194 performed two simulation studies (plus various sensitivity analyses; see Discussion). The purpose
195 of the simulations was to estimate the percentage of NMA estimates that would be expected to
196 be statistically significantly different under effect modification compared to no effect

197 modification, due to refractory status and LOT. Figure 1 shows a cartoon that illustrates the
198 design of these studies.
199

Insert Figure 1 about here

200
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
206 modification. All networks had the same topology as the network for PFS. Simulated effect sizes
207 (log HRs) and their standard errors were drawn from distributions that matched those for the PFS
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
213 estimates of RHR. We fitted random-effects component-NMA models[25] to each pair of simulated
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

219 version 3.5.2[26] with component NMAs performed using the netmeta[27] 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***

232 Almost all trials could have reported stratified estimates, but only 17 (40%) publications,
233 representing 8364 patients, did report estimates stratified by refractory status (Table 1).
234 Similarly, 18 (43%) publications, representing 7,503 patients, did report estimates stratified by
235 LOT (Table 1). Within-trial evidence for effect modification of HR for PFS by refractory status and
236 LOT is weak (Figure 2 and Figure 3). Only one test for equality of stratified HRs was statistically
237 significant with respect to refractory status ($p < 0.01$ for the comparison Kd vs Vd[28]) and another
238 with respect to LOT ($p = 0.01$ for the comparison DVd vs Vd[29]).

239

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; $p < 0.005$; $I^2 = 0\%$) for refractory status and 1.19 (95% CI 1.09 to 1.29; $p < 0.01$; $I^2 = 0\%$) 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; $p = 0.03$; $I^2 = 0\%$) for refractory status and 1.09 (95% CI 0.98 to 1.20; $p = 0.12$; $I^2 = 0\%$) for LOT (Supplementary Materials). No statistical 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 LOT did not introduce undue heterogeneity.

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

Insert Table 2 about here

271

272 **DISCUSSION**

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

280

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
284 than 1% of NMA estimates are likely to be statistically different than they would be if effect
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
291 Absence of evidence is not evidence of absence, and we may simply not have sufficient data to
292 detect the impact of effect modification. Still, if effect modification does occur, we would expect
293 to see consistent patterns which support effect modification, which we do not. In some cases,
294 estimates increase with refractory status or LOT, in others it is opposite, but in most cases the
295 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,

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
319 While we systematically reviewed evidence of effect modification with respect to refractory
320 status and number of lines of treatment, we did not systematically review other variables that
321 also may be effect modifiers. However, we did look at all stratified estimates when extracting
322 data and did not notice any variables that appeared to consistently demonstrate convincing
323 evidence of effect modification.
324

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

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.

Quantities such as ratios of hazard ratios, as used in meta-research,[30] are likely challenging to interpret, and we suspect that few will have an intuitive understanding of what constitutes a "large" or "important" RHR with respect to effect modification in RRMM. A major strength of this work is that having estimated RHRs, we then used simulations to investigate how many NMA estimates would be expected to be statistically significantly different under the estimated degree of effect modification. We hope this helps readers understand the likely impact of any effect 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 necessarily translate to fixed-effects NMAs, as used in some systematic reviews on treatments

347 for RRMM.[4,2] It is important to note that fixed- and random-effects NMAs make fundamentally
348 different assumptions about transitivity. Fixed-effects NMAs assume that "trial-level" treatment
349 effects can be added and subtracted to make indirect estimates. Random-effects NMAs assume
350 that *average* treatment effects can be added and subtracted. Use of average treatment effects
351 explicitly accounts for differences between trials, including different distributions of effect
352 modifiers.
353
354 Finally, we also performed sensitivity analyses to investigate the implications of the assumptions
355 we made in the simulations. For example, because RHR discards direction of modification, we
356 assumed that direction of modification is consistent within treatment comparison but may vary
357 between comparisons. This may not be true, so we performed a sensitivity analysis in which
358 direction is assumed consistent within and between comparisons. The result of this analysis
359 suggests that about half as many estimates would differ statistically compared to the main
360 analysis (i.e., that the main result reflects the worst-case).
361
362 ***Implications for research***
363 Understandably, the transitivity assumption underpinning NMA is often highly simplified in
364 articles aimed at non-statisticians. For example, articles tend to use arguments about "similarity"
365 of patients.[8] Given this oversimplification, it is unsurprising there are concerns about using NMA
366 in RRMM and other areas. The transitivity assumption does not concern patient similarity, nor
367 whether treatment effect *estimates* can be added or subtracted, it concerns whether target
368 parameters can be linearly combined. Patient similarity is a good place to start thinking about

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

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

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

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

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

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

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

AUTHOR CONTRIBUTIONS

Study conception and design: CJR

Acquisition, analysis, or interpretation of data: CJR, IKO, LG, GEN, AF

Drafting the manuscript: CJR, IKO, GEN, AF

Critical revision of the manuscript for important intellectual content: IKO, LG, GEN, AF

Statistical analysis: CJR

Supervision: AF

All authors approved the published version and agreed to be accountable for all aspects of the work.

DATA AND SOFTWARE AVAILABILITY

Data and software are publicly available at: https://github.com/multinormal/fhi.rrmm-em.2022

The specific version used to generate the results presented herein is archived at Zenodo: https://doi.org/10.5281/zenodo.6545584

DISCLOSURE OF CONFLICTS OF INTEREST AND FUNDING

None of the authors have any competing interests with respect to this work.

The authors conducted this research under the employ of the Norwegian Institute of Public Health (Folkehelseinstituttet). The work was funded via Norway's National System for Managed Introduction of New Health Technologies within the Specialist Health Service (Nye Metoder). The funder had no role in the conduct of this work.

SUPPLEMENTARY MATERIAL

The following supplementary material accompanies this manuscript:

1. Supplementary Materials.docx

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.

450 3. Botta C, Ciliberto D, Rossi M, et al. Network meta-analysis of randomized trials in multiple
451 myeloma: efficacy and safety in relapsed/refractory patients. *Blood Advances*.
452 2017;1(7):455-466.

453 4. Luo XW, Du XQ, Li JL, Liu XP, Meng XY. Treatment options for refractory/relapsed multiple
454 myeloma: an updated evidence synthesis by network meta-analysis. *Cancer Management*
455 *and Research*. 2018;10:2817.

456 5. Maiese EM, Ainsworth C, le Moine JG, Ahdesmäki O, Bell J, Hawe E. Comparative efficacy
457 of treatments for previously treated multiple myeloma: a systematic literature review and
458 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.

462 7. Weisel K, Sonneveld P, Spencer A, et al. A comparison of the efficacy of
463 immunomodulatory-free regimens in relapsed or refractory multiple myeloma: a network
464 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-

486 dose dexamethasone versus high-dose dexamethasone in relapsed/refractory multiple
487 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.
498 20. Moreau P, Joshua D, Chng WJ, et al. Impact of prior treatment on patients with relapsed
499 multiple myeloma treated with carfilzomib and dexamethasone vs bortezomib and
500 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,
509 bortezomib, and dexamethasone in previously treated multiple myeloma. *Journal of*
510 *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.

528 30. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of
529 methodological 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.
533 32. Attal M, Richardson PG, Rajkumar SV, et al. Isatuximab plus pomalidomide and low-dose
534 dexamethasone versus pomalidomide and low-dose dexamethasone in patients with
535 relapsed and refractory multiple myeloma (ICARIA-MM): a randomised, multicentre,
536 open-label, phase 3 study. *The Lancet*. 2019;394(10214):2096-2107.
537 33. Bahlis NJ, Dimopoulos MA, White DJ, et al. Daratumumab plus lenalidomide and
538 dexamethasone in relapsed/refractory multiple myeloma: extended follow-up of POLLUX,
539 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):2123-
542 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.
559 40. Garderet L, Iacobelli S, Moreau P, et al. Superiority of the triple combination of
560 bortezomib-thalidomide-dexamethasone over the dual combination of thalidomide-
561 dexamethasone in patients with multiple myeloma progressing or relapsing after
562 autologous transplantation: the MMVAR/IFM 2005-04 Randomized Phase III Trial from the
563 Chronic Leukemia Working Party of the European Group for Blood and Marrow
564 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.
581 46. Kropff M, Vogel M, Bisping G, et al. Bortezomib and low-dose dexamethasone with or
582 without continuous low-dose oral cyclophosphamide for primary refractory or relapsed
583 multiple myeloma: a randomized phase III study. *Annals of Hematology*.
584 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.

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
596 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.
602 52. Moreau P, Masszi T, Grzasko N, et al. Oral ixazomib, lenalidomide, and dexamethasone for
603 multiple myeloma. *New England Journal of Medicine*. 2016;374(17):1621-1634.
604 53. Moreau P, Dimopoulos MA, Mikhael J, et al. Isatuximab, carfilzomib, and dexamethasone
605 in relapsed multiple myeloma (IKEMA): a multicentre, open-label, randomised phase 3
606 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, placebo-
612 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.
621 58. Palumbo A, Chanan-Khan A, Weisel K, et al. Daratumumab, bortezomib, and
622 dexamethasone for multiple myeloma. *New England Journal of Medicine*.
623 2016;375(8):754-766.
624 59. Raje NS, Moreau P, Terpos E, et al. Phase 2 study of tabalumab, a human anti-B-cell
625 activating factor antibody, with bortezomib and dexamethasone in patients with
626 previously treated multiple myeloma. *British Journal of Haematology*. 2017;176(5):783-
627 795.
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.

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.

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.
664 70. Weber DM, Chen C, Niesvizky R, et al. Lenalidomide plus dexamethasone for relapsed
665 multiple myeloma in North America. *New England Journal of Medicine*.
666 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
671

672 **TABLES**

673 ***Table 1: Overview of included publications***

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

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

674

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 categories, that effect modification consistently increases or decreases with LOT, and that effect modification compounds over categories (see Methods).

677

678

679 **FIGURE LEGENDS**

680 ***Figure 1: Cartoon of the simulation study***

681 Each panel shows one of the 1000 pairs of simulated networks. Within a pair, the estimates of
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
686 and effect modification affect the magnitudes of the estimates, and in extreme cases, their
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
691 (sign). Corresponding estimates within a pair are tested for equality, and those that differ
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

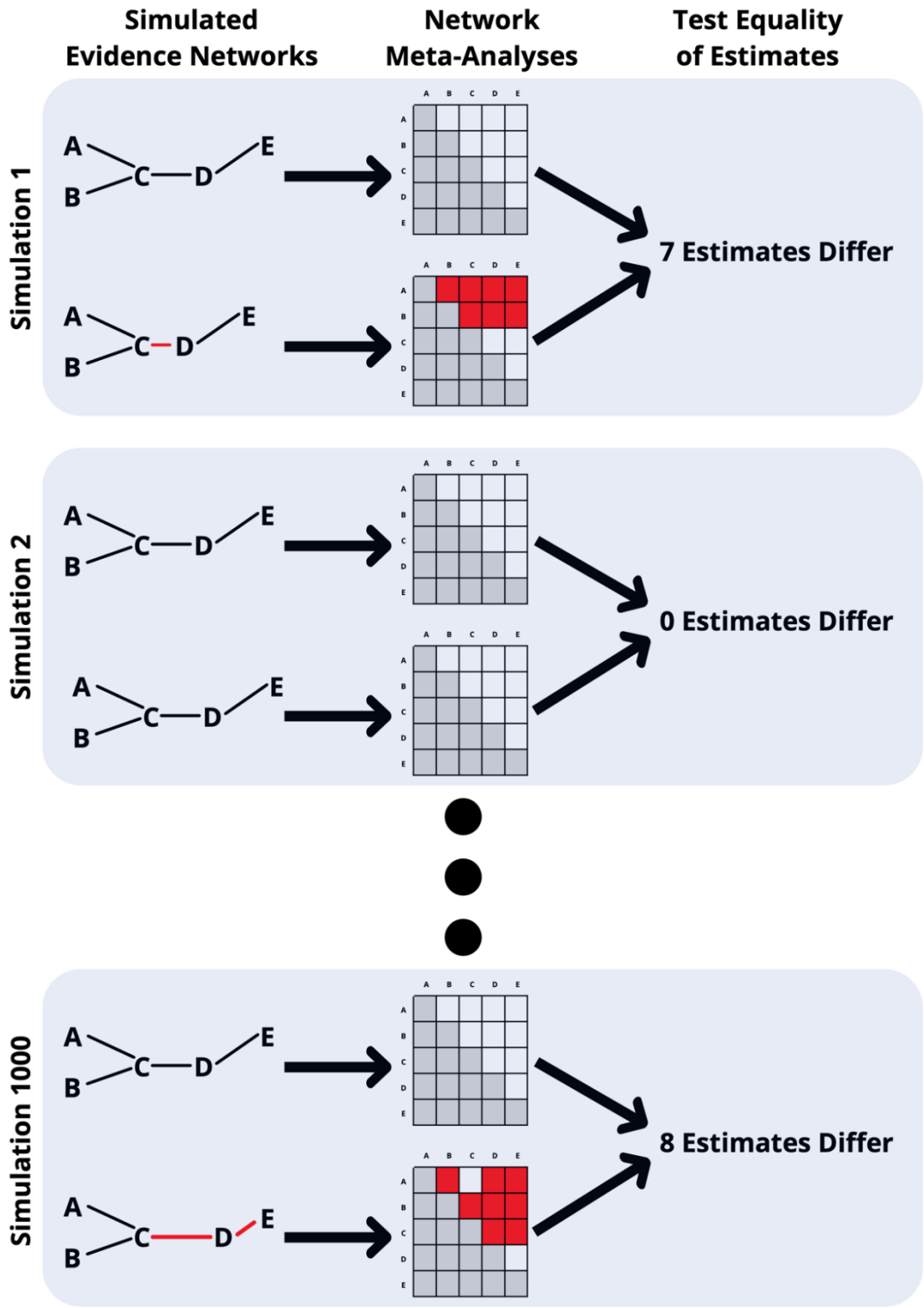
698 ***Figure 2: Hazard ratios for PFS stratified by refractory status***

699 Statistically significant stratified estimates of HR indicate likely treatment effect in specific patient
700 subgroups. Effect modification would be demonstrated by unequal stratified HRs within trial.

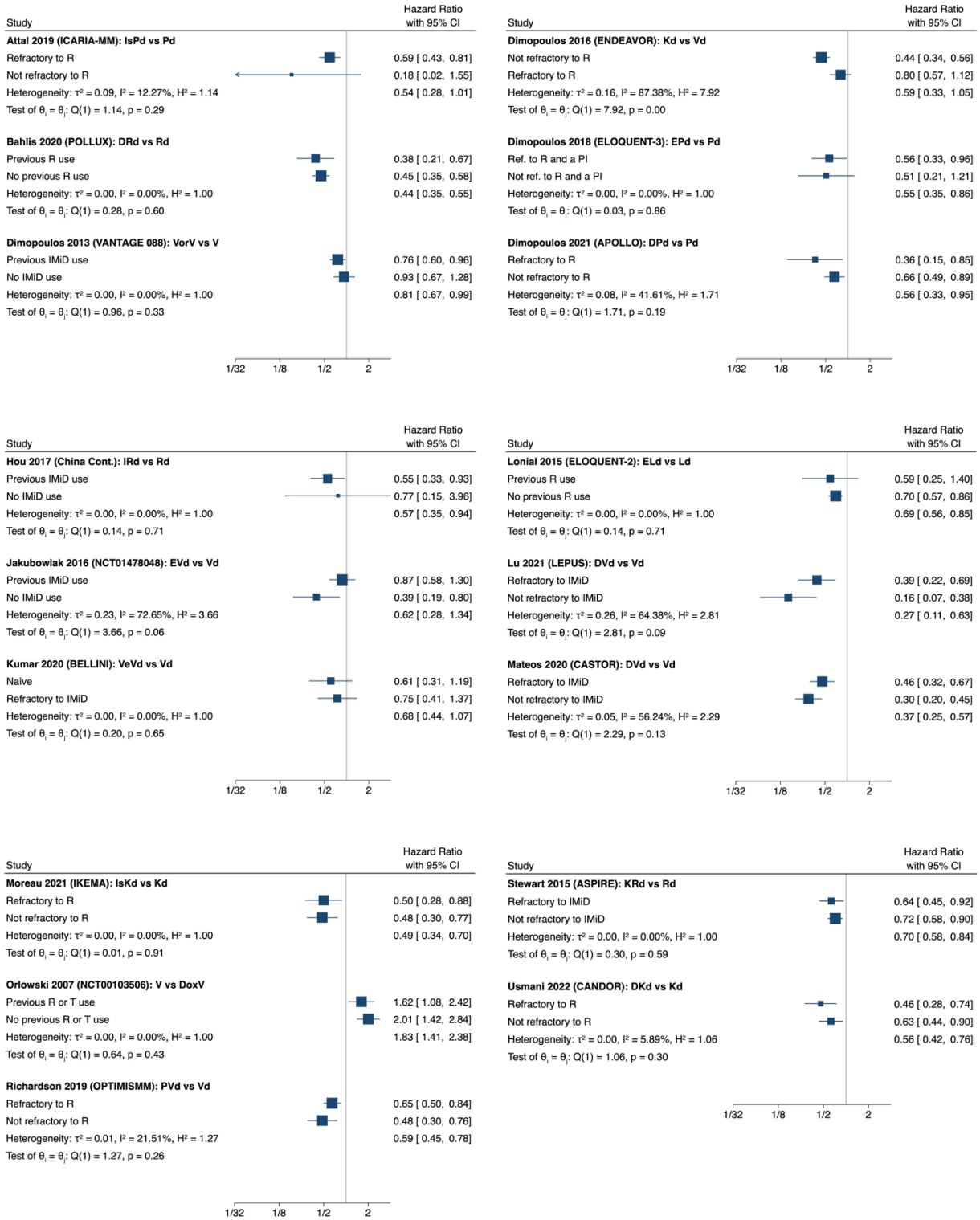
701 Only one of the 17 within-trial tests for equality of stratified HRs give statistically significant
702 results at the 95% significance level. Note the lack of a consistent pattern in the estimates across
703 trial that would lend face-validity to the effect modification hypothesis.
704
705 ***Figure 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
707 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
709 results at the 95% significance level. Note the lack of a consistent pattern in the estimates across
710 trial that would lend face-validity to the effect modification hypothesis.
711
712 ***Figure 4: Estimates of ratios of hazard ratios for***
713 The panels show estimates of ratios of hazard ratios (RHRs) constructed under conditions that
714 favor the effect modification hypothesis. The top panel shows RHRs for refractory status and the
715 bottom panel shows RHRs for number of lines of treatment. RHR = 1 corresponds to no effect
716 modification.
717

FIGURES

Figure 1: Cartoon of the simulation study



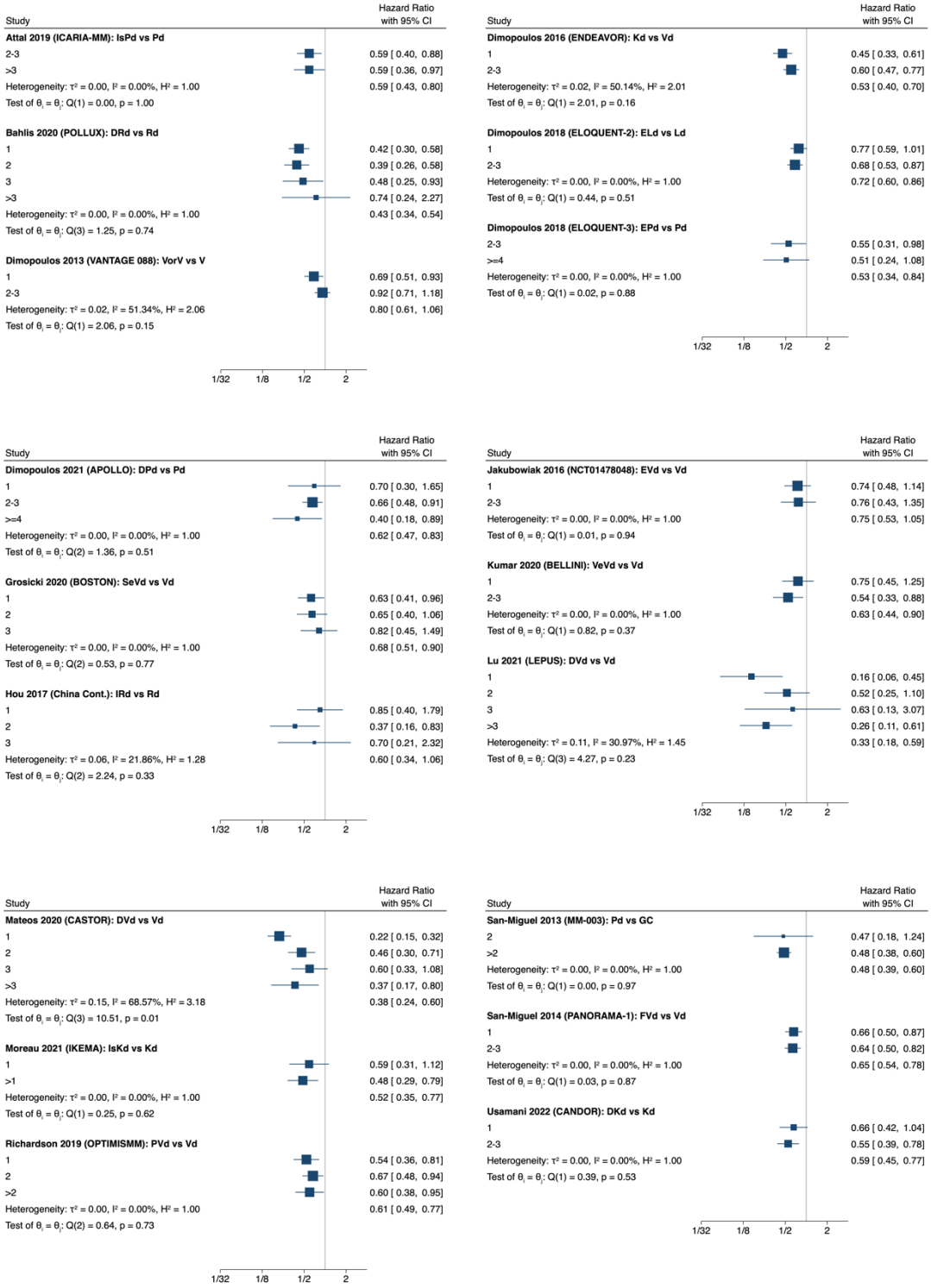
723 ***Figure 2: Hazard ratios for PFS stratified by refractory status***



724

725

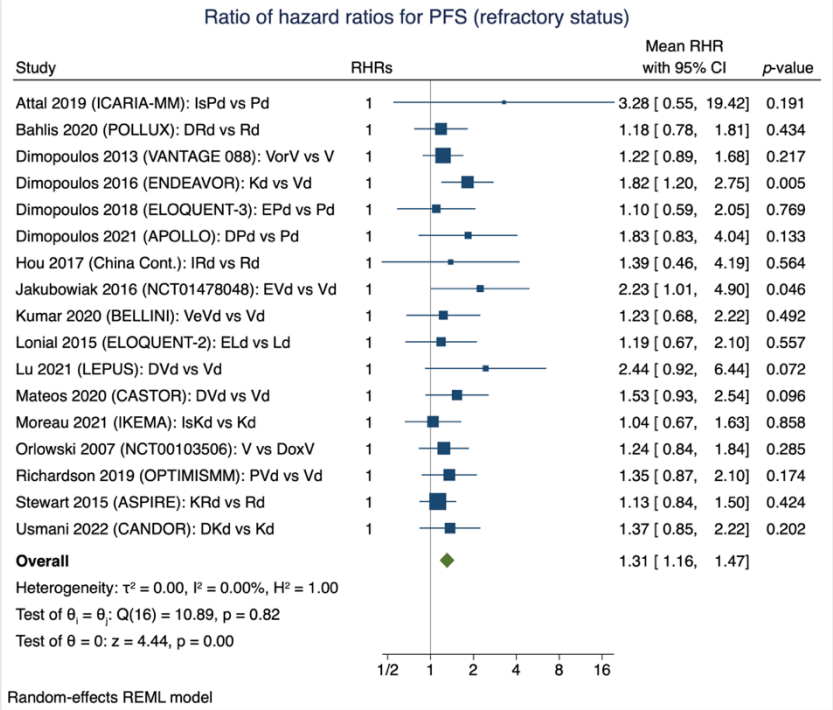
726 ***Figure 3: Hazard ratios for PFS stratified by number of lines of treatment***

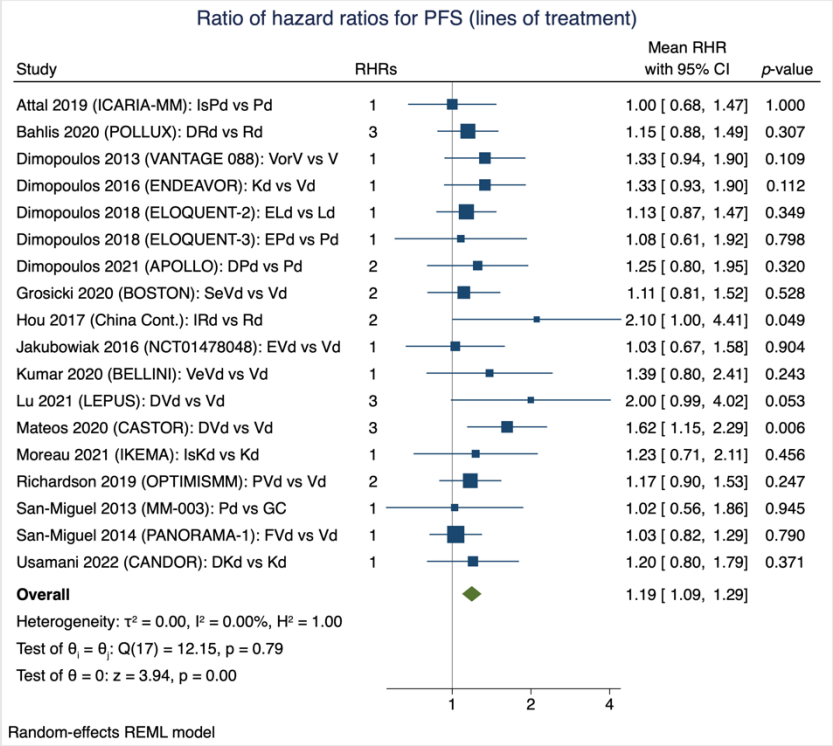


727

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

729





730