NIPH
Norwegian Institute of Public Health

REPORT

2023

HEALTH TECHNOLOGY ASSESSMENT:

Treatments for relapsing and/or refractory multiple myeloma

Utgitt av	Norwegian Institute of Public Health Division for Health Services
Title	Treatments for relapsing and/or refractory multiple myeloma. A health technology assessment
Norwegian title	Behandlinger av relapserende og/eller refraktær myelomatose: En fullstendig metodevurdering
Responsible	Camilla Stoltenberg, Director-General
Authors	Desser, Arna S., *project leader, Health economist, Norwegian Institute of Public Health* Ohm, Ingrid Kristine, *Researcher, Norwegian Institute of Public Health* Rose, Christopher James, *Statistican, Norwegian Institute of Public Health* Chaudhry, Fawaz, *Health economist, Norwegian Institute of Public Health* Næss, Gunn Eva, *Information specialist, Norwegian Institute of Public Health* Giske, Liv, *Senior researcher, Norwegian Institute of Public Health* Fretheim, Atle, *Research director, Norwegian Institute of Public Health*
ISBN	978-82-8406-364-5
Project number	ID2019_072
Type of report	Health Technology Assessment
No. of pages	130 (248 inkludert vedlegg)
Client	Bestillerforum for nye metoder
Subject heading (MeSH)	Multiple Myeloma, Bortezomib, Lenalidomide, Panobinostat, Pomalidomide, Thalidomide
Citation	Desser AS, Ohm IK, Rose CJ, Chaudhry F, Næss GE, Giske L, Fretheim A, "Treatments for relapsing and/or refractory multiple myeloma. A health technology assessment" 2023. Oslo: Norwegian Institute of Public Health, 2023.

Contents

Key messages

Multiple myeloma is the second most common type of blood cancer in Norway, with approximately 450 new cases diagnosed annually. There is currently no cure for multiple myeloma, so the goals of treatment are to prolong life, to achieve as strong a response as possible without unacceptable side effects, and to maintain the patient's quality of life at as high a level as possible throughout treatment.

We carried out a health technology assessment of different treatment regimens for patients with myltiple myeloma who are refractory to treatment, or experience relapse. We included data for five outcomes: overall survival, quality of life, severe adverse events, progression-free survival and discontinuation due to adverse events, with overall survival being our main primary outcome. Data were not available for all treatment regimens and outcomes. We inferred that there is not a single treatment regimen that is superior with respect to all outcomes. The following six triplet combinations are examples of treatment regimens relevant for non-refractory patients in a Norwegian setting, with clearly favorable hazard ratios for overall survival, that are also ranked highly with respect to other outcomes:

- [EP + d]: elotuzumab (E), pomalidomide (P) and dexamethasone (d)
- [IsP + d]: isatuximab (Is), pomalidomide (P), and dexamethasone (d)
- [DK + d]: daratumumab (D), carfilzomib (K), and dexamethasone (d)
- [KR + d]: carfilzomib (K), lenalidomide (R), and dexamethasone (d)
- [DR + d]: daratumumab (D), lenalidomide (R), and dexamethasone (d)
- [DV + d]: daratumumab (D), bortezomib (V), and dexamethasone (d)

The health economic analysis relied on a partitioned survival analysis to estimate total costs (in NOK), health gains (in quality-

Title:
Treatments for relapsing and/or refractory multiple myeloma

Type of publication:
Health technology assessment
Health technology assessment (HTA) is a multidisciplinary process that summarizes information about the medical, social, economic, and ethical issues related to the use of a health technology in a systematic, transparent, unbiased, robust manner. Its aim is to inform the development of safe, effective health policies that are patient focused and that seek to achieve best value.

Doesn't answer everything:
We do not address ethical, legal, or social aspects related to pharmacological treatment of relapsing and/or refractory multiple myeloma

---

Publisher:
Norwegian Institute of Public Health

Updated:
Last updated search for studies: January 2022

Internal peer reviewers:
Kjetil Gundro Brurberg, *department director*, Norwegian Institute of Public Health
Vida Hamadi, *health economist*, Norwegian Institute of Public Health

---

adjusted life years; QALYs), and incremental cost-effectiveness ratios (ICER) for treatments of interest. We grouped treatments into three separate groups, each based on the relevant reference treatment in order to perform the analysis. We used hazard ratios from the network meta-analysis to inform the overall and progression-free survival curves used in the model. The results of the health economic analyses were:

- In the [R + d] group, [R + d] had costs of NOK [redacted], with 2.90 QALYs gained. Only two other treatments were not dominated by other treatments: [IR + d] had costs of NOK [redacted], 3.82 QALYs, and an ICER of NOK [redacted] compared to [R + d]. [DR + d] had costs of NOK [redacted], 4.31 QALYs, and an ICER of NOK [redacted] compared to [IR + d].
- In the [V + d] group, [V + d] had costs of NOK [redacted] and 2.24 QALYs. [DV + d] was the only other treatment that was not dominated by other treatments, with costs of NOK [redacted], 3.63 QALYs, and an ICER of NOK [redacted] compared to [V + d].
- In the [P + d] group, [P + d] had costs of [redacted], and 0.81 QALYS. [EP + d] was the only other treatment that was not dominated by other treatments, with costs of NOK [redacted], 1.39 QALYs, and an ICER of NOK [redacted] compared to [P + d].

It is important to note the substantial uncertainty in the evidence underlying these results. We suggest putting more weight on comparisons from direct estimates from randomized trials, when possible. The results of the cost-effectiveness analysis should be considered highly uncertain because of lack of access to patient level data and the large degree of uncertainty in the network meta-analysis results which informed our work. Our results cannot be used to determine the best treatment sequencing among relapsed and/or refractory multiple myeloma patients.

External peer reviewers:

Øyvind Hjertner, *Chief physician,* Department of Hematology, St. Olav's University Hospital and *Associate Professor*, Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology
Lars Asphaug, *researcher,* Oslo University Hospital and Department of Health Management and Health Economics, University of Oslo
Rhiannon Owen, *Associate Professor,* Health Data Science, Swansea University, UK

Executive summary

Background

Multiple myeloma is the second most common type of blood cancer, with approximately 450 new cases diagnosed annually in Norway. The median age at diagnosis is approximately 70 years, and incidence is rare among individuals under age 30. Multiple myeloma affects plasma cells in the bone marrow. Because there is currently no cure for multiple myeloma, the goal of treatment is to achieve as strong a response as possible without unacceptable side effects, and to maintain the patient's quality of life at as high a level as possible throughout treatment.

Objective

To determine the clinical efficacy, safety, and cost-effectiveness of disease modifying treatments for relapsed and/or refractory multiple myeloma (RRMM) in a Norwegian context.

Efficacy and safety

Method

We have systematically collected and reviewed the evidence for clinical efficacy and safety for disease modifying treatments for relapsed and/or refractory multiple myeloma according to the PRISMA rules. We identified relevant publications from randomised, controlled trials (RCTs) through systematic reviews from our previous mapping review, as well as through systematic searches. The inclusion criteria were individuals over 18 years diagnosed with multiple myeloma, who either were refractory to at least one previous line of treatment or had experienced one or more relapses. The treatment (intervention) was any of the drugs listed below, alone or in combination with each other, and/or with a glucocorticosteroid such as dexamethasone, compared with any intervention-drug alone, or in combination with each other, or in combination with other drugs.[1]

Treatments named in the commission and listed in the project plan (protocol):

Abb.	Full drug name
D	Daratumumab
E	Elotuzumab
F	Panobinostat (Farydak)
I	Ixazomib
Is	Isatuximab
K	Carfilzomib
P	Pomalidomide
R	Lenalidomide (Revlimid)
V	Bortezomib (Velcade)

[1] *Treatment regimens can consist of up to three different drugs in combination, often with dexamethasone (d), and are presented in the text as abbreviations, e.g., the triple regimen of [DK + d] consists of daratumumab (D), carfilzomib (K) and dexamethsone (d). Other glucocorticoids than dexamethasone can also be used*

The primary outcomes were overall survival, health related quality of life, and severe adverse events, with overall survival being our main primary outcome. Secondary outcomes were progression free survival, adverse events, and discontinuation due to adverse events. We used the critical appraisal of one systematic review from which we included several studies; two researchers critically appraised the remaining included studies. All outcomes were analysed by component network meta-analyses. We present results for the treatment regimens relevant for Norway, as determined by the Norwegian guideline for multiple myeloma, with results for all included treatment regimens in the appendix. We assessed the certainty of evidence for all outcomes using the GRADE approach (Grading of Recommendations Assessment, Development and Evaluation), expressing the certainty as high, moderate, low, or very low, depending on the level of confidence we have in the effect estimates.

Results

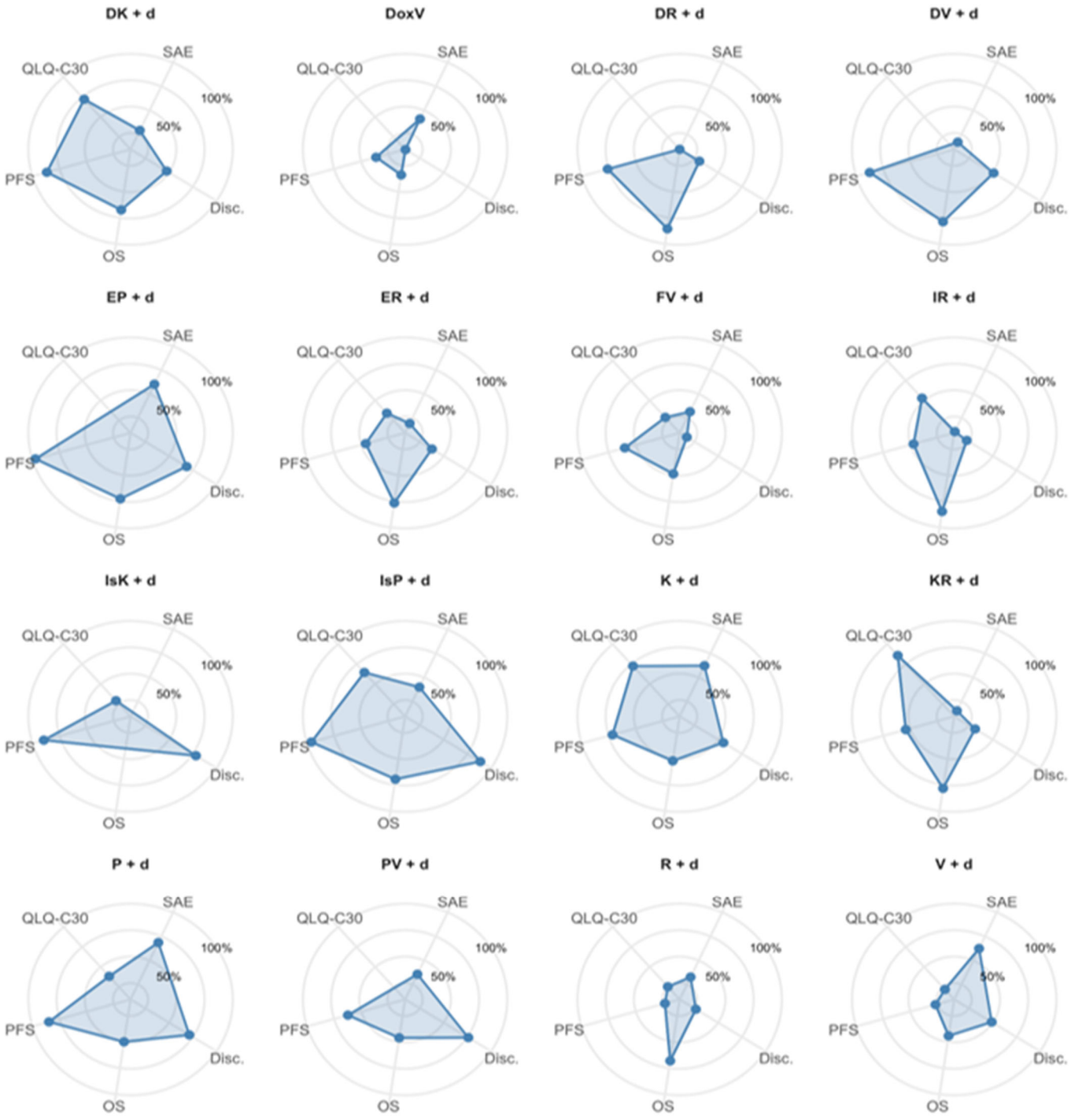
We included in total 72 articles from 50 RCTs that studied the effects of various treatment regimens containing one to three disease modifying drugs. We performed component network meta-analyses on up to 34 randomised, controlled trials, comprising 12 873 randomized patients, and 31 treatments.

The radar plots below illustrate the overall safety and efficacy of each treatment regimen across all outcomes. Each individual radar plot presents the available P-scores for the different outcomes, for each treatment regimen, as a polygon (shaded). A P-score expresses the mean extent of certainty that a given treatment regimen is superior to all other regimens included in the underlying meta-analysis (i.e., with respect to a single outcome such as overall survival). They are informally interpreted as the probability that a treatment is "best". A treatment with a higher P-score (closer to 100%) could be interpreted to be superior (i.e., longer survival, better quality of life, fewer severe adverse events, longer progression-free survival or fewer discontinuations due to adverse events) to a treatment with a lower P-score (closer to 0%). In the radar plots, treatment regimens with polygons with larger areas tend to be superior to those with smaller areas. However, this interpretation can be misleading because data was not available for all treatments and outcomes: it is therefore possible for a highly effective treatment to have a polygon with small area due to a lack of data. When comparing results for different treatment regimens, one should be careful not to interpret effect based solely on polygon area.

We inferred that there is not a single treatment regimen that is superior with respect to all outcomes. Radar plots for the double combination [P + d] exemplify treatment regimens that have polygon with large area. This would indicate better efficacy and safety than treatment regimens with smaller area polygons, e.g., [DR + d]. However, when looking closer at the individual P-scores, we find that [P + d] have lower P-score for overall survival than [DR + d]. As such, we would expect longer survival by treatment with [DR + d] than [P + d]. While radar plots may be useful for understanding tradeoffs between efficacy and safety, they should not be interpreted in isolation. Furthermore, the radar plots do not reflect assessments of the certainty of evidence or results of the health economic analysis. We assessed the certainty of evidence (GRADE) for one treatment being better than another, to be mainly low or very low, with a few exceptions.

The six triplet combinations [EP + d], [IsP + d], [DK + d], [KR + d], [DR + d] and [DV + d] are examples of treatment regimens relevant for non-refractory patients[2] that have clearly favorable hazard ratios for overall survival, that also are ranked highly with respect to other outcomes.

[2] We also present results for patients who are refractory for lenalidomide (R) and/or bortezomib (V) in the report.

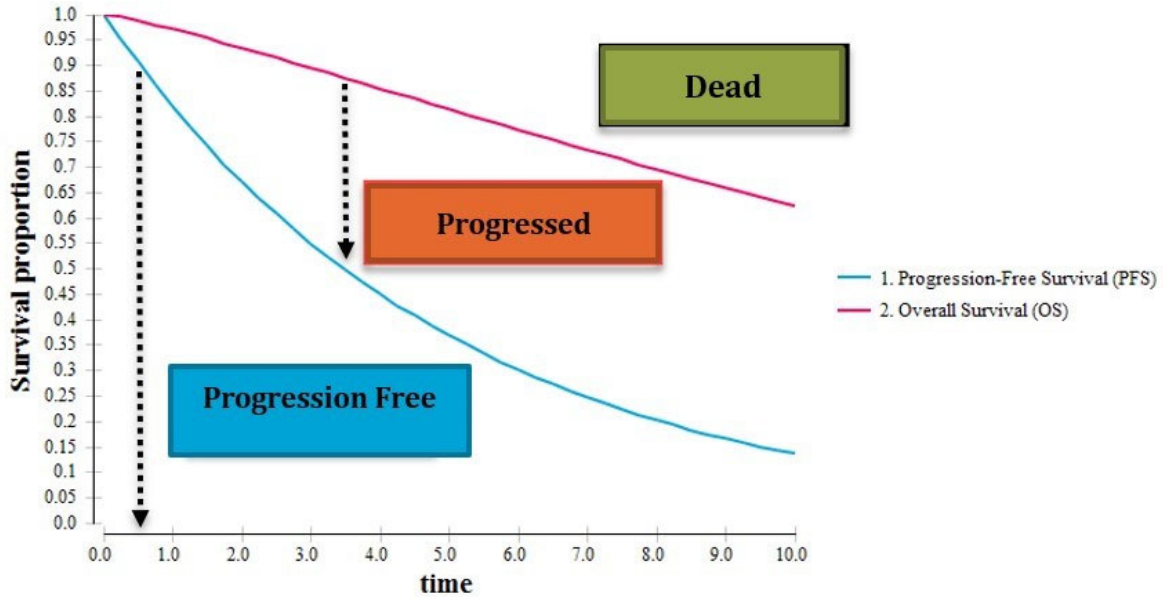


Radar plot of treatment regimens relevant for non-refractory patients
Radar plots show treatment regimens relevant to Norwegian clinical practice. D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), Dox: doxorubicin, E: elotuzumab, F: panobinostat, I: ixazomib, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), R: lenalidomide, SAE: severe adverse events (incidence rate ratio), V: bortezomib. The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.

Health economic evaluation

Method

We conducted a cost-utility analysis of 13 treatments for patients with relapsed and/or refractory multiple myeloma in which health effects were measured in quality-adjusted life-years (QALYs), costs in Norwegian kroner, and results were presented as incrementals cost-effectiveness ratios (ICERs). We chose to use a partitioned survival analysis model implemented in TreeAge to perform the analysis. Partitioned survival analyses are frequently used to model the cost-effectiveness of cancer treatments because Kaplan-Meier plots of overall and progression-free survival curves from clinical trials can be used to track patients through three health states: Progression-free, Progressed, and Dead.



Survival Curves and Health States in Partition Survival Analysis
OS – PFS = Progressed, PFS = Progression Free, OS = Alive (Progressed + Progression Free)

Without access to patient level trial data it is not possible to generate well-fitted survival curves for each treatment in the model. Instead, we used a technique, common in cost-effectiveness analyses, in which a survival curve is generated for the comparator treatment in the analysis and the corresponding curves for interventions of interest are generated by applying the relevant hazard ratios from a meta-analysis to the comparator's survival curve. As the hazard ratios for overall and progression-free survival were taken from the network meta-analysis in the clinical effect section of this report, there was not a "comparator" in the normal sense. In a network meta-analysis any treatment can be designated as a "reference treatment" since the matrix of results generated by the analysis provides hazard ratios for each intervention relative to all other interventions. Based on expert advice, we subdivided our economic model into three treatment groups, each based on one of three reference treatments: lenalidomide (Revlimid) + dexamethasone [R + d], bortezomib (Velcade) + dexamethasone [V + d], and pomalidomide + dexamethasone [P + d].

The [R + d] group included: [R + d], [DR + d], [RK + d], [ER + d], and [IR + d].
The [V + d] group included: [V + d], [DK + d], [K + d], [FV + d], and [DV + d].
The [P + d] group included: [P + d], [EP + d], and [IsP + d].

Costs for the analysis included: 1) cost of medications, 2) time costs for pharmacy and nursing staff for preparation and administration of medications given by injection or infusion, 3) time costs for doctor visits and tests at regular check-ups, and 4) patients' travel and time costs associated with treatment. We were unable to include costs of

severe adverse events, as they were not reported consistently in published trial results, but as these are quite small in relation to medication costs, they would not have resulted in meaningful changes in the results. To account for uncertainty associated with the variables included in the model (hazard ratios, utility values for capturing quality-of-life and treatment costs) we ran the model as a probabilistic sensitivity analysis with 10,000 random draw Monte Carlo iterations. The model also allowed us to calculate absolute shortfall, the variable used to determine the severity of a disease. We also conducted one-way sensitivity analysis to determine which variables had the largest impact on the results.

Results

We report the cost-effectiveness results for treatments in each reference group that were not dominated by another treatment. A treatment is considered dominated if it has a higher total cost and lower health effect than another treatment. In the [R + d] group, [R + d] had costs of NOK [redacted], with 2.90 QALYs gained. Only two other treatments were not dominated by other treatments: [IR + d] had costs of NOK [redacted], 3.82 QALYs, and an ICER of NOK [redacted] compared to [R + d]. [DR + d] had costs of NOK [redacted], 4.31 QALYs, and an ICER of NOK [redacted] compared to [IR + d]. In the [V + d] group, [V + d] had costs of NOK [redacted] and 2.24 QALYs. [DV + d] was the only other treatment that was not dominated by other treatments, with costs of NOK [redacted], 3.63 QALYs, and an ICER of NOK [redacted] compared to [V + d]. In the [P + d] group, [P + d] had costs of [redacted], and 0.81 QALYS. [EP + d] was the only other treatment that was not dominated by other treatments, with costs of NOK [redacted], 1.39 QALYs, and an ICER of NOK [redacted] compared to [P + d].

In addition to providing cost-effectiveness results, the health economic analysis provided estimates of average absolute shortfall for treatments in each of the reference groups. The values were similar across groups and ranged from 12.46 to 14.95 lost healthy life-years.

Discussion

Efficacy and safety

Our report is limited in having few RCTs for each treatment regimen. Because the included studies defined disconnected networks, we had to use component network meta-analysis, and were unable to formally test the assumption that the treatments can be modelled in this way. Including immature survival data in our analysis may have introduced bias in our report due to selective reporting of findings. It is likely that more mature data would lead to different meta-analysis results and more certain judgements about the evidence.

Most of the included RCTs are international studies, predominately conducted in North America and Europe. However, our overall survival results for both [IR + d] and [DV + d] are based on one small Chinese study and one large international study, where there were differences in the type of treatment offered to patients who progressed, making it difficult to interpret the findings. There were also ethnic and other differences in the study populations. There have been reported ethnic differences regarding incidence of multiple myeloma and aggressiveness of disease. In addition, ethnicity has also been shown to affect drug response in cancer treatment in some cancers. However, we do not know to what extent this does or does not apply to RRMM.

Ideally, future studies should directly compare the more effective triple regimens, rather than using less effective double regimens as controls. It may not be possible to power superiority trials comparing treatments of very similar effectiveness, but noninferiority trials may be a useful approach when relatively small effect sizes are expected.

Health economic evaluation

There is a great deal of uncertainty connected with the results of the health economic analysis. Some of this uncertainty reflects the fact that we needed to rely on the results of the component network meta-analysis performed for the clinical effect section of this report to perform the health economic analysis. The sources of uncertainty in the network meta-analysis results have been explained in the discussion of clinical effects. Additional uncertainty in the cost-effectiveness analysis can be the result of the methods we used to derive the underlying survival curves for the reference treatments. The lack of access to patient level data meant that we were unable to account correlation between progression-free and overall survival. This is particularly important because it limits our ability to accurately fit the portions of the survival curves beyond what is captured in the trial period. To the extent that we were unable to accurately account for dose reductions during treatment, our results may over-estimate treatment costs.

Conclusions

It is not possible to draw clear, brief conclusions for several reasons including a high degree of uncertainty across most results, the need to consider different outcomes simultaneously, and different considerations across subgroups of patients (e.g., those who are refractory to different drugs). We infer that there is no single treatment regimen that is superior with respect to all outcomes. The six triplet combinations [EP + d], [IsP + d], [DK + d], [KR + d], [DR + d] and [DV + d] are examples of treatment regimens relevant for non-refractory patients that have clearly favorable hazard ratios for overall survival, that also are ranked highly with respect to other outcomes. However, it is important to note the substantial uncertainty in the evidence underlying these results. We suggest putting more weight on comparisons from direct estimates from randomized trials, where they exist.

Cost-effectiveness results need to be viewed in the context of each reference group, as comparisons of treatments were not made across reference groups. In total, only seven treatments were not dominated by other treatments. [R + d], [IR + d] and [DR + d] in the [R + d]-group; [V + d] and [DV + d] in the [V +d]-group; and [P + d] and [EP + d] in the [P + d]-group. The ICERs for these treatments ranged from NOK [redacted] to NOK [redacted]. Absolute shortfall for patients with RRMM probably ranges from 12 to 15 healthy life-years lost. There is a high degree of uncertainty in the cost-effectiveness results.

Hovedbudskap

Myelomatose er den nest vanligste blodkrefttypen, med omtrent 450 nye tilfeller årlig i Norge. Det finnes ingen kur for myelomatose, og behandlingsmålet er derfor å oppnå så god effekt som mulig, uten uakseptable bivirkninger, samt at pasientens livskvalitet opprettholdes gjennom behandlingen.

Vi inkluderte data fra fem utfallsmål: totaloverlevelse, livskvalitet, alvorlige uønskede hendelser, progresjonsfri overlevelse, og avbrutt behandling som følge av uønskede hendelser, hvor totaloverlevelse er det primære utfallsmålet. Ettersom det ikke var tilgjengelige data for alle behandlingsregimer og utfallsmål, mener vi at det ikke er mulig å fastslå om ett behandlingsregime er overlegen de andre med hensyn på alle utfallsmål. De følgende seks trippelkombinasjonene er eksempler på behandlingsregimer med gunstige hasard ratioer for totaloverlevelse, og som også er høyt rangert med hensyn på andre utfallsmål:

- [EP + d]: elotuzumab (E), pomalidomid (P) og deksametason (d)
- [IsP + d]: isatuksimab (Is), pomalidomid (P), og deksametason (d)
- [DK + d]: daratumumab (D), karfilzomib (K), og deksametason (d)
- [KR + d]: karfilzomib (K), lenalidomid (R), og deksametason (d)
- [DR + d]: daratumumab (D), lenalidomid (R), og deksametason (d)
- [DV + d]: daratumumab (D), bortezomib (V), og deksametason (d)

Den helseøkonomiske analysen baserte seg på en "partitioned survival analysis" for å beregne totale kostnader (i NOK), helsegevinst (i QALYs), og kostnadseffektbrøk (incremental-cost-effectiveness-ratio, ICER) for de aktuelle behandlingsalternativene. Vi grupperte behandlingsregimene i tre separate grupper, hver basert på det relevante referanseregimet, for å kunne utføre analysen. Vi brukte hasard ratioer fra nettverksmetaanalysen for å utarbeide kurver for totaldødelighet og progresjonsfri overlevelse,

Tittel:
Behandlinger av relapserende og/eller refraktær myelomatose

Publikasjonstype:
Metodevurdering
En metodevurdering er resultatet av å innhente, kritisk vurdere, og sammenfatte relevante forskningsresultater ved hjelp av forhåndsdefinerte og eksplisitte metoder.

Svarer ikke på alt:
Vi har ikke sett på etiske, juridiske eller sosiale aspekter ved farmakologisk behandling av relapserende og/eller refraktær myelomatose

---

Hvem står bak denne publikasjonen?
Folkehelseinstituttet har gjennomført oppdraget etter forespørsel fra Bestillerforum for Nye Metoder

---

Når ble litteratursøket utført?
Søk etter studier ble sist oppdatert januar 2022

Intern fagfelle:
Kjetil Gundro Brurberg, *avdelingsdirektør*, Folkehelseinstituttet
Vida Hamadi, *helseøkonom*, Folkehelseinstituttet

Eksterne fagfeller:

som inngikk i modellen. Resultatene fra den helseøkonomiske analysen var:

- I [R + d] gruppen var [R + d] det behandlingsregimet som hadde lavest kostnader, NOK [redacted], med en helsegevinst på 2,90 QALYs . Bare to andre behandlingsregimer var ikke dominert av andre behandlingsregimer: [IR + d] hadde en kostnad på NOK [redacted], 3,82 QALYs, og ICER på NOK [redacted], sammenliknet med [R + d]. [DR + d] hadde en kostnad på NOK [redacted], 4,31 QALYs, og en ICER på NOK [redacted], sammenliknet med [IR + d].
- I [V + d] gruppen var både [V + d] og [DV + d] ikke dominert av andre behandlingsregimer: [DV + d] hadde en kostnad på NOK [redacted], 3,63 QALYs, og en ICER på NOK [redacted], sammenliknet med [V + d].
- I [P + d] gruppen var både [P + d] og [EP + d] ikke dominert av andre behandlingsregimer: [P + d] hadde en kostnad på NOK [redacted], og 0,81 QALYs. [EP + d] hadde en kostnad på NOK [redacted], 1,39 QALYs, og en ICER på NOK [redacted].

Det svært stor usikkerhet knyttet til disse resultatene, og vi har svært lav tillit til rangeringer av disse behandlingsregimene. Vi foreslår derfor å vektlegge sammenlikninger fra direkte estimater fra randomiserte studier, hvor slike finnes.

Resultatene av kost-nytteanalysen bør regnes som høyst usikre på grunn av mangel på tilgang til individdata samt usikkerheten som hefter ved resultatene fra nettverksmetaanalysen som inngår i de helseøkonomiske beregningene. Våre resultater kan ikke brukes til å fastslå den beste behandlingsrekkefølgen for pasienter med relapserende og/eller refraktær myelomatose.

Øyvind Hjertner, *overlege,* Avdeling for blodsykdommer, St. Olavs hospital og *førsteamanuensis,* Institutt for klinisk og molekylær medisin, NTNU
Lars Asphaug, *forsker,* Oslo universitetssykehus (OUS) og Avdeling for helseledelse og helseøkonomi, UiO
Rhiannon Owen, *førsteamanuensis*, Health Data Science, Swansea University, UK

Sammendrag

Innledning

Myelomatose er den nest vanligste typen blodkreft, med nesten 450 nye tilfeller i Norge hvert år. Median alder ved diagnose er ca. 70 år, og sykdommen er svært sjelden blant personer under 30 år. Myelomatose kjennetegnes av ukontrollert vekst av plasmacellene i benmargen. Det er foreløpig ingen kur for myelomatose, og behandlingsmålet er å oppnå en forbedret overlevelse, fortrinnsvis med høy livskvalitet, og uten sykdomsprogresjon eller uakseptable bivirkninger.

Hensikt

Å undersøke klinisk effekt, sikkerhet, samt kostnadseffektivitet av sykdomsmodifiserende legemidler til behandling av myelomatose med tilbakefall (relapserende myelomatose), og/eller som er behandlingsrefraktær for minst en tidligere behandling (RRMM), satt i en norsk kontekst.

Effekt og sikkerhet

Metode

Vi har systematisk samlet og oppsummert evidens for klinisk effekt og sikkerhet for sykdomsmodifiserende behandling av RRMM, i henhold til PRISMA-reglene. Relevante publikasjoner fra randomiserte kontrollerte studier ble identifisert gjennom systematiske oversikter fra vår tidligere kartleggingsrapport, samt gjennom systematiske søk. Inklusjonskriterinene var personer over 18 år diagnostisert med myelomatose som enten var behandlingsrefraktære eller som hadde erfart ett eller flere tilbakefall. Behandlingen (intervensjonen) var medikamentene som er listet opp nedenfor; alene eller i kombinasjon med hverandre og/eller med et glukokortikoid som for eksempel deksametason, sammenliknet med ett av intervensjonsmedikamentene alene, eller i kombinasjon med hverandre, eller i kombinasjon med andre medikamenter. [3]

Legemiddelbehandlinger som er nevnt i oppdraget og i prosjektplan:

Fork.	Virkestoff
D	Daratumumab
E	Elotuzumab
F	Panobinostat (Farydak)
I	Iksazomib
Is	Isatuksimab
K	Karfilzomib
P	Pomalidomid
R	Lenalidomid (Revlimid)
V	Bortezomib (Velcade)

[3] *Behandlingsregimer kan bestå av opptil tre ulike legemidler i kombinasjon, ofte sammen med deksametason (d), og presenteres i teksten som forkortelser, f.eks. trippelregimet [DK + d] består av daratumumab (D), karfilzomib (K) and deksametason (d). Andre glukokortikoider enn deksamentason kan også brukes.*

De primære utfallsmålene var totaloverlevelse, helserelatert livskvalitet og alvorlige uønskede hendelser der totaloverlevelse var det viktigste primære utfallsmålet. Sekundære utfallsmål var progresjonsfri overlevelse, uønskede hendelser og avbrutt behandling på grunn av uønsket hendelse. Vi brukte vurderingen av risiko for systematiske skjevheter (risk of bias) fra den systematiske oversikten for studiene som vi hadde inkludert fra denne, og gjennomførte vår egen vurdering for de resterende studiene. Alle utfallsmål ble analysert i en nettverksmetaanalyse. Vi har presentert data for legemiddelbehandlinger nevnt i bestillingen, i tillegg til legemiddelbehandlinger som har blitt brukt som komparator i de inkluderte studiene. Vi vurderte tilliten til resultatene ved hjelp av GRADE-tilnærmingen (Grading of Recommendations Assessment, Development and Evaluation) som uttrykkes som høy, middels, lav, og svært lav, avhengig av hvor stor tillit vi har til effektestimatene.

Resultat

Vi inkluderte 72 artikler fra 50 randomiserte, kontrollerte studier, som alle undersøkte effekt av ulike behandlingsregimer som inneholdt ett til tre sykdomsmodifiserende legemidler, og gjennomførte en komponent metaanalyse på opp til 34 randomiserte, kontrollerte studier, 12 873 randomiserte pasienter og 31 behandlinger.

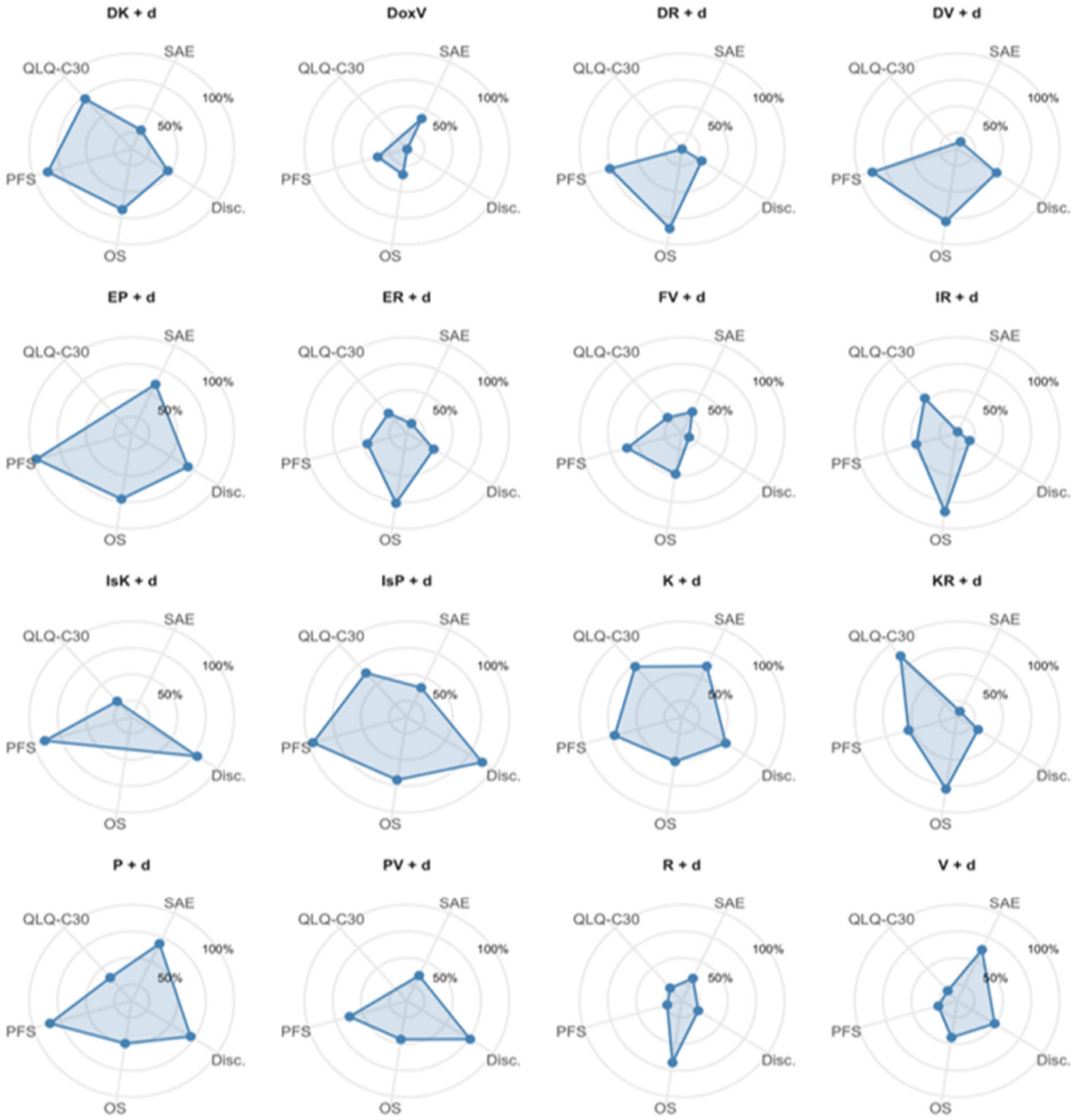
Radardiagrammene under illustrerer de samlede resultatene med hensyn på effekt og sikkerhet for hvert av behandlingsregimene på tvers av alle utfallsmål. Hvert enkelt radardiagram presenterer tilgjengelig P-score for de ulike utfallsmålene, for hvert behandlingsregime, som en polygon (farget). P-score uttrykker med gjennomsnittlig grad av sikkerhet at et behandlingsregime er overlegent alle andre inkluderte behandlingsregimer i den underliggende metaanalysen (det vil si med hensyn på ett enkelt utfallsmål, for eksempel totaloverlevelse). Uformelt tolkes dette som sannsynligheten for at et behandlingsregime er «best». Et behandlingsregime med en P-score nær 100% kan tolkes som å være overlegent (med hensyn på lenger overlevelse, bedre livskvalitet, færre alvorlige uønskede hendelser, lenger progresjonsfri overlevelse, eller færre avbrutte behandlinger som følge av uønskede hendelser) et behandlingsregime med lav P-score, nær 0%. I radardiagrammene vil behandlingsregimer med polygoner med større areal generelt sett være bedre enn behandlinger med polygoner med mindre areal. Denne tolkningen kan imidlertid være misvisende ettersom data ikke er tilgjengelige for alle behandlinger og utfallsmål. Det er derfor mulig for svært effektive behandlingsregimer å ha små areal i diagrammene som følge av manglende data. Ved sammenlikning av resultater for ulike behandlingsregimer bør man være varsom med å tolke effekt kun ut fra arealene i diagrammene.

Det er vanskelig å kunne slå fast om ett behandlingsregime overgår de andre når en tar hensyn til alle utfallsmål. Radardiagrammene for [P + d] er eksempler på legemiddelregimer som har store polygonareal i diagrammene. Dette burde indikere bedre effekt og sikkerhet enn behandlingsregimer med mindre areal, f.eks. [DR + d]. Når man ser nøyere på de individuelle P-scorene, finner man imidlertid at [P + d] har lavere P-score for totaloverlevelse enn [DR + d]. Basert på dette vil vi forvente lenger overlevelse ved behandling med [DR + d] enn ved [P + d]. Mens radardiagram kan være nyttige for å forstå balansen mellom effekt og sikkerhet, bør de ikke tolkes isolert, uten resultattabellene. Sammenlikning av de ulike behandlingsregimene vanskeliggjøres ytterligere som følge av at vår tiltro til resultatene varierer stort på tvers av behandlingsregimer og utfallsmål. Vi vurderte stort sett tiltroen til evidensen for at én behandling er bedre enn en annen, til å være lav eller svært lav, med enkelte unntak.

De seks legemiddelbehandlingene [EP + d], [IsP + d], [DK + d], [KR + d], [DR + d] og [ER + d] er eksempler på relevante behandlingsregimer for ikke-refraktære pasienter[4], som

[4] Vi presenterer også resultater for pasienter som er refraktære mot lenalidomid (R) og/eller bortezomib (V) i rapporten.

har gunstige hasard ratioer for totaloverlevelse, og som også er høyt rangert med hensyn på andre utfallsmål.

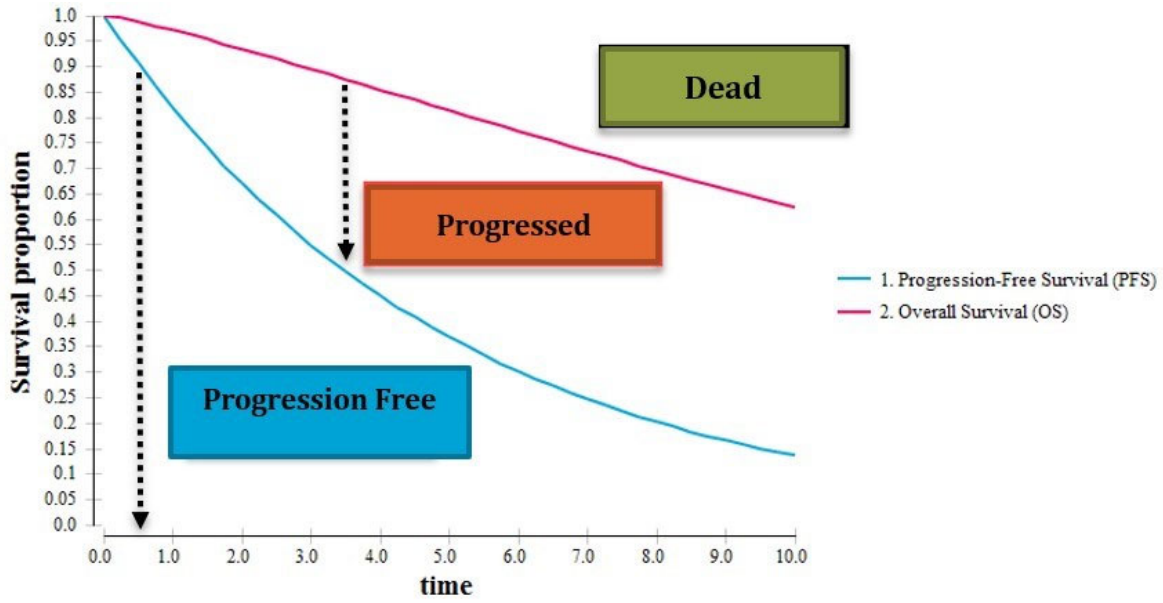


Radar diagram for behandlingsregimer relevante for ikke-refraktære pasienter
Radardiagrammene viser behandlingsregimer som er relevante for Norge. D: daratumumab, d: deksametason, Disc.: avsluttet behandling som følge av uønskede hendelser (risikoratio), Dox: doxorubicin, E: elotuzumab, F: panobinostat, I: ixazomib, Is: isatuximab, K: carfilzomib, OS: totaloverlevelse (hasardratio), P: pomalidomid, PFS: progresjonsfri overlevelse (hasardratio), QLQ-C30: livskvalitet (forskjell i gjennomsnittsscore), R: lenalidomide, SAE: alvorlige uønskede hendelser (insidensrateratio), V: bortezomib. Radardiagrammene oppsummerer effekt og sikkerhet, men reflekterer ikke vurdering av tiltro til evidensen eller resultater av den helseøkonomiske analysen.

Helseøkonomisk evaluering

Metode

Vi utførte en kost-nytteanalyse av 13 behandlingsregimer for pasienter med RRMM, hvor helsegevinst ble målt i form av kvalitetsjusterte leveår (QALYs), kostnader i norske kroner (NOK) og resultatene ble presentert som kostnadseffektbrøker (incremental-cost-effectiveness-ratios, ICERs). Vi valgte å benytte en «partitioned survival analysis»-modell i TreeAge for å utføre analysen. Paritioned survival anlaysis brukes gjerne for å modellere kost-nytteforholdet for kreftbehandlinger fordi Kaplan-Meierplot av kurver for total dødelighet og progresjonsfri overlevelse fra kliniske studier kan brukes for å følge pasienter gjennom tre helsetilstander: Progresjonsfri, progrediert og død.



Overvelseskurver og helsetilstander i partition survival analysis
OS – PFS = Progressed, PFS = Progression Free, OS = Alive (Progressed + Progression Free)

Uten tilgang til individdata er det ikke mulig å framstille godt tilpassede overlevelseskurver for hvert behandlingsregime i modellen. I stedet benyttet vi en vanlig teknikk for kost-nytteanalyser der vi framstiller en overlevelseskurve for komparator-regimet i analysen for deretter å framstille tilsvarende kurver for de aktuelle behandlingsregimene ved å anvende hasard ratioer fra en metaanalyse på komparatorens overlevelseskurve.

Ettersom hasard ratioer fra total- og progresjonsfri-overlevelse ble hentet fra nettverksmetaanalysen i den kliniske effektdelen av denne rapporten, hadde vi ingen «komparator» i konvensjonell forstand. I en nettverksmetaanalyse kan alle behandlingsregimene ses på som «referansebehandlinger», fordi matrisen av resultater fra analysene genererer en hasard ratio for hvert behandlingsregime i forhold til alle andre behandlingsregimer som inngår i analysen. Basert på råd fra klinisk ekspertise, delte vi den økonomiske modellen inn i tre undergrupper, hver basert på ett av tre referanseregimer: lenalidomid (Evlimid) + deksametason [R + d], bortezomib (Velcade) + deksametason [V + d], og pomalidomid + deksametason [P + d].

[R + d] gruppen omfattet: [R + d] , [DR + d], [RK + d], [ER + d] og [IR + d].
[V + d]-gruppen omfattet: [V + d], [DK + d], [K + d], [FV + d] og [DV + d].
[P + d]-gruppen omfattet: [P + d], [EP + d] og [IsP + d].

Kostnader som inngikk i analysen inkluderte: 1) legemiddelkostnader, 2) tidsbruk for apotek- og pleiepersonale for forberedelse and administrering av medikamenter ved

injeksjon eller infusjon, 3) tidsbruk for legebesøk og prøver tatt ved rutinekontroller, og 4) pasientenes tids- og reisekostnader knyttet til behandlingen. Vi var ikke i stand til å inkludere kostnader som følge av alvorlige uønskede hendelser ettersom disse ikke var konsekvent rapportert fra de kliniske studiene. Dette hadde neppe vesentlig innflytelse på resultatene fordi disse kostnadene er lave sammenliknet med legemiddelkostnadene.

For å ta høyde for usikkerhet assosiert med variablene som inngikk i modellen (hasard ratioer, nyttevekt for livskvalitet og behandlingskostnader) kjørte vi modellen som en probabilistisk sensitivitetsanalyse med 10 000 tilfeldige trekk ved Monte Carlo-repetisjoner. Modellen gjorde det også mulig å beregne absolutt prognosetap – variabelen som brukes til å fastsette en sykdoms alvorlighetsgrad. Vi gjennomførte også en enveis sensitivitetsanalyse for å avklare hvilke av variablene som hadde størst innflytelse på resultatene.

Resultater

Vi rapporterer kost-nytte-resultater for behandlingsregimer i hver referansegruppe, som ikke var dominert av et annet behandlingsregime. En behandling regnes som dominert hvis den har høyere totalkostnader og gir lavere helsegevinst en annen behandling.

I [R + d] gruppen var [R + d] det behandlingsregimet som hadde lavest kostnader, NOK [redacted], med en helsegevinst på 2,90 QALYs . Bare to andre behandlingsregimer var ikke dominert av andre behandlingsregimer: [IR + d] hadde en kostnad på NOK [redacted], 3,82 QALYs, og ICER på NOK [redacted], sammenliknet med [R + d]. [DR + d] hadde en kostnad på NOK [redacted], 4,31 QALYs, og en ICER på NOK [redacted], sammenliknet med [IR + d]. I [V + d] gruppen var både [V + d] og [DV + d] ikke dominert av andre behandlingsregimer: [DV + d] hadde en kostnad på NOK [redacted], 3,63 QALYs, og en ICER på NOK [redacted], sammenliknet med [V + d]. I [P + d] gruppen var både [P + d] og [EP + d] ikke dominert av andre behandlingsregimer: [P + d] hadde en kostnad på NOK [redacted], og 0,81 QALYs. [EP + d] hadde en kostnad på NOK [redacted], 1,39 QALYs, og en ICER på NOK [redacted]. I tillegg til kost-nytteresultatene gav den helseøkonomiske analysen estimater for gjennomsnittlig absolutt prognosetap for behandlingsregimene i hver referansegruppe. Disse var relativt like i hver referansegruppe: fra 12,46 til 14,95 tapte friske leveår.

Diskusjon

Effekt og sikkerhet

Rapporten har begrenset evidens for hvert av behandlingsregimene. Ettersom de inkluderte studiene definerte frakoblede nettverk, måtte vi bruke komponent nettverksmetaanalyse, og vi kunne derfor ikke formelt teste antakelsen om at behandlingsregimene kan modelleres på denne måten. Inklusjon av umodne data i analysene kan potensielt ha introdusert bias i rapporten som følge av selektiv rapportering. Det er sannsynlig at modnere data vil kunne føre til andre metaanalyse resultater, og sikrere antakelser om evidensen.

De fleste inkluderte RCTer er internasjonale studier, hovedsakelig er gjennomført i Nord-Amerika og Europa. Resultatene for totaloverlevelse ved behandling med behandlingsregimene [IR + d] og [DV + d] er imidlertid kun basert på én liten, kinesisk studie, og én stor internasjonal studie, hvor det var forskjeller i behandlingen som ble tilbudt pasienter med sykdomsprogresjon, noe som gjør det vanskelig å tolke resultatene. Det var også etniske og andre forskjeller i studiepopulasjonene. Det er rapportert etniske forskjeller i insidens av myelomatose og sykdomsaggressivitet. I tillegg har etnisitet vist seg å påvirke legemiddelrespons i kreftbehandling ved andre kreftformer. Vi vet imidlertid ikke om eller i hvilken grad dette også gjelder for RRMM. På bakgrunn av dette bør resultatet vårt for totaloverlevelse tolkes med forsiktighet.

Ideelt sett bør fremtidige studier direkte sammenlikne mer effektive trippelregimer fremfor å bruke mindre effektive dobbelregimer som kontroll. Det kan være vanskelig å gjennomføre overlegenhetsstudier med tilstrekkelig styrke dersom man skal sammenlikne behandlingsregimer med veldig lik effekt, men ikke-underlegenhetsstudier kan være nyttige.

Helseøkonomisk evaluering

Det er betydelig usikkerhet knyttet til resultatene fra den helseøkonomiske analysen. Noe av denne usikkerheten skyldes at vi måtte basere oss på resultatene fra nettverksmetaanalysen for beregning av de kliniske effektene av de forskjellige behandlingsregimene. Kildene til usikkerhet i nettverksmetaanalysen er beskrevet over. I tillegg er det usikkerhet knyttet til metodene vi benyttet for å beregne overlevelseskurvene for referansebehandlingene. Mangelen på tilgang til individdata medførte at vi ikke kunne ta høyde for korrelasjon mellom progresjonsfri og total overlevelse. Dette er viktig fordi det begrenser vår mulighet til å tilpasse de delene av overlevelseskurvene som går utover studieperiodene. I den grad vi ikke var i stand til å ta høyde for dosereduksjoner i løpet av behandlingen kan vi også ha overestimert behandlingskostnadene.

Konklusjon

Det er ikke mulig å trekke klare, korte konklusjoner, av flere årsaker, inkludert høy grad av usikkerhet på tvers av de fleste resultatene; behov for samlet vurdering av flere ulike utfallsmål; samt ulike hensyn på tvers av subgrupper av pasienter (f.eks. pasienter som er refraktære mot ulike legemidler). Ingen av behandlingsregimene utpeker seg som bedre enn de andre med hensyn på alle utfallsmål. De seks legemiddelbehandlingene [EP + d], [IsP + d], [DK + d], [KR + d], [DR + d] og [ER + d] er eksempler på relevante behandlingsregimer for ikke-refraktære pasienter, som ser ut til å ha gunstige hasard ratioer for totaloverlevelse, og som også er høyt rangert med hensyn på andre utfallsmål. Det er imidlertid viktig å merke seg den store usikkerheten i kunnskapsgrunnlaget for disse resultatene. Vi foreslår derfor å vektlegge sammenlikninger fra direkte estimater fra randomiserte studier, hvor slike finnes.

Kost-nytteresultatene må vurderes innenfor rammene av hver referansegruppene, ettersom sammenlikning av behandlingsregimer ikke ble gjort på tvers av referansegruppe. I alt var det kun sju behandlingsregimer som ikke var dominert av andre behandlingsregimer: [R + d], [IR + d] og [DR + d] i [R + d]-gruppen; [V + d] og [DV + d] i [V +d]-gruppen; og [P + d] og [EP + d] i [P + d]-gruppen. ICER for disse behandlingsregimene varierte mellom NOK [redacted] og NOK [redacted]. Absolutt prognosetap for pasienter med RRMM var mellom 12 og 15 tapte friske leveår. Det er høy grad av usikkerhet knyttet til resultatene av kost-nytteanalysen.

Preface

This Health Technology Assessment (HTA) was commissioned by the Regional Health Authorities (RHA) forum: The National System for Managed Introduction of New Health Technologies within the Specialist Health Service in Norway (Bestillerforum for Nye Metoder).

The Norwegian Institute of Public Health (NIPH) was commissioned by the RHA forum to perform a full health technology assessment on the pharmacological treatment of multiple myeloma. The commission was based on a proposal submitted by NIPH, the Norwegian Medicines Agency and Sykehusinnkjøp HF. Upon starting the work, it became evident that the commission was too vague, especially in terms of which multiple myeloma population should be included in the HTA (e.g., newly diagnosed patients eligible for stem cell transplantation, newly diagnosed patients not eligible for stem cell transplantation, or patients with relapsed and/or refractory multiple myeloma). Based on a recommendation by NIPH, the RHA forum gave NIPH a sub-commission to map the evidence of systematic reviews for the different multiple myeloma populations. The original commission was subsequently revised to a full HTA on the pharmacological treatment of patients with relapsed and/or refractory multiple myeloma, with a focus on drugs and drug combinations that are relevant in Norway. This HTA includes an assessment of clinical efficacy and safety, and a health-economic analysis of the pharmacological treatment of patients with relapsed and/or refractory multiple myeloma.

The internal working group consisted of:

- Ingrid Kristine Ohm, *researcher – effect and safety*
- Liv Giske, *senior researcher – effect and safety*
- Christopher James Rose, *statistician*
- Arna S. Desser, *health economist, project leader*
- Fawaz Tariq Chaudhry, *health economist*
- Gunn Eva Næss, *information specialist*
- Atle Fretheim, *research director*

In addition, the following have also contributed to the work:

External group members:

- Fredrik Schjesvold, Head of Oslo Myeloma Center, Oslo University Hospital.
- Mats Irgen Olsen, Department of Hematology, University Hospital of North Norway.
- Einar Haukås, Department of Blood and Cancer Diseases, Stavanger University Hospital.
- Olav Ljosne, patient representative, Blodkreftforeningen.

Reviewers – internal (at NIPH):

- Kjetil Gundro Brunberg, *unit director (reviewed the whole report)*
- Elisabet Vivianne Hafstad, *information specialist (reviewed the search strategy)*
- Vida Hamidi, *health economist (reviewed the health-economic analyses)*

Reviewers – external:

- Øyvind Hjertner, *Chief physician,* Department of Hematology, St. Olav's University Hospital and *Associate Professor*, Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology
- Lars Asphaug, *researcher,* Oslo University Hospital and Department of Health Management and Health Economics, University of Oslo
- Rhiannon Owen, *Associate Professor,* Health Data Science, Swansea University, UK

Acknowledgements
We would like to thank Anna Lien Espeland for her work on health economical analysis.

Conflicts of interest
All authors, external group members and reviewers have completed a conflicts of interest form, and no conflicts of interest have been reported.

We will emphasise that although the external experts and reviewers have contributed with valuable input and comments, NIPH is solely responsible for the content of this report.

Kåre Birger Hagen	Atle Fretheim	Arna S. Desser
Department director	*Research director*	*Project coordinator*

Introduction

Multiple Myeloma

Multiple myeloma, the second most common type of blood cancer, affects plasma cells in the bone marrow. Plasma cells are a type of white blood cell that produce immunoglobulins, which are complex proteins also known as antibodies. Myeloma cells (malignant plasma cells) produce abnormal M-proteins (monoclonal proteins) rather than normal, functioning antibodies. Clinical important manifestations of multiple myeloma include anemia, osteolytic bone lesions, kidney failure and hypercalcemia.Age and previous monoclonal gammopathy of undetermined significance (MGUS) are the most important risk factors for the disease (1). Approximately 450 new cases of myeloma are diagnosed annually in Norway. The median age at diagnosis is approximately 70 years, and incidence is rare among individuals under age 30 (2).

Multiple myeloma is often first suspected when patients experience skeletal pain, anemia, frequent respiratory or other infections, poor kidney function, or elevated calcium levels in the blood (hypercalcemia). Diagnosing multiple myeloma that will require treatment[5] involves a bone marrow or tumor biopsy to confirm the presence of malignant plasma cells and tests to confirm incidence of one or more CRAB criteria: C: hypercalcemia, R: kidney (renal) damage, A: anemia, and B: bone damage involving bone lesions or low bone density (3).

Because there is currently no cure for myeloma, the goals of initial and subsequent treatments are to prolong life, to achieve as strong a response as possible without unacceptable side effects, and to maintain the patient's quality of life at as high a level as possible throughout treatment. Norwegian guidelines recommend the most effective available treatment, given at the dose recommended in supporting clinical studies, based on a patient's age, overall health, and response to earlier rounds of treatment. Dosages can be adjusted downwards in response to side effects. Patients generally require continuous treatment over their remaing lifetime, with a possibility of short periods without treatment (3).

The initial choice of treatment for multiple myeloma is based on patient age. Individuals under age 70 are usually offered high-dose chemotherapy with an autologous stem cell transplant (ASCT, transplant with patient's own stem cells). The treatment consists of five phases: 1) induction, i.e., treatment with myeloma directed

[5] In addition to myeloma requiring treatment, there are two categories of myeloma ('smoldering multiple myeloma' and MGUS) in which no CRAB criteria are present, but where there is evidence of monoclonal proteins in bone marrow. MGUS (monoclonal gammopathy of undetermined significance) is a non-cancerous condition that is considered a precursor of myeloma. There is a 1% annual risk that MGUS will progress to multiple myeloma. Smoldering multiple myeloma (SMM) is diagnosed when levels of monoclonal protein in bone marrow exceed the MGUS cut-off level but are below the level at which multiple myeloma is diagnosed. SMM patients have an annual 10% risk of developing active MM within the first five years, and an annual risk that declines to 3% over the next five years, and to 1%-2% over the 10-year period after that. [2] In Norway, regular follow-up is recommended for MGUS and SMM patients.

drugs intended to achieve maximum response without negatively effecting stem cell harvesting, 2) stem cell harvesting, 3) ASCT, 4) consolidation, often a repeat of the induction treatment, and 5) maintenance. Treatment involving stem cell transplantation may also be appropriate for very healthy and motivated individuals over age 70 (3).

Patients over age 70 or individuals who cannot tolerate or do not wish to undergo the ASCT process are treated with chemotherapy and combinations of medications that inhibit plasma cell division in bone marrow. Patients under active treatment can often be treated at home with regular outpatient follow-up.

Pharmacological treatment of multiple myeloma has advanced greatly during the past decades, as several new disease modifying agents with various mechanisms of action have been developed and approved (4). The variety of these new agents have drastically increased myeloma patients' survival rate, but also pose a challenge given the complexity of the treatment strategies (4, 5). Therapy often relies on treatment regimens containing one or more disease modifying agents, often in combination with glucocorticoids such as dexamethasone (3). The following drug groups and specific disease modifying agents are used to treat relapsed, refractory multiple myeloma (RRMM):

- Proteasome inhibitors (PI)
    - Bortezomib, ixazomib, carfilzomib
    - Ending: -*zomib*
- Immunomodulatory imid drugs (IMiD)
    - Lenalidomide, pomalidomide
    - Ending: - *lidomide*
- Monoclonal antibodies (mAb)
    - Daratumumab, elotuzumab, isatuximab
    - Ending: -*mab*
- Histone deacetylace inhibitors
    - Panobinostat
    - Ending: -*inostat*

Both groups of newly diagnosed patients can experience disease remission, symptom relief, and increased survival. Patients experiencing side effects from treatments receive medications to relieve symptoms and pain. Radiation therapy can also be used either therapeutically or to control pain.

The large and growing number of potential treatment options, involving either a single drug or combinations of multiple drugs, has resulted in complex decisions about treatment paths when patients experience a relapse or become resistant to the current treatment (refractory disease). The objective of this project is to determine the clinical efficacy, safety, and cost-effectiveness of treatments for patients with relapsed, refractory multiple myeloma in a Norwegian context in order to facilitate decision-making within the context of priority setting criteria in the Norwegian healthcare system.

Priority setting in Norwegian health care

Principles for priority setting in the Norwegian health care sector have evolved over the past 30 years based on recommendations from a series of government

commissions.[6] Currently, there are three priority criteria for use in HTA at the group level[7]: the benefit criterion, the resource criterion, and the severity criterion (6).

- The benefit criterion: The greater the health benefit of an intervention, the greater its priority in the health sector. Health benefit is measured in healthy life years and expressed in terms of quality-adjusted life-years (QALYs).
- The resource criterion: The fewer resources an intervention requires, the greater the priority of the intervention.
- The severity criterion: The greater the severity of a disease/condition, the greater the priority of the intervention(s) used to treat it. Severity is quantified as "absolute shortfall", which is the number of health life years lost if a suggested treatment is not provided. In the context of a health technology assessment it is the difference between expected QALYs remaining for a healthy individual who is currently the same age as the mean age of patients undergoing treatment and the prognosis for expected QALYs at diagnosis of patients who don't receive treatment.

In addition to the three priority criteria, decision-makers considering health technology assessments have the discretion to consider other information, particularly the quality and uncertainty of evidence, and the potential budget consequences of an intervention (6).

Although this report is divided into separate sections—one examining clinical efficacy and safety results and the other providing the health economic analysis – it should be noted that these sections must be considered together in order to evaluate the treatments under consideration in terms of the three priority criteria. The clinical efficacy and safety section of this report provides the necessary information for establishing the clinical benefit of treatments in terms of gains in overall survival (OS) and progression-free survival (PFS), and safety considerations. The health economic evaluation section includes that information in the health economic model, along with the cost of resources used in treatments, to determine health gains measured in terms of quality-adjusted life-years, and severity, measured in terms of absolute shortfall.

Patient perspectives

Multiple myeloma is a disease that impacts heavily on the quality of life of patients who are diagnosed with it, both due to the illness itself, and its treatment. Coming to terms with a shortened life expectancy and the change in daily life is challenging, although there is much variation in how patients handle it. Most people have barely heard of the disease before, and the uncertainties around prognosis and treatment can be difficult for both the patients and the surroundings.

Information is key for myeloma patients, which means that much effort should be put into ensuring that they receive and understand the information that is presented by health care staff. The patient organization in Norway (Blodkreftforeningen) recommends that patients avoid attending hospital and treatment appointments alone, to increase the chance that all information is received and understood.

The patient representative in our expert group (Olav Ljøsne, Blodkreftforeningen) has pointed out that many multiple myeloma patients are concerned about the time it takes for new treatments to be available in Norway, i.e., that it takes too long time for a drug

[6] A report by the Norwegian Ministry of Health and Care Services (ref) provides an excellent summary of the development of priority setting in the Norwegian health care sector and the rationale behind current priority setting criteria.

[7] The same priority criteria for decisions at the clinical (individual patient) level, but the definitions for what constitutes benefits and severity are textual descriptions rather than quantitatively defined criteria that apply for group level decisions. (ref – same as footnote 6)

that has been developed to be made available for patient treatment. Patients are also concerned about the "*focus by authorities on cost and not on the individual benefit to the patients*", and that the variation in medical needs across individuals is not taken sufficiently into consideration. Furthermore, although myeloma patients "*are aware of challenges and uncertainty with new treatments*", the risk with trying new treatments is often seen as less of a problem than the disease itself.

Since the only known effective treatment for multiple myeloma is medication, access to effective treatment is of utmost importance for the patients, also with regards to quality of life. Our patient representative has emphasized the need for flexible approaches, e.g. allowing the use of "*new treatment at an earlier line than defined initially (often new medicine is approved for 4th or 5th line, while it may have good known/better effect at an earlier stage in the treatment process …)*".

Clinical efficacy and safety

METHODS

This health technology assessment (HTA) has been conducted in accordance with our project plan (*Appendix 2*) and the National Institute of Public Health's handbook, "Slik oppsummerer vi forskning" (7). We also used the PRISMA NMA checklist for reporting a systematic review involving a network meta-analysis, in our work outlining this HTA (*Appendix 4*).

Objective

To statistically characterize the relative efficacy and safety of the included treatments and to assess certainty of evidence.

Inclusion criteria

The inclusion criteria are presented in *Table 1* with OS being our main primary outcome.

Table 1: Inclusion criteria

PICOS	Inclusion
Population	Relapsed and/or refractory multiple myeloma (RRMM), i.e., individuals over 18 years, diagnosed with multiple myeloma who *either* are refractory to treatment, *or* have experienced one or more relapses
Intervention	Treatment with any of the following drugs, alone or in combination with each other and/or with glucocorticosteroids (e.g., dexamethasone, or prednisone): • Bortezomib (Velcade) • Carfilzomib (Kyprolis) • Daratumumab (Darzalex) • Elotuzumab (Empliciti) • Ixazomib (Ninlaro) • Isatuximab (Sarclisa) • Lenalidomide (Revlimid) • Panobinostat (Farydak) • Pomalidomide (Imnovid)
Comparison	• All intervention-drugs alone or in combination with each other, or in combination with other drugs • Placebo • Standard treatment • Glucocorticosteroids, e.g., dexamethasone, and prednisone
Outcome	Primary • Overall survival (OS) • Quality of life (QoL) • Severe adverse events (SAE) Secondary • Progression-free survival (PFS) • Adverse events (AE) • Discontinuation due to adverse events

Study design	• Systematic reviews based on randomised, controlled trials (RCTs) • RCTs

Literature search

From systematic reviews

We used the systematic reviews included in our published mapping review of March 2020 (8) as the basis to identify randomized, controlled trials (RCT) that were considered relevant for our HTA. In the mapping review, the search for systematic reviews were executed by a research specialist in February 2020 and was performed in Epistemonikos, using only the search term "myeloma".

Search for RCTs

In addition to RCTs identified through systematic reviews, we performed separate searches for eligible, published, and ongoing RCTs. An information specialist defined and processed the search terms in collaboration with the researchers of the team and put terms together in a search strategy (*Appendix 3*). The search strategies used a combination of controlled terms, i.e., Medical Subject Headings (MeSH), Emtree terms, as well as free-text terms with various synonyms that reflect the concepts of the population: "multiple myeloma", and the generic names of relevant pharmaceuticals as listed in our inclusion criteria (*Table 1*). The search was conducted in MEDLINE (Ovid) and Embase (Ovid) and included filters for study design (RCT) and publication year (published after 2017). The search was first conducted in February 2020 ("search 1") and then updated in September 2020 ("search 2"), March 2021 ("search 3"), and January 2022 ("search 4"). Before the third search, we contacted relevant firms to confirm that there were no additional publications that met our inclusion criteria. The search strategies are detailed in *Appendix 3*.

Searching for ongoing studies

In the database clinicaltrials.gov we limited the search strategy to relapsed or refractory Multiple Myeloma. We limited the recruitment status to "not yet recruiting," "recruiting," "enrolling by invitation," and "active, not recruiting." We limited the trial phase to phase two or three. In the WHO ICTRP Search Portal database, we also limited the search strategy to relapsed or refractory Multiple Myeloma. However, we selected ALL for recruiting status because of the limited options provided by the database, but then excluded the studies from clinicaltrials.gov (based on their NCT-number). There was also a trial limitation here for phase two or three. Finally, in the database EU Clinical Trials Register, there was also a limitation made in the search strategy for relapsed or refractory Multiple Myeloma. In the trial status, we selected only the ongoing studies. In this final database, we also limited the trial phase to phase two or three. The search for ongoing studies was updated in May 2022. In this update, we searched in the WHO ICTRP Search Portal database with equal limitations as above, the search period being from June 2021 until May 2022. The update also included a search in the Clinical Trials database. The limitation to relapsed or refractory Multiple Myeloma was maintained, but we did some minor changes in the limit selection on the left side menu in the database. We limited to the following filters under Recruitment: Not yet recruiting, recruiting, enrolling by invitation, active not recruiting, unknown status. We also limited the Study type to Interventional. The search strategies are detailed in *Appendix 4*.

Article selection

The studies included in this HTA were selected in a two-step process. In both steps, two persons worked independently, assessing articles against the inclusion criteria (*Table 1*). In the first step, two persons read all titles and abstracts retrieved either through the systematic reviews of our mapping review, or by the literature search. In the second step, all selected references were read in full text by the same two persons to decide which should be included in the HTA. Any disagreements throughout this work were resolved either through discussion or by consulting a third researcher.

References from systematic reviews

One researcher extracted all publications listed in the 13 systematic reviews (9-21) included in our mapping review (8). These references were then imported to Covidence (22) for screening of title and abstract (step one) and full text (step two).

References from search

References from our searches were imported to Rayyan (23) for screening of title and abstract (step one). The selected references were then imported to Covidence (22) for full text screening (step two).

Risk of bias assessment in included studies

References from systematic reviews

One researcher went through all the systematic reviews from which we included RCTs with regards to risk of bias assessment. We found that one of the systematic reviews: Maiese et al (13) had assessed risk of bias using the Cochrane risk of bias tool (24) for several of our included studies. As such, we have used the authors' risk of bias assessment for these studies.

References from search

For those references where risk of bias was *not* assessed in Maiese et al (13), we performed our own risk of bias assessment. Two researchers independently assessed the studies using the Cochrane risk of bias tool (24). Each study was rated as being at low, unclear, or high risk of bias on seven domains: selection bias (random sequence generation and allocation bias), performance bias, detection bias, attrition bias, reporting bias, and other bias. Any disagreements were resolved through discussion or by consulting a third researcher.

Data extraction

One researcher extracted relevant data from full-text articles to Covidence (22). The extracted data were then verified by a second researcher. Disagreements were resolved through discussion, or by consultation with a third researcher or other members of the project team. For two studies providing quality of life (Qol) data (25, 26), we extracted data points from figures showing changes over time in the reported QoL index domains using WebPlotDigitizer (27). Data exported from Covidence for meta-analysis were also checked by the statistician. We planned that a second researcher would check any data that needed to be converted by the statistician; in practice, rather than checking individual data, two researchers reviewed statistical analyses for obvious omissions and problems.

Table 2 shows which data were extracted (where possible) for each included study:

Table 2: Data extracted from included studies

About	Information extracted
The study	Authors, publication year, study design, country, clinical trial identification number, eligibility criteria, follow-up time, funding source (industry or non-industry).
The participants	For each trial arm and each outcome: numbers of participants randomized; numbers of participants included in analyses; average age; percentage of participants who were female; percentage of participants who were Caucasian; diagnosis; disease severity at baseline; percentage of participants who had received previous treatment (including stem cell treatment); average number of relapses.
The treatments*	For each trial arm: name of treatment (including combinations); posology (incl. dose level, frequency, duration, and route of administration).
The outcomes	For each pairwise comparison and each outcome: name of relative treatment effect estimate (e.g., HR, RR, OR); point estimate; name of measure of precision (e.g., 95% CI, SE, SD); precision (e.g., limits of the 95% CI).
The analyses	For each pairwise comparison and each outcome: analysis method (e.g., Cox regression, GLM); for cross-over and cluster studies, whether a unit of analysis error was made.

** Interventions and comparators.*
CI: confidence interval, HR: hazard ratio, OR: odds ratio, RR: risk ratio, SD: standard deviation, SE: standard error

For studies reporting on the same participants (e.g., extension studies), we extracted data with the longest follow-up, to avoid "double counting." In the case of cross-over studies with unit of analysis errors or failure to account for carry-over effects, we extracted data for the first period only.

To preserve the randomization used in the included studies and to model the policy of recommending a particular treatment (rather than adhering to it), we extracted and analyzed data following the intention-to-treat (ITT) principle, i.e., all randomized patients were counted and analyzed in the trial arms to which they were allocated. We extracted results for "modified intention-to-treat analyses" if we judged that the method used would provide similar estimates to an ITT analysis. As planned, we excluded "per protocol" results.

Where relative treatment effect estimates were not provided by the included studies or could not be extracted, we imputed relative treatment effects as described in the next section.

While we prespecified the scales of measurement on which we anticipated we would perform each meta-analysis, we planned to use alternatives if necessary. This was the case for the grade 3 and 4 severe adverse events outcome (SAE). We anticipated measuring relative treatment effect for this outcome using risk ratios (RRs) of one or more SAE. However, we chose to use incidence rate ratios (IRRs) instead. We chose this approach because there was more complete data available on numbers of SAEs (rather than numbers of patients experiencing SAEs, as required for RR), and because duration of follow-up varied sufficiently among the included studies that we doubted risk ratios would have similar interpretations across trials. The Analyses section describes how we imputed IRRs for SAE and the possible limitations of the approach. We treated data for any adverse events (AEs) in the same way, though chose not to present a meta-analysis for this outcome, as explained in the following section. Relative treatment effects were extracted or imputed from data presented in primary studies; we did not use data from systematic reviews or published meta-analyses in our meta-analyses, to avoid "double counting" patients.

Analyses

Treatment definition

To perform meta-analysis, it is necessary to define the treatments (i.e., the interventions and comparators). This is especially important in NMAs because these definitions determine the network topology (i.e., the treatments that have been directly compared by the included studies). We planned to define a treatment as a unique combination of a nonproprietary active drug name (e.g., bortezomib) and its posology (dose level, frequency, and route of administration), or, for combinations of active drugs, a unique combination of such.

Because we included relatively few studies, it was not possible to define the treatments as planned, as this would have resulted in a network with few if any connections. Instead, we followed an approach similar to those used in several other meta-analyses (12, 18) and defined a treatment as a unique combination of cytostatic drugs, i.e., we did not distinguish between different dosages, routes of administration etc. Further, to form connected networks and facilitate multiple treatment comparison, we chose to define treatments as combinations of components, as described in the next paragraph. This facilitates NMA of disconnected networks, as described in the meta-analysis section below.

We defined a component to be either a unique combination of specific cytostatic drugs (similar to our original plan) or a glucocorticosteroid (i.e., dexamethasone or methylprednisolone). This leads to a distinction between the treatments bortezomib (used alone) and bortezomib plus a glucocorticosteroid (used together), for example. As a treatment component can comprise multiple drugs, we use abbreviated names as is common in the literature. *Table 3* specifies all treatments (i.e., component combinations) identified after data extraction, along with the abbreviations we use. Note that some treatments or components may not be included in a specific NMA, for example, due to insufficient or incomplete data.

Table 3: Definition of treatment regimens and abbreviations

Treatment regimens	Abbreviation†
Observation	Observation
Bevacizumab and bortezomib	BevV
Bortezomib, cyclophosphamide and dexamethasone	CyV + d
Bortezomib and dexamethasone	V + d
Bortezomib	V
Bortezomib and pegylated liposomal doxorubicin	DoxV
Bortezomib, thalidomide and dexamethasone	VT + d
Bortezomib and vorinostat	VorV
Carfilzomib and dexamethasone	K + d
Carfilzomib, lenalidomide and dexamethasone	KR + d
Carfilzomib	K
Dexamethasone	d
Daratumumab and methylprednisone	D + d
Daratumumab, bortezomib and dexamethasone	DV + d
Daratumumab, carfilzomib and dexamethasone	DK + d
Daratumumab, lenalidomide and dexamethasone	DR + d
Daratumumab, pomalidomide and dexamethasone	DP + d
Daratumumab	D
Elotuzumab, bortezomib and dexamethasone	EV + d
Elotuzumab, lenalidomide and dexamethasone	ER + d
Elotuzumab, pomalidomide and dexamethasone	EP + d
Isatuximab	Is
Isatuximab and dexamethasone	Is + d
Isatuximab, carfilzomib, and dexamethasone	IsK + d
Isatuximab, pomalidomide and dexamethasone	IsP + d
Ixazomib, lenalidomide and dexamethasone	IR + d
Lenalidomide, cyclophosphamide and dexamethasone	CyR + d
Lenalidomide and dexamethasone	R + d
Panobinostat, bortezomib and dexamethasone	FV + d
Pembrolizumab, pomalidomide and dexamethasone	PemP + d
Pomalidomide, bortezomib and dexamethasone	PV + d
Pomalidomide, cyclophosphamide and dexamethasone	CyP + d
Pomalidomide and dexamethasone	P + d
Pomalidomide	P
Siltuximab and bortezomib	SV
Selinexor, bortezomib and dexamethasone	SeV + d
Tabalumab, bortezomib and dexamethasone	TabV + d
Thalidomide and dexamethasone	T + d
Venetoclax, bortezomib and dexamethasone	VenV + d

That we do not distinguish treatments by posology means that we were unable to include in the meta-analysis studies that *only* compared posologies of the same drug or drug combination. For example, a two-arm study that compared low versus high doses of the same drug could not be included in meta-analysis because, from the perspective of our treatment definition, it is a comparison of two identical treatments. While our meta-analyses are therefore unable to provide synthesized evidence about dose and other aspects of posology, it is possible to narratively summarize such differences.

Imputation

Point estimates of treatment effect and statements of precision (e.g., standard errors or confidence intervals) are necessary for meta-analysis but were not published or could not always be extracted for all outcomes. We therefore imputed these where necessary, as described below. Estimates of hazard ratios (HR) for OS and PFS were usually available, as these were typically prespecified endpoints of the included studies. We did not plan to impute HRs from statements of median survival and did not do so.

If standard error was not available, we imputed it where possible from an extracted confidence interval (after transforming to an appropriate scale, such as log HR) or from a *p*-value using standard Cochrane methods (28). If a study included arms that compared different posologies of the same drug or a drug combination with distinct drugs or drug combinations, we followed the recommended approach from the Cochrane handbook and combined arms to create the necessary comparisons (29). For the QoL outcome, we meta-analyzed difference in mean EORTC QLQ-C30 score (a composite of several QoL indicators). Reporting on this outcome was relatively poor compared to the "hard" outcomes such as OS. Usable point estimates were available for arm-wise means, but statements of precision (e.g., standard errors or confidence intervals) were often lacking or unsuitable for use in meta-analysis. Where practicable, we imputed standard error on difference in mean score by assuming a common standard deviation for all included studies. This assumed value was taken from Stewart 2015 (30). To evaluate whether this approach is reasonable, we used the imputed standard errors to compute statements of precision corresponding to those published in some of the included studies on QoL. For example, we used an imputed standard error to compute the *p*-value that the study would have reported under the assumption that our standard error was the same as in the study, and then compared that value to the published *p*-value. Where it was possible to make these comparisons, there was reasonable agreement. However, our approach can only provide approximate standard errors on difference in mean EORTC QLQ-C30 score, so our meta-analytical estimates for this outcome should be interpreted cautiously with this limitation in mind.

As described in the data extraction section, we meta-analyzed SAEs and AEs using IRR. IRRs were not reported by the included studies, so it was necessary to impute using the available information. An incidence rate is expressed in units of number of events per unit time per patient (e.g., SAEs per 1000 patient-years). An IRR is simply the ratio of two such rates (and is therefore unitless). IRR can be estimated from numbers of events (which were published and extracted) and total exposure times (which were not published). It is not appropriate to impute total exposure as the product of sample size and trial duration because some patients die or are otherwise lost to follow-up during a trial: this approach could result in IRR being heavily biased in favor of an ineffective treatment on which few events are observed because patients are not alive to experience them. We imputed expected total exposure (and hence point estimates of IRR) as follows. Let e_1 and e_2 denote the number of events in arms 1 and 2, n_1 and n_2 denote the number of patients randomized to arms 1 and 2, and $\mathrm{E}[T_1]$ and $\mathrm{E}[T_2]$ denote expected OS in arms 1 and 2. The total exposure times under the two treatments can be approximated by $n_1\,\mathrm{E}[T_1]$ and $n_2\,\mathrm{E}[T_2]$. A point estimate on IRR can then be imputed using:

$$IRR \approx \frac{e_1/n_1E[T_1]}{e_2/n_2E[T_2]} = \frac{e_1n_2E[T_2]}{e_2n_1E[T_1]}$$

The expected OS are unknown, but the ratio $\mathrm{E}[T_2]/\mathrm{E}[T_1]$ can be imputed using a HR for OS under the assumption that survival times follow distributions that are approximately exponential. The distributions are unlikely to be exactly exponential, so the approach can be expected to introduce some bias to estimates of IRR. However, we judged the imputation necessary because, given that SAE is a patient-important

outcome, we considered it preferable to perform a limited evidence synthesis than no evidence synthesis. Standard error on IRR (used to obtain confidence intervals in the meta-analysis) is a function of event counts e_1 and e_2. That is to say it is independent of the sample sizes and the imputed ratio of mean survival times and is not subject to bias due to the assumptions made above. Because our estimates of HR for OS are subject to uncertainty (e.g., due to the sampling error in the original studies), we propagated this error to the standard error on IRR to account for the additional uncertainty (i.e., to ensure that confidence intervals on IRR are not excessively precise). This likely also has a mitigating effect on the assumption that survival times are approximately exponentially distributed.

We used the same approach to impute IRRs for AEs. As this outcome is less relevant for the health economic analysis, and AEs are, by definition, less important than SAEs, we chose not to report this data to limit the degree to which the imputation described above may lead to misinterpretation.

Finally, we imputed risk ratios for discontinuation due to adverse events following standard Cochrane procedures (28).

Missing data and unit of analysis errors

We planned not to impute missing outcome data (i.e., for patients lost to follow-up) and did not do so. We planned to account for possible bias that missing outcomes may have introduced into the data extracted from the included studies in our risk of bias and GRADE assessment.

With respect to the outcomes whose relative treatment effects had to be imputed from extracted data (i.e., QoL, SAE, AE, and discontinuation), we did not attempt to account for possible unit of analysis errors to adjust for randomization above the level of patient), which would require additional assumptions about intraclass correlations.

Statistical analyses

Except as explained, we performed all statistical analyses as planned in our protocol (*Appendix 2*) (31).

Studies included in the statistical analyses

Some studies with more than two arms could not be included in meta-analysis because they compared different posologies of the same treatment (i.e., were comparisons of identical treatments under our treatment definition). The remaining studies with more than two arms were reduced to single comparisons using the method described in the Imputation section.

Assessment of possible publication bias

We planned to assess the possibility of publication bias for each primary outcome using funnel plots, however this was not possible due to the small number of included studies and lack of support for "comparison-adjusted" funnel plots under the meta-analysis model and software used (detailed below). We are therefore unable to comment on the degree to which studies may have been published or withheld because they showed favorable or unfavorable results.

Network meta-analysis

We planned to use contrast- or arm-wise network meta-analysis (NMA), or component NMA if we were unable to define treatments as planned. The latter was indeed the case. Further, the networks of evidence for all outcomes were disconnected and the use of component NMA allowed us to address this issue, as explained below in the next section.

NMA is a nontrivial area of statistics, but widely used within health technology assessment and the multiple myeloma literature. The following discussion attempts to provide a useful non-statistical summary of the methods we used. However, it is not

possible to provide a complete and unambiguous introduction without using mathematical notation and assuming a reasonably high level of statistical competence. Readers are referred to the statistical literature.

A set of included studies that each compare two or more of a set of multiple interventions (e.g., treatments or comparators named A, B, C, etc.) can be considered as a network. For example, if Study 1 compared treatments A and B, and Study 2 compared treatments B and C, these studies form the simple network A↔B↔C. In general, networks of evidence are more complex. Further, if relative treatment effect estimates are available from Studies 1 and 2, then the relative treatment effect of A versus C — which no study has estimated via direct comparison — can be estimated via a NMA model under appropriate assumptions (discussed further below).

Importantly, in addition to providing point estimates for all treatment comparisons, NMA models account for uncertainty arising from the included studies and the network topology (e.g., they provide confidence intervals for all treatment comparisons). In other words, if two treatments that have not been directly compared are "far apart" in the network, and multiple estimates must be combined to estimate a treatment effect for them, the confidence interval provided by the NMA will reflect the totality of the uncertainty. All else being equal, confidence intervals will necessarily be wider for treatments that are "far apart" than for treatments that are "closer".

The key assumption under a *fixed effects* NMA model is that relative treatment effects can be combined linearly (i.e., added or subtracted) on an appropriate scale (e.g., log HR, on which the multiplicative effect of HRs becomes additive on the log HR scale). The equivalent assumption under a *random effects* model (as used in this report, see below) is that *average* relative treatment effects can be combined linearly. This distinction is important — it should not be assumed that we are simply combining the estimates from studies without accounting for heterogeneity (differences in treatment effects due to differences in populations, study methods, etc.).

Dealing with disconnected networks of evidence

The approach described above assumes that the network is fully connected, as in the A↔B↔C example. However, if this is not true, as in the disconnected network A↔B C↔D, it is not possible to estimate the relative treatment effect for the A versus C comparison. This is because at least three treatment effect estimates are needed to estimate the six possible pairwise comparisons, but only two estimates are available. There are therefore infinitely many solutions to the underlying mathematical problem.

The problem of disconnected networks can be addressed using component NMA (32) providing:

1. the treatments can be considered to be combinations of components; and
2. the disconnected networks have components in common; and
3. the effect of a combination of components can be modelled as linear.

As explained in the Treatment definition section, we modelled treatments as combinations of "active" drug and glucocorticoid components. For example, the treatment daratumumab (D) and bortezomib (V) and dexamethasone (d) is denoted [DV + d]. This treatment is modelled using two components: a daratumumab and bortezomib component (DV), and dexamethasone (a glucocorticoid component) (d). The use of the plus (+) symbol in treatment names is used to indicate which parts of the treatment are modelled as distinct components. The notation also hints that the "additional effect" of each component is modelled linearly on an appropriate scale (log HR, say).

The assumption that the effects of components can be modelled linearly must be made conceptually and ideally tested when possible. Such an assumption should be

considered carefully in the face of disconnected networks, which may be explained by important differences between populations. Here, the word "population" is used in its statistical rather than lay sense.

A disconnected network could arise because the definition of the population used to identify the studies included in meta-analysis is incorrect in the sense that it would not be possible — *in principle* — to design a study that could make comparisons across the disconnected parts of the network. (Many studies can be designed *in principle* but cannot be run *in practice* — e.g., because they would be too expensive, or too few patients could be recruited to achieve the necessary power.)

We planned to use component NMA if the network of evidence was disconnected. Upon finding this was the case we reevaluated our inclusion criteria and the included studies to judge whether we had incorrectly defined the population (i.e., to determine whether it would be more appropriate to perform meta-analyses of each fully connected network). We based our judgement on a reconsideration of the original PICO, input from our clinical expert, the Norwegian guidelines for choosing among the available treatments for myeloma, and other systematic reviews and NMAs on myeloma. Based on this reevaluation we considered all included studies as part of a single disconnected network (for each outcome).

Unfortunately, it was not possible to formally test the assumption of additivity that underpins the component approach due to the network topologies. However, before adopting the component approach we compared estimates from regular (non-component) NMAs for the subnetworks to those from component NMAs for the entire (disconnected) network. The estimates were essentially identical. This provides some assurance that the component model is appropriate. However, it is not possible to test the assumption between the subnetworks. Results of the meta-analyses should be interpreted with this limitation in mind.

Modelling heterogeneity

As planned, we performed a component NMA for disconnected networks using the model proposed by Rücker et al. (32) (this paper provides full mathematical details on the model we used), and as implemented in version 1.3 of the netmeta add-on package for R (33). We assumed random effects for all outcomes, i.e., we estimated *average* relative treatment effects (rather than common effects, as in a fixed effects analysis) and the extent to which these vary across studies (heterogeneity). This is of relevance to the treatment components model and the choice to use a single glucocorticoid component (rather than to model dexamethasone and methylprednisolone as distinct components): our model estimates an average effect of a glucocorticoid component, which can vary by study, hence accounting for possible differences in effect of the two glucocorticoids and their possibly differing effects when used in combination with "active" drugs. Where possible, we summarized heterogeneity using the I^2 statistic and 95% confidence intervals on that statistic, as reported by the netymeta package.

Assessment of transitivity

NMAs assume that treatment effects are transitive. In plain language, this means that treatment effects can be added and subtracted to calculate other treatment effects, providing this is done on an appropriate scale (e.g., log HR rather than HR). It is the transitivity assumption that allows indirect treatment effects (i.e., those not studied by any trial) to be estimated. Note that there is an important difference between *treatment effects* (i.e., estimation targets, which are unknown and can only be estimated) and *treatment effect estimates* (i.e., trial results). The transitivity assumption must therefore be assessed conceptually. It is also necessary to consider differences in the distributions of *effect modifiers* across the included studies.

An earlier version of this report was criticized by our clinical expert regarding differences in *patient characteristics* between the networks, particularly with respect to refractory status and the number of lines of treatments they have received. Our

understanding of this argument is that an NMA is threatened if the included studies differ in patient characteristics. Papers on NMA aimed at non-statistical readers tend to frame discussion about threats to the assumptions that underpin NMAs in terms of differences in patient characteristics, so it is not surprising that we received this criticism. However, for NMAs of RCTs, the key issue is not whether the included studies have different distributions of *patient characteristics*, but whether they have different distributions of *effect modifiers*.

Briefly, a patient characteristic such as having received many previous lines of treatment may be a *risk factor* for OS (if for no other reason than such patients are likely to be older), but on average risk factors cancel out across trial arms in an RCT. Risk factors are therefore not an issue for an NMA of well-conducted RCTs. An *effect modifier*, however, is a variable that is associated with treatment effect. Effect modification is about stratification — i.e., that treatment effects (comparisons between treatments) are different in distinct patient subgroups. Effect modification can be a problem for NMA because, if the included studies have different distributions of effect modifiers, each trial result will have a different interpretation. This issue can be addressed by using random effects NMA, which attempts to account for heterogeneity. However, to address our clinical advisor's concerns, we performed a systematic review, meta-analysis, and simulation study, looking at the evidence for effect modification of HR for OS and PFS with respect to the purported effect modifiers refractory status and number of lines of treatment. The result of this work is available as a preprint (34). Briefly, we found very weak evidence in support of the effect modification hypothesis. Even under a scenario that strongly favored the effect modification hypothesis, simulations suggest that no more than about 5% of random effects NMA estimates would differ under effect modification versus no effect modification. With some caveats (see the preprint), we conclude that random-effects NMA can probably be used safely as in this report.

For the purpose of the analysis, we consider all patients as belonging to the same population. We believe this is sensible since we are not aware of convincing evidence that treatment effect differs with respect to variables such as whether patients are refractory to other treatments, or have received more lines of treatment, or with respect to other variables. Other recent systematic reviews on the same topic have used a similar approach (12, 13, 17, 35). We also note that it is common practice in randomized trials of myeloma treatments to include patients who are refractory to different types of medication, and in many cases, not to report results stratified by the purported effect modifiers, as should be done if it is found that treatment effect differs by patient subgroup.

In addition to the above, we also performed prespecified analyses for each outcome (except QoL, as explained below), as follows, to assess the appropriateness of the transitivity assumption that underpins NMA.
We inspected plots of the distributions of study and baseline characteristics (potential treatment effect modifiers) among the included studies. Where possible, we then performed statistical testing of null hypotheses that each characteristic is not associated with treatment comparison. Such an associations would threaten the NMA if the variable is an effect modifier (or, more specifically, if the magnitude of effect modification is sufficiently large). Where possible, we studied the following study and baseline patient characteristics:

1. Study setting (e.g., North America, Europe, multinational)
2. Funding source (industry or non-industry)
3. Average patient age
4. Percentage of patients who were female
5. Percentage of patients who were Caucasian
6. Average number of previous lines of therapy
7. Average time since diagnosis

8. Percentage of patients refactory to immunomodulatory drugs. (There was insufficient data to perform these analyses at the level of specific drugs, or to perform similar analyses for proteasome inhibitors.)
9. Average performance status
10. Average disease stage

Due to insufficient data, we could not perform the planned analyses for the quality-of-life outcome, nor could we analyze data on diagnosis at baseline or average number of relapses for any of the outcomes.

While we studied many possible effect modifiers, effect modification was not clearly demonstrated in any of the stratified analyses in the original trial publications. It is possible that effect modification is not being reported, but assuming this is not the case, we are not overly concerned about the transitivity assumption.

Assessment of inconsistency

If, despite our efforts to identify threats to the transitivity assumption described above, the transitivity assumption does not hold, or if there are other threats to the assumptions underpinning NMA, inconsistency may be observed between estimates from direct comparisons (i.e., results of the individual RCTs) and NMA estimates. We performed prespecified analyses to identify such inconsistency, as follows. For each outcome and possible pairwise comparison, we statistically compared pairwise meta-analytical estimates (representing direct evidence from the included RCTs) to NMA estimates (representing all available evidence) and to indirect NMA estimates. An indirect NMA estimate for a particular comparison is obtained by omitting all studies that make the comparison from a NMA on the remaining studies. Discrepancies between these estimates indicate possible inconsistency in the network of evidence. We were unable to perform other planned analyses of inconsistency. In particular, we could not perform loop-based analyses of inconsistency because there were no closed loops in any of the networks analyzed.

Presentation of results

We present estimates of relative treatment effect as point estimates (means) and 95% confidence intervals using forest plots and tables. We did not compute and so cannot present prediction intervals — i.e., intervals in which estimates from future studies are likely to fall — because this is not supported by the software we used. Readers are therefore cautioned that future studies may report estimates that differ from the estimates we report and we are unable to predict what these may be.

We had planned to re-express relative treatment effect estimates in "absolute" terms in the summary of findings tables to aid understanding and facilitate comparisons, following Cochrane methods. For example, we planned to re-express HRs for OS as an assumed median OS time under a reference treatment and a corresponding median for the other treatments. However, an earlier draft of this report was criticized by our clinical advisor, who misunderstood this presentation. Our planned presentation assumed a single reference population, rather than different references corresponding to patients who have been treated with and are refractory to specific treatments. To prevent misunderstanding and to avoid having to present re-expressed results for multiple populations, which would make a complex report even harder to read, we chose not to re-express treatment effect estimates.

We used P-scores (36) to quantify the extent of evidence that each treatment is superior to all other treatments with respect to a given outcome and used these values to rank the treatments. Because RRMM patients may be refractory to one or more treatments, we present P-scores for four populations: non-refractory patients; patients refractory to lenalidomide (Revlimid; R); patients refractory to bortezomib (Velcade; V); and patients refractory to both lenalidomide and bortezomib. However, these are only a small number of the possible combinations of refractory status; it is not feasible to study and present all combinations. We caution against using these rankings in

isolation. We also made radar plots to show P-scores (i.e., treatment rankings) for all included outcomes and treatments, allowing the relative advantages and disadvantages of the treatments, as characterized by our meta-analyses, to be visualized simultaneously. However, note that the radar plots do not incorporate our GRADE assessments of certainty of evidence or an economic assessment.

Figure 1 illustrates how treatment effect estimates can be obtained for patients who are refractory to specific treatments, and how we computed P-scores for the various populations. The left-most panel shows hypothetical point estimates of HRs for an NMA of four hypothetical treatments (named A, B, C, and D). Because each trial provides a treatment effect estimate that is conditional on patients not being refractory at randomization to the treatments studied — otherwise the trial would presumably have been deemed unethical and could not have been conducted — the NMA estimates also have this interpretation.

The center panel shows that all estimates involving treatment A are invalid for patients who are refractory to A. Assuming that being refractory to A is not a treatment modifier for the other comparisons (e.g., treatment A is in a different class to the other treatments, or has a sufficiently different mechanism of action), the estimates that involve A can be discarded and the remaining estimates used as is. These remaining estimates can be used to compute P-scores for patients who are refractory to A. These will likely differ from the P-scores for non-refractory patients and may differ in their rank order (see *Table 12* for an example).

Similarly, the right-most panel shows how estimates involving treatment B can be discarded in addition to those for A, to obtain treatment effect estimates and P-scores for patients who are refractory to treatments A and B (again, under the assumption that being refractory to A or B is not an effect modifier for the remaining estimates). For example, if being refractory to treatment A is known to be an effect modifier for treatment effects involving B (perhaps treatments A and B are in the same class or have the same mechanism of action), estimates involving A and B can be discarded.

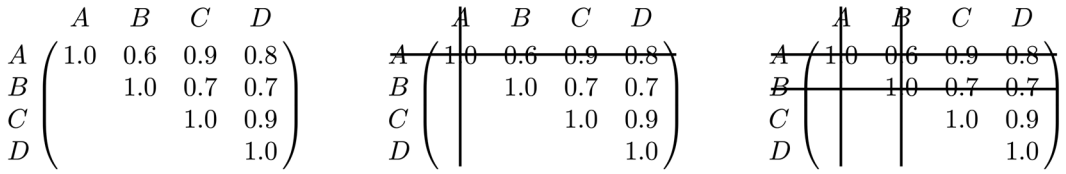


Figure 1: Treatment effect estimates for non-refractory patients (left) and patients who are refractory to specific treatments (center and right). See text for details.

Sensitivity analysis — China Continuation Study

After reviewing the results of the NMAs on effect and safety, one of our clinical experts criticized our inclusion of the China Continuation Study (37) on the basis that the included patients, who were all Chinese, are not representative of Norwegian patients due to differences in genetics, treatment history and treatment follow up. These differences might explain the difference between the results for OS in this study, and the larger TOURMALINE-MM1 study (37, 38). Both were comparisons of [IR + d] and [R + d], and both meet our prespecified inclusion criteria. Similar concerns were raised about our inclusion of the Chinese LEPUS-study ([DV + d] vs. [V + d]) (39), which also had OS results that differed from those of a larger study (40), although the difference was less marked.

To address the expert's concerns, we performed a non-prespecified sensitivity analysis in which we repeated all NMAs but excluded the China Continuation Study. We summarize the results with respect to OS and PFS by showing how P-scores for the treatments change when the China Continuation Study is excluded. We present summaries for non-refractory patients, patients refractory to lenalidomide (R), patients

refractory to bortezomib (V), and patients refractory to lenalidomide (R) and bortezomib (V). Note that P-scores and treatment rankings will not change for patients refractory to lenalidomide (R), or lenalidomide (R) and bortezomib (V) because [IR + d] is not an admissible treatment for such patients.

GRADE: assessing the certainty of evidence

The certainty of evidence for our outcomes was assessed using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach in accordance with the GRADE handbook, as well as the GRADE Working Group guidance of rating estimates from NMAs (41, 42). In the GRADE approach, RCTs are as a starting point, considered to provide high quality evidence. The subsequent rating of the certainty of evidence may be reduced after further assessment, thereby reducing the confidence of the effect estimate (41, 42). As all the included studies in our HTA are RCTs, our outcomes were set to start out at high certainty of evidence for each treatment regimen. The certainty was then further assessed with regards to the following factors: 1) study limitations (risk of bias), 2) inconsistency, 3) indirectness, 4) imprecision (see below for further information), and 5) publication bias (41).

We assessed certainty of evidence of all NMA estimates for treatment regimens relevant for Norway, for all outcomes (see *How we present the findings*). In addition, we assessed the certainty of evidence for all direct evidence, i.e., estimates that were extracted directly from the included studies, for OS and PFS. Certainty of evidence is classified as in *Table 4* (43). Two researchers assessed certainty of evidence, and any disagreements were resolved through discussion.

Assessing imprecision

To assess imprecision in a fair and reproducible manner, we made forest plots as a visual aid to better judge the width and placement of the 95% confidence intervals according to our set thresholds of no effect and assumed important effect, respectively. We graded one down if the 95% confidence interval crossed the threshold of no effect (either 0 or 1 depending on the outcome), as well as *either* the upper or lower threshold of assumed important effect. We graded two down if the 95% confidence interval crossed the threshold of no effect, in addition to *both* upper and lower threshold of assumed important effect, and/or if the 95% confidence interval was particularly large.

Assessing indirect evidence

We only assessed certainty of indirect evidence *within* the same network, not across networks, and we based our approach on the guidance from the GRADE Working Group for rating effect estimates from NMAs (42). In brief, we first assessed all direct evidence contributing to the NMA-estimate separately. We then chose the lowest of the ratings from the direct evidence and used this to further assess imprecision for the indirect NMA-estimate, as well as intransitivity between the contributing studies. Based on these assessments, we set a final rating for the total certainty evidence for the selected treatment regimens, for each outcome.

Table 4: Certainty of evidence classification according to GRADE handbook and GRADEpro *(41, 43)*

GRADE	Definition
High certainty ⊕⊕⊕⊕	We are very confident that the true effect lies close to that of the estimate of the effect. Further research is very unlikely to change our confidence in the estimate of effect
Moderate certainty ⊕⊕⊕	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially

	different. Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low certainty ⊕⊕	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect. Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low certainty ⊕	Any estimate of effect is very uncertain, and we have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

Assumed important effect

We chose to set thresholds for what could be considered as an important effect for patients for each outcome, as an aid in assessing certainty of evidence for imprecision.

For survival data

In consulting with clinical experts, they pointed out that interventions that lead to improvements of less than 3 months survival in OS and PFS could be defined as not clinically important. However, as we could not translate the direct HR effect estimates to assumed and corresponding survival times expressed in months, we chose to GRADE all effect estimates, i.e., both the non-NMA estimates and NMA-estimates (direct and indirect) in the same manner. We therefore set liberal thresholds of assumed important effect for both OS and PFS at HR=0.8 and HR=1.25.

For quality of life

For QoL, the effect estimates were expressed as difference in QLQ-30 global health score. We performed a simple literature search and found a paper suggesting that an absolute change in the EORTC QLQ-C30 score of 8-12 could be considered important by patients with multiple myeloma in Norway (44). Based on this, we used the lower absolute score change of 8 (i.e., +8 and -8) as the threshold for assumed important effect.

For severe adverse events and discontinuation due to adverse events

A simple literature search revealed no information regarding assumed important effect for SAE and discontinuation due to adverse events. As such, we assessed the importance of effects and the precision of the estimates based on how likely it seemed that patients would make different decisions if the true effect was near the lower or upper end of the 95% confidence interval. Based on this, we assumed it to be an important effect if the effect estimates of differences in events (i.e., SAE, and discontinuations, respectively), were below -50 or above 50.

Ethical aspects

Ethics was not assessed for this health technology assessment.

Legal aspects

Legal aspects was not assessed for this health technology assessment.

RESULTS

Literature search and article selection

The article selection is presented in *Figure 2*.

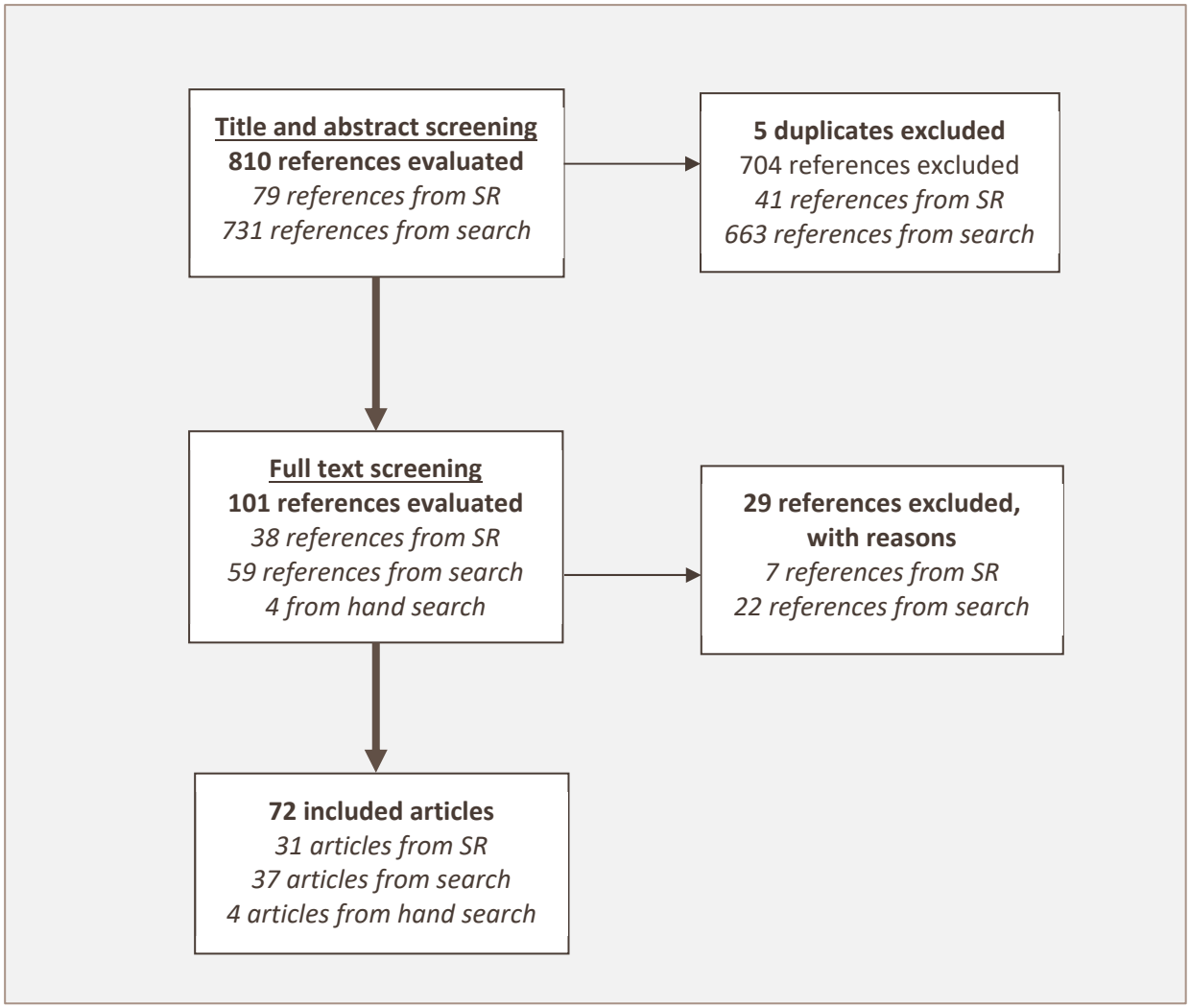


Figure 2: Flow chart of article selection.
SR: systematic reviews.

References from systematic reviews

We identified 79 references from the 13 systematic reviews (9-21) included in our previous mapping review (8). First, 41 references were excluded based on screening of title and abstract, and then further seven references were excluded based on full text screening. In total 31 articles were included from the systematic reviews (*Table 5*).

References from search

We identified 731 references from four searches performed in the period of February 2020 to January 2022. First, 661 references were excluded based on screening of title and abstract, and then further 22 references were excluded based on full text screening. In total, 37 articles were included from the searches (*Table 5*).

References from hand search

We performed a hand search for articles from studies already identified from systematic reviews, that present data on QoL. In total, four articles were included from the hand search (*Table 5*).

Description of studies

Excluded studies

A full list of articles excluded through the full text screening (seven from systematic reviews, and 22 from searches), with reasons for why they were excluded, is presented in *Appendix 5*. In brief, exclusions were mainly due to wrong study design, and wrong population, intervention, or outcome. In addition, several articles were also excluded when we had included other articles from the same study with longer follow-up.

Included studies

All included studies are briefly presented in *Table 5*. In total, we included 72 articles from 50 RCTs, with a total number of 28 339 participants, ranging from 15 to 465 in the different trials. All participants had multiple myeloma and had experienced at least one relapse and/or were treatment refractory to at least one previous line of treatment. The included studies are presented in further detail in *Appendix 6*. Some of the included articles or studies were however omitted from the analysis, either because they compared different doses or administration methods of the same treatment regimen, or because they did not present data in a manner that was compatible with the meta-analysis. A list of these omitted studies and articles with reasons for why, is presented in *Appendix 7*.

Ongoing studies

The list detailing relevant ongoing clinical trials is found in *Appendix 8*. In brief, we found 28 ongoing trials that represent 8616 planned participants and include various treatment regimens with different drugs. Five of the ongoing studies have results that have been included in this HTA (*Appendix 8*).

Table 5: Included studies

Study name	Publication – year – (ref)	Publications identified from search or SR	Treatments
1703	Richardson 2015 (45)	SR (19)	[ER + d], n=36 [ER + d], n=37
AMBER	White 2013 (46)	SR (9, 12, 13)	[BevV], n=49 [PboV], n=53
APEX	Richardson 2005 (47)	SR (9, 12, 13, 17)	[V], n=333 [d], n=336
	Richardson 2007 (48)	SR (9, 12, 13, 17)	
	Lee 2008 (49)	Hand search	
APOLLO	Dimopoulos 2021 (50)	Search 4	[DP + d], n=151 [P + d], n=153
	Terpos 2022† (51)	Hand search	
ARROW	Moreau 2018 (52)	Search 1	[K + d], n=240 [K + d], n=238
	Moreau 2019 (53)	Search 1	
ASPIRE	Stewart 2015 (30)	SR (9, 12, 13, 15-17, 21)	[KR + d], n=396 [R + d], n=396
	Siegel 2018 (54)	Search 1	
BELLINI	Kunar 2021 (55)	Search 3	[VenV + d], n=194 [PboV +]d, n=97
BOSTON	Grosecki 2020 (56)	From search 4	[SeV + d], n=195 [V + d], n=207
CANDOR	Dimopoulos 2020 (57)	Search 2	[DK + d], n=312 [K + d], n=154
	Usmani 2022 (58)	Search 4	
	Siegel 2021 (59)	Search 4	
CASTOR	Palumbo 2016 (60)	SR (9, 12, 13, 17-20)	[DV + d], n=251 [V + d], n=247
	Mateos 2020 (40)	Search 2	
	Hungria 2021 (61)	Search 3	
COLUMBA	Mateos 2020 (62)	Search 2	[D], n=263 [D], n=259
CREST	Jagannath 2004 (63)	SR (13)	[V + d], n=28 [V + d], n=26
DOXIL-MMY-3021	Orlowski 2007 (64)	SR (9, 12, 13)	[DoxV], n=324 [V], n=322
	Orlowski 2016 (65)	SR (9, 12, 13)	
ELOQUENT-2*	Cella 2018 (66)	Search 1	[ER + d], n=321 [R + d], n=325
	Dimopoulos 2020 (67)	Search 3	
ELOQUENT-3	Dimopoulos 2018 (68)	Search 1	[EP + d], n=60 [P + d], n=154
ENDEAVOR	Dimopoulos 2016 (69)	SR (9, 12, 13, 15, 17, 18)	[K + d], n=464 [V + d], n=465
	Ludwig 2019 (70)	Search 1	
	Orlowski 2019 (71)	Search 1	
FOCUS	Hájek 2017 (72)	SR (12, 15, 21)	[K], n=157 [cs (± Cy)], n=158
ICARIA-MM	Attal 2019 (73)	Search 1	[IsP + d], n=154 [P + d], n=153
IFM 2009-02	Leleu 2013 (74)	SR (10, 13, 14, 21)	[P + d], n=43 [P + d], n=41
IKEMA	Moreau 2021 (75)	Search 4	[IsK + d], n=179 [K + d], n=123

KEYNOTE-183	Mateos 2019 (76)	Search 1	[PemP + d], n=125 [P + d], n=124
LEPUS	Lu 2021 (39)	Search 4	[DV + d], n=141 [V + d], n=70
MM-002	Richardson 2014 (77)	SR (10, 12-14, 21)	[P], n=108 [P + d], n=113
MM-003	San Miguel 2013 (78) Song 2015† (79) Weisel 2015† (80)	SR (10, 12, 13, 17, 21) Hand search Hand search	[P + d], n=302 [d], n=153
MM-009	Weber 2007 (81)	SR (9, 12, 13, 17)	[R + d], n=177 [Pbo + d], n=176
MM-010	Dimopoulos 2007 (82)	SR (9, 12, 13, 17)	[R + d], n=176 [Pbo + d], n=175
MMVAR/IFM 2005-04	Garderet 2012 (83)	SR (9, 12, 16-18)	[VT + d], n=135 [T + d], n=134
MMY-3033	Terpos 2018 (84)	Search 1	[V], n=53 [V], n=27
Nordic Myeloma Study	Hjorth 2012 (25)	SR (9, 12, 17, 18)	[T + d], n=67 [V + d], n=64
OPTIMISMM	Richardson 2019 (85) Weisel 2020 (86)	Search 1 Search 2	[PV + d], n=281; [V + d], n=278
PANORAMA-1	San Miguel 2014 (87) San Miguel 2016 (88) Richardson 2018 (89)	SR (9, 11-13, 16-18, 20) SR (9, 11-13, 16-18, 20) 2018	[FV + d], n=387 [V + d], n=396
PANORAMA-3	Laubach 2021 (90)	Search 3	[FV + d], n=82 [FV + d], n=83
POLLUX	Dimopoulos 2016 (91) Bahlis 2020 (92) Plesner 2021 (93)	SR (9, 12, 13, 17, 19, 20) Search 1 Search 4	[DR + d], n=281 [R + d], n=276
SIRIUS	Lonial 2016 (94)	SR (19)	[D + mp], n=106 [D + mp], n=18
The China Continuation Study	Hou 2017 (37)	SR (12, 13)	[IR + d], n=57 [R + d], n=58
TOURMALINE-MM1	Moreau 2016 (95) Leleu 2018 (26) Richardson 2021 (38)	SR (9, 12, 13, 16, 17) Search 1 Search 4	[IR + d], n=360 [R + d], n=465
VANTAGE-088	Dimopoulos 2021 (96)	SR (9, 12, 13, 17, 20)	[VorV], n=317 [V], n=320
	Ailawadhi 2020 (97)	Search 2	[K + d], n=64 [K + d], n=57
	Baz 2016 (98)	SR (10, 13, 21)	[CyP + d], n=34 [P + d], n=36
	Dimopoulos 2021 (99)	Search 4	[Is], n=109 [Is + d], n=55
	Iida 2018 (100)	Search 1	[T + d], n=22 [V + d], n=22
	Jakubowiak 2016 (101)	SR (9, 12, 18, 19)	[EV + d], n=77

			[V + d], n=75
	Kropff 2017 (102)	SR (18)	[CyV + d], n=46 [V + d], n=47
	Mikhael 2020 (103)	Search 2	[Is], n=23; [Is], n=25; [Is], n=24; [Is], n=23
	Mina 2020 (104)	Search 2	[V + d], n=15 [V + d], n=23 [Std.care], n=20
	Montefusco 2020 (105)	Search 2	[CyV + d], n=76 [CyR + d], n=79
	Moreau 2011 (106)	SR (9, 13)	[V + d], n=148 [V + d], n=74
	Orlowski 2015 (107)	SR (9, 12, 13, 17)	[SV], n=142; [PboV], n=144
	Raje 2017 (108)	Search 1	[TabV + d], n=74 [TabV + d], n=74 [PboV + d], n=72
	Sehgal 2015 (109)	SR (13)	[P + d], n=19 [P + d], n=20

Bev: bevacizumab; cs.: corticosteroids; Cy: cyclophosphamide; d: dexamethasone; D: daratumumab; Dox: doxorubicin; E: elotuzumab; F: panobinostat (Farydak); I: ixazomib; Is: isatuximab; K: carfilzomib (Kyprolis); mp: methylprednisolone; P: pomalidomide; p: prednisone; Pbo: placebo; Pem: pembrolizumab; R: lenalidomide (Revlimid); S: siltuximab; Se: Selinexor; SR: systematic review; T: thalidomide; Tab: tabalumab; V: bortezomib (Velcade); Vor: vorinostat

Shaded rows: included studies that could not be used in the statistical analysis. †Separate article of included study, that could not be used in the statistical analysis.

**The ELOQUENT-2 study was first identified through several systematic reviews (9, 12, 13, 16, 17, 19, 20), though we have only included articles identified through our searches as these had longer follow-up.*

Risk of bias in included studies

For 22 of our included RCTs we used the risk of bias assessment made by Maiese et al (13) (*Table 6*, *Figure 3*, *Figure 5A*). For the remaining 23 RCTs we made our own risk of bias assessment (*Table 6*, *Figure 4*, *Figure 5B*).

Table 6: List of studies where risk of bias was assessed by Maiese et al (13) and by us

Studies where RoB was assessed by Maiese et al (13)	Studies where RoB was assessed by us
AMBER: *White 2013 (46)*	**1703:** *Richardson 2015 (45)*
APEX: *Richardson 2005 (47); Richardson 2007 (48); Lee 2008 (49)*	**Ailawadhi 2020** (97)
ASPIRE: *Stewart 2015 (30); Siegel 2018 (54)*	**APOLLO:** *Dimopoulos 2021 (50); Terpos 2022 (51)*
Baz 2016 (98)	**ARROW:** *Moreau 2018 (52); Moreau 2019 (53)*
CASTOR: *Palumbo 2016 (60); Hungria 2020 (61); Mateos 2020 (40)*	**BELLINI:** *Kumar 2020 (55)*
CREST: *Jagannath 2004 (63)*	**BOSTON:** *Grosicki 2020 (56)*
DOXIL-MMY-3021: *Orlowski 2007 (64); Orlowski 2016 (65)*	**CANDOR:** *Dimopoulos 2020 (57); Siegel 2021 (59); Usmani 2022 (58)*
ELOQUENT-2: *Cella 2018 (66); Dimopoulos 2020 (67)*	**COLOMBA:** *Mateos 2020 (62)*
ENDEAVOR: *Dimopoulos 2016 (69); Ludwig 2019 (70); Orlowski 2019 (71)*	**Dimopoulos 2021** (99)
IFM 2009-02: *Leleu 2013 (74)*	**ELOQUENT-3:** *Dimopoulos 2018 (68)*

MM-002: *Richardson 2014 (77)*	**FOCUS:** *Hájek 2017 (72)*
MM-003: *San Miguel 2013 (78); Song 2015 (79); Weisel 2015 (80)*	**ICARIA-MM:** *Attal 2019 (73)*
MM-009: *Weber 2007 (81)*	**Iida 2018** (100)
MM-010: *Dimopoulos 2007*	**IKEMA:** *Moreau 2021 (75)*
Moreau 2011 (106)	**Jakubowiak 2016** (101)
Orlowski 2015 (107)	**KEYNOTE-183:** *Mateos 2019 (76)*
PANORAMA-1: *San Miguel 2014 (87); San Miguel 2016 (88); Richardson 2018 (89)*	**Kropff 2017** (102)
POLLUX: *Dimopoulos 2016 (91); Bahlis 2020 (92); Plesner 2021 (93)*	**LEPUS:** *Lu 2021 (39)*
Sehgal 2015 (109)	**Mikhael 2020** (103)
The China Continuation Study: *Hou 2017 (37)*	**Mina 2020** (104)
TOURMALINE-MM1: *Moreau 2016 (95); Leleu 2018 (26); Richardson 2021 (38)*	**MMVAR/IFM 2005-04:** *Garderet 2012 (83)*
VANTAGE-088: *Dimopoulos 2013 (96)*	**MMY-3033:** *Terpos 2018 (84)*
	Montefusco 2020 (105)
	Nordic Myeloma Study: *Hjorth 2012 (25)*
	OPTIMISMM: *Richardson 2019 (85); Weisel 2020 (86)*
	PANORAMA-3: *Laubach 2021 (90)*
	Raje 2017 (108)
	SIRIUS: *Lonial 2016 (94)*

Our risk of bias assessment was based on having OS as our main outcome. As such, we considered studies with no or unclear blinding of the personnel who carried out the outcome assessment to have low risk of bias, as we were confident this would likely not have any effect on OS. We were unsure if OS may be affected by blinding of participants and the personnel involved in the treatment of the patients. We therefore assessed the risk of bias as uncertain when there was no blinding, or the degree of blinding was unclear. However, when assessing an *overall* risk of bias for each of these studies, we gave little weight to our initial assessment of this domain. When used in GRADEing certainty of evidence, we assessed the overall risk of bias specifically for each outcome. Thus, overall risk of bias assessment for one study could differ between outcomes.

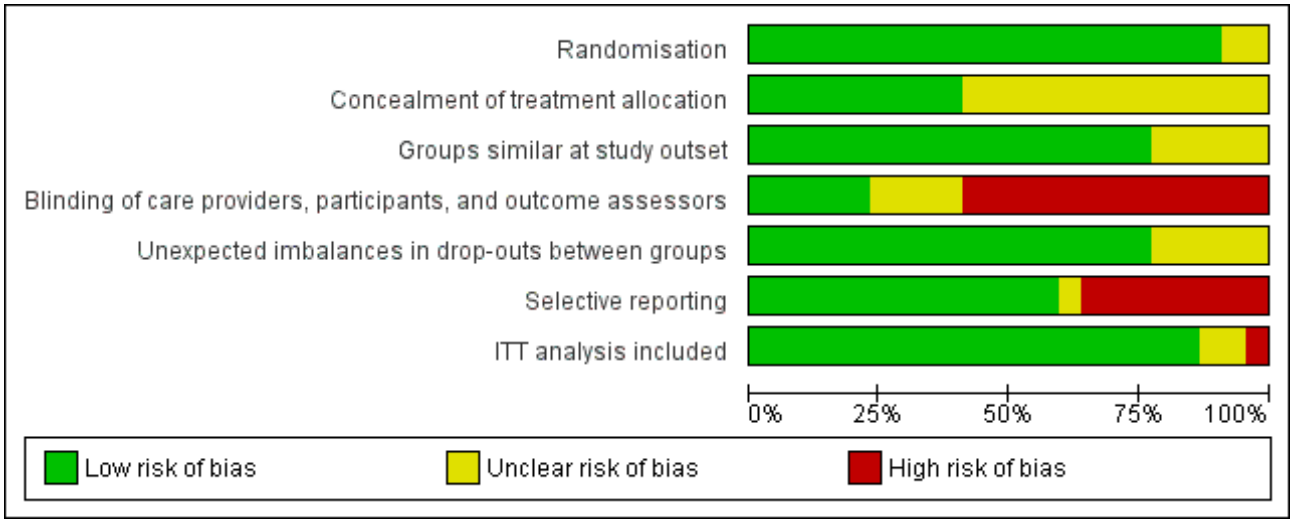


Figure 3: Risk of bias graphs across included studies*. Assessments made by Maiese et al (13).*

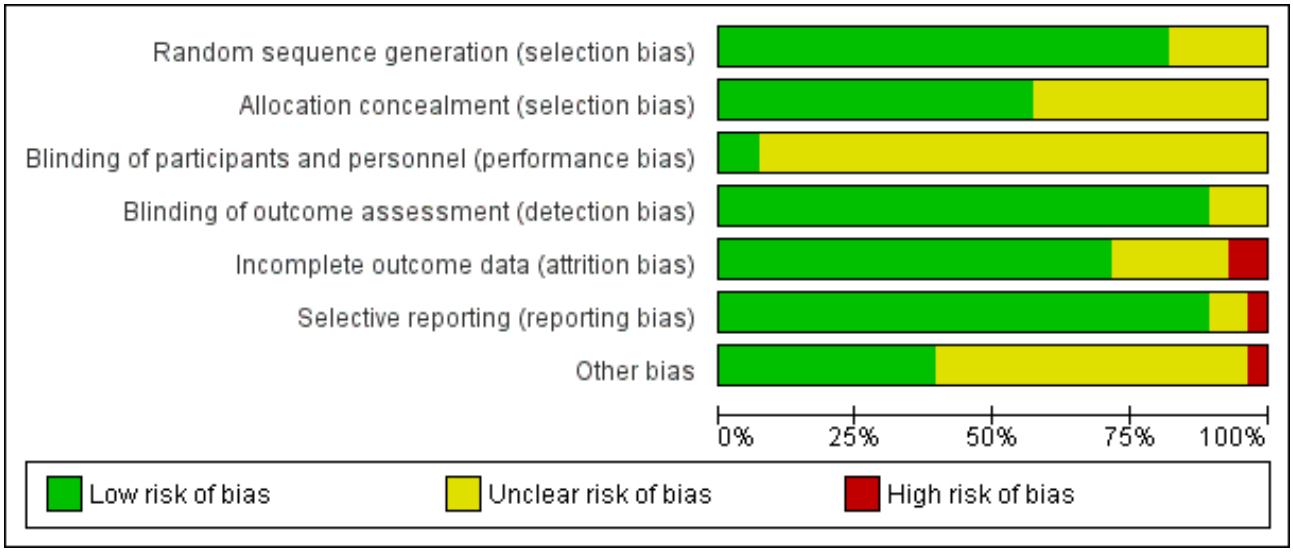


Figure 4: Risk of bias graphs across included studies. *Assessments made by us.*

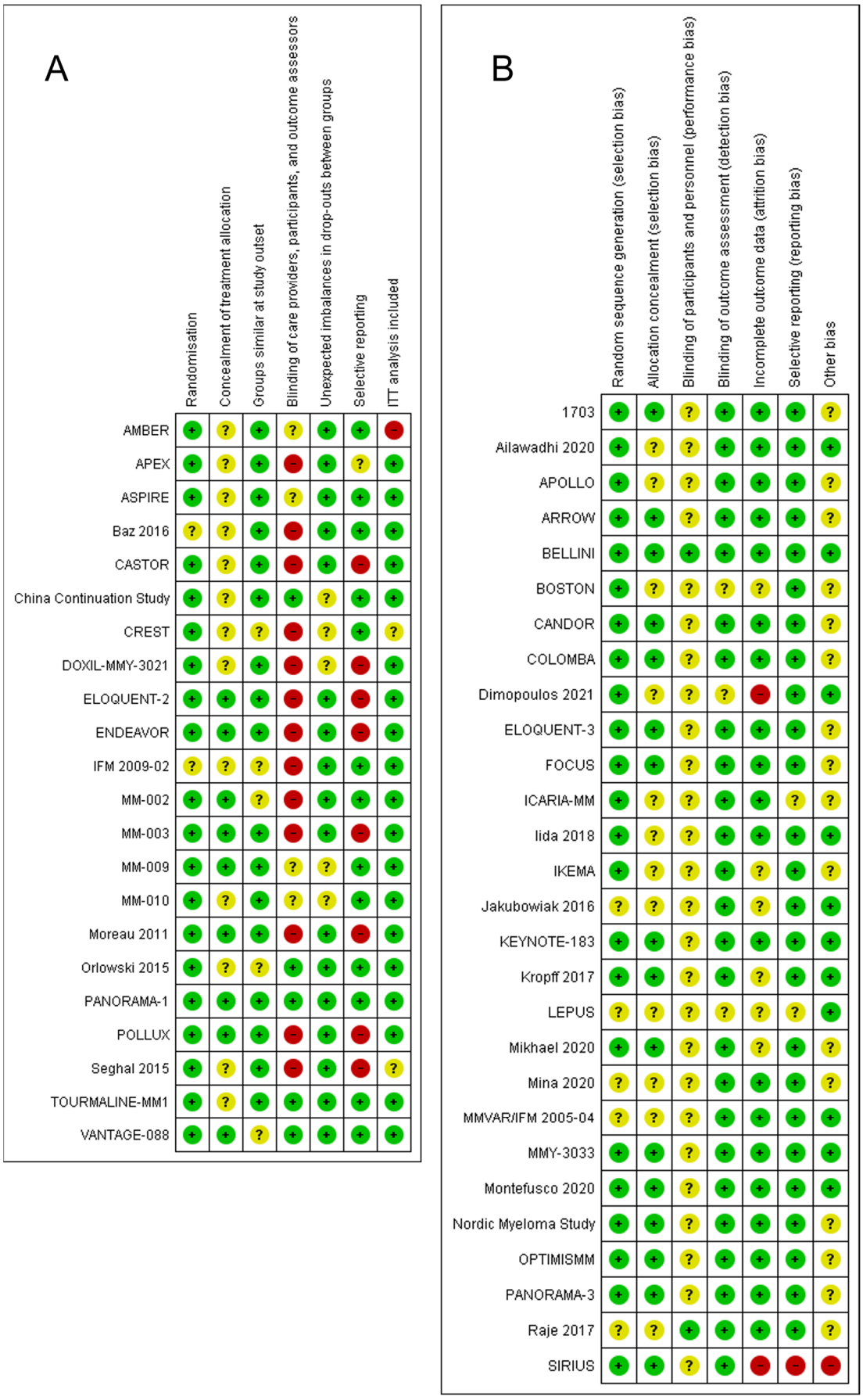


Figure 5: Risk of bias tables for each included study. *Risk of bias assessment made by A) Maiese et al (13), and B) us.*

How we present the findings

The Method chapter explains how we defined treatments and how meta-analysis was performed. We suggest readers familiarize themselves with this chapter before trying to read the results.

While only some of the treatments reported on were pre-specified and listed in our protocol, we also included treatments used as comparators (which are necessary for obtaining treatment effect estimates and hence performing a NMA), and therefore present results for all treatments included in the meta-analysis. We present results of treatment regimens that are relevant for Norway, i.e., listed as first, second or third treatment choice in the Norwegian guideline for treatment of multiple myeloma (3). Results for all treatments from all of the included studies, are presented in *Appendix 10*.

The calculation and interpretation of a relative treatment effect such as HR necessitates a reference treatment. In a NMA, all treatments are comparators and any one of them can be used as a reference without affecting the underlying estimates.
For each outcome, we present:

- A figure that presents the topology of the evidence network resulting from the included studies. In all cases evidence networks were disconnected. Readers are directed to the Method chapter for an explanation of how we addressed this issue.
- A matrix of all possible pairs of relative treatment effect estimate provided by the NMA, for treatments relevant in Norway, as well as our assessment of the certainty of the evidence for each comparison. While "statistically significant" results are highlighted, this simply means that there is insufficient evidence to precisely estimate the relative effects of other pairs of treatments, and does not preclude superiority of some treatments over others being demonstrated in future studies
- A summary of findings table ranked by P-score for non-refractory patients, along with P-scores for three other populations (see Methods), for treatment regimens relevant for Norway. A P-score can be interpreted as the probability that a given treatment is superior, with respect to the outcome, to all other treatments, accounting for the modeled uncertainty of the relative treatment effects. P-scores are not P-values. P-scores are not anticipated to sum to 100%.
- A forest plot of the direct treatment effect estimates (HRs with 95% CIs) published in the included studies for the two routinely-studied efficacy outcomes OS and PFS. (Direct treatment effect estimates, including statements of precision, were often not published for the other outcomes.)

Furthermore, in *Appendix 10*, we present additional results for each outcome:

- A figure that shows the treatments, studies, designs, and sample sizes that could be included in the meta-analysis.
- A forest plot showing the extracted or imputed data and the corresponding estimates provided by meta-analysis.
- A matrix of all possible pairs of relative treatment effect estimate provided by the NMA, for all treatments of the included studies
- A summary of findings table ranked by P-score for non-refractory patients, along with P-scores for three other populations (see Methods), for all included treatment regimens. A P-score can be interpreted as the probability that a given treatment is superior, with respect to the outcome, to all other treatments, accounting for the modeled uncertainty of the relative treatment effects. P-scores are not P-values. P-scores are not anticipated to sum to 100%.

Finally, we summarize all included treatments with respect to all outcomes studied by presenting radar plots of P-scores. However, readers are discouraged from drawing

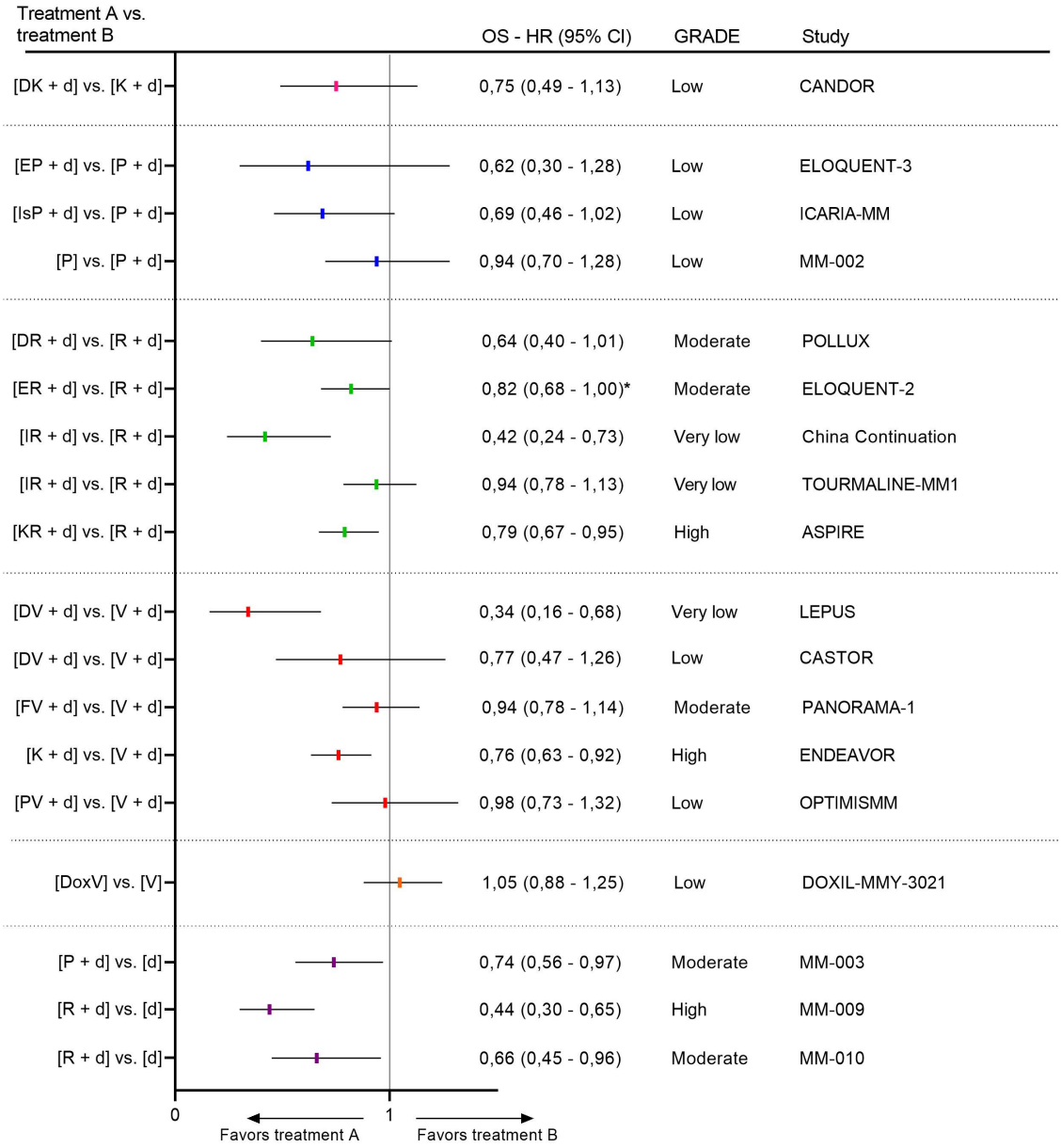
conclusions based on these plots alone but should consider all available evidence. While it may be tempting to favor treatments with "larger" polygons in these figures, note that some outcomes are missing for some treatments, which can distort the size of the polygons and lead to potentially erroneous conclusions.

Certainty of evidence

We evaluated the certainty of the estimates of the primary outcomes using the GRADE-NMA approach. Readers are referred to the Method chapter for more information on how this was done (*GRADE: assessing the certainty of evidence*). Our GRADE judgements are presented in the matrix plots for all outcomes, as well as in *Appendix 9*.

Direct evidence

Evidence extracted directly from the included studies of HRs with 95% confidence intervals for OS for treatment regimens relevant for Norway, was set into a forest plot, along with assessments of certainty of evidence (GRADE), and study name (Figure 6: Forest plot of direct evidence - overall survival*Figure 6*). Within the forest plot, the direct evidence is organized after the common treatments (i.e., treatment B), such as [K + d], [P + d], etc. Details of our assessment of GRADE is presented in *Appendix 9 -*



Detailed GRADE: overall survivalDirect estimates.

Figure 6: Forest plot of direct evidence - overall survival
** 95,4% confidence interval. Only shown treatment regimens that are relevant for Norway. CI: confidence interval, D: daratumumab, d: dexamethasone, Dox: doxurubicin, E: elotuzumab, F: panobinostat, HR: hazard ratio, I: ixazomib, Is: isatuximab, K: carfilzomib, OS: overall survival, P: pomalidomide, R: lenalidomide, V: bortezomib. HR<1 favours treatment A, HR>1 favours treatment B.*

NMA results

The NMA of OS resulted in three disconnected networks (*Figure 7*) with a total of 31 treatment regimens. An NMA was performed using data on the 31 treatment regimens from 31 RCTs that enrolled a total of 12 279 patients (*Appendix 10 - Additional results: overall survival*).

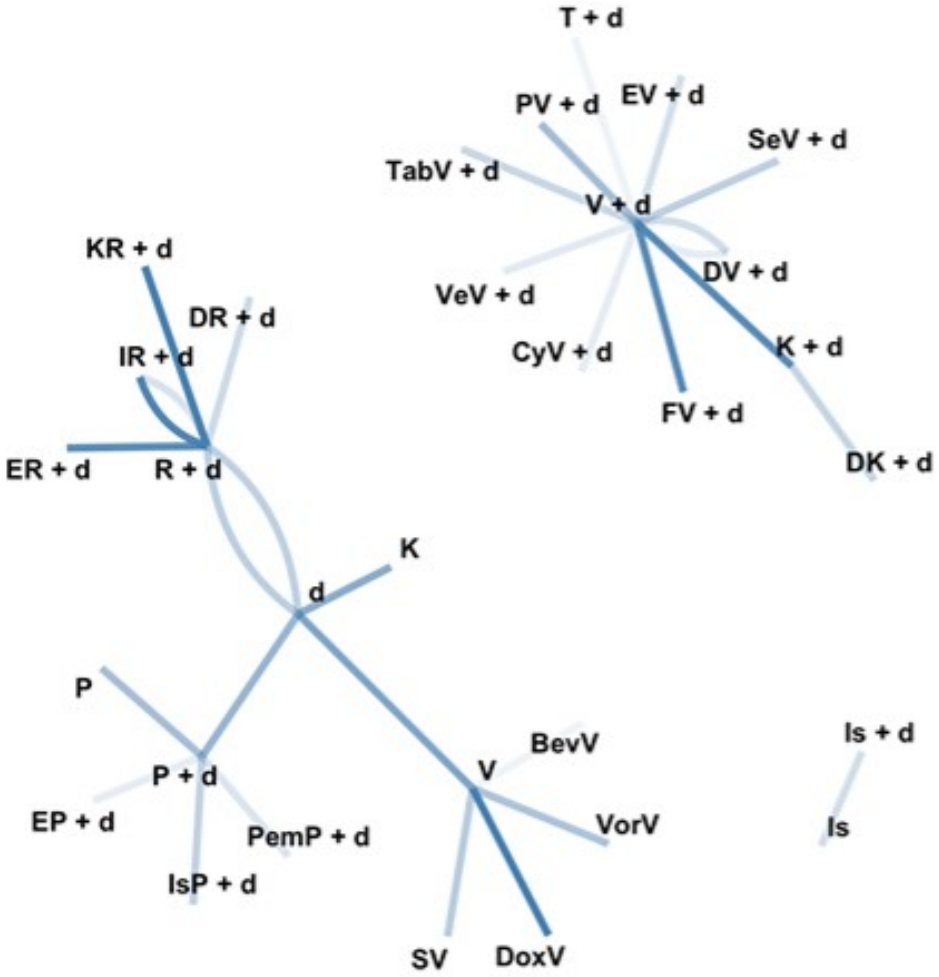


Figure 7: Network topology for overall survival
Each vertex represents a treatment, and each edge (line) represents a direct treatment comparison. More precise estimates (e.g., those supported by studies with larger sample sizes) are indicated by darker edges. Bev: bevacizumab, Cy: cyclophosphamide, d: dexamethasone, D: daratumumab, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, Pem: pembrolizumab, R: lenalidomide, S: siltuximab, Se: selinexor, T: thalidomide, Tab: tabalumab, V. bortezomib, Ve: venetoclax, Vor: vorinostat

We used random effects component NMA to estimate average HRs to account for between-study heterogeneity. An average treatment effect estimate should not be misinterpreted as being equivalent to an estimate from a single study (which cannot account for between-study heterogeneity). Consequently, confidence intervals on average treatment effect estimates may be wider that those arising from individual studies. The matrix plot (*Table 7*) presents estimates of HRs and 95% confidence intervals, in addition to our assessment of the certainty of evidence, for treatment regimens relevant for Norway. Heterogeneity (I^2) was estimated to be 74.9% (95% CI 43.1% to 88.9%).

The summary of findings table (*Table 8****Feil! Fant ikke referansekilden.***) presents treatment regimens relevant for Norway ranked by P-score for non-refractory patients along with assessments of certainty of evidence (GRADE). It also presents ranked treatments for patients who are refractory to lenalidomide and/or bortezomib.

Table 7: Matrix plot - overall survival

Treatment A	Treatment B													
	DK + d	**DoxV**	**DR + d**	**DV + d**	**EP + d**	**ER + d**	**FV + d**	**IR + d**	**IsP + d**	**K + d**	**KR + d**	**P + d**	**PV + d**	**V + d**
R + d	1.00 (0.32 – 3.18) *NP*	0.56 (0.22 – 1.47) *Very low*	1.56 (0.73 – 3.34) *Low*	1.22 (0.42 – 3.56) *NP*	1.18 (0.33 – 4.14) *Very low*	1.22 (0.65 – 2.29) *Low*	0.72 (0.24 – 2.13) *NP*	1.44 (0.87 – 2.37) *Very low*	1.06 (0.35 – 3.19) *Very low*	0.75 (0.31 – 1.83) *NP*	1.27 (0.68 – 2.37) *Low*	0.73 (0.32 – 1.68) *Very low*	0.69 (0.23 – 2.09) *NP*	0.67 (0.28 – 1.64) *NP*
DK + d		0.56 (0.17 – 1.89) *NP*	1.56 (0.39 – 6.21) *NP*	1.22 (0.41 – 3.58) *Very low*	1.17 (0.25 – 5.53) *NP*	1.22 (0.33 – 4.54) *NP*	0.71 (0.24 – 2.14) *Very low*	1.44 (0.41 – 5.05) *NP*	1.06 (0.25 – 4.42) *NP*	0.75 (0.36 – 1.56) *Low*	1.26 (0.34 – 4.70) *NP*	0.73 (0.21 – 2.49) *NP*	0.69 (0.22 – 2.10) *Very low*	0.67 (0.27 – 1.65) *Very low*
DoxV			2.77 (0.81 – 9.44) *Very low*	2.16 (0.79 – 5.91) *NP*	2.09 (0.51 – 8.57) *Very low*	2.16 (0.68 – 6.83) *Very low*	1.27 (0.46 – 3.54) *NP*	2.55 (0.86 – 7.54) *Very low*	1.88 (0.53 – 6.75) *Very low*	1.33 (0.51 – 3.49) *NP*	2.25 (0.71 – 7.08) *Very low*	1.29 (0.45 – 3.70) *Very low*	1.22 (0.43 – 3.48) *NP*	1.19 (0.53 – 2.68) *NP*
DR + d				0.78 (0.21 – 2.90) *NP*	0.75 (0.17 – 3.27) *Very low*	0.78 (0.29 – 2.10) *Very low*	0.46 (0.12 – 1.73) *NP*	0.92 (0.37 – 2.28) *Very low*	0.68 (0.18 – 2.59) *Very low*	0.48 (0.15 – 1.55) *NP*	0.81 (0.30 – 2.17) *Very low*	0.47 (0.15 – 1.44) *Very low*	0.44 (0.11 – 1.69) *NP*	0.43 (0.13 – 1.39) *NP*
DV + d					0.97 (0.22 – 4.29) *NP*	1.00 (0.29 – 3.48) *NP*	0.59 (0.25 – 1.40) *Very low*	1.18 (0.36 – 3.86) *NP*	0.87 (0.22 – 3.40) *NP*	0.62 (0.28 – 1.36) *Very low*	1.04 (0.30 – 3.60) *NP*	0.60 (0.19 – 1.90) *NP*	0.56 (0.23 – 1.39) *Very low*	0.55 (0.30 – 1.01) *Very low*
EP + d						1.04 (0.25 – 4.24) *Very low*	0.61 (0.14 – 2.74) *NP*	1.22 (0.32 – 4.73) *Very low*	0.90 (0.28 – 2.96) *Very low*	0.64 (0.16 – 2.51) *NP*	1.08 (0.26 – 4.39) *Very low*	0.62 (0.24 – 1.59) *Low*	0.58 (0.13 – 2.67) *NP*	0.57 (0.15 – 2.24) *NP*
ER + d							0.59 (0.17 – 2.07) *NP*	1.18 (0.53 – 2.64) *Very low*	0.87 (0.24 – 3.10) *Very low*	0.62 (0.21 – 1.84) *NP*	1.04 (0.43 – 2.53) *Low*	0.60 (0.21 – 1.70) *Very low*	0.56 (0.16 – 2.03) *NP*	0.55 (0.19 – 1.64) *NP*
FV + d								2.01 (0.61 – 6.67) *NP*	1.48 (0.37 – 5.88) *NP*	1.05 (0.46 – 2.38) *Very low*	1.77 (0.50 – 6.23) *NP*	1.02 (0.32 – 3.29) *NP*	0.96 (0.38 – 2.41) *Very low*	0.94 (0.50 – 1.77) *Low*
IR + d									0.74 (0.22 – 2.47) *Very low*	0.52 (0.19 – 1.45) *NP*	0.88 (0.40 – 1.96) *Very low*	0.51 (0.19 – 1.34) *Very low*	0.48 (0.14 – 1.62) *NP*	0.47 (0.17 – 1.30) *NP*
IsP + d										0.71 (0.21 – 2.41) *NP*	1.19 (0.34 – 4.24) *Very low*	0.69 (0.33 – 1.41) *Very low*	0.65 (0.16 – 2.61) *NP*	0.63 (0.19 – 2.16) *NP*
K + d											1.68 (0.57 – 5.01) *NP*	0.97 (0.36 – 2.61) *NP*	0.91 (0.39 – 2.14) *Very low*	0.90 (0.53 – 1.51) *Low*

		Treatment B													
		DK + d	DoxV	DR + d	DV + d	EP + d	ER + d	FV + d	IR + d	IsP + d	K + d	KR + d	P + d	PV + d	V + d
	KR + d												0.58 (0.20 – 1.63) *Very low*	0.54 (0.15 – 1.95) *NP*	0.53 (0.18 – 1.58) *NP*
	P + d													0.94 (0.29 – 3.11) *NP*	0.92 (0.34 – 2.48) *NP*
	PV + d														0.98 (0.50 – 1.92) *Low*

D: daratumumab, d: dexamethasone, Dox: doxorubicin,, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib.

NP: not possible to GRADE because the comparison is across networks.

Only listed treatment regimens relevant for Norway

Table 8: Summary of findings for overall survival

Treatments are ordered by overall rank relevant for non-refractory patients

Treatment*	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant for people who are:		
		Refractory to R	Refractory to V	Refractory to R & V
DR + d	1 (0.86)	NA	1 (0.84)	NA
IR + d	2 (0.85)	NA	2 (0.83)	NA
DV + d	3 (0.79)	1 (0.78)	NA	NA
KR + d	4 (0.78)	NA	3 (0.75)	NA
ER + d	5 (0.77)	NA	4 (0.73)	NA
EP + d	6 (0.73)	2 (0.75)	5 (0.67)	1 (0.77)
IsP + d	7 (0.70)	3 (0.72)	6 (0.63)	2 (0.74)
R + d	8 (0.68)	NA	7 (0.63)	NA
DK + d	9 (0.67)	4 (0.70)	8 (0.60)	3 (0.71)
K + d	13 (0.52)	8 (0.57)	9 (0.45)	4 (0.58)
P + d	14 (0.50)	9 (0.56)	10 (0.44)	5 (0.57)
FV + d	15 (0.49)	10 (0.54)	NA	NA
PV + d	17 (0.46)	12 (0.52)	NA	NA
V + d	19 (0.44)	14 (0.51)	NA	NA
DoxV	24 (0.34)	19 (0.41)	NA	NA

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, NA: not applicable, P: pomalidomide, R: lenalidomide, V: bortezomib

** The Methods chapter describes how treatments are defined.*

§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Only listed treatment regimens relevant for Norway. Ranks are calculated for all treatments included in the NMA (i.e., ranks may not begin at 1 and ranks for treatments not relevant to Norway are not shown).

NOTE: We have limited confidence in the specific order in which these treatments are ranked, and therefore caution against using this ranking uncritically.

Among treatment regimens relevant for Norway, the three highest ranked treatments for non-refractory patients for OS are [DR + d], [IR + d] and [DV + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 86%, 85% and 79%, respectively (*Table 8*).

For patients who are refractory to lenalidomide (R) (i.e., not responding to lenalidomide), the three highest ranked relevant treatment regimens are [DV + d], [EP + d] and [IsP + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 78%, 75% and 72%, respectively (*Table 8*).

For patients who are refractory to bortezomib (V) (i.e., not responding to bortezomib), the three highest ranked relevant treatment regimens are [DR + d], [IR + d] and [KR + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 84%, 83% and 75%, respectively (*Table 8*).

For patients who are refractory to both lenalidomide (R) and bortezomib (V) (i.e., not responding to either lenalidomide or bortezomib), the three highest ranked relevant treatment regimens are [EP + d], [IsP + d] and [DK + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 77%, 74% and 71%, respectively (*Table 8*).

The matrix plot (*Table 7*) presents the NMA effect estimates for OS, for all possible comparisons between the treatment regimens relevant to Norway. As shown for OS, none of the 105 NMA effect estimates were statistically significant (*Table 7*).

Judgements about the ranking of the various treatment regimens should not be made without also taking the certainty of the evidence into consideration. We were not able to assess the certainty of evidence for 54 of 105 of the comparisons between the relevant treatment regimens, because the GRADE method has not yet been extended to address disconnected networks (*Figure 7*). Of the 51 comparisons we could assess however, the majority was set as very low, mostly due to very wide 95% confidence intervals (*Table 7*). Details of our assessment of GRADE is presented in *Appendix 9 - Detailed GRADE: overall survival*. As such, for OS we cannot confidently say that one treatment regimen is better than another. In assessing the results of OS, it would be reasonable, if possible, to put more weight on the direct estimates from the individual studies, rather than on the NMA results.

Even though our results show that the treatment regimens [IR + d] and [DV + d] are ranked second and third highest for patients that are non-refractory to any drug with 85% and 79% probability of being the best, respectively, we have very low confidence in all their effect estimates (*Table 7*, *Appendix 9 - Detailed GRADE: overall survival*). The NMA effect estimate of the treatment regimen [IR + d] (compared with [R + d]), is based on two studies: one large international study (TOURMALINE-MM1) which showed little or no difference on OS between the two groups, and one very small regional (Chinese) study, which showed that the intervention group had a substantial improvement of OS (37, 38). As shown in *Figure 6,* the 95% confidence intervals of these two studies do not overlap. There are several factors that could explain this discrepancy, including genetic differences between the study populations, and differences in the follow up treatments after progression. This, in addition to inconsistent results in the two studies, and the wide 95% confidence interval in the direct NMA effect estimate, caused us to rate down the certainty of evidence to very low (*Appendix 9 - Detailed GRADE: overall survival*).

Similarly, the NMA effect estimate of the treatment regimen [DV + d] (compared with [V + d]), is also based on two studies: one large international study (CASTOR), and a smaller study conducted exclusively in China (LEPUS) (39, 60). The smaller LEPUS study showed a point estimate that indicated better effect on OS than the larger CASTOR study (*Figure 6*). Contrary to [IR + d] though, effect estimates of [DV + d] vs [V + d] were more consistent as the 95% confidence intervals overlap, suggesting that [DV + d] treatment in the two studies may have similar effect on OS (*Figure 6*). Similar to [IR + d], the effect discrepancies for [DV + d] in the LEPUS and CASTOR studies may be caused by several factors, including differences (genetic and others) between the study populations, as well as potential differences in the follow-up treatments after progression. Based on this, we rated down the certainty of evidence to very low (*Appendix 9 - Detailed GRADE: overall survival*).

Transitivity analysis for overall survival

In analyses performed to assess the transitivity assumption we found statistically significant ($P<0.05$) associations between treatment comparison and the following study and baseline patient characteristics: study setting, funding source, number of previous treatment lines, and the percentage of patients with stage I and III disease. However, there is substantial variability across all variables studied.

We judge that associations with study setting and funding source are unlikely to threaten the transitivity assumption. These associations are “significant” because there are a small number of specific treatment comparisons that were either not multinational studies or that were not industry-funded. I.e., while the p-value is significant, there is almost no variation with respect to treatment comparison.

Due to lack of data, it was not possible to perform these analyses for the percentage of patients who were Caucasian, who were refractory to immunomodulatory drugs, or had performance status ≥ 3. We do not have strong evidence that transitivity may be threatened with respect to these characteristics, but it is possible.

Of the characteristics studied, there is only convincing evidence of potentially important associations with respect to number of previous lines of treatment and disease stage, and these could in principle threaten transitivity. Our systematic review, meta-analysis, and simulation study (34) did not identify that these variables are important effect modifiers, though some uncertainty remains, and we did not find any evidence of inconsistency between direct, indirect, and NMA estimates. We therefore conclude that any violations of the assumptions underpinning the NMA do not manifest as detectable inconsistency and the estimates are probably trustworthy. Ideally, we would have also used loop-based methods to assess inconsistency, but this was not possible because there were no closed loops.

Sensitivity analysis — China Continuation Study

Figure 8 shows how P-scores differ when the China Continuation Study is removed from the NMA for OS. For non-refractory patients, the P-score for [IR + d] changes from 0.85 to 0.73. For patients who are refractory to V, the P-score for [IR + d] changes from 0.83 to 0.69. These differences result in a marked drop in rankings for [IR + d]; this is addressed further in *Discussion*.

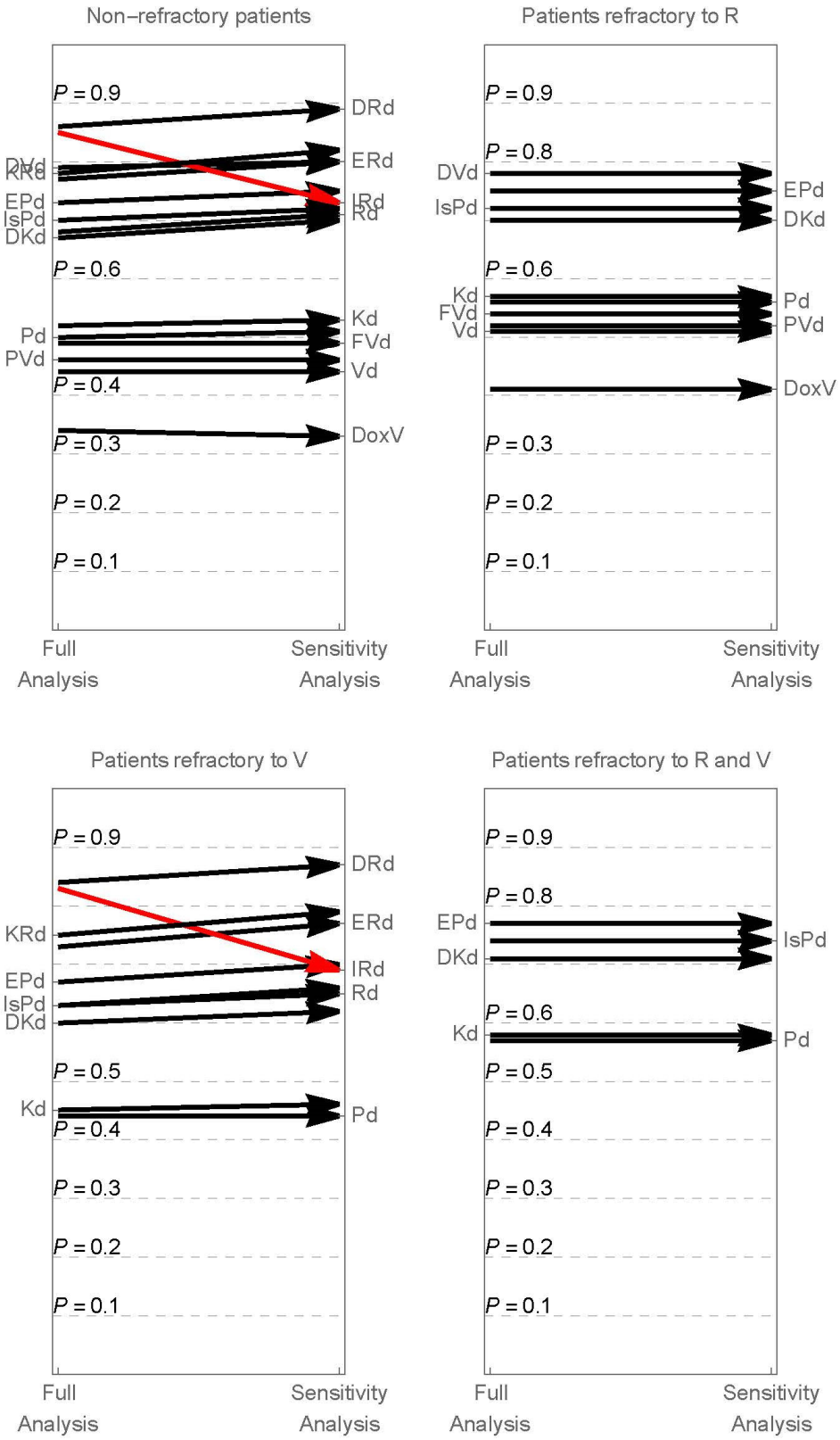


Figure 8: Sensitivity analysis results showing the effect of removing the China Continuation Study on P-scores for overall survival
D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, R: lenalidomide, V: bortezomib

Results – Quality of life

NMA results

The NMA of QoL resulted in three disconnected networks (*Figure 9*) with a total of 12 treatment regimens. A NMA was performed using data on the 12 treatments from nine RCTs that enrolled a total of 5063 patients (*Appendix 10 - Additional results: quality of life*).

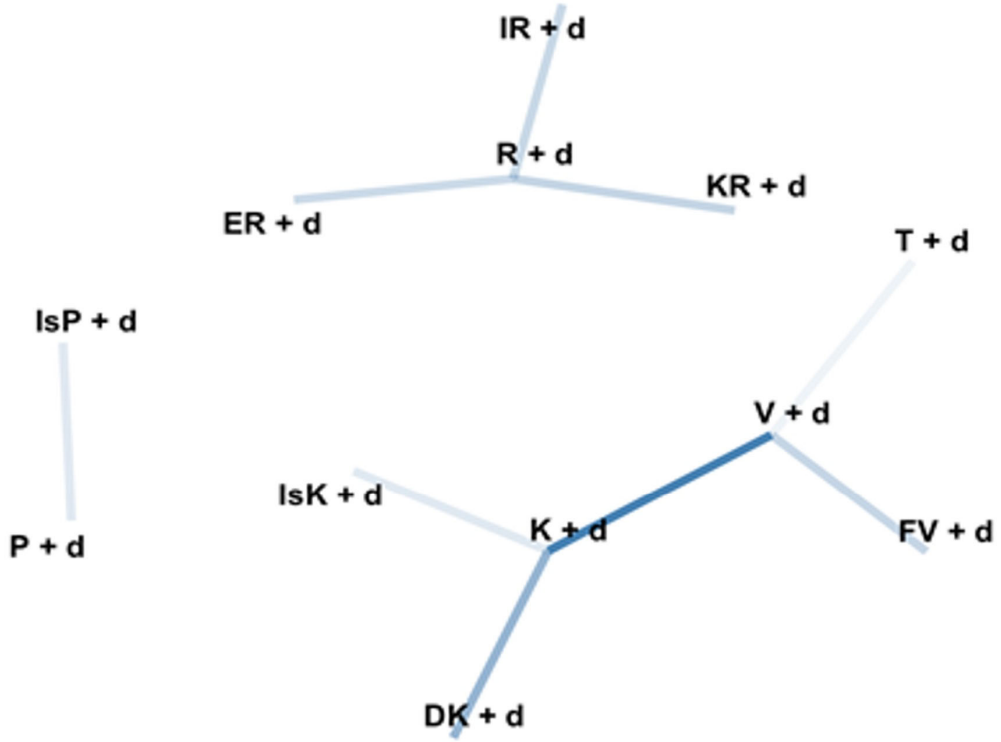


Figure 9: Network topology for quality of life
Each vertex represents a treatment, and each edge (line) represents a direct treatment comparison. More precise estimates (e.g., those supported by studies with larger sample sizes) are indicated by darker edges. D: daratumumab, d: dexamethasone, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, R: lenalidomide, T: thalidomide, V: bortezomib

We used random effects component NMA to estimate average difference in mean QLQ-C30 Global Health Status score and account for between-study heterogeneity. A positive difference indicates higher (better) QoL relative to a comparator. The variance component modelling heterogeneity, and hence I^2, could not be estimated, so the resulting model may misestimate the degree of uncertainty that would exist if this could have been estimated. The matrix plot (*Table 9*) presents estimates of average differences and 95% confidence intervals, in addition to our assessment of the certainty of evidence, for treatment regimens relevant for Norway.

The summary of findings table (*Table 10*) presents treatment regimens relevant for Norway ranked by P-score for non-refractory patients, along with assessments of certainty of evidence (GRADE). It also presents ranked treatments for patients who are refractory to lenalidomide and/or bortezomib.

Table 9: Matrix plot - quality of life

Treatment A	Treatment B: DK + d	ER + d	FV + d	IR + d	IsK + d	IsP + d	K + d	KR + d	P + d	V + d
R + d	-3.31 (-6.80 – 0.19) *NP*	-0.46 (-4.83 – 3.91) *High*	-0.02 (-4.37 – 4.33) *NP*	-1.90 (-6.03 – 2.23) *High*	0.30 (-5.73 – 6.33) *NP*	-2.79 (-6.43 – 0.85) *NP*	-3.25 (-6.12 – -0.38) *NP*	-4.81 (-8.75 – -0.87) *Moderate*	-0.79 (-4.43 – 2.85) *NP*	0.26 (-2.61 – 3.13) *NP*
DK + d		2.85 (-1.81 – 7.51) *NP*	3.29 (-1.65 – 8.23) *Moderate*	1.41 (-3.14 – 5.96) *NP*	3.61 (-3.33 –10.55) *Low*	0.52 (-3.84 – 4.88) *NP*	0.06 (-2.38 – 2.50) *High*	-1.50 (-5.97 – 2.97) *NP*	2.52 (-1.84 – 6.88) *NP*	3.57 (0.68 – 6.46) *High*
ER + d			0.44 (-4.89 – 5.77) *NP*	-1.44 (-7.45 – 4.57) *High*	0.76 (-6.02 – 7.54) *NP*	-2.33 (-7.11 – 2.44) *NP*	-2.79 (-7.00 – 1.42) *NP*	-4.35 (-10.23 – 1.53) *Low*	-0.33 (-5.11 – 4.44) *NP*	0.72 (-3.49 – 4.93) *NP*
FV + d				-1.88 (-7.12 – 3.36) *NP*	0.32 (-7.47 – 8.11) *Low*	-2.77 (-7.84 – 2.30) *NP*	-3.23 (-7.52 – 1.06) *High*	-4.79 (-9.96 – 0.38) *NP*	-0.77 (-5.84 – 4.30) *NP*	0.28 (-3.72 – 4.28) *High*
IR + d					2.20 (-4.50 – 8.90) *NP*	-0.89 (-5.56 – 3.78) *NP*	-1.35 (-5.44 – 2.74) *NP*	-2.91 (-8.62 – 2.80) *Low*	1.11 (-3.56 – 5.78) *NP*	2.16 (-1.93 – 6.25) *NP*
IsK + d						-3.09 (-9.66 – 3.48) *NP*	-3.55 (-10.05 – 2.95) *Moderate*	-5.11 (-11.76 – 1.53) *NP*	-1.09 (-7.66 – 5.48) *NP*	-0.04 (-6.72 – 6.64) *Moderate*
IsP + d							-0.46 (-4.33 – 3.42) *NP*	-2.02 (-6.60 – 2.57) *NP*	2.00 (-4.33 – 8.33) *Low*	3.05 (-0.82 – 6.93) *NP*
K + d								-1.56 (-5.56 – 2.44) *NP*	2.46 (-1.42 – 6.33) *NP*	3.51 (1.96 – 5.05) *High*
KR + d									4.02 (-0.57 – 8.60) *NP*	5.07 (1.07 – 9.07) *NP*
P + d										1.05 (-2.82 – 4.93) *NP*

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib.

NP: not possible to GRADE because the comparison is across networks. Statistically significant effect estimates are highlighted in either pink (favours treatment A) or blue (favours treatment B).
Only listed treatment regimens relevant for Norway

Table 10: Summary of findings for difference in mean QLQ-C30 Global Health Status score

Treatments are ordered by overall rank relevant for non-refractory patients

Treatment*	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant to people who are:		
		Refractory to R	Refractory to V	Refractory to R & V
KR + d	1 (0.86)	NA	1 (0.81)	NA
K + d	3 (0.73)	2 (0.71)	3 (0.66)	3 (0.63)
DK + d	4 (0.72)	3 (0.71)	4 (0.66)	2 (0.63)
IsP + d	5 (0.65)	4 (0.64)	5 (0.59)	4 (0.55)
IR + d	6 (0.53)	NA	6 (0.48)	NA
P + d	7 (0.38)	5 (0.38)	7 (0.33)	5 (0.27)
ER + d	8 (0.34)	NA	8 (0.30)	NA
IsK + d	9 (0.29)	7 (0.29)	9 (0.25)	6 (0.21)
FV + d	10 (0.29)	6 (0.29)	NA	NA
R + d	11 (0.25)	NA	10 (0.19)	NA
V + d	12 (0.22)	8 (0.23)	NA	NA

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, NA: not applicable, P: pomalidomide, R: lenalidomide, V: bortezomib

** The Methods chapter describes how treatments are defined.*

§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment.

Only listed treatment regimens relevant for Norway. Ranks are calculated for all treatments included in the NMA (i.e., ranks may not begin at 1 and ranks for treatments not relevant to Norway are not shown).

NOTE: We have limited confidence in the specific order in which these treatments are ranked, and therefore caution against using this ranking uncritically.

Among treatment regimens relevant for Norway, the three highest ranked treatments for non-refractory patients for QoL are [KR + d], [K + d] and [DK + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 86%, 73% and 72%, respectively (*Table 10*).

For patients who are refractory to lenalidomide (R) (i.e., not responding to lenalidomide), the three highest ranked relevant treatment regimens are [K + d], [DK + d] and [IsP + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 71%, 71% and 64%, respectively (*Table 10*).

For patients who are refractory to bortezomib (V) (i.e., not responding to lenalidomide), the three highest ranked relevant treatment regimens are [KR + d], [K + d] and [DK + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 81%, 66% and 66%, respectively (*Table 10*).

For patients who are refractory to both lenalidomide (R) and bortezomib (V) (i.e., not responding to either lenalidomide or bortezomib), the three highest ranked relevant treatment regimens are [DK + d], [K + d] and [IsP + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 63%, 63% and 55%, respectively (*Table 10*).

The matrix plot (*Table 9*) presents the NMA effect estimates for QoL, for all possible comparisons between the treatment regimens relevant to Norway. For QoL, only five of 55 NMA effect estimates were statistically significant (*Table 9*):

- [K + d] vs. [R + d] – indirect comparison across networks
- [K + d] vs. [V + d] – direct comparison
- [KR + d] vs. [R + d] – direct comparison

- [KR + d] vs. [V + d] – indirect comparison across networks
- [DK + d] vs. [V + d] – indirect comparison within the same network

Judgements about the ranking of the various treatment regimens should not be made without also taking the certainty of the evidence into consideration. We were not able to assess the certainty of evidence for 38 of 55 of the comparisons between the relevant treatment regimens, because the GRADE method has not yet been extended address disconnected networks (*Figure 9*). However, the 17 NMA effect estimates that we could assess, were set between high and low certainty of evidence (*Table 9*). For the five significant NMA estimates, we could only assess certainty of evidence for three (*Table 9*):

- [K + d] vs. [V + d]: high certainty of evidence
- [DK + d] vs. [V + d]: high certainty of evidence
- [KR + d] vs. [R + d]: moderate certainty of evidence

Details of our assessment of GRADE is presented in *Appendix - Detailed GRADE: quality of life*.

Transitivity analysis for quality of life

We were unable to perform the planned analyses of study and baseline patient characteristics, or the analysis of inconsistency between the direct, indirect, and network evidence, for this outcome due to insufficient data. We therefore rely on a conceptual evaluation of the validity of the transitivity assumption.

Results – Severe adverse events

NMA results

The NMA of SAE resulted in three disconnected networks (*Figure 10*) with a total of 30 treatment regimens. A NMA was performed using data on the 30 treatments from 31 RCTs that enrolled a total of 12 124 patients (*Appendix 10 - Additional results: severe adverse events*).

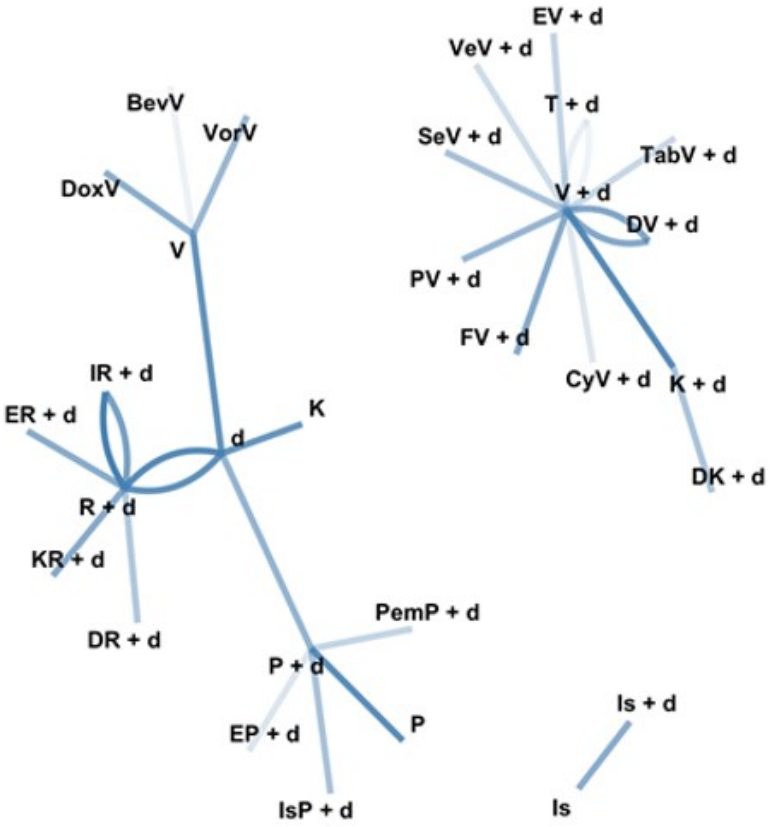


Figure 10: Network topology for severe adverse events
Each vertex represents a treatment, and each edge (line) represents a direct treatment comparison. More precise estimates (e.g., those supported by studies with larger sample sizes) are indicated by darker edges. Bev: bevacizumab, Cy: cyclophosphamide, d: dexamethasone, D: daratumumab, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, Pem: pembrolizumab, R: lenalidomide, Se: selinexor, Tab: tabalumab, V: bortezomib, Ve: venetoclax, Vor: vorinostat

We used random effects component NMA to estimate average incidence rate ratios (IRRs) and account for between-study heterogeneity. An average treatment effect estimate should not be misinterpreted as being equivalent to an estimate from a single study (which cannot account for between-study heterogeneity). Consequently, confidence intervals on average treatment effect estimates may be wider that those arising from individual studies. The matrix plot (*Table 11*) presents estimates of IRRs and 95% confidence intervals, in addition to our assessment of the certainty of evidence, for treatment regimens relevant for Norway. Heterogeneity (I^2) was estimated to be 4.3% (95% CI 0% to 72.1%).

The summary of findings table (*Table 12*) presents treatment regimens relevant for Norway ranked by P-score for non-refractory patients, along with assessments of certainty of evidence (GRADE). It also presents ranked treatments for patients who are refractory to lenalidomide and/or bortezomib.

Table 11: Matrix plot - severe adverse events

Treatment A	Treatment B													
	DK + d	**DoxV**	**DR + d**	**DV + d**	**EP + d**	**ER + d**	**FV + d**	**IR + d**	**IsP + d**	**K + d**	**KR + d**	**P + d**	**PV + d**	**V + d**
R + d	0.92 (0.32 – 2.70) *NP*	1.23 (0.50 – 3.06) *Very low*	0.50 (0.23 – 1.09) *Moderate*	0.70 (0.29 – 1.68) *NP*	1.98 (0.54 – 7.20) *Very low*	0.70 (0.36 – 1.36) *Moderate*	0.99 (0.37 – 2.67) *NP*	0.56 (0.38 – 0.84) *Low*	1.21 (0.40 – 3.63) *Very low*	1.94 (0.92 – 4.12) *NP*	0.63 (0.33 – 1.20) *Moderate*	2.35 (1.06 – 5.21) *Very low*	1.08 (0.39 – 3.03) *NP*	1.92 (0.91 – 4.07) *NP*
DK + d		1.34 (0.41 – 4.40) *NP*	0.54 (0.14 – 2.03) *NP*	0.76 (0.28 – 2.06) *Very low*	2.14 (0.44 –10.41) *NP*	0.76 (0.22 – 2.68) *NP*	1.07 (0.35 – 3.23) *Very low*	0.61 (0.19 – 1.91) *NP*	1.31 (0.31 – 5.46) *NP*	2.10 (0.98 – 4.51) *Moderate*	0.68 (0.19 – 2.38) *NP*	2.54 (0.76 – 8.54) *NP*	1.17 (0.38 – 3.65) *Very low*	2.08 (0.85 – 5.08) *Low*
DoxV			0.40 (0.12 – 1.33) *Very low*	0.57 (0.23 – 1.41) *NP*	1.60 (0.37 – 7.01) *Very low*	0.57 (0.19 – 1.75) *Very low*	0.80 (0.29 – 2.23) *NP*	0.46 (0.17 – 1.23) *Very low*	0.98 (0.26 – 3.63) *Very low*	1.57 (0.63 – 3.94) *NP*	0.51 (0.17 – 1.55) *Very low*	1.90 (0.65 – 5.54) *Very low*	0.88 (0.31 – 2.53) *NP*	1.56 (0.71 – 3.44) *NP*
DR + d				1.41 (0.43 – 4.56) *NP*	3.98 (0.88 –18.05) *Very low*	1.42 (0.51 – 3.94) *Very low*	1.99 (0.56 – 7.05) *NP*	1.13 (0.47 – 2.73) *Very low*	2.44 (0.63 – 9.39) *Very low*	3.91 (1.32 –11.58) *NP*	1.26 (0.46 – 3.50) *Very low*	4.73 (1.55 –14.47) *Low*	2.19 (0.60 – 7.95) *NP*	3.87 (1.31 –11.45) *NP*
DV + d					2.83 (0.66 –12.16) *NP*	1.01 (0.34 – 3.01) *NP*	1.41 (0.64 – 3.13) *Very low*	0.81 (0.31 – 2.11) *NP*	1.73 (0.48 – 6.28) *NP*	2.78 (1.45 – 5.32) *Low*	0.90 (0.30 – 2.68) *NP*	3.36 (1.19 – 9.55) *NP*	1.55 (0.67 – 3.58) *Very low*	2.75 (1.75 – 4.34) *Low*
EP + d						0.36 (0.08 – 1.52) *Very low*	0.50 (0.11 – 2.31) *NP*	0.28 (0.07 – 1.10) *Very low*	0.61 (0.17 – 2.18) *Very low*	0.98 (0.25 – 3.93) *NP*	0.32 (0.07 – 1.35) *Very low*	1.19 (0.43 – 3.29) *Low*	0.55 (0.12 – 2.59) *NP*	0.97 (0.24 – 3.89) *NP*
ER + d							1.40 (0.43 – 4.62) *NP*	0.80 (0.37 – 1.73) *Very low*	1.72 (0.48 – 6.19) *Very low*	2.76 (1.02 – 7.50) *NP*	0.89 (0.35 – 2.25) *Very low*	3.34 (1.19 – 9.39) *Low*	1.54 (0.46 – 5.22) *NP*	2.73 (1.01 – 7.42) *NP*
FV + d								0.57 (0.19 – 1.66) *NP*	1.23 (0.31 – 4.82) *NP*	1.97 (0.89 – 4.37) *Low*	0.63 (0.19 – 2.08) *NP*	2.38 (0.76 – 7.46) *NP*	1.10 (0.42 – 2.86) *Very low*	1.95 (1.02 – 3.73) *High*
IR + d									2.15 (0.67 – 6.93) *Very low*	3.45 (1.47 – 8.11) *NP*	1.11 (0.52 – 2.40) *Very low*	4.18 (1.71 –10.21) *Very low*	1.93 (0.64 – 5.81) *NP*	3.42 (1.46 – 8.02) *NP*
IsP + d										1.60 (0.48 – 5.36) *NP*	0.52 (0.14 – 1.86) *Very low*	1.94 (0.91 – 4.13) *Very low*	0.90 (0.22 – 3.61) *NP*	1.59 (0.48 – 5.30) *NP*
K + d											0.32 (0.12 – 0.87) *NP*	1.21 (0.47 – 3.10) *NP*	0.56 (0.24 – 1.29) *Low*	0.99 (0.62 – 1.57) *High*

		Treatment B													
		DK + d	DoxV	DR + d	DV + d	EP + d	ER + d	FV + d	IR + d	IsP + d	K + d	KR + d	P + d	PV + d	V + d
	KR + d												3.75 (1.34 –10.49) *Low*	1.73 (0.51 – 5.84) *NP*	3.07 (1.14 – 8.28) *NP*
	P + d													0.46 (0.14 – 1.49) *NP*	0.82 (0.32 – 2.09) *NP*
	PV + d														1.77 (0.88 – 3.57) *Moderate*

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib.
NP: not possible to GRADE because the comparison is across networks. Statistically significant effect estimates are highlighted in either pink (favours treatment A) or blue (favours treatment B).
Only listed treatment regimens relevant for Norway

Table 12: Summary of findings for severe adverse events
Treatments are ordered by overall rank relevant for non-refractory patients

Treatment*	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant to people who are:		
		Refractory to R	Refractory to V	Refractory to R & V
P + d	8 (0.69)	8 (0.62)	7 (0.61)	7 (0.43)
V + d	11 (0.62)	13 (0.53)	NA	NA
K + d	12 (0.62)	12 (0.53)	10 (0.54)	10 (0.33)
EP + d	13 (0.60)	11 (0.54)	9 (0.55)	8 (0.37)
DoxV	17 (0.41)	18 (0.33)	NA	NA
IsP + d	18 (0.40)	17 (0.33)	11 (0.37)	11 (0.17)
PV + d	20 (0.36)	20 (0.29)	NA	NA
R + d	22 (0.33)	NA	12 (0.31)	NA
FV + d	23 (0.32)	22 (0.25)	NA	NA
DK + d	24 (0.29)	23 (0.23)	13 (0.28)	12 (0.09)
ER + d	26 (0.19)	NA	14 (0.18)	NA
DV + d	27 (0.17)	25 (0.13)	NA	NA
KR + d	28 (0.15)	NA	15 (0.14)	NA
IR + d	29 (0.11)	NA	16 (0.10)	NA
DR + d	30 (0.09)	NA	17 (0.08)	NA

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, NA: not applicable, R: lenalidomide, V: bortezomib
** The Methods chapter describes how treatments are defined.*
§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Only listed treatment regimens relevant for Norway. Ranks are calculated for all treatments included in the NMA (i.e., ranks may not begin at 1 and ranks for treatments not relevant to Norway are not shown).

NOTE: We have limited confidence in the specific order in which these treatments are ranked, and therefore caution against using this ranking uncritically.

Perhaps unsurprisingly, the treatment regimens ranked highest in terms of OS, were ranked lowest for SAE, and vice versa. That is to say the more "aggressive" treatments appear to be associated with higher rates of SAEs.

Among treatment regimens relevant for Norway, the three highest ranked treatments for non-refractory patients for SAE are [P + d], [V + d] and [K + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 69%, 62% and 62%, respectively (*Table 12*). These treatment regimens were among the lowest ranked treatment regimens in terms of OS.

For patients who are refractory to lenalidomide (R) (i.e., not responding to lenalidomide), the three highest ranked relevant treatment regimens are [P + d], [V + d] and [K + d]. The estimated probabilities that these are the best of all analysed treatment regimens are 62%, 53% and 53%, respectively (*Table 12*).

For patients who are refractory to bortezomib (V) (i.e., not responding to lenalidomide), the three highest ranked relevant treatment regimens are [P + d], [EP + d] and [K + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are 61%, 55% and 54%, respectively (*Table 12*).

For patients who are refractory to both lenalidomide (R) and bortezomib (V) (i.e., not responding to either lenalidomide or bortezomib), the three highest ranked relevant treatment regimens are [P + d], [EP + d] and [K + d]. The estimated probabilities that

these are the best of all analysed treatment regimens, are 43%, 37% and 33%, respectively (*Table 12*).

The matrix plot (*Table 11*) presents the NMA effect estimates for SAE, for all possible comparisons between the treatment regimens relevant to Norway. For SAE, only 18 of 105 NMA effect estimates were statistically significant (*Table 11*).

Judgements about the ranking of the various treatment regimens should not be made without also taking the certainty of the evidence into consideration. We were not able to assess the certainty of evidence for 54 of 105 of the comparisons between the relevant treatment regimens, because the GRADE method has not yet been extended address disconnected networks (*Figure 10*). However, the 51 NMA effect estimates that we could assess, were set between high and very low certainty of evidence, with the majority set as low or very low (*Table 11*). Of the 18 significant NMA estimates, we could only assess certainty of evidence for nine NMA estimates, in which one were set as high, six were set as low, and two were set as very low (*Table 11*). Details of our assessment of GRADE is presented in *Appendix 9 - Detailed GRADE: severe adverse events.*

Transitivity analysis for severe adverse events

Results of analyses performed to assess the transitivity assumption were consistent with those for OS.

Direct evidence

Evidence extracted directly from the included studies of HRs with 95% confidence intervals for PFS for treatment regimens relevant for Norway, ,was set into a forest plot, along with assessments of certainty of evidence (GRADE), and study name (*Figure 11*). Within the forest plot, the direct evidence is organized after the common treatments (i.e., treatment B), such as [K + d], [P + d], etc.

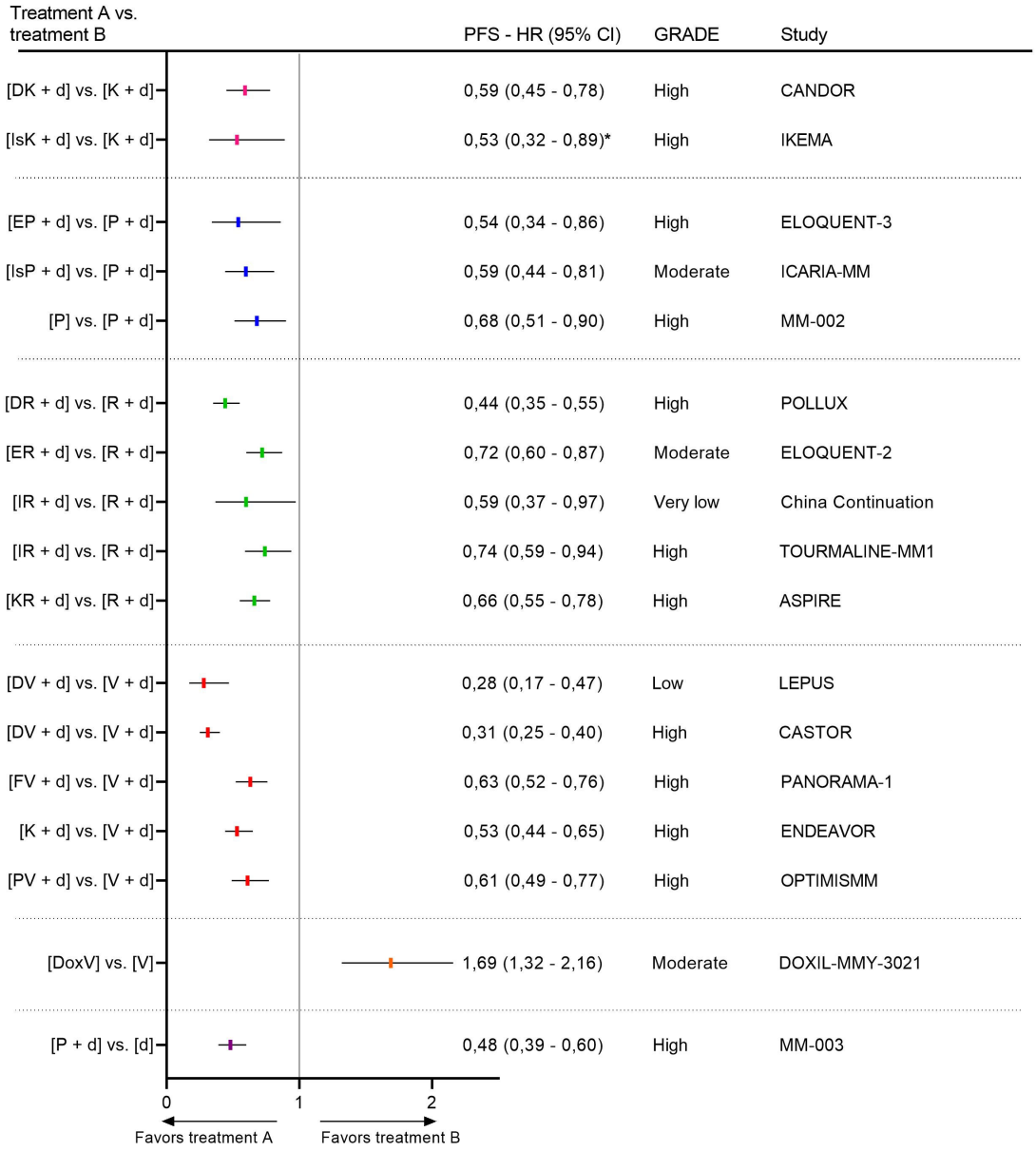


Figure 11: Forest plot of direct evidence – progression-free survival
** 95,4% confidence interval. Only shown treatment regimens that are relevant for Norway. CI: confidence interval, D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, HR: hazard ratio, I: ixazomib, Is: isatuximab, K: carfilzomib, PFS: progression-free survival, P: pomalidomide, R: lenalidomide, V: bortezomib. HR<1 favours treatment A, HR>1 favours treatment B*

NMA results

The NMA of PFS resulted in five disconnected networks (*Figure 12*) with a total of 35 treatment regimens. A NMA was performed using data on the 35 treatments from 32 RCTs that enrolled a total of 11 936 patients (*Appendix 10 - Additional results: progression-free survival*).

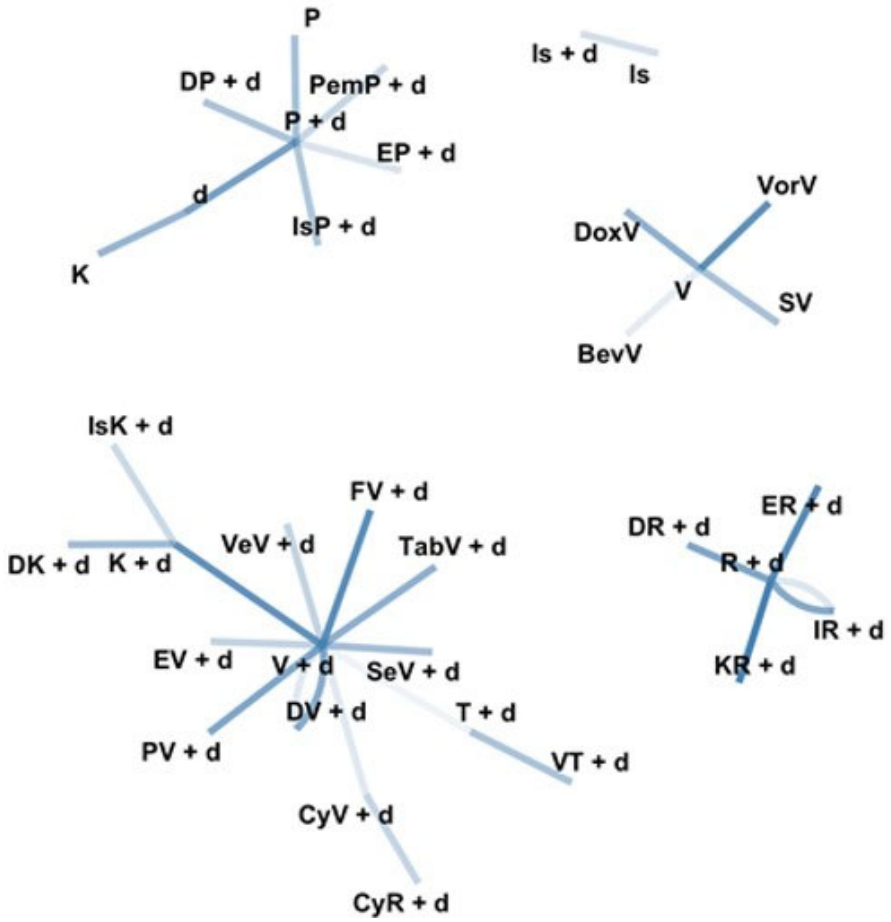


Figure 12: Network topology for progression-free survival
Each vertex represents a treatment, and each edge (line) represents a direct treatment comparison. More precise estimates (e.g., those supported by studies with larger sample sizes) are indicated by darker edges. Bev: bevacizumab, Cy: cyclophosphamide, d: dexamethasone, D: daratumumab, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, Pem: pembrolizumab, R: lenalidomide, Se: selinexor, T: thalidomide, Tab: tabalumab, V. bortezomib, Ve: venetoclax, Vor: vorinostat

We used random effects component NMA to estimate average HRs and account for between-study heterogeneity. The variance component was estimated to be zero, so the resulting model is equivalent to a fixed (common) effects model and may underestimate the degree of uncertainty that would exist if more studies could have been included. The matrix plot (*Table 13*) presents estimates of HRs and 95% confidence intervals, in addition to our assessment of the certainty of evidence, for treatment regimens relevant for Norway. Heterogeneity (I^2) was estimated to be 0% (95% CI 0% to 36.7%).

The summary of findings table (*Table 14*) presents treatment regimens relevant for Norway ranked by P-score for non-refractory patients, along with assessments of certainty of evidence (GRADE). It also presents ranked treatments for patients who are refractory to lenalidomide and/or bortezomib.

Table 13: Matrix plot - progression-free survival

Treatment A	Treatment B														
	DK + d	**DoxV**	**DR + d**	**DV + d**	**EP + d**	**ER + d**	**FV + d**	**IR + d**	**IsK + d**	**IsP + d**	**K + d**	**KR + d**	**P + d**	**PV + d**	**V + d**
R + d	3.35 (2.13 – 5.26) *NP*	1.20 (0.79 – 1.82) *NP*	2.27 (1.81 – 2.85) *High*	3.44 (2.17 – 5.45) *NP*	5.65 (3.37 – 9.48) *NP*	1.39 (1.15 – 1.67) *High*	1.66 (1.06 – 2.61) *NP*	1.41 (1.14 – 1.73) *Low*	3.73 (2.20 – 6.33) *NP*	5.12 (3.49 – 7.50) *NP*	1.98 (1.38 – 2.83) *NP*	1.52 (1.27 – 1.80) *High*	3.05 (2.42 – 3.84) *NP*	1.72 (1.08 – 2.74) *NP*	1.05 (0.70 – 1.58) *NP*
DK + d		0.36 (0.22 – 0.58) *NP*	0.68 (0.42 – 1.10) *NP*	1.03 (0.69 – 1.53) *Very low*	1.69 (0.86 – 3.32) *NP*	0.41 (0.26 – 0.67) *NP*	0.50 (0.34 – 0.73) *High*	0.42 (0.26 – 0.68) *NP*	1.11 (0.69 – 1.79) *Low*	1.53 (0.85 – 2.73) *NP*	0.59 (0.45 – 0.78) *High*	0.45 (0.28 – 0.73) *NP*	0.91 (0.56 – 1.49) *NP*	0.51 (0.34 – 0.77) *High*	0.31 (0.22 – 0.44) *High*
DoxV			1.89 (1.21 – 2.96) *NP*	2.86 (1.91 – 4.28) *NP*	4.70 (2.45 – 9.03) *NP*	1.16 (0.75 – 1.79) *NP*	1.38 (0.94 – 2.04) *NP*	1.17 (0.75 – 1.82) *NP*	3.10 (1.78 – 5.40) *NP*	4.26 (2.45 – 7.39) *NP*	1.64 (1.11 – 2.44) *NP*	1.26 (0.82 – 1.95) *NP*	2.54 (1.60 – 4.02) *NP*	1.43 (0.95 – 2.15) *NP*	0.87 (0.62 – 1.23) *NP*
DR + d				1.51 (0.92 – 2.48) *NP*	2.49 (1.44 – 4.29) *NP*	0.61 (0.46 – 0.82) *High*	0.73 (0.45 – 1.19) *NP*	0.62 (0.45 – 0.84) *Low*	1.64 (0.94 – 2.86) *NP*	2.25 (1.48 – 3.43) *NP*	0.87 (0.58 – 1.30) *NP*	0.67 (0.50 – 0.89) *High*	1.34 (1.01 – 1.79) *NP*	0.76 (0.46 – 1.24) *NP*	0.46 (0.30 – 0.72) *NP*
DV + d					1.64 (0.83 – 3.25) *NP*	0.40 (0.25 – 0.65) *NP*	0.48 (0.36 – 0.64) *Low*	0.41 (0.25 – 0.67) *NP*	1.08 (0.67 – 1.76) *Very low*	1.49 (0.83 – 2.68) *NP*	0.57 (0.43 – 0.77) *Low*	0.44 (0.27 – 0.71) *NP*	0.89 (0.54 – 1.47) *NP*	0.50 (0.37 – 0.68) *Low*	0.30 (0.25 – 0.38) *Low*
EP + d						0.25 (0.14 – 0.42) *NP*	0.29 (0.15 – 0.58) *NP*	0.25 (0.14 – 0.43) *NP*	0.66 (0.32 – 1.37) *NP*	0.91 (0.52 – 1.58) *Very low*	0.35 (0.19 – 0.65) *NP*	0.27 (0.16 – 0.46) *NP*	0.54 (0.34 – 0.86) *High*	0.30 (0.15 – 0.60) *NP*	0.19 (0.10 – 0.36) *NP*
ER + d							1.20 (0.75 – 1.92) *NP*	1.01 (0.76 – 1.34) *Very low*	2.68 (1.55 – 4.65) *NP*	3.68 (2.45 – 5.54) *NP*	1.42 (0.97 – 2.09) *NP*	1.09 (0.85 – 1.41) *Moderate*	2.20 (1.67 – 2.88) *NP*	1.24 (0.76 – 2.01) *NP*	0.75 (0.49 – 1.16) *NP*
FV + d								0.85 (0.52 – 1.37) *NP*	2.24 (1.39 – 3.61) *High*	3.08 (1.72 – 5.50) *NP*	1.19 (0.91 – 1.56) *Moderate*	0.91 (0.57 – 1.46) *NP*	1.83 (1.12 – 3.00) *NP*	1.03 (0.77 – 1.39) *Low*	0.63 (0.52 – 0.76) *High*
IR + d									2.65 (1.52 – 4.61) *NP*	3.64 (2.40 – 5.51) *NP*	1.41 (0.95 – 2.08) *NP*	1.08 (0.82 – 1.42) *Very low*	2.17 (1.64 – 2.87) *NP*	1.22 (0.74 – 2.00) *NP*	0.74 (0.48 – 1.16) *NP*
IsK + d										1.37 (0.72 – 2.61) *NP*	0.53 (0.36 – 0.78) *High*	0.41 (0.24 – 0.70) *NP*	0.82 (0.46 – 1.44) *NP*	0.46 (0.28 – 0.75) *High*	0.28 (0.18 – 0.43) *NP*
IsP + d											0.39 (0.23 – 0.64) *NP*	0.30 (0.20 – 0.44) *NP*	0.60 (0.44 – 0.81) *Moderate*	0.34 (0.19 – 0.61) *NP*	0.20 (0.12 – 0.35) *NP*

		Treatment B														
		DK + d	DoxV	DR + d	DV + d	EP + d	ER + d	FV + d	IR + d	IsK + d	IsP + d	K + d	KR + d	P + d	PV + d	V + d
	K + d												0.77 (0.52 – 1.13) *NP*	1.54 (1.02 – 2.33) *NP*	0.87 (0.64 – 1.17) *Moderate*	0.53 (0.44 – 0.64) *High*
	KR + d													2.01 (1.54 – 2.63) *NP*	1.13 (0.70 – 1.84) *NP*	0.69 (0.45 – 1.06) *NP*
	P + d														0.56 (0.34 – 0.94) *NP*	0.34 (0.22 – 0.54) *NP*
	PV + d															0.61 (0.49 – 0.76) *High*

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib.

NP: not possible to GRADE because the comparison is across networks. Statistically significant effect estimates are highlighted in either pink (favours treatment A) or blue (favours treatment B).

Only listed treatment regimens relevant for Norway

Table 14: Summary of findings for progression-free survival
Treatments are ordered by overall rank relevant for non-refractory patients

Treatment*	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant to people who are:		
		Refractory to R	Refractory to V	Refractory to R & V
EP + d	1 (0.97)	1 (0.96)	1 (0.95)	1 (0.93)
IsP + d	2 (0.96)	2 (0.95)	2 (0.93)	2 (0.90)
IsK + d	4 (0.90)	4 (0.87)	4 (0.83)	4 (0.75)
DV + d	5 (0.87)	5 (0.85)	NA	NA
DK + d	6 (0.87)	6 (0.84)	5 (0.79)	5 (0.70)
P + d	7 (0.85)	7 (0.81)	6 (0.76)	6 (0.66)
DR + d	8 (0.76)	NA	7 (0.64)	NA
K + d	9 (0.71)	9 (0.66)	9 (0.57)	9 (0.45)
PV + d	12 (0.62)	11 (0.58)	NA	NA
FV + d	13 (0.60)	12 (0.56)	NA	NA
KR + d	15 (0.54)	NA	11 (0.40)	NA
IR + d	19 (0.47)	NA	14 (0.33)	NA
ER + d	21 (0.46)	NA	15 (0.31)	NA
DoxV	23 (0.36)	19 (0.36)	NA	NA
R + d	28 (0.23)	NA	18 (0.11)	NA

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, NA: not applicable, P: pomalidomide, R: lenalidomide, V: bortezomib
** The Methods chapter describes how treatments are defined.*
§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Only listed treatment regimens relevant for Norway. Ranks are calculated for all treatments included in the NMA (i.e., ranks may not begin at 1 and ranks for treatments not relevant to Norway are not shown).

NOTE: We have limited confidence in the specific order in which these treatments are ranked, and therefore caution against using this ranking uncritically.

Among treatment regimens relevant for Norway, the three highest ranked treatments for PFS are [EP + d], [IsP + d] and [IsK + d], regardless of whether patients are refractory to lenalidomide (R), bortezomib (V), or both. The estimated probabilities that these are the best of all analysed treatment regimens, are as follows (*Table 14*):

- For non-refractory patients: 97%, 96% and 90%, respectively
- For patients who are refractory to lenalidomide (R) (i.e., not responding to lenalidomide): 96%, 95% and 87%, respectively
- For patients who are refractory to bortezomib (V) (i.e., not responding to bortezomib): 96%, 93% and 83%, respectively
- For patients who are refractory to both lenalidomide (R) and bortezomib (V) (i.e., not responding to either lenalidomide or bortezomib): 93%, 90% and 75%, respectively

The matrix plot (*Table 13*Table 11) presents the NMA effect estimates for PFS, for all possible comparisons between the treatment regimens relevant to Norway. For PFS, 74 of 120 NMA effect estimates were statistically significant (*Table 13*).

Judgements about the ranking of the various treatment regimens should not be made without also taking the certainty of the evidence into consideration. We were not able to assess the certainty of evidence for 85 of 105 of the comparisons between the relevant treatment regimens, because the GRADE method has not yet been extended to

address disconnected networks (*Figure 12*). However, the 35 NMA effect estimates that we could assess, were set between high and very low certainty of evidence (*Table 13*). Of the 74 significant NMA estimates, we could only assess certainty of evidence for 23 NMA estimates, in which 16 were set as high, one was set as moderate, and six were set as low certainty of evidence (*Table 13*). Details of our assessment of GRADE is presented in *Appendix 9 - Detailed GRADE: progression-free survival.*

Transitivity analysis for progression-free survival

Results of analyses performed to assess the transitivity assumption were consistent with those for OS. There is no evidence of inconsistency between direct, indirect, and NMA estimates.

Sensitivity analysis — China Continuation Study

Figure 13 shows how P-scores differ when the China Continuation Study is removed from the NMA for PFS. For non-refractory patients, the P-score for [IR + d] changes from 0.47 to 0.44. For patients who are refractory to V, the P-score for [IR + d] changes from 0.33 to 0.30. These differences change the rankings for [IR + d] and [ER + d], but the P-scores are so similar for these two treatments in the full and sensitivity analyses that it would be unwise to conclude that one of the treatments is definitively better than the other with respect to PFS.

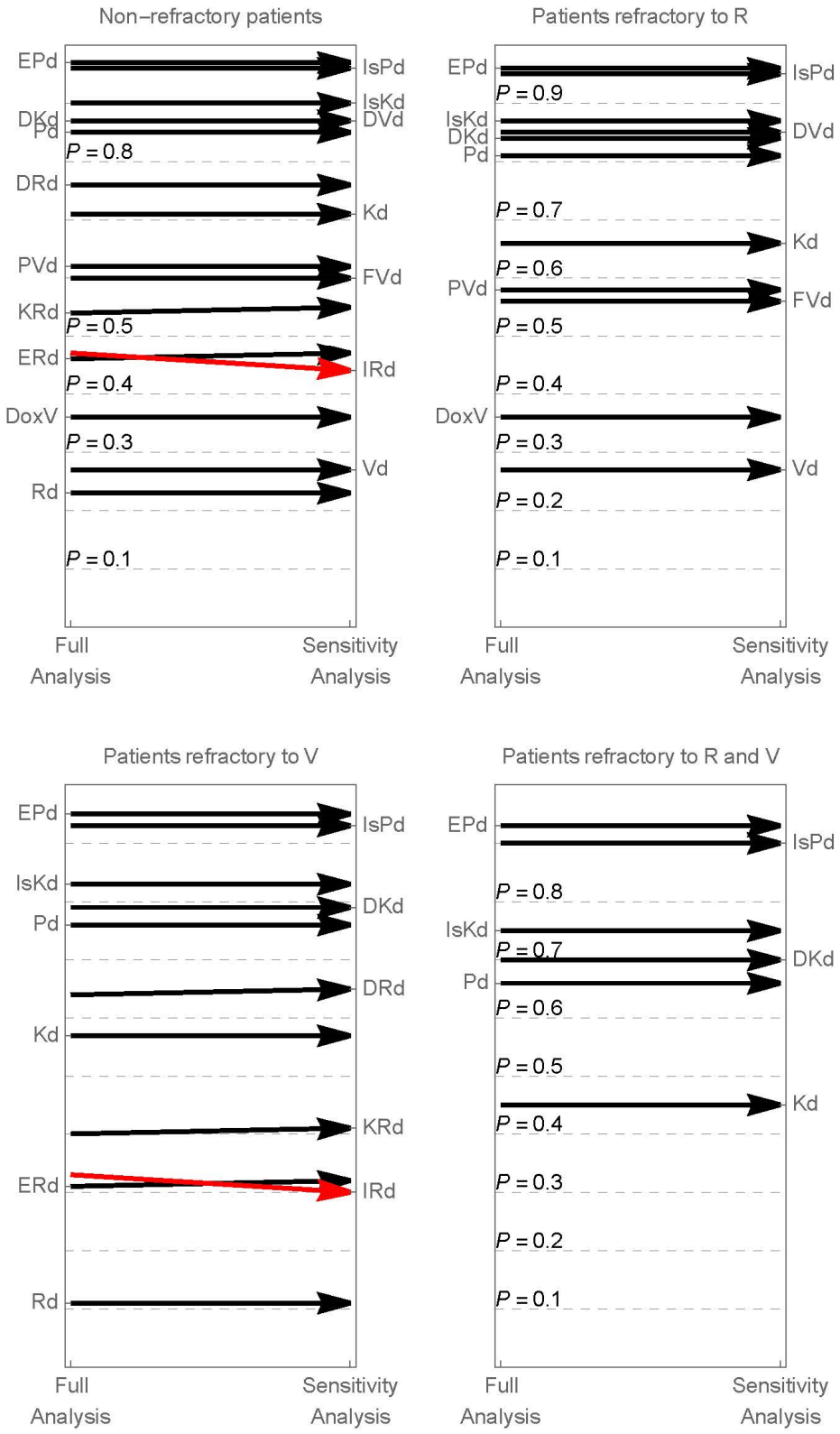


Figure 13: Sensitivity analysis results showing the effect of removing the China Continuation Study on P-scores for progression-free survival

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, R: lenalidomide, V: bortezomib

Results – Discontinuation due to adverse events

NMA results

The NMA of discontinuation due to adverse events resulted in three disconnected networks (*Figure 14*) with a total of 35 treatment regimens. A NMA was performed using data on the 35 treatments from 34 RCTs that enrolled a total of 12 873 patients (*Appendix 10 - Additional results: discontinuation due to adverse events*).

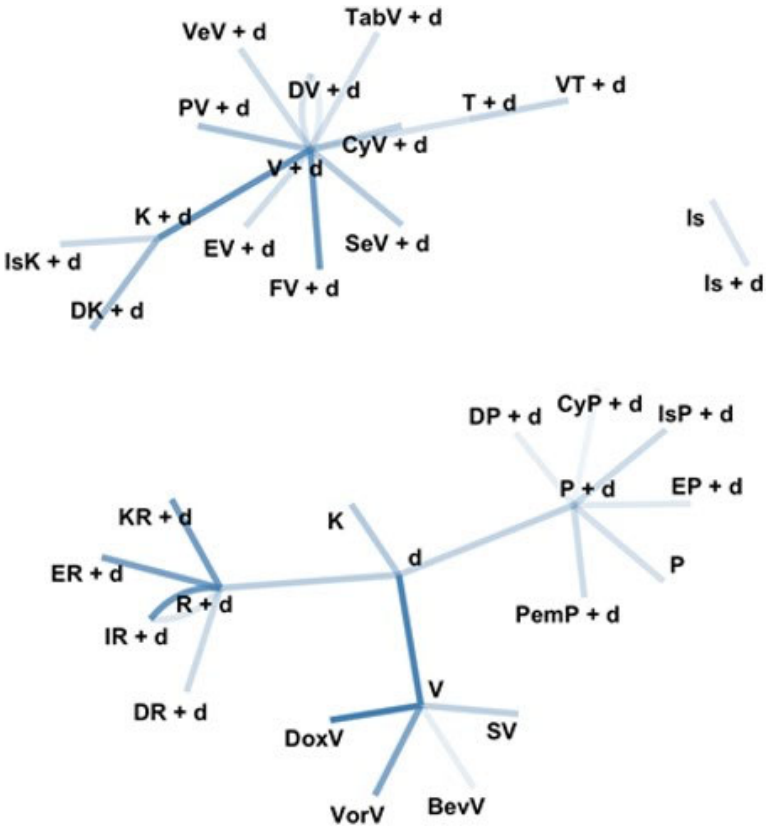


Figure 14: Network topology for risk of discontinuation due to adverse events
Each vertex represents a treatment, and each edge (line) represents a direct treatment comparison. More precise estimates (e.g., those supported by studies with larger sample sizes) are indicated by darker edges. Bev: bevacizumab, Cy: cyclophosphamide, d: dexamethasone, D: daratumumab, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, P: pomalidomide, Pem: pembrolizumab, R: lenalidomide, Se: selinexor, T: thalidomide, Tab: tabalumab, V. bortezomib, Ve: venetoclax, Vor: vorinostat

We used random effects component NMA to estimate average risk ratios (RRs) and account for between-study heterogeneity. An average treatment effect estimate should not be misinterpreted as being equivalent to an estimate from a single study (which cannot account for between-study heterogeneity). Consequently, confidence intervals on average treatment effect estimates may be wider that those arising from individual studies. The matrix plot (*Table 15*) presents estimates of HRs and 95% confidence intervals, in addition to our assessment of the certainty of evidence, for treatment regimens relevant for Norway. Heterogeneity (I^2) was estimated to be 31.1% (95% CI 0% to 73.6%).

The summary of findings table (*Table 16*) presents treatment regimens relevant for Norway ranked by P-score for non-refractory patients, along with assessments of certainty of evidence (GRADE). It also presents ranked treatments for patients who are refractory to lenalidomide and/or bortezomib.

Table 15: Matrix plot - discontinuation due to adverse events

Treatment A \ Treatment B	DK + d	DoxV	DR + d	DV + d	EP + d	ER + d	FV + d	IR + d	IsK + d	IsP + d	K + d	KR + d	P + d	PV + d	V + d
R + d	1.49 (0.45 – 4.91) *NP*	0.64 (0.26 – 1.54) *Low*	1.02 (0.46 – 2.27) *Low*	1.58 (0.42 – 5.86) *NP*	2.47 (0.51 –11.96) *Very low*	1.21 (0.75 – 1.95) *High*	0.79 (0.26 – 2.46) *NP*	0.91 (0.59 – 1.40) *Low*	2.81 (0.76 –10.31) *NP*	4.08 (1.16 –14.37) *Very low*	1.70 (0.59 – 4.89) *NP*	1.08 (0.67 – 1.73) *High*	2.35 (0.89 – 6.16) *Very low*	2.54 (0.78 – 8.26) *NP*	1.54 (0.55 – 4.31) *NP*
DK + d		0.43 (0.15 – 1.24) *NP*	0.68 (0.16 – 2.86) *NP*	1.06 (0.37 – 3.05) *Very low*	1.66 (0.29 – 9.48) *NP*	0.81 (0.22 – 2.93) *NP*	0.53 (0.23 – 1.21) *Moderate*	0.61 (0.17 – 2.16) *NP*	1.88 (0.73 – 4.81) *Very low*	2.74 (0.63 –11.82) *NP*	1.14 (0.66 – 1.98) *High*	0.72 (0.20 – 2.60) *NP*	1.57 (0.46 – 5.33) *NP*	1.70 (0.70 – 4.13) *Moderate*	1.03 (0.52 – 2.03) *Moderate*
DoxV			1.60 (0.49 – 5.25) *Very low*	2.47 (0.78 – 7.88) *NP*	3.88 (0.82 –18.28) *Very low*	1.90 (0.70 – 5.17) *Low*	1.24 (0.48 – 3.21) *NP*	1.42 (0.53 – 3.80) *Very low*	4.40 (1.34 –14.47) *NP*	6.41 (1.88 –21.79) *Very low*	2.67 (1.07 – 6.66) *NP*	1.69 (0.62 – 4.59) *Low*	3.69 (1.47 – 9.25) *Low*	3.98 (1.46 –10.87) *NP*	2.41 (1.06 – 5.50) *NP*
DR + d				1.55 (0.33 – 7.21) *NP*	2.43 (0.41 –14.24) *Very low*	1.19 (0.47 – 3.02) *Very low*	0.78 (0.19 – 3.12) *NP*	0.89 (0.36 – 2.21) *Very low*	2.76 (0.60 –12.71) *NP*	4.01 (0.90 –17.83) *Very low*	1.67 (0.44 – 6.29) *NP*	1.06 (0.42 – 2.68) *Very low*	2.31 (0.66 – 8.09) *Very low*	2.49 (0.60 –10.38) *NP*	1.51 (0.41 – 5.57) *NP*
DV + d					1.57 (0.25 – 9.78) *NP*	0.77 (0.19 – 3.11) *NP*	0.50 (0.20 – 1.28) *Very low*	0.57 (0.14 – 2.29) *NP*	1.78 (0.55 – 5.81) *Very low*	2.59 (0.54 –12.38) *NP*	1.08 (0.44 – 2.67) *Very low*	0.68 (0.17 – 2.76) *NP*	1.49 (0.39 – 5.69) *NP*	1.61 (0.59 – 4.36) *Very low*	0.98 (0.43 – 2.20) *Very low*
EP + d						0.49 (0.09 – 2.54) *Very low*	0.32 (0.06 – 1.76) *NP*	0.37 (0.07 – 1.88) *Very low*	1.14 (0.18 – 7.02) *NP*	1.65 (0.37 – 7.30) *Very low*	0.69 (0.13 – 3.60) *NP*	0.44 (0.08 – 2.26) *Very low*	0.95 (0.27 – 3.31) *Low*	1.03 (0.18 – 5.83) *NP*	0.62 (0.12 – 3.21) *NP*
ER + d							0.66 (0.19 – 2.24) *NP*	0.75 (0.39 – 1.43) *Very low*	2.32 (0.58 – 9.30) *NP*	3.38 (0.88 –12.99) *Very low*	1.41 (0.44 – 4.49) *NP*	0.89 (0.45 – 1.75) *Moderate*	1.94 (0.66 – 5.71) *Very low*	2.10 (0.59 – 7.51) *NP*	1.27 (0.41 – 3.97) *NP*
FV + d								1.14 (0.34 – 3.84) *NP*	3.54 (1.33 – 9.40) *Low*	5.15 (1.25 –21.22) *NP*	2.15 (1.16 – 3.96) *High*	1.36 (0.40 – 4.63) *NP*	2.96 (0.93 – 9.48) *NP*	3.20 (1.53 – 6.71) *High*	1.94 (1.22 – 3.09) *High*
IR + d									3.10 (0.79 –12.21) *NP*	4.51 (1.19 –17.04) *Very low*	1.88 (0.60 – 5.88) *NP*	1.19 (0.62 – 2.26) *Very low*	2.59 (0.90 – 7.46) *Very low*	2.80 (0.80 – 9.85) *NP*	1.70 (0.55 – 5.19) *NP*
IsK + d										1.46 (0.31 – 6.89) *NP*	0.61 (0.28 – 1.30) *Moderate*	0.38 (0.10 – 1.53) *NP*	0.84 (0.22 – 3.16) *NP*	0.90 (0.32 – 2.54) *Very low*	0.55 (0.23 – 1.29) *Very low*
IsP + d											0.42 (0.11 – 1.62) *NP*	0.26 (0.07 – 1.01) *Very low*	0.58 (0.26 – 1.29) *Low*	0.62 (0.15 – 2.66) *NP*	0.38 (0.10 – 1.43) *NP*

		Treatment B														
		DK + d	DoxV	DR + d	DV + d	EP + d	ER + d	FV + d	IR + d	IsK + d	IsP + d	K + d	KR + d	P + d	PV + d	V + d
Treatment A	K + d												0.63 (0.20 – 2.01) *NP*	1.38 (0.46 – 4.10) *NP*	1.49 (0.74 – 2.99) *Moderate*	0.90 (0.61 – 1.34) *High*
	KR + d													2.18 (0.74 – 6.40) *Very low*	2.36 (0.66 – 8.42) *NP*	1.43 (0.46 – 4.45) *NP*
	P + d														1.08 (0.32 – 3.62) *NP*	0.65 (0.23 – 1.90) *NP*
	PV + d															0.61 (0.34 – 1.08) *High*

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib.

NP: not possible to GRADE because the comparison is across networks. Statistically significant effect estimates are highlighted in either pink (favours treatment A) or blue (favours treatment B).

Only listed treatment regimens relevant for Norway

Table 16: Summary of findings for discontinuation due to adverse events
Treatments are ordered by overall rank relevant for non-refractory patients

Treatment*	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant to people who are:		
		Refractory to R	Refractory to V	Refractory to R & V
IsP + d	1 (0.89)	1 (0.83)	1 (0.83)	1 (0.78)
IsK + d	2 (0.78)	2 (0.70)	2 (0.70)	3 (0.63)
PV + d	3 (0.76)	5 (0.67)	NA	NA
P + d	7 (0.71)	7 (0.65)	6 (0.65)	7 (0.56)
EP + d	8 (0.69)	8 (0.64)	7 (0.64)	6 (0.56)
K + d	16 (0.56)	16 (0.51)	12 (0.50)	12 (0.40)
DV + d	19 (0.51)	19 (0.49)	NA	NA
V + d	20 (0.49)	20 (0.46)	NA	NA
DK + d	21 (0.48)	21 (0.45)	13 (0.44)	13 (0.34)
ER + d	25 (0.38)	NA	15 (0.32)	NA
KR + d	26 (0.31)	NA	17 (0.26)	NA
DR + d	27 (0.30)	NA	18 (0.25)	NA
R + d	30 (0.27)	NA	19 (0.20)	NA
IR + d	32 (0.23)	NA	20 (0.16)	NA
FV + d	34 (0.17)	29 (0.21)	NA	NA
DoxV	35 (0.10)	30 (0.13)	NA	NA

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: pomalidomide, I: ixazomib, Is: isatuximab, K: carfilzomib, NA: not applicable, P: pomalidomide, R: lenalidomide, V: bortezomib
** The Methods chapter describes how treatments are defined.*
§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment.
Only listed treatment regimens relevant for Norway. Ranks are calculated for all treatments included in the NMA (i.e., ranks may not begin at 1 and ranks for treatments not relevant to Norway are not shown).

NOTE: We have limited confidence in the specific order in which these treatments are ranked, and therefore caution against using this ranking uncritically.

Among treatment regimens relevant for Norway, the three highest ranked treatments for patients who are non-refractory, as well as patients who are refractory to lenalidomide (R) (i.e., not responding to lenalidomide), in terms of discontinuation due to adverse events are [IsP + d], [IsK + d] and [PV + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are as follows (*Table 16*):

- For non-refractory patients: 89%, 78%, and 76%, respectively
- For patiens refractory to lenalidomide (R): 83%, 70%, and 67%, respectively

For patients who are either refractory to bortezomib (V) (i.e., not responding to bortezomib) or refractory to both lenalidomide (R) and bortezomib (V) (i.e., not responding to either lenalidomide or bortezomib), the three highest ranked relevant treatment regimens are [IsP + d], [IsK + d] and [P + d]. The estimated probabilities that these are the best of all analysed treatment regimens, are as follows (*Table 16*):

- For patients refractory to bortezomib (V): 83%, 70% and 65%, respectively
- For patients refractory to both lenalidomide (R) and bortezomib (V): 78%, 63%, and 56%, respectively

The matrix plot (*Table 15*) presents the NMA effect estimates for discontinuation due to adverse events, for all possible comparisons between the treatment regimens relevant

to Norway. For discontinuation due to adverse events, only 13 of 120 NMA effect estimates were statistically significant (*Table 15*).

Judgements about the ranking of the various treatment regimens should not be made without also taking the certainty of the evidence into consideration. We were not able to assess the certainty of evidence for 63 of 105 of the comparisons between the relevant treatment regimens, because the GRADE method has not yet been extended address disconnected networks (*Figure 14*). However, the 57 NMA effect estimates that we could assess, were set between high and very low certainty of evidence, with the majority being very low (*Table 13*). Of the 13 significant NMA estimates, we could only assess certainty of evidence for eight NMA estimates, in which three were set as high, two were set as low, and three were set as very low certainty of evidence (*Table 15*). Details of our assessment of GRADE is presented in *Appendix 9 - Detailed GRADE: discontinuation due to adverse events*.

Transitivity analysis for discontinuation due to adverse events

Analyses performed to assess the transitivity assumption were consistent with those for OS. There is no evidence of inconsistency between direct, indirect, and NMA estimates.

Radar plots of P-scores for all included treatments and outcomes

All NMAs are summarized in sets of radar plots of P-scores: *Figure 15* for non-refractory patients, *Figure 16* for patients who are refractory to lenalidomide (R), *Figure 17* for patients who are refractory to bortezomib (V), and *Figure 18* for patients who are refractory to both lenalidomide (R) and bortezomib (V). Each radar plot shows the available P-scores for a given treatment as a polygon (shaded). A P-score quantifies the extent of evidence that a given treatment is superior to all other included treatments with respect to a given outcome, and accounts for uncertainty. We would expect on average that patients given a treatment with a P-score for OS closer to 100% would survive longer, and that patients given a treatment with a P-score for SAE closer to 100% would experience fewer SAEs. In general, treatments with polygons with larger areas are likely superior to those with smaller areas. P-scores are not expected to sum to 100%. The radar plots do not incorporate our GRADE assessments of the certainty of the evidence or economic analysis.

Comparing the overall efficacy and safety of the various treatment regimens across outcomes is challenging, as not all treatment regimens have data (and therefore no P-score) on all outcomes. When comparing results between treatment regimens, one should be careful not to interpret effect solely based on polygon area. Radar plots of the double combination [P + d] exemplify treatment regimens that have a large polygon area. This would indicate better efficacy and safety than treatment regimens with smaller polygons, e.g., [ER + d]. However, when looking more closely at the individual P-scores, we find that [P + d] has lower P-score for OS than [ER + d], and we would expect longer survival by treatment with [ER + d] than with [P + d]. As such, although radar plots may be useful for understanding tradeoffs between efficacy and safety, they should not be interpreted in isolation. In addition, the radar plots do not reflect assessments of the certainty of evidence or results of the health economic analysis.

As shown in *Figure 15*, the six triplet combinations [EP + d], [IsP + d], [DK + d], [KR + d], [DR + d] and [DV + d] are examples of treatment regimens relevant for non-refractory patients that have clearly favorable HRs for OS, that also ranked highly with respect to other outcomes.

As shown in *Figure 16*, *Figure 17*, and *Figure 18*, the three triplet combinations [DK + d], [EP + d] and [IsP + d] are examples of treatment regimens that have clearly favorable HRs for OS, and also ranked highly with respect to other outcomes, and that are relevant for patients who are refractory to lenalidomide (R), bortezomib (V), and lenalidomide (R) and bortezomib (V), respectively.

Remember that P-scores, and hence the shapes of the polygons, would be different for patients who are refractory to specific treatments: it is not possible to simply ignore polygons for treatments that are not competitors on the basis of refractoriness and assume that the shapes of the remaining polygons would be unchanged.

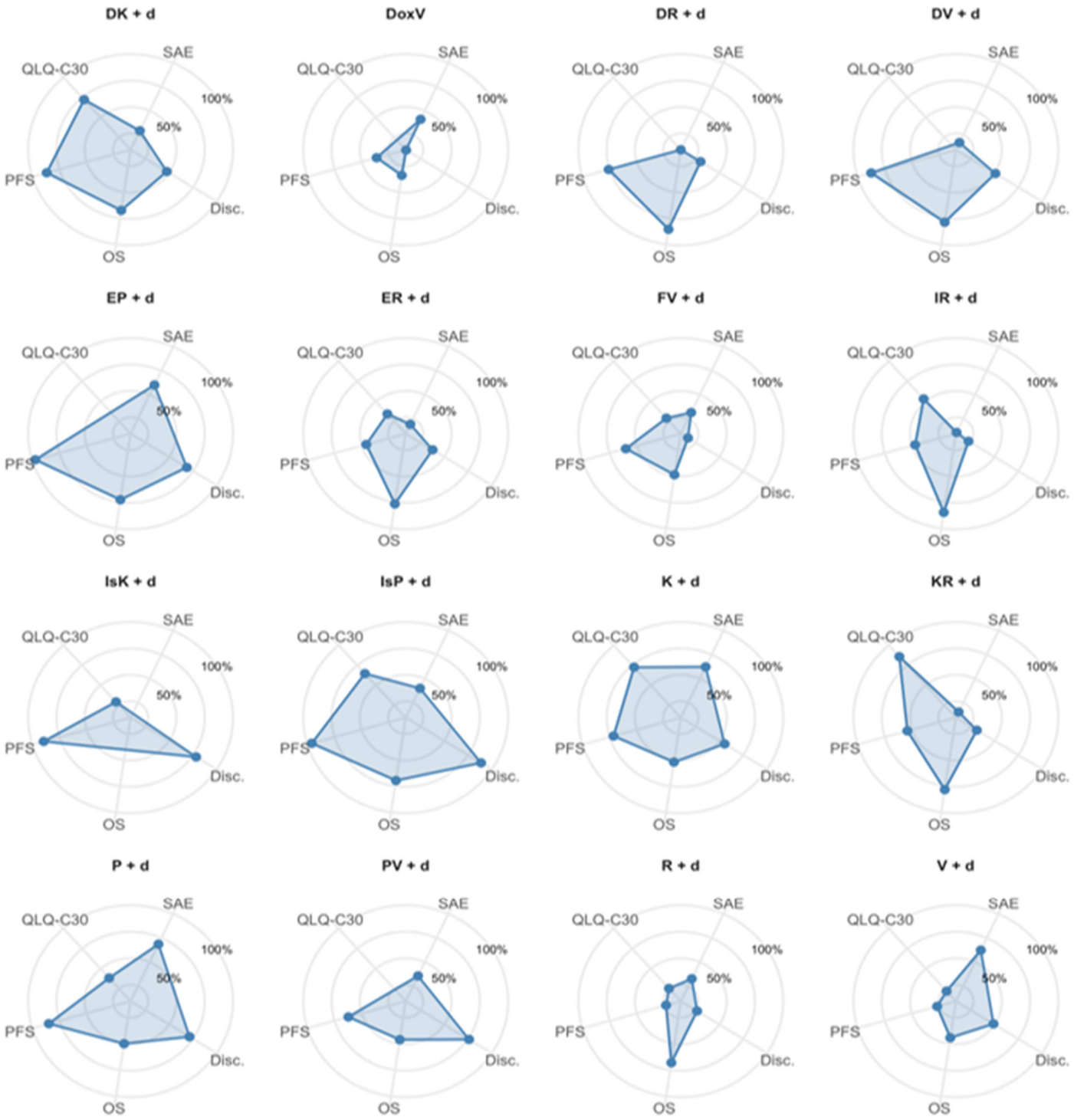


Figure 15: Radar plot of treatment regimens relevant for non-refractory patients

Radar plots show treatment regimens relevant to Norwegian clinical practice. D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), Dox: doxorubicin, E: elotuzumab, F: panobinostat, I: ixazomib, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), R: lenalidomide, SAE: severe adverse events (incidence rate ratio), V: bortezomib. The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.

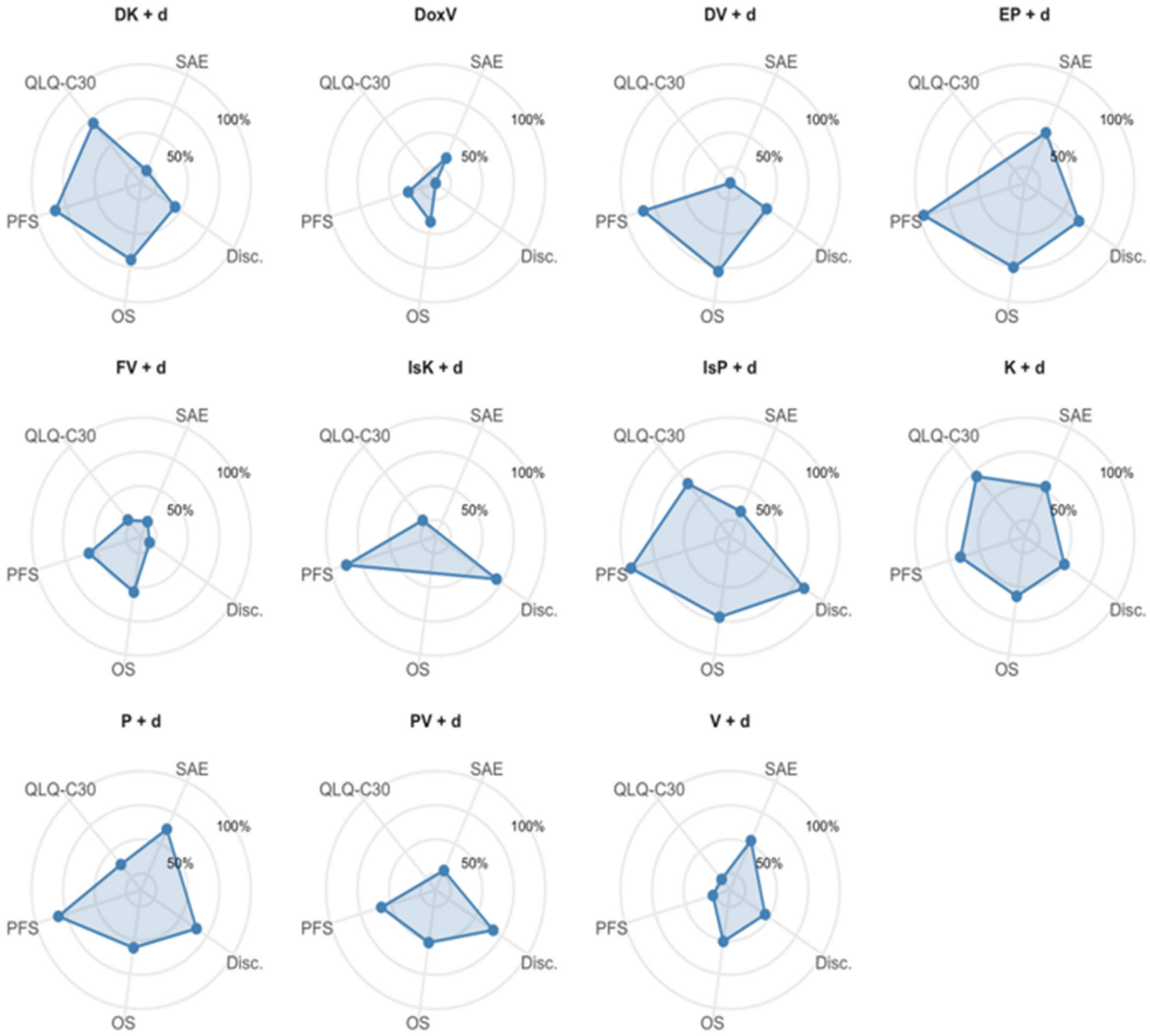


Figure 16: Radar plot of treatment regimens relevant for patients refractory to lenalidomide (R)
Radar plots show treatment regimens relevant to Norwegian clinical practice. D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), Dox: doxorubicin, E: elotuzumab, F: panobinostat, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), SAE: severe adverse events (incidence rate ratio), V: bortezomib. The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.

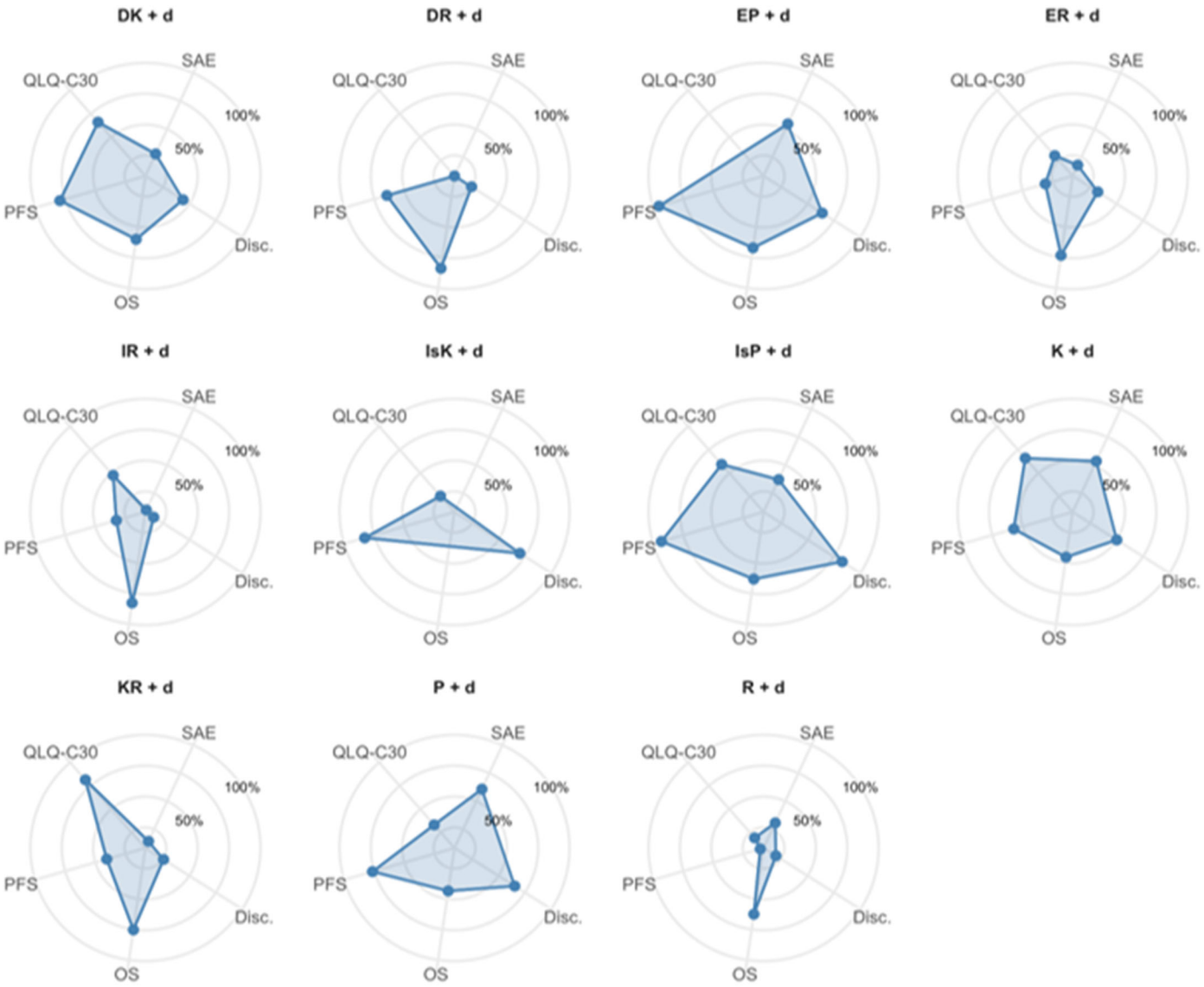


Figure 17: Radar plot of treatment regimens relevant for patients refractory to bortezomib (V)
Radar plots show treatment regimens relevant to Norwegian clinical practice. D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), E: elotuzumab, I: ixazomib, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), R: lenalidomide, SAE: severe adverse events (incidence rate ratio). The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.

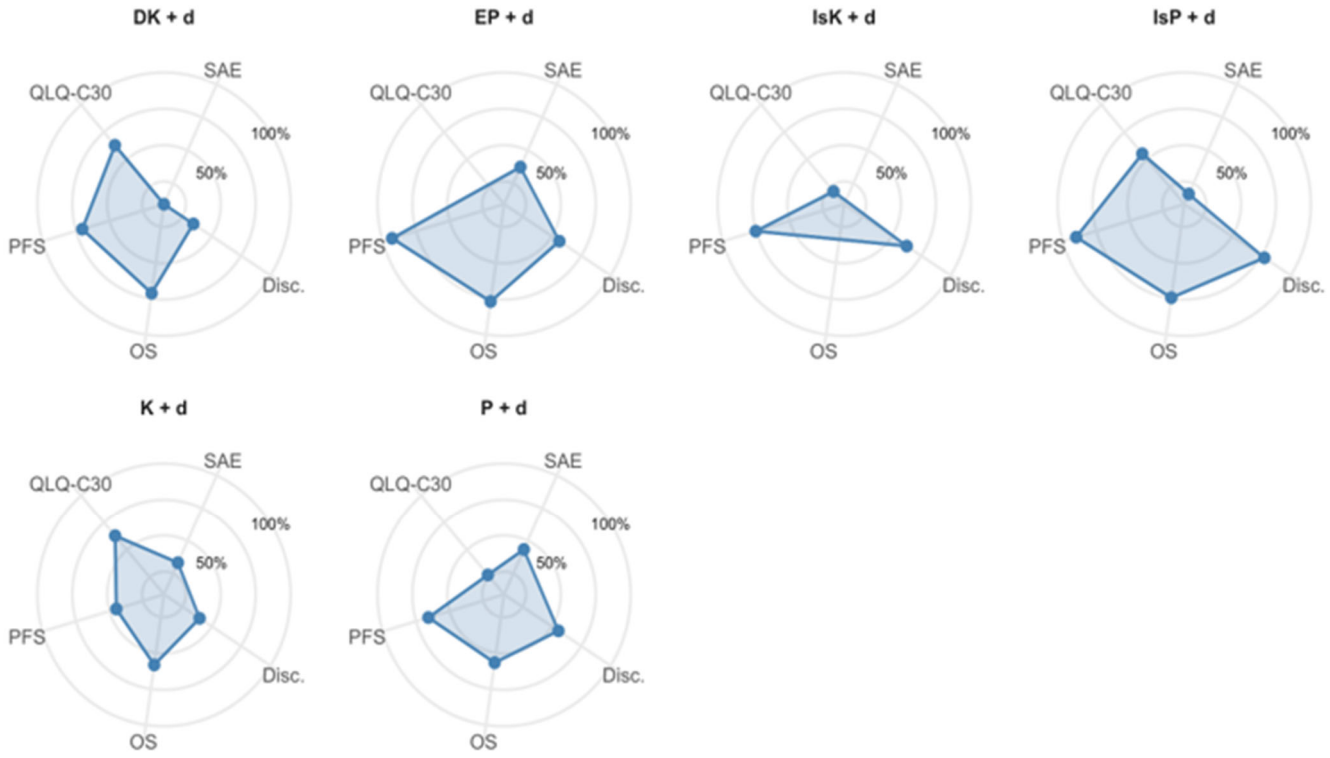


Figure 18: Radar plot of treatment regimens relevant for patients refractory to lenalidomide (R) and bortezomib (V)

Radar plots show treatment regimens relevant to Norwegian clinical practice. D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), E: elotuzumab, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), SAE: severe adverse events (incidence rate ratio). The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.

Health economic evaluation

METHODS

Objectives

To conduct a cost-effectiveness analysis of treatments for relapsed and/or refractory multiple myeloma and provide an estimate of disease severity, in accordance with priority setting rules for the Norwegian health care system.

Literature Search

Over the course of the project, we performed three simple literature searches (May 2020, May 2021, and March 2022) in PubMed for cost-effectiveness analyses of treatments for relapsed and/or refractory multiple myeloma using the terms "myeloma" and "cost-effectiveness". Unlike searches for clinical effect and safety results, a search for cost-effectiveness analyses is not intended to identify studies that could be used directly to determine the cost-effectiveness in Norway of treatments for relapsed and/or refractory multiple myeloma. Health systems can vary widely across countries in terms of organizational structure and costs, making it impossible to apply published results of a cost-effectiveness analysis conducted in one country to a different country.[8] Instead, the goal of the search is to find relevant cost-effectiveness analyses to help inform our choices about what type of economic model to use and alternative ways to structure the model; as potential sources for utility weights needed to calculate quality-adjusted life-years (QALYs); and to identify strengths and of weaknesses of our results by comparing them with results in other studies.

The searches returned three systematic reviews—Aguiar 2016 (110), Asra 2021 (111) and Seefat 2022 (112), which included 8, 17, and 13 articles, respectively—and a total of 61 individual articles. After discarding duplicates, we examined the remaining articles and excluded those that focused on treatments for newly diagnosed patients rather than relapsed and/or refractory patients, those that did not present results in terms of costs per QALYs or only presented a budget impact analysis, and those that provided incomplete or limited information about methods, e.g., letters, abstracts, reviews. In total we included fourteen relevant cost-effectiveness analyses.[9]

Information about the treatments examined, the type of analysis conducted, utility weights used for calculating QALYs, and the cost-effectiveness results from each included analysis, is provided in *Appendix 12*.

[8] Although it is sometimes possible to obtain permission to apply Norwegian data to an existing economic model, none of the models we examined were suitable.

[9] We did not include "grey literature", e.g., single technology assessments performed in other countries because confidential pricing agreements in individual countries resulted in either "empty" or highly redacted cost-effectiveness analyses by sources such as NICE or CADTH.

Three different types of cost-effectiveness models were used in the included studies: 1) four studies (113-116) performed discrete event simulations, made possible by access to patient level data from trials with a single intervention and a relevant comparator, 2) four studies, also with access to patient level data, used either a Markov (117-119) or a semi-Markov model (120), 3) the remaining ten studies performed partitioned survival analyses.

Nine of the cost-effectiveness analyses examined a single intervention (comprising either two or three medications) with a single comparator (113-117, 121-124). Four analyses compared either several interventions with a single comparator or a single intervention with different comparators (118-120, 125). The largest number of treatments compared in a single analysis examined the cost-effectiveness of six potential "triplet" treatments relative to one of two standard comparators used as either second or third-line treatments (126).

Health economic model

Model Basics

We used TreeAge (127) to perform a cost-utility analysis of 13 treatments for relapsed and/or refractory multiple myeloma patients in which relevant costs were expressed in 2022 Norwegian kroner (NOK)[10] and effects were expressed in QALYs. The analysis was performed from an extended health sector perspective, which includes all direct reatment costs and the time and travel costs for patients during treatment. Both costs and effects were discounted at an annual rate of 4%, in accordance with Norwegian guidelines for health economic analyses (128). We assumed a starting age for treatment of 65 years, based on the mean age of patients included in trials used for the NMA, and a 20-year time horizon for the model. An important characteristic of our analysis is that we are only able to determine the cost-effectiveness of each treatment compared to other treatments without any consideration of treatment sequencing, beyond the assumption that all patients have a relapsed and/or refractory status. It is also important to note that we had no access to patient level data in conducting this analysis.

Based on the information from the cost-effectiveness analyses from our literature search, the large number of treatments, and the lack of patient level data, we decided to use a partitioned survival analysis approach for the economic model. We also relied on the NICE Technical Document on the appropriate use of partitioned survival analyses for decision modeling in health care to help inform our choice (129).

Partitioned survival analysis

Partitioned survival models, often referred to as "Area Under the Curve" models, are a common choice for determining cost-effectiveness for cancer treatments. As with other types of cost-effectiveness models, patients are tracked through different pre-defined health states. A typical partitioned survival model for cancer treatments includes three states, 1) progression-free, 2) progressed, and 3) dead, and is characterized by two survival curves, PFS and OS (*Figure 19*). State membership is derived from the survival curves at each model cycle. Note that the area under the OS curve includes all patients who are alive, but some are in the progression-free state (area under the PFS curve) and others are in the progressed state (area between the PFS and OS curves), i.e., the states are not mutually exclusive. State membership is determined as follows: the percent dead at any time is 1 minus the OS curve at each point in time. Similarly,

[10] All drug costs are based on 2022 price agreements. Other costs, e.g., patient travel, doctor/nurse time, drug administration costs, etc. were calculated earlier using 2021 prices. We decided not to update these prices in the most recent model update, following changes to clinical effect results, as they represent a very small portion of total treatment cost and would not change the overall results of our analysis.

membership in the progressed state is the difference between the OS and PFS curves at each time point (129). While it is possible to include additional states in a partitioned survival analysis to capture, for example, a state that identifies a second treatment after progression on the first, the information requirement generally makes that impossible without access to patient level data.

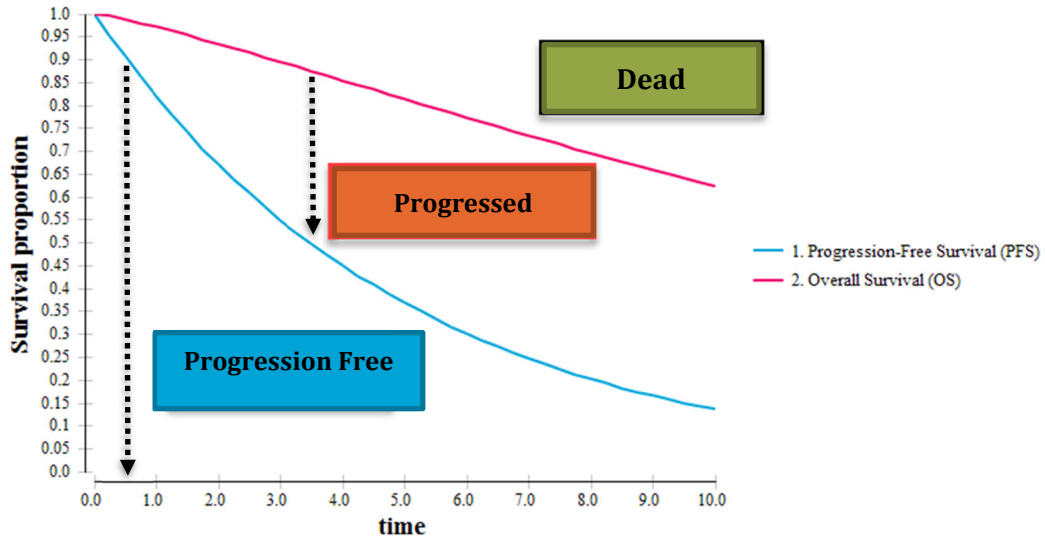


Figure 19: Survival Curves and Health States in Partition Survival Analysis
OS – PFS = Progressed, PFS = Progression Free, OS = Alive (Progressed + Progression Free)

Reporting of results

We report model results as incremental cost-effectiveness ratios (ICERs), which measure the additional cost of gaining one additional QALY, relative to the current treatment:

$$ICER = \frac{Cost_{intervention} - Cost_{comparator}}{Effect_{intervention} - Effect_{comparator}}$$

An intervention is considered cost-effective if the ICER is below the willingness-to-pay, λ, for an additional QALY gained, compared to the next best treatment, as expressed in the following decision rule:

$$\Delta C / \Delta E < \lambda$$

Because ICERs have poor statistical properties, they are often rearranged to express either incremental net monetary benefit (INMB) or incremental net health benefit (INHB). A treatment is then considered cost-effective based on the following decision rules:

$$INMB: \lambda \cdot \Delta E - \Delta C > 0$$
$$INHB: \Delta E - (\Delta C / \lambda) > 0$$

Absolute shortfall

While there is no official Norwegian threshold value for willingness-to-pay for an additional QALY, the Magnussen group's proposal for how to operationalize the severity criteria suggests that threshold values for willingness-to-pay should increase with increases in disease severity, measured as absolute shortfall – the number of healthy life-years lost without treatment (6). There has been acceptance for linking willingness-to-pay to disease severity, but no general agreement how increases in severity should effect willingness-to-pay. As part of the health-economic analysis, we

compute absolute shortfall in order to provide decision-makers with a basis for applying the severity criterion. We calculated absolute shortfall, the measure of disease severity recommended in Norwegian priority setting guidelines for the health sector (130), using the following formula:

$$QALY(A) - Prognosis\ (A) = Absolute\ shortfall\ (AS)$$

We used undiscounted QALYs for Prognosis (i.e., QALYs remaining for patients with standard of care in absence of intervention at mean diagnosed age) and QALY (A), which refers to the total amount of remaining QALYs for a healthy population at the mean diagnosed age (130). We recognize that the ultimate decision about the relevant willingness-to-pay for different levels of absolute shortfall rests with the decision-makers who evaluate this report.

Addressing uncertainty

We addressed uncertainty in model parameters by conducting a probabilistic sensitivity analysis with 10,000 random draw Monte Carlo iterations. Probabilistic sensitivity analysis accounts for parameter uncertainty in a model by defining confidence intervals and a relevant statistical distribution for each parameter in the model. For each Monte Carlo iteration a value is drawn from the distribution describing each parameter, resulting in new estimates of the benefits and costs of each treatment.
Results from the probabilistic sensitivity analysis are characterized in a scatter plot of all resulting combinations of costs and health benefits, and as cost-effectiveness acceptability curves (CEACs) which show "*the probability that a given intervention is the most cost-effective option given the observed data … the CEAC can show a decision-maker the probability that, if the intervention is funded or reimbursed, this will be the correct decision*" (131). Finally we presented the cost-effectiveness acceptability frontiers (CEAF) for a range of willingness-to-pay thresholds to show which strategies are cost effective at a given willingness-to-pay, given its probability of cost effectiveness (131). In comparison to CEACs, which present all strategies and their probability for cost effectiveness, the CEAFs only present strategies that could be considered cost-effective over a given willingness-to-pay range, and excludes the other strategies.

We performed one-way sensitivity analyses in which model parameters were varied individually to determine the variables that had the largest impact on the deterministic cost-effectiveness results. Variables included in the one-way sensitivity analyses were drug prices; utility values for the progression-free and progressed health states; time costs for doctors, nurses, and patients associated with infusions or injections or regular follow-up appointments at the hospital; patient transportation costs; and end-of-life costs. Results of one-way sensitivity analyses were presented as Tornado diagrams.

Because of the high prices of myeloma drugs, we planned to conduct a scenario analysis to capture the reduction in drug costs attributable to dose reductions resulting from treatment discontinuation because of toxicity or other problems. By subtracting the accumulated costs at median time-to-discontinuation from accumulated costs at median time-to-progression we planned to estimate differences in total drug costs resulting from dose reductions.

Model structure

We used TreeAge (127) to develop a partitioned survival analysis model with three states: 1) Progression-free, 2) Progressed, and 3) Dead, to examine the cost-effectiveness of 13 different treatments for patients with relapsed and/or refractory multiple myeloma.[11] The model relies on monthly cycles in order to accumulate costs

[11] We were unable to include in the health economic analyses one treatment, doxorubicin + bortezomib [DoxV]), that was included in the list of treatments relevant for Norway used in the clinical effect section of

based on standard treatment protocols. The health outcome, QALY are accumulated on an annual basis.

In cost-utility analyses based on a single trial, the survival curves for the treatment group are often modeled by multiplying the survival curves for the comparator group by the relevant HRs reported in the trial. When a NMA is used to compare multiple treatments-based results from several different trials, there is no "comparator" as such, but comparisons within the network can be made by designating a "reference treatment" and using the relevant HRs from the matrix of results to define survival curves for all treatments of interest relative to the reference treatment.

We originally intended to use [R + d] as the reference treatment for the model. However, when we generated the OS and PFS curves by applying the appropriate HRs from the NMA to the OS and PFS curves for [R + d], five of the interventions showed PFS curves that were entirely above the OS curve – an invalid outcome. This problem may be explained by a variety of factors, including but not limited to, use of mean HRs rather than study-level HRs, undetected inconsistency in the NMA, possible violations of the additivity assumption necessary for a component NMA, violations of the proportional hazards assumption within one or more of the included studies, limitations in the approach used to impute reference survivor functions, and the fact that the included studies, and hence the NMA and partitioned survival analysis, did not account for correlations between PFS and OS. Issues such as these are discussed in more detail in the NICE Technical Support Document (129).

Table 17: Treatments included in the health economic model

Treatments by Reference Group
Reference treatment: **R + d**
R + d
DR + d
RK + d
ER + d
IR + d
Reference treatment: **V + d**
V + d
DK + d
K + d
FV + d
DV + d
Reference treatment: **P + d**
P + d
EP + d
IsP + d

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib (Velcade).

The problem of invalid survival curves resolved when, as had been suggested by our clinical expert, we divided the model into three disconnected reference groups, each with its own reference treatment (*Table 17*). We fitted parametric survival curves for OS and OFS for each of the reference treatments based on Kaplan-Meier plots of the

the report. Because in the clinical trial comparing [DoxV] to bortezomib (Velcade), dexamethasone was not included in either arm, we would have needed to include a new reference treatment and generate new survival curves in order to evaluate it. It was not one of the medications of interest mentioned in our original commission.

respective outcomes in clinical trials. Each reference treatment serves as the "comparator" for the other treatments included in the group. The survival curves for the remaining treatments in each group were generated by applying the relevant HRs from the NMA to the parametric survival curves for the reference treatments. In essence, the complete model comprises three models, each tracking state membership and the accumulation of QALY and treatment costs over time for the treatments within a specific reference group (*Figure 20*).

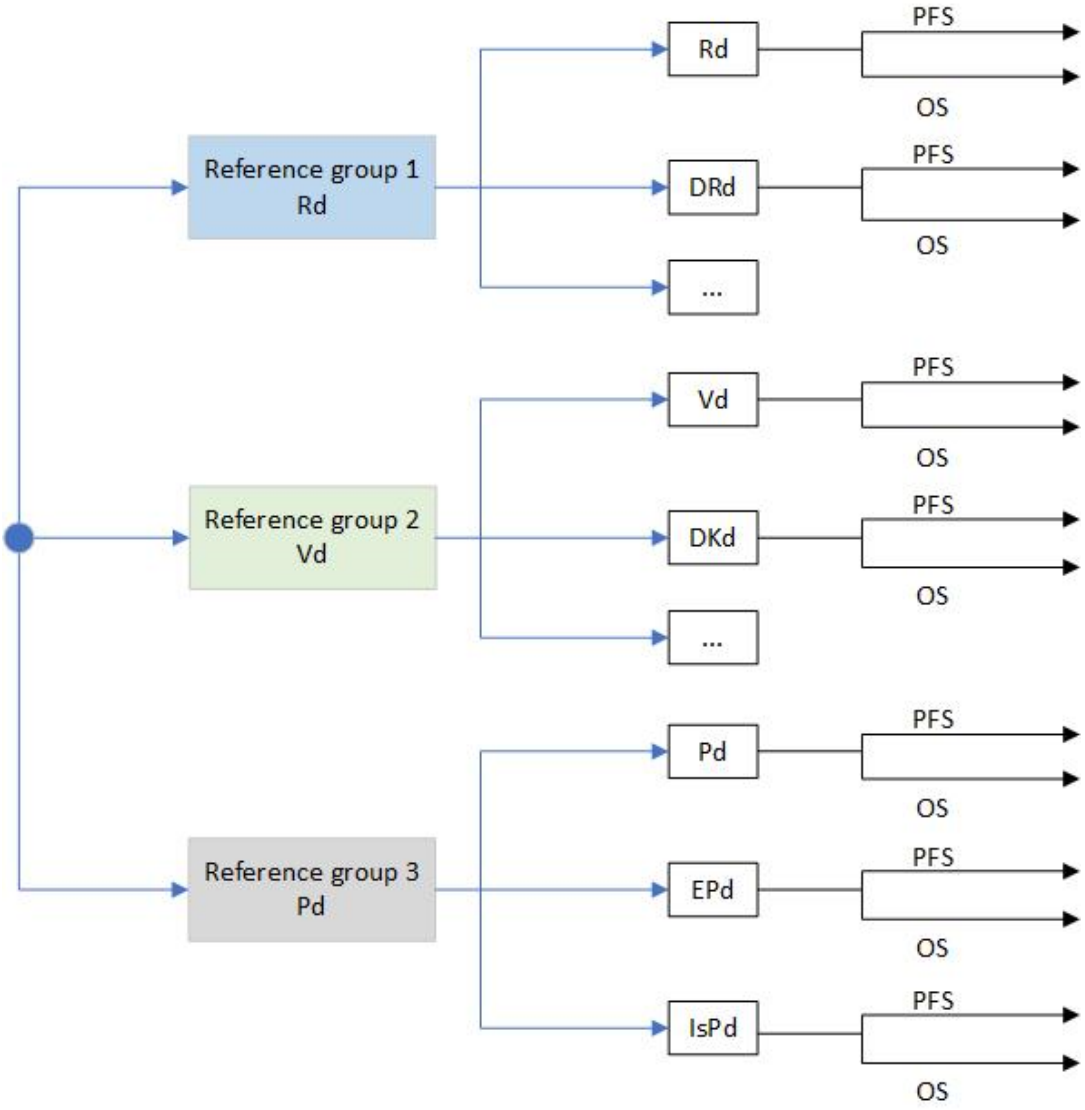


Figure 20: Model schematic
For the [R + d] and [V + d] reference groups, the boxes with three dots indicate that other treatments, each with a set of PFS and OS survival curves, are included in the analysis. In the [R+d] group the additional treatments are [KR + d], [ER + d], and [IR + d]. In the [V + d] group the additional treatments are [K + d], [FV + d] and [DV + d].
D: daratumumab, d: dexamethasone, E: elotuzumab, Is: isatuximab, K: carfilzomib, OS: overall survival, P: pomalidomide, PFS: progression-free survival, R: lenalidomide, V: bortezomib

Fitting survival curves for reference treatments

To specify the OS and PFS curves for each of the reference treatments, we relied on a survival fitting technique and associated Excel spreadsheet proposed by Hoyle and Henley, which can be used to reconstruct the underlying patient-level data for number of events and censored patients from each trial (132). This technique makes it possible to more accurately determine the position of points on Kaplan-Meier plots so that they can be expressed as parametric survival curves for use in the cost-utility model.

1. For each of the three reference treatments, we selected the relevant study with the largest patient populations that reported results for both OS and PFS. We then used WebPlotDigitizer (27) to extract data from the Kaplan-Meier plots that defined the survival curves by recording the percentage of patients on each survival curve at each time point where that information was shown on the Kaplan-Meier plot.
2. Because it is difficult to determine the exact position of points on published survival plots using WebPlotDigitizer, we entered this data into the Excel worksheet from the Hoyle and Henley article (132) to determine more precisely the number of events and censored patients in a given time interval. To ensure that the point estimates were extracted accurately, we checked for the censored patients and events between intervals, and callibrated the points manually to adjust for any corrections[12] (*Appendix 13*).
3. The patient level data extracted in steps 1 and 2 were imported to R Studio (133) to estimate parametric survival functions and parameters for OS and PFS using Lognormal, Exponential and Weibull distributions. Based on the Akaike Information Criterion (AIC) and visual inspection we decided that the Weibull distribution provided the best fit for the reference curves, and the values for the shape and scale variables to parametize the survival functions in TreeAge (127).

Model parameters

Table 18 provides the point estimates and confidence intervals for all parameters used in the model. The parameters are grouped by variable type and include: 1) HRs for OS and PFS from the NMA detailed in the clinical effect section of the report, 2) utility weights used to capture QoL gains (or losses) and are used to derive QALYs, and 3) the likelihood of SAEs.

Costs accrued on a monthly basis but were calculated in intervals to account for differences in frequency of dosage over different treatment cycles for each intervention. Utility values for PFS and OS were assumed to be constant across all treatments and were applied based on progressed vs. non-progressed status. A 95% confidence interval was provided for all parameters to account for variable uncertainty in the probabilistic sensitivity analysis.

A log-normal distribution was used for clinical effect of treatments expressed as HRs and SAE (presented as incidence rates per person months), Weibull distribution was used for reference survival curves in each subgroup, Gamma distribution for cost of drugs and Beta distribution for utility values and disutility values (134).

For each treatment, we imputed expected numbers of SAEs as follows. First, we reconstructed the survivor function for the treatment's reference ([R + d], [V + d], or [P + d]) as described in "Fitting survival curves for reference treatments". We then imputed the total exposure time for the patients allocated to the reference on the trial from which the survivor function was imputed. Total exposure time was imputed as the product of mean OS time (area under the reconstructed survivor function; $E[T_{\text{ref}}]$) and the number of patients allocated to the reference treatment (N_{ref}). We then imputed SAE rate for the reference as the ratio of the number of SAEs among patients allocated to the reference (n_{ref}) to the corresponding imputed total exposure time ($N_{\text{ref}} \times E[T_{\text{ref}}]$). The imputed SAE rate for a given reference is therefore $R_{\text{ref}} \approx n_{\text{ref}}/(N_{\text{ref}} \times E[T_{\text{ref}}])$. Monthly numbers of SAEs for non-reference treatments were then imputed by applying IRRs (taken from the NMA) to the rates imputed for the

[12] The table for summary of point calibration process on the Kaplan Meier plots for survival curves under *Appendix 13* summarizes the process to determine the number of events and censored patients based on the Kaplan Meier data available for the reference curve from the trials.

references. For example, the SAE rate for treatment A was imputed as $R_A \approx \text{IRR}_{A,\,ref} \times R_{ref}$, where $\text{IRR}_{A,\,ref} = \text{IR}_A/\text{IR}_{ref}$ is the estimated ratio of the incidence rate for treatment A (IR_A) to that for the reference treatment (IR_{ref}). The 95% confidence intervals on the IRRs were used to obtain confidence intervals on the imputed SAE rates.

The incidence rate for SAE was reported in per person months for the economic model, details of IRR are provided in the NMA section of the report (see *Results – Severe adverse events*). The utility (for PFS and OS) was adjusted using the utility decrement and multiplying them by the Incidence Risk (IR) for each of the respective treatments.

Table 18: Parameters in the health economic model

Parameters	Treatment (Reference)	Base-case values	Relevant Distribution	Range for CI (95%)	Source
Hazard ratio					
KR + d (PFS)	R + d	0.663	LogNormal	(0.554 – 0.786)	NMA
KR + d (OS)	R + d	0.79	LogNormal	(0.4 – 1.407)	NMA
ER + d (PFS)	R + d	0.723	LogNormal	(0.598 – 0.867)	NMA
ER + d (OS)	R + d	0.821	LogNormal	(0.418 – 1.457)	NMA
IR + d (PFS)	R + d	0.714	LogNormal	(0.576 – 0.875)	NMA
IR + d (OS)	R + d	0.694	LogNormal	(0.397 – 1.131)	NMA
DR + d (PFS)	R + d	0.443	LogNormal	(0.351 – 0.552)	NMA
DR + d (OS)	R + d	0.638	LogNormal	(0.272 – 1.278)	NMA
DK + d (PFS)	V + d	0.315	LogNormal	(0.219 – 0.438)	NMA
DK + d (OS)	V + d	0.673	LogNormal	(0.245 – 1.496)	NMA
FV + d (PFS)	V + d	0.633	LogNormal	(0.521 – 0.762)	NMA
FV + d (OS)	V + d	0.942	LogNormal	(0.477 – 1.681)	NMA
PV + d (PFS)	V + d	0.614	LogNormal	(0.49 – 0.76)	NMA
PV + d (OS)	V + d	0.986	LogNormal	(0.476 – 1.818)	NMA
DV + d (PFS)	V + d	0.302	LogNormal	(0.247 – 0.365)	NMA
DV + d (OS)	V + d	0.55	LogNormal	(0.288 – 0.959)	NMA
K + d (PFS)	V + d	0.532	LogNormal	(0.439 – 0.639)	NMA
K + d (OS)	V + d	0.896	LogNormal	(0.512 – 1.46)	NMA
EP + d (PFS)	P + d	0.545	LogNormal	(0.333 – 0.843)	NMA
EP + d (OS)	P + d	0.618	LogNormal	(0.214 – 1.416)	NMA
IsP + d (PFS)	P + d	0.597	LogNormal	(0.435 – 0.8)	NMA
IsP + d (OS)	P + d	0.686	LogNormal	(0.309 – 1.325)	NMA
Utility weights					
Progression free	N.A.	0.82	Beta	(0.76 – 0.88)	Carlson 2018 (126)
Progressed	N.A.	0.65	Beta	(0.56 – 0.74)	Carlson 2018 (126)
Disutility of Adverse Events	N.A.	0.08	Beta	(0.076 – 0.083)	Carlson 2018 (126)
Incidence risk – Adverse Events					
KR + d	R + d	0.011	LogNormal	(0.006 – 0.02)	NMA
ER + d	R + d	0.009	LogNormal	(0.005 – 0.019)	NMA
IR + d	R + d	0.012	LogNormal	(0.008 – 0.018)	NMA
DR + d	R + d	0.014	LogNormal	(0.006 – 0.03)	NMA
DK + d	V + d	0.018	LogNormal	(0.007 – 0.044)	NMA
FV + d	V + d	0.017	LogNormal	(0.009 – 0.032)	NMA
PV + d	V + d	0.015	LogNormal	(0.007 – 0.031)	NMA
DV + d	V + d	0.024	LogNormal	(0.015 – 0.038)	NMA
K + d	V + d	0.008	LogNormal	(0.005 – 0.013)	NMA
EP + d	P + d	0.05	LogNormal	(0.018 – 0.13)	NMA
IsP + d	P + d	0.08	LogNormal	(0.038 – 0.17)	NMA

CI: confidence interval, D: daratumumab, d: dexamethasone, Dox: doxorubicin E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), N.A.: not applicable, NMA: network meta-analysis, OS: overall survival, P: pomalidomide, PFS: progression-free survival, R: lenalidomide (Revlimid), V: bortezomib (Velcade)

Costs

Costs were valued in NOK for 2022[13] using an extended healthcare perspective. The base-case included all direct medical cost of treatment (drug cost, infusion, injections, and including nursing hours), doctor visits, patient cost for drug administration, visits to doctor and the patient's travel cost for a return journey for drugs administration.

Drug prices were based on confidential, negotiated prices (LIS price), when an agreement existed. If no negotiated price existed, we used the maximum pharmacy sales price (AUP) based on information from the Norwegian Medicines Agency (*Table 19*) (135). The information regarding the dosages and treatment protocols for different drugs and combinations of drugs was taken from The Norwegian Pharmaceutical Product Compendium (Felleskatalogen) summary of product characteristics (136). We assumed patients received medications in tablet form when available. For drugs that could be administered either intravenously or subcutaneously, we have used the subcutaneous mode, the preferred method based on expert advice, unless the intravenous was specified. For the dosing schedules, we assumed a body weight of 72,1 kg and body surface area of 1,82 m^2. Most drug regimens were based on cycles of four weeks, three weeks or both depending on the specific drugs included in the combination therapy.

To calculate costs for each treatment, we created an Excel sheet to track weekly costs:

1. For each treatment, we determined the treatment protocols for each included medication based on dosing information in Felleskatalogen. These are the protocols used in the trial on which a treatment was approved for use by the European Medicines Agency. Most treatments are based on four-week cycles, but some are based on three-week cycles.
2. For each week in a cycle we noted, for each medication, the required dose (in mg) and the days on which the medication was taken. This allowed us to calculate the total doses per medication per week.
3. We then calculated the total cost of each medication by applying the cost per dose. Because medications are often available in packages of varying sizes, and contain tablets, powder for infusions, or liquid for injections, we selected the packaging option that resulted in the lowest price per dose.
4. To the medication costs, we added the relevant costs for regular doctor visits and testing, costs associated with attendance at the clinic for injections or infusions, and patient travel and time costs.
5. For treatment protocols that change after a specified number of cycles, we adjusted the dosage information to reflect the new number of doses per week for each medication in the treatment for which a revised dose was specified.
6. If a treatment protocol limits the number of weeks during which patients receive a specific medication, we stopped accruing those drug costs at that point, but continued to include the costs for regular doctor visits, tests, and patient travel expenses and time.

We converted all weekly cycle costs into monthly costs for our analysis. For drug regimens that followed a 3-week cycle we multiplied total treatment costs during a cycle by 0.75 to convert the cost to a monthly cycle (4 week).

[13] All drug costs are based on 2022 price agreements, or AUP (maximum pharmacy sales price) if no price agreement existed. Other costs, e.g., patient travel, doctor/nurse time, drug administration costs, etc. were calculated earlier using 2021 prices. We decided not to update these prices in the most recent model update, following changes to clinical effect results, as they represent a very small portion of total treatment cost and would not change the overall results of our analysis.

Table 19: Drug prices excluding VAT in NOK

Drug name	Admin mode	Active ingredient per unit	Units per package	Package Cost (NOK)	Source
Lenalidomide (R)	tablet	25 mg/tb.	21 tb.	[redacted]	(LIS)
Bortezomib (V)	inj. s.c.	3.5 mg/vial	3.5 mg	[redacted]	(LIS)
Carfilzomib (K)	i.v. inf.	60 mg/vial	60 mg powder	[redacted]	(LIS)
Daratumumab (D)	inj. s.c.	120 mg/ml	15 ml=1800mg/vial	[redacted]	(LIS)
Elotuzumab (E)	i.v. inf.	400 mg/vial	400 mg/vial	[redacted]	(NOMA)
Ixazomib (I)	tablet	4 mg/tb.	3 tb.	[redacted]	(LIS)
Isatuximab (Is)	i.v. inf.	20 mg/ml	25 ml/vial	[redacted]	(NOMA)
Dexamethasone (d)	tablet	4 mg/tb.	100 tb.	[redacted]	(NOMA)
Panobinostat (F)	tablet	20 mg/tb.	6 tb.	[redacted]	(LIS)
Pomalidomide (P)	tablet	4 mg/tb.	21 tb.	[redacted]	(LIS)

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat, I: ixazomib, inf.: infusion, inj.: injection, Is: isatuximab, i.v.: intravenous, K: carfilzomib (Kyprolis), tb.: tablets, s.c.: subcutaneous, LIS: negotiated prices from Sykehusinnkjøp (Norwegian Hospital Procurement Trust), NOK: Norwegian kroner, NOMA: Norwegian Medicines Agency (Maximum Pharmacy Sales price; AUP), P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib (Velcade). ***Note: Dose protocols for each medication included in a treatment are available in The Norwegian Pharmaceutical Product Compendium (Felleskatologen).***

Administration costs are based on the unit cost database developed by the Norwegian Medicines Agency (*Table 20*) (137). The rate provided for intravenous administration is for 30–60 minutes infusion and includes nurse time, consumables, work costs of the hospital pharmacy, additives, and overhead cost. We have assumed 10 minutes extra nurse time for each extra hour of infusion time. The rate provided for subcutaneous injection includes nurse time, consumables, work cost for the hospital pharmacy, additives, and overhead costs. Based on expert input, we excluded the costs of pre-medication, but have assumed 10 minutes extra nurse time for treatment. The infusion time for drug treatments is based on expert input.

We did not include the cost of adverse events because of large variation in which events were reported across trials and lack of information about when adverse events were most likely to occur. We considered including the costs of reported incidents of sepsis, because according to our clinical expert this is most likely to result in hospitalizations. However, the number of sepsis events was relatively small and seemed to affect a similar percent of patients across treatments. We estimated the potential impact of sepsis hospitalizations on total cost to be quite small relative to the cost of the treatment medications and decided not to include sepsis hospitalization costs in the model.

The cost of doctor visits (out-patient appointment with doctor and blood tests) per treatment cycle are based on expert opinion and charged as monthly visits in the model in accordance with the type of treatment. The costs include consultation time for the doctor and the cost of blood tests. If a treatment regimen is based on oral medications, patient hospital visits occur once per treatment cycle.

In accordance with existing Norwegian guidelines (137), we used the value of free time, as provided in the Norwegian Medicines Agency unit cost database, to account for the time patients spend related to treatment. We also included patients' transportation costs based on the rate in the NOMA unit cost database.

Patients who progress, stop receiving treatment medications, but continue to have doctor visits including necessary blood tests, at the same intervals as before progression, until they begin a new treatment. Because it is difficult to know the amount of time between progression and beginning of a new treatment, we included the costs of doctor visits and blood tests for progressed patient but excluded the costs of travel time and patient time to avoid over-estimating costs for progressed patients.

Lastly, we included the cost of death (end-of-life costs) based on the diagnosis related groups (DRG) reimbursement system assumed for 14 days as an exit cost when patients die (135).

Table 20: Unit costs for hospital resources and patient travel and time, 2021 NOK

Resource use	Unit cost (NOK)	Description (cost calculation)	Source
Doctor consultation	924	1 hour doctor time + blood tests. Once every treatment cycle*.	Expert Opinion from Myeloma center, and Lovdata, 2021 (138)
Drug administration (i.v.)	3,078	30-60 minutes infusion based on average cost.	NOMA (137)
Drug administration (s.c.)	226	15 minutes nurse time for subcutaneous injection.	NOMA (137)
Extra cost nurse (i.v.)	74	Each extra hour of nurse time includes an extra 10 minutes of nurse time.	Expert Opinion from Myeloma center, and NOMA (137)
Transport	1,240	Cost of return journey to the hospital.	NOMA (137)
Patient time	225	Patient time cost for opportunity cost of time lost due to treatment based on average hourly salary in Norway.	NOMA (137)
Cost of Death	59,001		NOMA (137)

** For treatments based on infusions or injections, the frequency of resource use follows the treatment protocols described in The Norwegian Pharmaceutical Product Compendium (Felleskatologen). For tablet-based treatments, resource costs and patient travel and time costs are incurred based on the frequency of doctor visits. i.v.: intravenous, s.c.: subcutaneous, NOK: Norwegian kroner, NOMA: Norwegian medicines agency*

RESULTS

Absolute shortfall

We calculated absolute shortfall in order to determine disease severity. Because our health economic model used three reference treatments rather than a single comparator, we used the reference treatments as a basis for the subgroup analysis to calculate absolute shortfall for each subgroup analysis individually, and then calculated the mean absolute shortfall for all reference groups (*Table 21*) (6).

Table 21: Absolute shortfall for reference treatments

Prognosis	Remaining QALYS	Reference	Absolute shortfall
3.34	15.8	Rd	12.46
0.85	15.8	Pd	14.95
2.50	15.8	Vd	13.30
			Average: 13.57

Average age at diagnosis is assumed to be 65 years old.

The severity calculation shows that based on a mean age at diagnosis of 65, the mean loss of "healthy life-years" for a person receiving any of the reference treatments would range from 12.46 to 14.95 QALYs compared to a healthy individual with no disease. Although the calculations of absolute shortfall were not based on a single common comparator, the differences in absolute shortfall among the three reference groups are relatively small. The ultimate decision about an appropriate threshold based on the computed levels of absolute shortfall lies with the decision-makers.

Total costs and effects based on negotiated drug prices

The results for the reference groups [R + d], [V + d] and [P + d] calculated using probabilistic values in the model are presented in *Table 22*, *Table 23*,and *Table 24*, respectively, for a 20 year-time horizon. We also presented the deterministic base-case results of cost-effectiveness results for all subgroups in the *Appendix 14: Deterministic results for all subgroups*.

To calculate incremental costs and effects for treatments in each reference group, the treatments were ranked in TreeAge from the lowest cost to the highest cost strategy. A treatment is considered "dominated" if it has a combination of higher costs and lower health benefits than another treatment, i.e., it lies to the left and above another treatment on a graph that measures costs on the vertical axis and health benefits on the horizontal axis.[14] If one treatment has both higher costs and higher health benefits than another treatment, then the decision about whether it should be approved, will depend on whether the incremental cost-effectiveness ratio associated with the costlier, more effective treatment relative to the less expensive, less effective treatment is above or below the willingness-to-pay of the decision makers.

Within each reference group (Rd, Vd, and Pd), we report total costs and total benefits for each treatment within the group, but only report incremental cost-effectiveness ratios (ICERs) for treatments not dominated by other treatments. ICERs reported in the tables were calculated based on excluding the dominated strategies.

[14] In some cases, a treatment can be dominated with respect to a combination of two other treatments. This is referred to as being "extended dominated".

In the [R + d] reference group (*Table 22*), Rd had the lowest cost and [DR + d] the highest cost (*Appendix 14: Cost effectiveness frontier [R + d]*). Rd was the treatment with the lowest health effect, measured in quality-adjusted life-years (QALYs). [DR + d] had the highest health effect as measured in QALYs. (see *Appendix 14: Cost effectiveness frontier [R + d]*).

In the [V + d] reference group [V + d] had the lowest treatment cost while [DK + d] had the highest costs *Table 23*. [V + d] had the lowest effect (QALYs) and [DV + d] had the highest effect (QALYs) in the model (see *Appendix 14: Cost effectiveness frontier [V + d]*).

In the [P + d] reference group (*Table 24*), treatment costs were lowest for [P + d] and highest for [IsP + d]. However, [P + d] had the lowest effect and [EP + d] had the highest effect of treatment in term of QALYs (see *Appendix 14: Cost effectiveness frontier [P + d]*).

For each subgroup in *Table 22* ([R + d]), *Table 23* ([V + d]) and *Table 24* ([P + d]) we also reported health effects in terms of life-year gains, however, these were not used to calculate incremental cost-effectiveness ratios.

Cost-Effectiveness

In the cost-effectiveness analysis for the [R + d] group only three strategies were not dominated by other treatments ([R + d], [IR + d], [DR + d], *Table 22*). The ICER for [IR + d] compared to [R + d] was NOK [redacted] per QALY, while the ICER for [DR + d] compared to [IR + d] was NOK [redacted] per QALY. In the [V + d] group, only [V + d] and [DV + d] were not dominated by other treatments. The ICER of [DV + d] compared to [V + d] was NOK [redacted] per QALY (*Table 23*). In the [P + d] group only [P + d] and [EP + d] were not dominated by other strategies. The ICER for [EP + d] compared to [P + d] was NOK [redacted] per QALY (*Table 24*).

Table 22: Cost-effectiveness analysis for reference group [R + d]

Strategy	Cost (NOK)	Incr Cost (NOK)	Effect (QALYs)	Incr. effect	ICER (NOK)	ICER vs	Effect (Life-years)
R + d	[redacted]		2.90				4.00
IR + d	[redacted]	[redacted]	3.82	0.93	[redacted]	R + d	5.35
KR + d	[redacted]	[redacted]	3.63	- 0.20	Dominated	IR + d	5.00
ER + d	[redacted]	[redacted]	3.51	- 0.31	Dominated	IR + d	4.87
DR + d	[redacted]	[redacted]	4.31	0.48	[redacted]	IR + d	5.83

ICER: incremental cost-effectiveness ratios, Incr.: incremental, NOK: Norwegian krone, QALY: quality adjusted life year, The ICER values were rounded to the nearest thousand.

Table 23: Cost-effectiveness analysis for reference group [V + d]

Strategy	Cost (NOK)	Incr Cost (NOK)	Effect (QALYs)	Incr effect	ICER (NOK)	ICER vs	Effect (Life-years)
V + d	[redacted]		2.24				3.25
FV + d	[redacted]	[redacted]	2.50	0.26	[redacted]	V + d	3.58
PV + d	[redacted]	[redacted]	2.46	- 0.04	Dominated	FV + d	3.50
K + d	[redacted]	[redacted]	2.59	0.09	[redacted]	FV + d	3.66
DV + d	[redacted]	[redacted]	3.63	1.39	[redacted]	V + d	5.13
DK + d	[redacted]	[redacted]	3.37	- 0.26	Dominated	DV + d	4.74

ICER: incremental cost-effectiveness ratios, Incr.: incremental, NOK: Norwegian krone, QALY: quality adjusted life year, The ICER values were rounded to the nearest thousand kroner.

Table 24: Cost-effectiveness analysis for reference group [P + d]

Strategy	Cost (NOK)	Incr Cost (NOK)	Effect (QALYs)	Incr effect	ICER (NOK)	ICER vs	Effect (Life-years)
P + d	[redacted]		0.81				1.15
EP + d	[redacted]	[redacted]	1.39	0.58	[redacted]	P + d	1.98
IsP + d	[redacted]	[redacted]	1.19	- 0.21	Dominated	EP + d	1.72

ICER: incremental cost-effectiveness ratios, Incr.: incremental, NOK: Norwegian krone, QALY: quality adjusted life year, The ICER values were rounded off the nearest thousand kroner.

Probabilistic sensitivity analysis

Results for the [R + d] reference group

The cost effectiveness acceptablility curves (CEACs) for [R + d] group (*Figure 21*) showed that for the willingness-to-pay values below NOK [redacted] per QALY only [R + d] had highest probability of cost-effectiveness (99% –100%). However, the probability of cost effectiveness for [R + d] decreased significantly for higher willingness-to-pay values such as above NOK [redacted] per QALY at which [IR + d] had highest probability of cost effectiveness (47%). The probability of [IR + d] being cost effective reached its highest level (50%) at the willingness-to-pay threshold of NOK [redacted] per QALY (*Figure 21*).

For the willingness-to-pay values above NOK [redacted] per QALY only [DR + d] was found to be cost effective with a lower probability cost effectiveness compared to [IR + d] (38% – 40%). The CEAFs presented in the *Appendix 14: Cost Effectiveness Acceptability Frontiers for all subgroups*, show the strategies that have the highest probability of cost effectiveness for a given range of willingness-to-pay threshold.

Figure 22 presents the uncertainty among all treatment strategies by showing the variation of combinations cost and effects in the probabilistic simulations. [ER + d] was had the highest uncertainty in costs and [DR + d] in effects.

Figure 21: Cost effectiveness acceptability curves for the [R + d] reference group
The graph presents all treatments and their respective probability of cost effectiveness for a range of willingness-to-pay thresholds based on results the probabilistic sensitivity analysis with 10,000 Monte Carlo simulations. D: daratumumab, d: dexamethasone, E: elotuzumab, I: ixazomib, K: carfilzomib, R: lenalidomide

Figure 22: Scatter plot of points representing combinations of costs and effects from the probabilistic sensitivity analysis for the [R + d] reference group
D: daratumumab, d: dexamethasone, E: elotuzumab, I: ixazomib, K: carfilzomib, R: lenalidomide

Results for [V + d] reference group

The CEACs for [V + d] subgroup indicated high probability of cost effectiveness for [V + d] for the willingness-to-pay values lower than NOK [redacted] per QALY (92% – 100%) and a decreasing probability of cost effectiveness for the willingness-to-pay greater than NOK [redacted] per QALY (*Figure 23*).

The probability of cost effectiveness for a willingness-to-pay value above NOK [redacted] per QALY favoured [DV + d], as it was found to have a high probability of cost effectiveness compared to all other strategies (ranging between 38% – 60% as willingness-to-pay increased). The CEAFs presented in the *Appendix 14: Cost Effectiveness Acceptability Frontiers for all subgroups* showed the same results as the CEACs.

Figure 24 presents the uncertainty among all strategies (i.e., treatments) by showing the spread of cost and effects in the probabilistic simulations. [DK + d] was found to have the highest variation of cost and effects, indicating highest uncertainty.

Figure 23: Cost effectiveness acceptability curve for the [V + d] reference group
The graph presents all treatments and their respective probability of cost effectiveness for a range of willingness-to-pay thresholds based on results the probabilistic sensitivity analysis with 10,000 Monte Carlo simulations. D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat., I: ixazomib, K: carfilzomib, P: pomalidomide, V: bortezomib

Figure 24: Scatter plot of points representing combinations of costs and effects from the probabilistic sensitivity analysis for the [V + d] reference group
D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat., I: ixazomib, K: carfilzomib, P: pomalidomide, V: bortezomib

Results for [P + d]

The results for [P + d] subgroup indicated [P + d] had highest probability of cost effectiveness for a willingness-to-pay values of less than NOK [redacted] per QALY (95% - 100%) (*Figure 25*), and the probability cost effectiveness changed in the favour for [EP + d] above the WTP values of NOK [redacted] per QALY, as [EP + d] was found to have 45% - 60% probability of cost effectiveness (the highest being at willingness-to-pay of NOK [redacted]).

The CEAFs presented in the *Appendix 14: Cost Effectiveness Acceptability Frontiers for all subgroups* clearly show the strategies that have the highest probability of cost effectiveness for a given range of willingness-to-pay threshold.

[EP + d] and [IsP + d] were both found to have high uncertainty for the variation in costs and effects in the probabilistic simulation (*Figure 26*).

Figure 25: Cost effectiveness acceptability curve [P + d]

The graph presents all treatments and their respective probability of cost effectiveness for a range of willingness-to-pay thresholds based the probabilistic sensitivity analysis with 10,000 Monte Carlo simulations. d: dexamethasone, E: elotuzumab, Is: isatuximab, P: pomalidomide

Figure 26: Scatter plot of points representing combinations of costs and effects in the probabilistic sensitivity analysis for [P + d] reference group
d: dexamethasone, E: elotuzumab, Is: isatuximab, P: pomalidomide

One-way sensitivity analysis

[R + d] Reference Group

Figure 27 shows the impact of changes in drug prices and utility values for the progressed and progression-free health states on the ICERs for the comparison of the treatments [DR + d] vs [IR + d].

Figure 28 shows the impact of changes in drug prices and utility values for the progressed and progression-free health states on the ICERs for the comparison of the treatments [IR + d] vs [R + d].

An attempt to include the HRs for OS and PFS (see *Table 18*) that were used to derive the survival curves for [DR + d] and [IR + d] resulted in "infinite variation" in the ICERS in each comparison, mostly likely because it could have led to intersecting survival curves in each case.

Figure 27: Tornado diagram for reference group [R + d] (ICER [DR + d] vs [IR + d])
Red: ICER with maximum variable value, Blue: ICER with minimum variable value. Parentheses (Base-case value, range based on impact on ICER). d: dexamethasone, ICER: incremental cost effectiveness ratios, p_D: price of daratumumab, P_Ix: price of ixazomib, PFSS: progression-free health state, PPSS: progressed health state, p_R: price of lenalidomide

Figure 28: Tornado diagram for reference group [R + d] (ICER [IR + d] vs [R + d])
Red: ICER with maximum variable value, Blue: ICER with minimum variable value. Parentheses (Base-case value, range based on impact to ICER). d: dexamethasone, ICER: incremental cost effectiveness ratios, P_Ix: price of ixazomib, PFSS: progression-free health state, PPSS: progressed health state, p_R: price of lenalidomide

[V + d] Reference group

Figure 29 shows the impact of changes in drug prices and utility values for the progressed and progression-free health states on the ICERs for the comparison of the treatments [DV + d] vs. [V +d].

An attempt to include the HRs for OS and PFS (see *Table 18*) that were used to derive the survival curves for [DV + d] resulted in "infinite variation" in the ICERS in for the comparison of [DV + d] vs. [V +d], mostly likely because it could have led to intersecting survival curves.

Figure 29: Tornado diagram for reference group [V + d] (ICER [DV + d] vs [V + d])
Red: ICER with maximum variable value, Blue: ICER with minimum variable value, ***Feil! Fant ikke referansekilden.*** *provides the variable descriptions, Parentheses (Base-case value, range based on impact to ICER). ICER: incremental cost effectiveness ratios. d: dexamethasone, ICER: incremental cost effectiveness ratios, p_D: price of daratumumab, PFSS: progression-free health state, PPSS: progressed health state, p_V: price of bortezomib*

[P + d] Model

Figure 30 shows the impact of changes in drug prices and utility values for the progressed and progression-free health states on the ICERs for the comparison of the treatments [EP + d] vs. [P +d].

An attempt to include the HRs for OS and PFS (see *Table 18*) that were used to derive the survival curves for [EP + d] resulted in "infinite variation" in the ICERS in for the comparison of [EP + d] vs. [P +d], mostly likely because it could have led to intersecting survival curves.

Figure 30: Tornado diagram for reference group [P + d] (ICER [EP + d] vs [P + d])
Red: ICER with maximum variable value, Blue: ICER with minimum variable value. Parentheses (Base-case value, range based on impact to ICER). C_iv: infusion cost, d: dexamethasone, ICER: incremental cost effectiveness ratios, p_E: price of elotuzumab, p_P: price of pomalidomide, PFSS: progression-free health state, PPSS: progressed health state

Scenario analyses (Dose Reduction)

Although we intended to perform an analysis to determine the effect of dose reductions resulting from treatment discontinuation by comparing median time-to-treatment-discontinuation with median time-to-progression, we found that in three instances time-to-discontinuation was longer than time-to-progression – a result that is inconsistent with clinical experience – so we discarded the analysis.

Discussion

Clinical effect and safety

Key findings of systematic review

We have systematically reviewed the literature on clinical efficacy and safety for disease modifying treatments of RRMM. We performed a NMAs to facilitate comparisons between treatments that have not been directly compared in any of the included studies. NMAs were performed on five pre-selected outcomes: OS, QoL, SAE, PFS, and discontinuation due to advers events. The evidence base for the analysis comprised of 51 articles reporting on 37 RCTs, all studying the effect of various treatment regimens containing at least one of the following drugs: bortezomib (V; Velcade), carfilzomib (K), daratumumab (D), elotuzumab (E), ixazomib (I), isatuximab (Is), lenalidomide (R; Revlimid), pomalidomide (P) and panobinostat (F).

For non-refractory patients, the highest ranked relevant treatment regimens for our selected outcomes were: [DR + d] for OS, [KR + d] for QoL, [V + d] for SAE, [EP + d] for PFS, and [IsP + d] for discontinuations due to adverse events, respectively.

Radar plots illustrate the relative efficacy for each relevant treatment regimen with respect to our pre-selected outcomes. For non-refractory patients, examples of relevant treatment regimens with favorable HRs for OS that also rank highly with respect to other outcomes include [EP + d], [IsP + d], [DK + d], [KR + d], [DR + d] and [DV + d] (*Figure 15*). For patients who are refractory to lenalidomide (R) (*Figure 16*), bortezomib (V) (*Figure 17*), and lenalidomide (R) and bortezomib (V) (*Figure 18*), examples of relevant treatment regimens with favorable HRs for OS that also rank highly with respect to other outcomes, include [DK + d], [EP + d] and [IsP + d].

However, comparing the overall effect of the various treatment regimens across outcomes is challenging, as not all treatment regimens have data on all outcomes, and some treatment regimens show better effect on one outcome, while having poorer effect on other outcomes, and vice versa. While radar plots such as *Figure 15*, *Figure 16*, *Figure 17*, and *Figure 18* may be useful for understanding tradeoffs between efficacy and safety, they should not be interpreted in isolation. Furthermore, comparing the different treatment regimens is complicated further because our confidence in the results varies considerably across treatment regimens and outcomes.

The certainty of evidence from the systematic review

As previously described, we used the GRADE approach in assessing the certainty of evidence. The main advantage of using GRADE is that it makes our judgements transparent and open to criticism. Even though the GRADE approach provides a framework to evaluate the certainty of evidence in a systematic manner, it still relies on subjective judgement. We therefore acknowledge that others may rate the evidence differently than we have.

We did not assess certainty of evidence of effect estimates calculated across networks, because the GRADE approach has not yet been extended to component NMA. The effect estimates are instead calculated based on an assumption that allows the available

evidence to be synthesized within a single model, e.g., common treatment components (see the *Method chapter: Dealing with disconnected networks of evidence* for a more detailed description). As such, we caution against putting too much weight on these effect estimates

Strengths and weaknesses of the method and evidence base

A general strength of this HTA is that the work has been performed in a systematic manner in accordance with our published project plan (*Appendix 2*). Throughout the process, at least two researchers have independently selected studies and extracted data, as well as independently assessed the methodological quality of the included studies (risk of bias), and our confidence in the results (GRADE). As such, we are confident that we have taken reasonable steps to produce a trustworthy health technology assessment. Furthermore, as we have regularly updated our systematic literature search throughout this process, we are confident that we have identified all relevant studies published prior to January 2022.

We identified very few RCTs for each specific treatment regimen. Because of this, we chose to pool all treatment regimens containing the same combinations of drugs independent of posology (i.e., dose, administration form, etc.), to ensure having sufficiently connected networks in the NMA. Also, we chose to regard all treatment regimens containing the same cytostatic drugs, but different glucocorticoids (e.g., dexamethasone and prednisone) as the same treatment regime (see the *Method chapter: Treatment definition* for a more thorough description). For example, if treatment regimen 1 contained drug A + drug B + dexamethasone, and treatment regimen 2 contained drug A + drug B + prednisone, we regarded these treatment regimens as the same, with the abbreviated denotation [AB + d] (d is dexamethasone or another glucocorticoid). Despite this, the effect estimates for the individual treatment regimens are still based on a limited pool of evidence, which is a clear limitation.

Potential biases in review process

Pooling all treatment regimens across dose, administration form, and dose interval, caused several studies to be reduced to "single arm" trials without a comparison. Consequently, these studies were left out of our statistical analysis (45, 52, 53, 62, 63, 74, 84, 90, 94, 97, 103, 104, 106, 109). In doing so, potential nuances in treatment effect due to posology may have been lost. In addition, a few articles were omitted from the statistical analysis of QoL, due to the required data not being available in a format that we needed, and/or not compatible with our planned meta-analysis (26, 49, 51, 61, 70, 79, 80, 86, 93). As such, the meta-analysis for QoL is based on very limited data from few studies (25, 30, 59, 66, 70, 73, 75, 89, 95), and the results should therefore be interpreted with extreme caution. All included studies that were omitted from our statistical analysis are presented with reason for omission in *Appendix 7*.

The main outcome in this report is OS, and our analysis included survival data published from 27 RCTs. Several ongoing studies are investigating OS, but as they have not presented their immature data (with HRs), we could not include these studies in our meta-analysis. However, some of the studies included in our analysis of OS present data (with HRs) that were immature at the time of publication. We cannot disregard the possibility that studies publishing immature survival data have found the results to be more promising than studies that did not publish immature data. Furthermore, the immature survival data may change over time with longer follow-up. As such, there are potential biases that could affect our meta-analysis of OS, and therefore we strongly advice to interpret these results with caution.

Strengths and weaknesses of the statistical analysis

The main strengths of the statistical analysis are that we used appropriate methods to synthesize the available evidence and provide estimates of relative treatment effect and reported the analyses thoroughly and transparently. In addition, based on input from

our clinical expert, we systematically reviewed the evidence for effect modification of HRs for OS and PFS with respect to refractory status and number of previous lines of treatment. This work is currently available as a preprint (34). Briefly, this work suggests that the evidence for effect modification is very weak, and simulations suggest that any effect modification, should it occur, is unlikely to substantially impact random effects component NMAs, as used herein. However, the following potential limitations should also be taken into consideration.

Due to the nature of the included studies, we were unable to use the planned treatment definitions that would have allowed us to characterize differences in relative treatment effect associated with differences in posology.

Because the included studies defined disconnected networks, we had to use component NMA. While we prespecified that we might do this, we were unable to formally test the assumption that the treatments can be modelled in this way. We therefore rely on a conceptual evaluation of the validity of this assumption across networks, however we did confirm that estimates are consistent between regular NMAs of each subnetwork and component NMA of the full disconnected networks.

In principle we could have taken the component approach further and treated each distinct drug as a component, such that triplet regimens would have been modelled as a linear combination of three components, for example. We chose not to do this because we judged the required assumption to be harder to justify — for example, it is possible that two “active” drugs used in combination may have a very different effect than the sum of the two used alone. We hope to explore this approach in subsequent research, but it was not possible within the scope of this project.

We combined trial arms as necessary to include some studies that had compared different posologies of the same drug or drug combination. The advantages of this approach are that it allows more evidence to be included in meta-analysis and it is the approach recommended by Cochrane. However, the method may lead to estimates of relative treatment effect that are biased toward posologies not used in clinical practice (e.g., toward a lower dose found not to be sufficiently effective, or toward a higher dose found to be poorly tolerated). If instead we had taken a maximalist stance with respect to the granularity of treatment definition, drawing hard distinctions between dose, schedule, and route of administration, it is unlikely that evidence synthesis would have been possible.

With respect to QoL (a primary outcome), we extracted and analysed difference in mean EORTC QLQ-C30 global health scores. QLQ-C30 is not the only tool used to assess QoL in this literature, however it was the one for which data were most complete. In principle we could have incorporated data from other tools by transforming to a common scale such as Standardized Mean Difference (SMD). However, we judged that this would not meaningfully increase the amount of data available and would substantially complicate the analysis and its interpretation.

QLQ-C30 data were typically not reported by the included studies or could not be extracted in a way that easily facilitates meta-analysis. While point estimates were often available, statements of precision were not. It is essential to characterize precision because this information is necessary to perform meta-analysis. It was therefore necessary to impute precision, which we did by assuming a common standard deviation (cf. standard error) for all included studies. While we evaluated this assumption where possible and found broad agreement with published results, the approach can only provide approximate standard errors. Our meta-analytical estimates for this outcome should therefore be interpreted cautiously.

We analysed SAE (i.e., adverse events grade 3 and 4) using incidence rate ratios (IRRs). Because information on total exposure was lacking, it was necessary to impute point estimates, using component network meta-analytical estimates of HRs for OS as a proxy for relative expected OS time. This is likely to introduce some bias. However, note that

we accounted for the uncertainty on our estimates of HR used in the imputation. We judged that it was reasonable to perform this imputation for SAEs (rather than fail to analyze this outcome) because SAE is a patient important outcome. However, we chose not to report the planned analysis of adverse events (which would require the same type of imputation) to limit possible overinterpretation of results. This was a "borderline" decision, and others may reasonably have chosen to also report estimates for adverse events, or not to have meta-analysed either outcome in this way.

The key assumption underpinning the use of IRR is that events (i.e., SAEs) occur independently of one another. This might not be true from the perspective of an individual patient: a patient who experiences one SAE might be at higher or lower risk of a subsequent SAEs, because of the previous one. However, the assumption is probably reasonable from the perspective of a health system, in which patients can be viewed as presenting with SAEs at an average rate. We note that HTA are generally best interpreted from this health system (i.e., population) perspective, rather than the perspective of an individual patient.

Due to limitations of the available data and software used, it was not possible to use funnel plots to assess possible publication bias, as we had planned to do. It is therefore possible that the literature on this disease does not include studies that estimated smaller or even opposite relative treatment effects compared to those we were able to include.

We were able to analyse inconsistency as planned by comparing direct, indirect, and network estimates (and found no evidence of inconsistency), but we were unable to perform other planned analyses because there were no closed loops.

While we planned to account for heterogeneity using random effects, it was not possible to estimate variance components for all outcomes, and in these cases the model used is equivalent to a fixed (common) effects model.

Finally, the software we used did not allow us to report prediction intervals, so we are unable to make predictions about what future studies may report.

Overall completeness and applicability of evidence from systematic review

Most of the included RCTs are international studies, predominately conducted in North America and Europe. Only a few of the included studies report the distribution of ethnicity/race within their study population. In these studies, the majority of participants are "white" or "Caucasian", and only a small number are of other ethnicities such as "black" or Asian. A few included studies have been conducted either regionally, e.g., in Scandinavia (25), or nationally, e.g., in China (37). The China Continuation Study is a small, local expansion study based on the large, international TOURMALINE-MM1 study (95), but with an entirely Chinese study population (37). The study participants in the China Continuation study are slightly younger, with shorter disease duration, but still with similar distribution regarding the ISS stage (multiple myeloma International Staging System), which suggest a more aggressive disease compared with the general study population in the TOURMALINE-MM1 study (37, 95). These disparities could be a result of ethnicity, as ethnic differences are evident in the epidemiology of multiple myeloma: "Blacks" have increased risk of multiple myeloma compared with "Whites", but possibly with less aggressive disease (139, 140). Asians on the other hand, have lower risk of multiple myeloma than "Whites", though the incidence seems to be increasing (140-142).

The China Continuation study also showed that the patients had a higher exposure to ixazomib (i.e., area under the curve = blood concentration over time), than patients in the TOURMALINE-MM1 study, despite having the same posology for the same treatment regimen (37, 95). Ethnic diversity in drug response has been shown in several studies, and Asian patients have been shown to be more susceptible to the effect of some chemotherapeutics agents (143-145).

The results for OS in the China Continuation Study and the LEPUS study seem to deviate from the findings in the larger international studies (TOURMALINE-MM1 and CASTOR) for the same comparisons. This can be interpreted as an indication that ethnicity is an effect modifier. Interestingly, when considering the PFS results, we see no discrepancies between the Chinese and international studies. Further research may reveal that ethnicity is indeed an effect modifier, and that different ethnicities should be considered separately. In our view there is currently not sufficient evidence for such an approach, and we have therefore considered all ethnicities together, as is current practice among RRMM trialists and systematic review authors.

In light of all this, we caution against overinterpreting our results regarding OS, as we are uncertain about how applicable they are to a Norwegian setting.

Still, one clinical expert had major concerns about the inclusion of the China Continuation Study, which compared [IR + d] and [R + d] in that 1) the results in this specific trial are not applicable to the Norwegian setting because all patients were Chinese and 2) patients had received care before and during the trial that likely differed from current Norwegian clinical practice. We therefore performed a non-prespecified sensitivity analysis to address these concerns. that The sensitivity analysis demonstrates that excluding the China Continuation Study (which meets our prespecified inclusion criteria) has a substantial impact on the results, leading to a marked drop in the ranking of [IR + d] for OS but not PFS (*Figure 8* and *Figure 13*).

This was not surprising, since our assessment of the certainty of the underlying evidence for the [IR + d] ranking in our main analysis was very low, largely due to the discrepancy in the findings from China Continuation Study and TOURMALINE-MM1 (which also compared [IR + d] and [R + d]). It may appear that excluding the China Continuation Study solves the discrepancy, but the result may still be misleading for at least two reasons. First, using trial results to justify deviating from protocol and excluding trials that meet prespecified inclusion criteria may lead to results that do not answer the original research question, and is usually cautioned against, e.g. in The Cochrane Handbook chapter on heterogeneity: "*In general it is unwise to exclude studies from a meta-analysis on the basis of their results as this may introduce bias*" (146). Second, there are aspects of both the TOURMALINE-MM1 and the China Continuation Study that may limit their applicability to the Norwegian context. As pointed out by the TOURMALINE-MM1 investigators: In the China Continuation Study only 50% of the patients who experienced disease progression received any subsequent therapy, and those that did lacked "*access to the broader range of approved or investigational agents and regimens available to patients in North America and Europe*" (38). On the other hand, like the China Continuation Study, TOURMALINE-MM1 was double blinded, and the blinding remained after disease progression, which meant that "*equal proportions of blinded patients received PI or non-PI treatment as next-line therapy in each arm*" (38). The TOURMALINE-MM1-investigators state that "*patients progressing on placebo-Rd*" … "*were more likely to remain PI-sensitive and therefore benefit from PI-based next-line therapy—representing a de facto crossover*" (38). In addition: "*Daratumumab became clinically available for the treatment of RRMM shortly after completion of enrollment to TOURMALINE-MM1. Among TOURMALINE-MM1 patients receiving subsequent daratumumab, there was an OS trend in favor of placebo-Rd (HR 1.15). We hypothesize that this could be because of placebo-Rd patients receiving daratumumab earlier and in larger numbers than ixazomib-Rd patients*" (38). In summary, there are problems with the interpretation of the estimates of HRs for OS in both these studies, and we believe the most appropriate option is to include both studies in our main analysis, in line with our protocol and Cochrane handbook guidance, while emphasising that the certainty of our findings on the effectiveness of [IR + d] on OS, is very low.

It is common for studies that compare the same treatments to provide differing results. This is called heterogeneity and is typically addressed in meta-analysis using random effects, as in this report, which estimate the average of the treatment effects that such

trials estimate. It is then important that estimates of average effect are not misinterpreted as being equivalent to study-level estimates, and that possible explanations for heterogeneity are considered (e.g., differences due to trial design and conduct, as well as the play of chance).

Generalisability to a clinical setting

The goal of systematic reviews is to summarise available evidence that meet a defined set of criteria. Regardless of the amount and quality of evidence that can be included in a systematic review, it is important to remember that systematic reviews and meta-analyses, as well as single studies, typically report treatment effects that do not necessarily reflect the treatment effect for an individual patient. In other words, our findings are probably most usefully interpreted at the health system, rather than individual patient level.

Consistency of systematic review with other reviews

We have identified four systematic reviews that have evaluated the efficacy of treatment regimens for relapsed and/or refractory multiple myeloma, published between 2017 and 2020 (12, 13, 17, 35). All studies reported on PFS (12, 13, 17, 35). In addition, Luo et al also reported on OS (12).

Similar to our results, Luo *et al* found [IR + d] and [DR + d] to be the two highest ranked treatment regimens for OS (12). In contrast, all four systematic reviews found the triple combination daratumumab + lenalidomide + dexamethasone [DR + d] to be the highest ranked treatment regimen for PFS (12, 13, 17, 35), whereas we found [EP + d] to be the highest ranked treatment regimen. However, four of the five highest ranked treatment regimens in our analysis of PFS were not included in the other systematic reviews, likely because the study results were not available. Also, while the systematic review of Dhakal *et al* was published in 2020, they searched for literature published up to July 2018 (35). As such, findings for several of treatment regimens we included had not yet been published when the earlier systematic reviews conducted their literature search. Our work therefore adds information not available in other similar reviews.

Similar to the approach we took, all four systematic reviews assumed that dose, administration methods, and dose intervals have little or no effect, and therefore pooled treatment regimens that contain the same drugs but differ in posology (12, 13, 17, 35). According to our clinical experts, this assumption is valid. These other systematic reviews made assumptions similar to the ones we made but they could not have used the component NMA model we employed because it was published after the four systematic reviews. Firstly, in order to connect all treatment options into one network, three of the systematic reviews assumed that the relative efficacy of [V + d] vs [d] is equal to that of [V] vs [d], and that [T + d] vs [d] is equal to that of [T] vs [d] (12, 17, 35). Whereas van Beurden-Tan *et al* states that this assumption seemed valid from a clinical perspective (17), our clinical experts advised against it. Our approach explicitly does not assume that regimens containing a glucocorticoid have the same effect as regimens that do not.

Secondly, three systematic reviews used time-to-progression as a proxy for PFS in cases where HRs and 95% confidence intervals were not available (13, 17, 35). We have not made this assumption, and any studies presenting data without HRs have therefore been omitted from the meta-analysis of PFS. However, we acknowledge that not imputing HRs from times-to-progression was a borderline decision and had there been more heterogeneity in reporting we would have probably chosen to do it.

Health economic evaluation

Key findings of health economic evaluation

We conducted a cost utility analysis of 13 treatments for patients with relapsed and/or refractory multiple myeloma using a partitioned survival analysis model. We expressed health outcomes in QALYs and costs in NOK. We subdivided the model into three treatment groups, each based on one of three reference treatments: lenalidomide (Revlimid) + dexamethasone [R + d], bortezomib (Velcade) + dexamethasone [V + d], and pomalidomide + dexamethasone [P + d]. After deriving OS and PFS curves for each of the reference treatments, we used HRs from the NMA, conducted as part of the clinical effect and safety portion of the project, to generate survival curves for the other treatments in each of the groups.

Costs included all direct medical cost associated with treatment (drug costs, infusion, injections, nursing hours), doctor visits and test, travel costs for patients, time associated with treatment and doctor visit, and costs related to end-of-life care. Model results were expressed as ICERs, which reflect the extra cost of one additional QALY gained. The deterministic model results, in which treatments were ranked in TreeAge from lowest cost to highest cost strategy within the treatment groups gave the following results, with results rounded to the nearest 1 000 kroner:

In the [R + d] group, [R + d] was the lowest cost treatment at NOK [redacted], with 2.90 QALYs gained. Only two other treatments were not dominated by other treatments: [IR + d], with costs of [redacted] and 3.71 QALYs and an ICER of NOK [redacted] compared to [R + d]. [DR + d] had costs of [redacted] and 4.08 QALYs, and an ICER of NOK [redacted] compared to [IR + d].

In the [V + d] group both [V + d] and [DV + d] were not dominated by other treatments, [DV + d] with costs of [redacted] and 3.48 QALYs, with an ICER of NOK [redacted] compared to [V + d].

In the [P + d] group both [P + d] and [EP + d] were not dominated by other treatments. [P + d] had costs of [redacted] and 0.81 QALYS and [EP + d] had costs of [redacted] and QALYs of 1.20 for an ICER of NOK [redacted].

The results for the probabilistic analysis were presented using Cost-Effectiveness Acceptability Curves (CEACs), which make it possible to determine which treatments had the highest probability of being cost effective for varying levels of willingness-to-pay.

In the [R + d] group, only [R + d] had the possibility of being cost-effective at willingness-to-pay levels below NOK [redacted]. IRd had a 47% probability of cost-effectiveness at willingness-to-pay levels of above NOK [redacted]. [DR + d] had a 38% - 40% probability of cost-effectiveness but only at willingness-to-pay levels of approximately [redacted] kroner and above.

In the [V + d] group, [V + d] was found to have a high probability of cost effectiveness below the willingness-to-pay of NOK [redacted]. For the [DV + d] probability of cost effectiveness was only higher than other treatments above a willingness-to-pay of NOK [redacted] but with a lower probability of just 38%.

In the [P + d] group, [P + d] had the highest probability of being cost effective over wide willingness-to-pay ranges. [EP + d]'s probability of being cost-effective was 45% - 60%, but only at a willingness-to-pay of more than NOK [redacted].

Treatments are generally considered cost-effective if the ICER is below the willingness-to-pay for an extra (QALY). Although Norway does not have an official threshold value for willingness-to-pay for an extra QALY, the current priority criteria show acceptance for the principle that severity of the disease should be considered when making this decision. We computed severity, measured as "absolute shortfall" for each of the

reference treatments and found that they ranged from 12.46 to 14.95 lost healthy life-years (average: 13.57). We leave it to the decision-makers to determine the appropriate level of willingness-to-pay based on severity.

Strengths and weaknesses of the health economic evaluation

The primary contribution of this health economic evaluation is that it constitutes the largest cost-effectiveness analysis of potential treatments for relapsed/refractory multiple myeloma ever conducted. We evaluated thirteen separate treatments for a disease, whose treatment pathways are extremely complex, without having access to patient level data. As our search for other cost-effectiveness analyses of treatments for multiple myeloma revealed, most other cost-effectiveness analyses of multiple myeloma treatments have examined a single intervention and an established comparator, a few made comparisons between two interventions and a single comparator, or between a single proposed intervention compared to different established treatments. The largest number of treatments included in a single analysis was in a cost-effectiveness analysis performed by Carlson, et al., which included only six treatments compared in a partitioned survival analysis in which the model was populated with HRs from a small NMA that included only those treatments (126).

The most important weakness of the health economic evaluation and, perhaps, of the clinical effect section of this project, is the high degree of uncertainty surrounding the results. Much of this uncertainty was unavoidable because our lack of access to patient level data limited our ability to derive parametric survival curves that captured relevant patient information. As the survival curves form the basis for estimating total costs and total health benefits accumulated during treatment, less certainty in the survival curves leads to more uncertainty in the results of the analysis. To investigate the cost-effectiveness of treatments of interest, we needed to use HRs from the NMA of OS and PFS to generate survival curves for each treatment relative to its relevant reference treatment (*Table 17*).

As noted in the discussion describing our decision to divide the model into three treatment groups, there are a variety of potential shortcomings in both the methods used to conduct the NMA and those to generate the survival curves needed for the economic model, that could have increased the general level of uncertainty around our results. These include, but are not limited to (1) use of mean HRs rather than study-level HRs, undetected inconsistency in the NMA, (2) violations of the proportional hazards assumption within one or more of the included studies, (3) limitations in the approach used to impute reference survivor functions, and (4) the fact that the included studies, and hence the NMA and partitioned survival analysis, did not account for correlations between PFS and OS.

While the HRs for PFS were all statistically significant, none of the HRs for OS were, and all HR had wide confidence intervals which introduced more uncertainty in the results. Although we tried to conduct a one-way sensitivity analysis to determine the impact of variation in the HRs on cost-effectiveness results, we were unable to do so, because that analyses yielded results showing “infinite variation”. This was likely by intersecting survival curves caused by wider confidence intervals around the HRs.

Because costs of myeloma medications are quite high, it is important to account for patients who either receive reduced doses or who discontinue treatment before disease progression. Not doing so would result in an over-estimation of treatment costs and increases in the incremental cost-effectiveness ratios. Had survival curves for time to treatment discontinuation been available for all treatments, we could have used that information to determine the point at which treatment costs of treatment should stop being “accumulated” in the model. However, that information was not available so instead we used information about median time-to-discontinuation and median time-to-progression to determine the period of time during which discontinued treatment occurred. While results of the dose reduction analysis seemed to provide reasonable

results for some treatments, for three treatments ([DR + d], [EP + d] and [IsP + d]) the median time-to-discontinuation was greater than the median time-to-progression – a result that is inconsistent with clinical experience – so we discarded the analysis. This was likely a reflection of other uncertainty in the model. As a result, it is possible that total costs are over-estimated for some treatments.

Estimating the five-year budget consequences of introducing new treatments in the health care system is normally an important part of a health-economic analysis. Unfortunately, the complexity of clinical decisions regarding which treatments are most appropriate for different patients makes it impossible to estimate how many patients might receive a given treatment.

Consistency of the health economic evaluation with other studies

It is difficult to compare the consistency of health economic results across countries because of variantions in the structure health systems and the costs associated with health care. Even if health systems were similar comparing calculations of incremental cost-effectiveness ratios across countries would require not just converting from one currency to another, but also taking into account differences in the inflation rates if studies were conducted in different years.

To the extent that different cost-effectiveness analyses report a decision about whether treatments were considered to be cost-effect, those decisions were sometimes, but not always consistent with what we might suspect the decision would be in Norway based on similar decisions that have been made in Norway. However, other factors beside the cost-effectiveness of a treatment might could account for differences in reimbursement choices made in different countries.

Need for further research

It would be useful if future randomised or perhaps registry-based studies were to directly compare the more effective and safe treatments, rather than to use one of the few commonly used controls (i.e., the treatments found at the centres of the networks). In addition to providing direct evidence that would be useful in its own right this could allow fully connected networks to be formed, alleviating the need to assume that treatments can be componentized (and allowing that assumption to be tested). This would also allow for inconsistency in the network of evidence to be assessed via closed loops. It may not be feasible to power superiority trials comparing two highly effective treatments, but noninferiority trials may be tractable and could provide useful information (147). However, treating relapsed and/or refractory multiple myeloma is complicated and depends on several factors including prior treatment regimens, as well as aggressiveness of the disease. As such, we do recognize that conducting an ideal study to compare the most effective treatment options may not be feasible.

Conclusion

It is not possible to draw clear, brief conclusions about the clinical effectiveness of the treatments we examined for several reasons including a high degree of uncertainty across most results; the need to consider different outcomes simultaneously; and different considerations across subgroups of patients (e.g., those who are refractory to different drugs). We infer that there is no single treatment regimen that is superior with respect to all outcomes. The six triplet combinations [EP + d], [IsP + d], [DK + d], [KR + d], [DR + d] and [DV + d] are examples of treatment regimens relevant for non-refractory patients that have clearly favorable HRs for OS, that also are ranked highly with respect to other outcomes. Similarly, the three triplet combinations [DK + d], [EP + d] and [IsP + d] are examples of treatment regimens that are relevant for patients who are refractory to lenalidomide (R) and/or bortezomib (V), and that have clearly favorable HRs for OS, and also ranked highly with respect to other outcomes. However, it is important to note the substantial uncertainty in the evidence underlying these results.

In the [R + d] group, [R + d] was the lowest cost treatment at NOK [redacted], with 2.90 QALYs gained. Only two other treatments were not dominated by other treatments: [IR + d], with costs of [redacted] and 3.71 QALYs and an ICER of NOK [redacted] compared to [R + d]. [DR + d] had costs of [redacted] and 4.08 QALYs, and an ICER of NOK [redacted] compared to [IR + d].

In the [V + d] group both [V + d] and [DV + d] were not dominated by other treatments, [DV + d] with costs of [redacted] and 3.48 QALYs, with an ICER of NOK [redacted] compared to [V + d].

In the [P + d] group both [P + d] and [EP + d] were not dominated by other treatments. [P + d] had costs of [redacted] and 0.81 QALYS and [EP + d] had costs of [redacted] and QALYs of 1.20 for an ICER of NOK [redacted].

References

1. Durie B. Pasient-håndbok: Myelomatose. International Myeloma Foundation; 2017.
2. Cancer Registry of Norway. Cancer in Norway 2020 - Cancer incidence, mortality, survival and prevalence in Norway. Report. Oslo, Norway: Cancer Registry of Norway; 2021.
3. Helsedirektoratet. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av maligne blodsykdommer. Nasjonal faglig retningslinje. Oslo, Norway: Helsedirektoratet; 2020 May 2020. Report No.: IS-2930 Contract No.: IS-2930.
4. Bazarbachi AH, Al Hamed R, Malard F, Harousseau JL, Mohty M. Relapsed refractory multiple myeloma: a comprehensive overview. Leukemia. 2019;33(10):2343-57.
5. Bobin A, Liuu E, Moya N, Gruchet C, Sabirou F, Lévy A, et al. Multiple Myeloma: An Overview of the Current and Novel Therapeutic Approaches in 2020. Cancers (Basel). 2020;12(10).
6. Norwegian Ministry of Health and Care Services. På ramme alvor - Alvorlighet og prioritering. Report. Oslo, Norway: Norwegian Ministry of Health and Care Services,; 2015 October 2015.
7. Folkehelseinstituttet. Slik oppsummerer vi forskning. Håndbok for Nasjonalt kunnskapssenter for helsetjenesten. Oslo: Folkehelseinstituttet; 2015.
8. Ohm IR, C; Hafstad, E; Zinöcker, S; Lund, UH; Desser, AS. Legemidler for behandling av myelomatose. Oslo, Norway: Folkehelseinstituttet; 2020.
9. Botta C, Ciliberto D, Rossi M, Staropoli N, Cuce M, Galeano T, et al. Network meta-analysis of randomized trials in multiple myeloma: efficacy and safety in relapsed/refractory patients. Blood Adv. 2017;1(7):455-66.
10. Chen R, Wang Y, Luan C, Gao C, Zhang X, Chen B. Effect of pomalidomide on relapsed/refractory multiple myeloma: a systematic review and meta-analysis. J Cancer. 2017;8(10):1801-8.
11. Gao X, Shen L, Li X, Liu J. Efficacy and toxicity of histone deacetylase inhibitors in relapsed/refractory multiple myeloma: Systematic review and meta-analysis of clinical trials. Exp Ther Med. 2019;18(2):1057-68.
12. 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 Manag Res. 2018;10:2817-23.
13. Maiese EM, Ainsworth C, Le Moine JG, Ahdesmaki O, Bell J, Hawe E. Comparative Efficacy of Treatments for Previously Treated Multiple Myeloma: A Systematic Literature Review and Network Meta-analysis. Clin Ther. 2018;40(3):480-94 e23.
14. Mushtaq A, Iftikhar A, Hassan H, Lakhani M, Sagar F, Kamal A, et al. Pomalidomide-Based Regimens for Treatment of Relapsed and Relapsed/Refractory Multiple Myeloma: Systematic Review and Meta-analysis of Phase 2 and 3 Clinical Trials. Clin Lymphoma Myeloma Leuk. 2019;19(7):447-61.

15. Shah C, Bishnoi R, Wang Y, Zou F, Bejjanki H, Master S, et al. Efficacy and safety of carfilzomib in relapsed and/or refractory multiple myeloma: systematic review and meta-analysis of 14 trials. Oncotarget. 2018;9(34):23704-17.
16. Sun Z, Zheng F, Wu S, Liu Y, Guo H, Liu Y. Triplet versus doublet combination regimens for the treatment of relapsed or refractory multiple myeloma: A meta-analysis of phase III randomized controlled trials. Crit Rev Oncol Hematol. 2017;113:249-55.
17. van Beurden-Tan CHY, Franken MG, Blommestein HM, Uyl-de Groot CA, Sonneveld P. Systematic Literature Review and Network Meta-Analysis of Treatment Outcomes in Relapsed and/or Refractory Multiple Myeloma. J Clin Oncol. 2017;35(12):1312-9.
18. Weisel K, Sonneveld P, Spencer A, Beksac M, Rizzo M, Xu Y, et al. A comparison of the efficacy of immunomodulatory-free regimens in relapsed or refractory multiple myeloma: a network meta-analysis. Leuk Lymphoma. 2019;60(1):151-62.
19. Zhang T, Wang S, Lin T, Xie J, Zhao L, Liang Z, et al. Systematic review and meta-analysis of the efficacy and safety of novel monoclonal antibodies for treatment of relapsed/refractory multiple myeloma. Oncotarget. 2017;8(20):34001-17.
20. Zheng Y, Shen H, Xu L, Feng J, Tang H, Zhang N, et al. Monoclonal Antibodies versus Histone Deacetylase Inhibitors in Combination with Bortezomib or Lenalidomide plus Dexamethasone for the Treatment of Relapsed or Refractory Multiple Myeloma: An Indirect-Comparison Meta-Analysis of Randomized Controlled Trials. J Immunol Res. 2018;2018:7646913.
21. Zou Y, Ma X, Yu H, Hu C, Fan L, Ran X. Carfilzomib/pomalidomide single-agent or in combination with other agents for the management of relapsed/refractory multiple myeloma: a meta-analysis of 37 trials. Oncotarget. 2017;8(24):39805-17.
22. Covidence systematic review software [Internet]. Available from: https://www.covidence.org/.
23. Ouzzani MH, Hossam; Fedorowicz, Zbys; Elmagarmid, Ahmed Rayyan — a web and mobile app for systematic reviews. Systematic Reviews. 2016;5:201.
24. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
25. Hjorth M, Hjertner Ø, Knudsen LM, Gulbrandsen N, Holmberg E, Pedersen PT, et al. Thalidomide and dexamethasone vs. bortezomib and dexamethasone for melphalan refractory myeloma: a randomized study. Eur J Haematol. 2012;88(6):485-96.
26. Leleu X, Masszi T, Bahlis NJ, Viterbo L, Baker B, Gimsing P, et al. Patient-reported health-related quality of life from the phase III TOURMALINE-MM1 study of ixazomib-lenalidomide-dexamethasone versus placebo-lenalidomide-dexamethasone in relapsed/refractory multiple myeloma. American journal of hematology. 2018.
27. Rohatgi A. WebPlotDigitizer: Version 4.4. 4.4 ed2020.
28. Higgins J, Li T, Deeks J, (editors). Chapter 6: Choosing effect measures and computing estimates of effect. Updated February 2021. In: Cochrane Handbook for Systematic Reviews of Interventions version 62 [Internet]. Available from: https://training.cochrane.org/handbook/current/chapter-06.
29. Higgins J, Eldridge S, Li T, (editors). Chapter 23: Including variants on randomized trials. Updated February 2021. In: Cochrane Handbook for Systematic Reviews of Interventions version 62 [Internet]. Cochrane. Available from: https://training.cochrane.org/handbook/current/chapter-23.
30. Stewart AK, Rajkumar SV, Dimopoulos MA, Masszi T, Špička I, Oriol A, et al. Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. N Engl J Med. 2015;372(2):142-52.

31. Desser A, Ohm I, Giske L, Hafstad E, Næss GE, Rose C, et al. Treatments for relapsing, refractory multiple myeloma: A health technology assessment protocol. Project plan. Norway: Norwegian Institute of Public Health; 2021.
32. Rücker G, Petropoulou M, Schwarzer G. Network meta-analysis of multicomponent interventions. Biometrical Journal. 2020;62(3):808-21.
33. Rücker G, Krahn U, König J, Efthimiou O, Schwarzer G. netmeta: Network Meta-Analysis using Frequentist Methods. R package version 1.3-0. 1.3-0 ed2021.
34. Rose CJ, Ohm IK, Giske L, Næss GE, Fretheim A. Effect modification in network meta-analyses of treatments for relapsing refractory multiple myeloma (RRMM): systematic review, meta-analysis, and simulation. medRxiv. 2022:2022.05.13.22275047.
35. Dhakal B, Narra RK, Giri S, Szabo A, Smunt TL, Ghose S, et al. Association of adverse events and associated cost with efficacy for approved relapsed and/or refractory multiple myeloma regimens: A Bayesian network meta-analysis of phase 3 randomized controlled trials. Cancer. 2020;126(12):2791-801.
36. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol. 2015;15:58.
37. Hou J, Jin J, Xu Y, Wu D, Ke X, Zhou D, et al. Randomized, double-blind, placebo-controlled phase III study of ixazomib plus lenalidomide-dexamethasone in patients with relapsed/refractory multiple myeloma: China Continuation study. J Hematol Oncol. 2017;10(1):137.
38. Richardson PG, Kumar SK, Masszi T, Grzasko N, Bahlis NJ, Hansson M, et al. Final Overall Survival Analysis of the TOURMALINE-MM1 Phase III Trial of Ixazomib, Lenalidomide, and Dexamethasone in Patients With Relapsed or Refractory Multiple Myeloma. Journal of Clinical Oncology. 2021;39(22):2430-42.
39. Lu J, Fu W, Li W, Hu J, An G, Wang Y, et al. Daratumumab, Bortezomib, and Dexamethasone Versus Bortezomib and Dexamethasone in Chinese Patients with Relapsed or Refractory Multiple Myeloma: Phase 3 LEPUS (MMY3009) Study. Clinical lymphoma, myeloma & leukemia. 2021;21(9):e699-e709.
40. Mateos MV, Sonneveld P, Hungria V, Nooka AK, Estell JA, Barreto W, et al. Daratumumab, Bortezomib, and Dexamethasone Versus Bortezomib and Dexamethasone in Patients With Previously Treated Multiple Myeloma: Three-year Follow-up of CASTOR. Clinical lymphoma, myeloma & leukemia. 2020;20(8):509-18.
41. Schünemann HJ, Brozek J, Guyatt G, Oxman AD, (editors). GRADE handbook for grading quality of evidence and strength of recommendation [The GRADE Working Group,]. 2013 [updated Updated October 2013. Available from: https://gdt.gradepro.org/app/handbook/handbook.html.
42. Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.
43. GRADEpro GDT. GRADEpro Guideline Development Tool. developed by Evidence Prime, Inc.: McMaster University; 2015.
44. Kvam AK, Fayers PM, Wisloff F. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. Eur J Haematol. 2011;87(4):330-7.
45. Richardson PG, Jagannath S, Moreau P, Jakubowiak AJ, Raab MS, Facon T, et al. Elotuzumab in combination with lenalidomide and dexamethasone in patients with relapsed multiple myeloma: final phase 2 results from the randomised, open-label, phase 1b-2 dose-escalation study. Lancet Haematol. 2015;2(12):e516-27.
46. White D, Kassim A, Bhaskar B, Yi J, Wamstad K, Paton VE. Results from AMBER, a randomized phase 2 study of bevacizumab and bortezomib versus

bortezomib in relapsed or refractory multiple myeloma. Cancer. 2013;119(2):339-47.
47. Richardson PG, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, Facon T, et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. N Engl J Med. 2005;352(24):2487-98.
48. Richardson PG, Sonneveld P, Schuster M, Irwin D, Stadtmauer E, Facon T, et al. Extended follow-up of a phase 3 trial in relapsed multiple myeloma: final time-to-event results of the APEX trial. Blood. 2007;110(10):3557-60.
49. Lee SJ, Richardson PG, Sonneveld P, Schuster MW, Irwin D, San Miguel JF, et al. Bortezomib is associated with better health-related quality of life than high-dose dexamethasone in patients with relapsed multiple myeloma: results from the APEX study. Br J Haematol. 2008;143(4):511-9.
50. Dimopoulos MA, Terpos E, Boccadoro M, Delimpasi S, Beksac M, Katodritou E, et al. Daratumumab plus pomalidomide and dexamethasone versus pomalidomide and dexamethasone alone in previously treated multiple myeloma (APOLLO): an open-label, randomised, phase 3 trial. Lancet Oncology. 2021;22(6):801-12.
51. Terpos E, Dimopoulos MA, Boccadoro M, Delimpasi S, Beksac M, Katodritou E, et al. Health-related quality of life in patients with relapsed/refractory multiple myeloma treated with pomalidomide and dexamethasone +/- subcutaneous daratumumab: Patient-reported outcomes from the APOLLO trial. Am J Hematol. 2022.
52. Moreau P, Mateos M-V, Berenson JR, Weisel K, Lazzaro A, Song K, et al. Once weekly versus twice weekly carfilzomib dosing in patients with relapsed and refractory multiple myeloma (A.R.R.O.W.): interim analysis results of a randomised, phase 3 study. The Lancet Oncology. 2018;19(7):953-64.
53. Moreau P, Kumar S, Boccia R, Iida S, Goldschmidt H, Cocks K, et al. Convenience, satisfaction, health-related quality of life of once-weekly 70 mg/m2 vs. twice-weekly 27 mg/m2 carfilzomib (randomized A.R.R.O.W. study). Leukemia. 2019;33(12):2934-46.
54. Siegel DS, Dimopoulos MA, Ludwig H, Facon T, Goldschmidt H, Jakubowiak A, et al. Improvement in Overall Survival With Carfilzomib, Lenalidomide, and Dexamethasone in Patients With Relapsed or Refractory Multiple Myeloma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2018;36(8):728-34.
55. Kumar SK, Harrison SJ, Cavo M, de la Rubia J, Popat R, Gasparetto C, et al. Venetoclax or placebo in combination with bortezomib and dexamethasone in patients with relapsed or refractory multiple myeloma (BELLINI): a randomised, double-blind, multicentre, phase 3 trial. Lancet Oncology. 2020;21(12):1630-42.
56. Grosicki S, Simonova M, Spicka I, Pour L, Kriachok I, Gavriatopoulou M, et al. Once-per-week selinexor, bortezomib, and dexamethasone versus twice-per-week bortezomib and dexamethasone in patients with multiple myeloma (BOSTON): a randomised, open-label, phase 3 trial. Lancet. 2020;396(10262):1563-73.
57. Dimopoulos M, Quach H, Mateos MV, Landgren O, Leleu X, Siegel D, et al. Carfilzomib, dexamethasone, and daratumumab versus carfilzomib and dexamethasone for patients with relapsed or refractory multiple myeloma (CANDOR): results from a randomised, multicentre, open-label, phase 3 study. Lancet. 2020;396(10245):186-97.
58. Usmani SZ, Quach H, Mateos MV, gren O, Leleu X, Siegel D, et al. Carfilzomib, dexamethasone, and daratumumab versus carfilzomib and dexamethasone for patients with relapsed or refractory multiple myeloma (CANDOR): updated outcomes from a randomised, multicentre, open-label, phase 3 study. Lancet Oncology. 2022;23(1):65-76.

59. Siegel D, Weisel K, Zahlten-Kumeli A, Medhekar R, Ding B, Leleu X. Health-related quality of life outcomes from the CANDOR study in patients with relapsed or refractory multiple myeloma. Leukemia and Lymphoma. 2021;62:3002-10.
60. Palumbo A, Chanan-Khan A, Weisel K, Nooka AK, Masszi T, Beksac M, et al. Daratumumab, Bortezomib, and Dexamethasone for Multiple Myeloma. N Engl J Med. 2016;375(8):754-66.
61. Hungria V, Beksac M, Weisel KC, Nooka AK, Masszi T, Spicka I, et al. Health-related quality of life maintained over time in patients with relapsed or refractory multiple myeloma treated with daratumumab in combination with bortezomib and dexamethasone: results from the phase III CASTOR trial. British Journal of Haematology. 2020;8:08.
62. Mateos MV, Nahi H, Legiec W, Grosicki S, Vorobyev V, Spicka I, et al. Subcutaneous versus intravenous daratumumab in patients with relapsed or refractory multiple myeloma (COLUMBA): a multicentre, open-label, non-inferiority, randomised, phase 3 trial. The Lancet Haematology. 2020;7(5):e370-e80.
63. Jagannath S, Barlogie B, Berenson J, Siegel D, Irwin D, Richardson PG, et al. A phase 2 study of two doses of bortezomib in relapsed or refractory myeloma. Br J Haematol. 2004;127(2):165-72.
64. Orlowski RZ, Nagler A, Sonneveld P, Bladé J, Hajek R, Spencer A, et al. Randomized phase III study of pegylated liposomal doxorubicin plus bortezomib compared with bortezomib alone in relapsed or refractory multiple myeloma: combination therapy improves time to progression. J Clin Oncol. 2007;25(25):3892-901.
65. Orlowski RZ, Nagler A, Sonneveld P, Bladé J, Hajek R, Spencer A, et al. Final overall survival results of a randomized trial comparing bortezomib plus pegylated liposomal doxorubicin with bortezomib alone in patients with relapsed or refractory multiple myeloma. Cancer. 2016;122(13):2050-6.
66. Cella D, McKendrick J, Kudlac A, Palumbo A, Oukessou A, Vij R, et al. Impact of elotuzumab treatment on pain and health-related quality of life in patients with relapsed or refractory multiple myeloma: results from the ELOQUENT-2 study. Annals of hematology. 2018;97(12):2455-63.
67. Dimopoulos MA, Lonial S, White D, Moreau P, Weisel K, San-Miguel J, et al. Elotuzumab, lenalidomide, and dexamethasone in RRMM: final overall survival results from the phase 3 randomized ELOQUENT-2 study. Blood Cancer Journal. 2020;10(9):91.
68. Dimopoulos MA, Dytfeld D, Grosicki S, Moreau P, Takezako N, Hori M, et al. Elotuzumab plus Pomalidomide and Dexamethasone for Multiple Myeloma. The New England journal of medicine. 2018;379(19):1811-22.
69. Dimopoulos MA, Moreau P, Palumbo A, Joshua D, Pour L, Hájek R, et al. Carfilzomib and dexamethasone versus bortezomib and dexamethasone for patients with relapsed or refractory multiple myeloma (ENDEAVOR): a randomised, phase 3, open-label, multicentre study. Lancet Oncol. 2016;17(1):27-38.
70. Ludwig H, Moreau P, Dimopoulos MA, Mateos M-V, Kaiser M, Hajek R, et al. Health-related quality of life in the ENDEAVOR study: carfilzomib-dexamethasone vs bortezomib-dexamethasone in relapsed/refractory multiple myeloma. Blood cancer journal. 2019;9(3):23.
71. Orlowski RZ, Moreau P, Niesvizky R, Ludwig H, Oriol A, Chng WJ, et al. Carfilzomib-Dexamethasone Versus Bortezomib-Dexamethasone in Relapsed or Refractory Multiple Myeloma: Updated Overall Survival, Safety, and Subgroups. Clinical lymphoma, myeloma & leukemia. 2019;19(8):522-30.e1.
72. Hájek R, Masszi T, Petrucci MT, Palumbo A, Rosiñol L, Nagler A, et al. A randomized phase III study of carfilzomib vs low-dose corticosteroids with

optional cyclophosphamide in relapsed and refractory multiple myeloma (FOCUS). Leukemia. 2017;31(1):107-14.
73. Attal M, Richardson PG, Rajkumar SV, San-Miguel J, Beksac M, Spicka I, et al. Isatuximab plus pomalidomide and low-dose dexamethasone versus pomalidomide and low-dose dexamethasone in patients with relapsed and refractory multiple myeloma (ICARIA-MM): a randomised, multicentre, open-label, phase 3 study. Lancet (London, England). 2019;394(10214):2096-107.
74. Leleu X, Attal M, Arnulf B, Moreau P, Traulle C, Marit G, et al. Pomalidomide plus low-dose dexamethasone is active and well tolerated in bortezomib and lenalidomide-refractory multiple myeloma: Intergroupe Francophone du Myélome 2009-02. Blood. 2013;121(11):1968-75.
75. Moreau P, Dimopoulos MA, Mikhael J, Yong K, Capra M, Facon T, et al. Isatuximab, carfilzomib, and dexamethasone in relapsed multiple myeloma (IKEMA): a multicentre, open-label, randomised phase 3 trial. Lancet. 2021;397(10292):2361-71.
76. Mateos M-V, Blacklock H, Schjesvold F, Oriol A, Simpson D, George A, et al. Pembrolizumab plus pomalidomide and dexamethasone for patients with relapsed or refractory multiple myeloma (KEYNOTE-183): a randomised, open-label, phase 3 trial. The Lancet Haematology. 2019;6(9):e459-e69.
77. Richardson PG, Siegel DS, Vij R, Hofmeister CC, Baz R, Jagannath S, et al. Pomalidomide alone or in combination with low-dose dexamethasone in relapsed and refractory multiple myeloma: a randomized phase 2 study. Blood. 2014;123(12):1826-32.
78. San-Miguel J, Weisel K, Moreau P, Lacy M, Song K, Delforge M, et al. Pomalidomide plus low-dose dexamethasone versus high-dose dexamethasone alone for patients with relapsed and refractory multiple myeloma (MM-003): a randomised, open-label, phase 3 trial. Lancet Oncol. 2013;14(11):1055-66.
79. Song KW, Dimopoulos MA, Weisel KC, Moreau P, Palumbo A, Belch A, et al. Health-related quality of life from the MM-003 trial of pomalidomide plus low-dose dexamethasone versus high-dose dexamethasone in relapsed and/or refractory multiple myeloma. Haematologica. 2015;100(2):e63-7.
80. Weisel K, Dimopoulos M, Song KW, Moreau P, Palumbo A, Belch A, et al. Pomalidomide and Low-Dose Dexamethasone Improves Health-Related Quality of Life and Prolongs Time to Worsening in Relapsed/Refractory Patients With Multiple Myeloma Enrolled in the MM-003 Randomized Phase III Trial. Clin Lymphoma Myeloma Leuk. 2015;15(9):519-30.
81. Weber DM, Chen C, Niesvizky R, Wang M, Belch A, Stadtmauer EA, et al. Lenalidomide plus dexamethasone for relapsed multiple myeloma in North America. N Engl J Med. 2007;357(21):2133-42.
82. Dimopoulos M, Spencer A, Attal M, Prince HM, Harousseau JL, Dmoszynska A, et al. Lenalidomide plus dexamethasone for relapsed or refractory multiple myeloma. N Engl J Med. 2007;357(21):2123-32.
83. Garderet L, Iacobelli S, Moreau P, Dib M, Lafon I, Niederwieser D, et al. Superiority of the triple combination of bortezomib-thalidomide-dexamethasone over the dual combination of thalidomide-dexamethasone 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. J Clin Oncol. 2012;30(20):2475-82.
84. Terpos E, Gobbi M, Potamianou A, Lahaye M, Couturier C, Cavo M. Retreatment and prolonged therapy with subcutaneous bortezomib in patients with relapsed multiple myeloma: A randomized, controlled, phase III study. European journal of haematology. 2018;100(1):10-9.
85. Richardson PG, Oriol A, Beksac M, Liberati AM, Galli M, Schjesvold F, et al. Pomalidomide, bortezomib, and dexamethasone for patients with relapsed or

refractory multiple myeloma previously treated with lenalidomide (OPTIMISMM): a randomised, open-label, phase 3 trial. The Lancet Oncology. 2019;20(6):781-94.
86. Weisel K, Dimopoulos M, Moreau P, Yagci M, Larocca A, Kanate AS, et al. Health-related quality-of-life results from the phase 3 OPTIMISMM study: pomalidomide, bortezomib, and low-dose dexamethasone versus bortezomib and low-dose dexamethasone in relapsed or refractory multiple myeloma. Leukemia & Lymphoma. 2020;61(8):1850-9.
87. San-Miguel JF, Hungria VT, Yoon SS, Beksac M, Dimopoulos MA, Elghandour A, et al. Panobinostat plus bortezomib and dexamethasone versus placebo plus bortezomib and dexamethasone in patients with relapsed or relapsed and refractory multiple myeloma: a multicentre, randomised, double-blind phase 3 trial. Lancet Oncol. 2014;15(11):1195-206.
88. San-Miguel JF, Hungria VT, Yoon SS, Beksac M, Dimopoulos MA, Elghandour A, et al. Overall survival of patients with relapsed multiple myeloma treated with panobinostat or placebo plus bortezomib and dexamethasone (the PANORAMA 1 trial): a randomised, placebo-controlled, phase 3 trial. Lancet Haematol. 2016;3(11):e506-e15.
89. Richardson PG, Schlossman RL, Roy AN, Panneerselvam A, Acharyya S, Sopala M, et al. Patient-reported outcomes of multiple myeloma patients treated with panobinostat after >=2 lines of therapy based on the international phase 3, randomized, double-blind, placebo-controlled PANORAMA-1 trial. British journal of haematology. 2018;181(5):628-36.
90. Laubach JP, Schjesvold F, Mariz M, Dimopoulos MA, Lech-Mar, a E, et al. Efficacy and safety of oral panobinostat plus subcutaneous bortezomib and oral dexamethasone in patients with relapsed or relapsed and refractory multiple myeloma (PANORAMA 3): an open-label, randomised, phase 2 study. Lancet Oncology. 2021;22(1):142-54.
91. Dimopoulos MA, Oriol A, Nahi H, San-Miguel J, Bahlis NJ, Usmani SZ, et al. Daratumumab, Lenalidomide, and Dexamethasone for Multiple Myeloma. N Engl J Med. 2016;375(14):1319-31.
92. Bahlis NJ, Dimopoulos MA, White DJ, Benboubker L, Cook G, Leiba M, 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.
93. Plesner T, Dimopoulos MA, Oriol A, San-Miguel J, Bahlis NJ, Rabin N, et al. Health-related quality of life in patients with relapsed or refractory multiple myeloma: treatment with daratumumab, lenalidomide, and dexamethasone in the phase 3 POLLUX trial. British Journal of Haematology. 2021;194(1):132-9.
94. Lonial S, Weiss BM, Usmani SZ, Singhal S, Chari A, Bahlis NJ, et al. Daratumumab monotherapy in patients with treatment-refractory multiple myeloma (SIRIUS): an open-label, randomised, phase 2 trial. Lancet. 2016;387(10027):1551-60.
95. Moreau P, Masszi T, Grzasko N, Bahlis NJ, Hansson M, Pour L, et al. Oral Ixazomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. N Engl J Med. 2016;374(17):1621-34.
96. Dimopoulos M, Siegel DS, Lonial S, Qi J, Hajek R, Facon T, et al. Vorinostat or placebo in combination with bortezomib in patients with multiple myeloma (VANTAGE 088): a multicentre, randomised, double-blind study. Lancet Oncol. 2013;14(11):1129-40.
97. Ailawadhi S, Sexton R, Lentzsch S, Abidi MH, Voorhees PM, Cohen AD, et al. Low-dose versus High-dose Carfilzomib with Dexamethasone (S1304) in Patients with Relapsed-Refractory Multiple Myeloma. Clinical Cancer Research. 2020;26(15):3969-78.

98. Baz RC, Martin TG, 3rd, Lin HY, Zhao X, Shain KH, Cho HJ, et al. Randomized multicenter phase 2 study of pomalidomide, cyclophosphamide, and dexamethasone in relapsed refractory myeloma. Blood. 2016;127(21):2561-8.
99. Dimopoulos M, Bringhen S, Anttila P, Capra M, Cavo M, Cole C, et al. Isatuximab as monotherapy and combined with dexamethasone in patients with relapsed/refractory multiple myeloma. Blood. 2021;137:1154-65.
100. Iida S, Wakabayashi M, Tsukasaki K, Miyamoto K, Maruyama D, Yamamoto K, et al. Bortezomib plus dexamethasone vs thalidomide plus dexamethasone for relapsed or refractory multiple myeloma. Cancer science. 2018;109(5):1552-61.
101. Jakubowiak A, Offidani M, Pégourie B, De La Rubia J, Garderet L, Laribi K, et al. Randomized phase 2 study: elotuzumab plus bortezomib/dexamethasone vs bortezomib/dexamethasone for relapsed/refractory MM. Blood. 2016;127(23):2833-40.
102. Kropff M, Vogel M, Bisping G, Schlag R, Weide R, Knauf W, 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. Ann Hematol. 2017;96(11):1857-66.
103. Mikhael J, Richter J, Vij R, Cole C, Zonder J, Kaufman JL, et al. A dose-finding Phase 2 study of single agent isatuximab (anti-CD38 mAb) in relapsed/refractory multiple myeloma. Leukemia. 2020;14:14.
104. Mina R, Belotti A, Petrucci MT, Zambello R, Capra A, Di Lullo G, et al. Bortezomib-dexamethasone as maintenance therapy or early retreatment at biochemical relapse versus observation in relapsed/refractory multiple myeloma patients: a randomized phase II study. Blood Cancer Journal. 2020;10(5):58.
105. Montefusco V, Corso A, Galli M, Ardoino I, Pezzatti S, Carniti C, et al. Bortezomib, cyclophosphamide, dexamethasone versus lenalidomide, cyclophosphamide, dexamethasone in multiple myeloma patients at first relapse. British Journal of Haematology. 2020;188(6):907-17.
106. Moreau P, Pylypenko H, Grosicki S, Karamanesht I, Leleu X, Grishunina M, et al. Subcutaneous versus intravenous administration of bortezomib in patients with relapsed multiple myeloma: a randomised, phase 3, non-inferiority study. Lancet Oncol. 2011;12(5):431-40.
107. Orlowski RZ, Gercheva L, Williams C, Sutherland H, Robak T, Masszi T, et al. A phase 2, randomized, double-blind, placebo-controlled study of siltuximab (anti-IL-6 mAb) and bortezomib versus bortezomib alone in patients with relapsed or refractory multiple myeloma. Am J Hematol. 2015;90(1):42-9.
108. Raje NS, Moreau P, Terpos E, Benboubker L, Grzasko N, Holstein SA, 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):783-95.
109. Sehgal K, Das R, Zhang L, Verma R, Deng Y, Kocoglu M, et al. Clinical and pharmacodynamic analysis of pomalidomide dosing strategies in myeloma: impact of immune activation and cereblon targets. Blood. 2015;125(26):4042-51.
110. Aguiar PM, Lima TM, Storpirtis S. Systematic review of the economic evaluations of novel therapeutic agents in multiple myeloma: what is the reporting quality? J Clin Pharm Ther. 2016;41(2):189-97.
111. Asrar MM, Lad DP, Prinja S, Bansal D. A systematic review of economic evaluations of treatment regimens in multiple myeloma. Expert Rev Pharmacoecon Outcomes Res. 2020:1-11.
112. Seefat MR, Cucchi DGJ, Dirven S, Groen K, Zweegman S, Blommestein HM. A Systematic Review of Cost-Effectiveness Analyses of Novel Agents in the Treatment of Multiple Myeloma. Cancers (Basel). 2021;13(22).

113. Borg S, Nahi H, Hansson M, Lee D, Elvidge J, Persson U. Cost effectiveness of pomalidomide in patients with relapsed and refractory multiple myeloma in Sweden. Acta Oncol. 2016;55(5):554-60.
114. Brown RE, Stern S, Dhanasiri S, Schey S. Lenalidomide for multiple myeloma: cost-effectiveness in patients with one prior therapy in England and Wales. Eur J Health Econ. 2013;14(3):507-14.
115. Fragoulakis V, Kastritis E, Psaltopoulou T, Maniadakis N. Economic evaluation of therapies for patients suffering from relapsed-refractory multiple myeloma in Greece. Cancer Manag Res. 2013;5:37-48.
116. Möller J, Nicklasson L, Murthy A. Cost-effectiveness of novel relapsed-refractory multiple myeloma therapies in Norway: lenalidomide plus dexamethasone vs bortezomib. J Med Econ. 2011;14(6):690-7.
117. Aceituno S, Gozalbo I, Appierto M, Lizán L. Cost-effectiveness of lenalidomide in combination with dexamethasone compared to bortezomib in combination with dexamethasone for the second-line treatment of multiple myeloma in Chile. Medwave. 2018;18(3):e7220.
118. Cai H, Zhang L, Li N, Zheng B, Liu M. Cost-effectiveness analysis on binary/triple therapy on the basis of ixazomib or bortezomib for refractory or relapsed multiple myeloma. Leuk Lymphoma. 2019;60(12):2951-9.
119. Pelligra CG, Parikh K, Guo S, Chandler C, Mouro J, Abouzaid S, et al. Cost-effectiveness of Pomalidomide, Carfilzomib, and Daratumumab for the Treatment of Patients with Heavily Pretreated Relapsed-refractory Multiple Myeloma in the United States. Clin Ther. 2017;39(10):1986-2005.e5.
120. Zhang TT, Wang S, Wan N, Zhang L, Zhang Z, Jiang J. Cost-effectiveness of Daratumumab-based Triplet Therapies in Patients With Relapsed or Refractory Multiple Myeloma. Clin Ther. 2018;40(7):1122-39.
121. Jakubowiak AJ, Campioni M, Benedict Á, Houisse I, Tichy E, Giannopoulou A, et al. Cost-effectiveness of adding carfilzomib to lenalidomide and dexamethasone in relapsed multiple myeloma from a US perspective. J Med Econ. 2016;19(11):1061-74.
122. Jakubowiak AJ, Houisse I, Májer I, Benedict Á, Campioni M, Panjabi S, et al. Cost-effectiveness of carfilzomib plus dexamethasone compared with bortezomib plus dexamethasone for patients with relapsed or refractory multiple myeloma in the United States. Expert Rev Hematol. 2017;10(12):1107-19.
123. Campioni M, Agirrezabal I, Hajek R, Minarik J, Pour L, Spicka I, et al. Methodology and results of real-world cost-effectiveness of carfilzomib in combination with lenalidomide and dexamethasone in relapsed multiple myeloma using registry data. Eur J Health Econ. 2020;21(2):219-33.
124. Wong XY, Chng WJ, Aziz MIA, Ng K. Cost-effectiveness of daratumumab in combination with lenalidomide and dexamethasone for relapsed and/or refractory multiple myeloma. Expert Rev Pharmacoecon Outcomes Res. 2021:1-9.
125. Hornberger J, Rickert J, Dhawan R, Liwing J, Aschan J, Löthgren M. The cost-effectiveness of bortezomib in relapsed/refractory multiple myeloma: Swedish perspective. Eur J Haematol. 2010;85(6):484-91.
126. Carlson JJ, Guzauskas GF, Chapman RH, Synnott PG, Liu S, Russo ET, et al. Cost-effectiveness of Drugs to Treat Relapsed/Refractory Multiple Myeloma in the United States. J Manag Care Spec Pharm. 2018;24(1):29-38.
127. TreeAge Pro. TreeAge Software. Williamstown, MA, USA: R1; 2021.
128. Helsedirektoratet. Økonomisk evaluering av helsetiltak – en veileder. In: finansiering Dho, editor. Oslo: Helsedirektoratet; 2012.
129. Woods B SE, Palmer S, Latimer N, Soares M. . NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. 2017.

130. Norwegian Ministry of Health and Care Services. Principles for priority setting in health care. Summary of a white paper on priority setting in the Norwegian health care sector. In: Services NMoHaC, editor. 2017.
131. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. USA: Oxford University Press; 2005.
132. Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. BMC Med Res Methodol. 2011;11:139.
133. RStudio Team. RStudio: Integrated Development for R. Boston, MA, USA: RStudio; 2020.
134. Briggs A, Sculpher M, Claxton K. Decision Modellling for Health Economic Evaluation. 1st ed. USA: Oxford University Press; 2006.
135. Norwegian Medicine Agency. Legemiddelsøk [Webpage]. Norway: Norwegian Medicine Agency; 2022 [Available from: https://www.legemiddelsok.no/.
136. Felleskatalogen AS. Felleskatalogen [Nettside]. Norway: Legemiddelindustrien; 2022 [Available from: https://www.felleskatalogen.no/medisin/.
137. Norwegian Medicine Agency. Enhetskostnadsdatabase [Webpage]. Oslo, Norway: Norwegian Medicine Agency,; 2022 [updated 16.08.2022. Available from: https://legemiddelverket.no/offentlig-finansiering/dokumentasjon-for-metodevurdering/enhetskostnadsdatabase.
138. Forskrift om stønad til dekning av utgifter til undersøkelse og behandling i private medisinske laboratorie- og røntgenvirksomheter, (2003).
139. Kazandjian D. Multiple myeloma epidemiology and survival: A unique malignancy. Semin Oncol. 2016;43(6):676-81.
140. Marinac CR, Ghobrial IM, Birmann BM, Soiffer J, Rebbeck TR. Dissecting racial disparities in multiple myeloma. Blood Cancer J. 2020;10(2):19.
141. Wang S, Xu L, Feng J, Liu Y, Liu L, Wang J, et al. Prevalence and Incidence of Multiple Myeloma in Urban Area in China: A National Population-Based Analysis. Front Oncol. 2019;9:1513.
142. Tang CH, Liu HY, Hou HA, Qiu H, Huang KC, Siggins S, et al. Epidemiology of multiple myeloma in Taiwan, a population based study. Cancer Epidemiol. 2018;55:136-41.
143. Afifah NN, Diantini A, Intania R, Abdulah R, Barliana MI. Genetic Polymorphisms and the Efficacy of Platinum-Based Chemotherapy: Review. Pharmgenomics Pers Med. 2020;13:427-44.
144. O'Donnell PH, Dolan ME. Cancer pharmacoethnicity: ethnic differences in susceptibility to the effects of chemotherapy. Clin Cancer Res. 2009;15(15):4806-14.
145. Phan VH, Moore MM, McLachlan AJ, Piquette-Miller M, Xu H, Clarke SJ. Ethnic differences in drug metabolism and toxicity from chemotherapy. Expert Opin Drug Metab Toxicol. 2009;5(3):243-57.
146. Deeks JJ, Higgins JP, Altman DG. 10.10.3 Strategies for addressing heterogeneity. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. 6.3 ed: Cochrane; 2022.
147. Hahn S. Understanding noninferiority trials. Korean J Pediatr. 2012;55(11):403-7.

Appendix 1

Glossary list

Abbreviation	Definition
AE	Adverse events
ASCT	Autologous stem cell transplant
Bev	Bevacizumab
CEAC	Cost Effectiveness Acceptability Curves
CEAF	Cost Effectiveness Acceptability Frontiers
CI	Confidence interval
CRAB	Hypercalcemia; Renal insufficiency; Anemia; and Bone lesions
Cy	Cyclophosphamide
D	Daratumumab
d	Dexamethasone (or another glucocorticoid)
Dox	Doxorubicin
E	Elotuzumab
EPOC	The Cochrane Effective Practice and Organisation of Care
F	Panobinostat (Farydak)
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
HR	Hazard ratio
HTA	Health Technology Assessment
I	Ixazomib
ICER	Incremental Cost-Effectiveness Ratios
INHB	Incremental Net Health Benefit
INMB	Incremental Net Monetary Benefit
IRR	Incidence Rate Ratio
Is	Isatuximab
ITT	Intention-to-treat
K	Carfilzomib (Kyprolis)

Abbreviation	Definition
KI	Konfidensintervall
MD	Median Difference
MeSH	Medical Subject Heading
MGUS	Monoclonal Gammopathy of Undetermined Significance
NA	Not Assessed
NMA	Network Meta-Analysis
NOK	Norwegian kroner
OS	Overall survival
P	Pomalidomide
Pem	Pembrolizumab
PFS	Progression-free survival
PICO	Population; Intervention; Comparator; Outcome
QALY	Quality-Adjusted Life-Years
QLQ-C30 (EORTC)	Quality of life questionnaire specifically for multiple myeloma, by the European Organisation for Research and Treatment of Cancer
QoL	Quality of Life
R	Lenalidomide (Revlimide)
RCT	Randomised Controlled Trial
RoB	Risk of Bias
RR	Risk Ratio
RRMM	Relapsing, refractory multiple myeloma
S	Siltuximab
SAE	Severe Adverse Events
SMM	Smouldering Multiple Myeloma
T	Thalidomide
Tab	Tabalumab
V	Bortezomib (Velcade)
Ven	Venetoclax
Vor	Vorinostat

Appendix 2

Project plan

The project plan (protocol) was published in August 2021 (31), and is found here:

https://www.fhi.no/contentassets/420cc15517aa44af92bc057ed1ab522b/prosjektplan-for-behandling-av-myelomatose_id2019_072_.pdf

Appendix 3

Detailed search strategy

Search strategy from February-August 2020

Database: Ovid MEDLINE® and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions® <1946 to September 03, 2020>

Search Date: 07.09.2020

1 Clinical Trial, Phase III/ or exp Randomized Controlled Trial/ (515980)
2 (cross-over or crossover or ((double or single or triple) adj blind*) or (phase adj ("3" or "III")) or placebo or random*).tw,kw,kf. (1310409)
3 1 or 2 (1408374)
4 exp Multiple Myeloma/ (41318)
5 (myeloma* or Kahler disease).tw,kw,kf. (54703)
6 4 or 5 (61942)
7 Bortezomib/ (5694)
8 bortezomib.tw,kw,kf. (8199)
9 carfilzomib.tw,kw,kf. (950)
10 daratumumab.tw,kw,kf. (632)
11 elotuzumab.tw,kw,kf. (259)
12 isatuximab.tw,kw,kf. (65)
13 ixazomib.tw,kw,kf. (330)
14 Lenalidomide/ (2668)
15 lenalidomide.tw,kw,kf. (4307)
16 Panobinostat/ (513)
17 panobinostat.tw,kw,kf. (711)
18 Pomalidomide/ (0)
19 pomalidomid*.tw,kw,kf. (693)
20 Thalidomide/ (8932)
21 thalidomid*.tw,kw,kf. (8148)
22 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 (21608)
23 3 and 6 and 22 (1033)
24 (exp Animals/ or exp Animal Experimentation/) not Humans/ (4731206)
25 (animal* or dog or dogs or "in vitro" or mouse or mice or rat or rats or rodent*).ti. (1863666)
26 24 or 25 (5212771)
27 23 not 26 (1026)
28 (202002* or 202003* or 202004* or 202005* or 202006* or 202007* or 202008*).dt. (850427)
29 27 and 28 (63)

Database: Embase <1974 to 2020 September 03>
Search Date: 07.09.2020

1 Phase 3 Clinical Trial/ or exp Randomized Controlled Trial/ or Crossover Procedure/ or Double-Blind Procedure/ or Single-Blind Procedure/ (711831)
2 (cross-over or crossover or ((double or single or triple) adj blind*) or (phase adj ("3" or "III")) or placebo or random*).tw,kw. (1781327)
3 1 or 2 (1891248)
4 Multiple Myeloma/ (76323)
5 (myeloma* or Kahler disease).tw,kw. (83278)
6 4 or 5 (98992)
7 Bortezomib/ (30750)
8 bortezomib.tw,kw. (18786)
9 Carfilzomib/ (4194)
10 carfilzomib.tw,kw. (2735)
11 Daratumumab/ (2487)
12 daratumumab.tw,kw. (1762)
13 Elotuzumab/ (1077)
14 elotuzumab.tw,kw. (619)
15 Isatuximab/ (314)
16 isatuximab.tw,kw. (153)
17 Ixazomib/ (1328)
18 ixazomib.tw,kw. (841)
19 Lenalidomide/ (19234)
20 lenalidomide.tw,kw. (12112)
21 Panobinostat/ (3806)
22 panobinostat.tw,kw. (1602)
23 Pomalidomide/ (3315)
24 pomalidomid*.tw,kw. (2101)
25 Thalidomide/ (27428)
26 thalidomid*.tw,kw. (13028)
27 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 (66213)
28 3 and 6 and 27 (3613)
29 (exp Animal/ or exp Animal Experiment/) not exp Human/ (4977847)
30 (animal* or dog or dogs or "in vitro" or mouse or mice or rat or rats or rodent*).ti. (2023123)
31 29 or 30 (5420566)
32 28 not 31 (3540)
33 (202002* or 202003* or 202004* or 202005* or 202006* or 202007* or 202008*).dc. (1274593)
34 32 and 33 (173)

Search strategy from August 2020 – March 2021

Database: Ovid MEDLINE® and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions® <1946 to March 12, 2021>

Search Date: 14.03.2021

1 Clinical Trial, Phase III/ or exp Randomized Controlled Trial/ (528869)
2 (cross-over or crossover or ((double or single or triple) adj blind*) or (phase adj ("3" or "III")) or placebo or random*).tw,kw,kf. (1362145)
3 1 or 2 (1461425)
4 exp Multiple Myeloma/ and (Relaps* or Refractory).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (4797)
5 ((relaps* or refractory) adj3 (myeloma* or Kahler* disease)).tw,kf,kw. (3022)

6 4 or 5 (5471)
7 Bortezomib/ (5908)
8 bortezomib.tw,kw,kf. (8517)
9 carfilzomib.tw,kw,kf. (1045)
10 daratumumab.tw,kw,kf. (788)
11 elotuzumab.tw,kw,kf. (288)
12 isatuximab.tw,kw,kf. (87)
13 ixazomib.tw,kw,kf. (369)
14 Lenalidomide/ (2811)
15 lenalidomide.tw,kw,kf. (4530)
16 Panobinostat/ (528)
17 panobinostat.tw,kw,kf. (750)
18 Pomalidomide/ (0)
19 pomalidomid*.tw,kw,kf. (759)
20 Thalidomide/ (9087)
21 thalidomid*.tw,kw,kf. (8279)
22 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 (22486)
23 3 and 6 and 22 (509)
24 (exp Animals/ or exp Animal Experimentation/) not Humans/ (4799583)
25 (animal* or dog or dogs or “in vitro” or mouse or mice or rat or rats or rodent*).ti. (1893870)
26 24 or 25 (5295514)
27 23 not 26 (507)
28 (2020083* or 202009* or 202010* or 202011* or 202012* or 202101* or 202102*).dt. (785905)
29 27 and 28 (22)

Database: Embase <1974 to 2021 March 12>
Search Date: 14.03.2021

1 Phase 3 Clinical Trial/ or exp Randomized Controlled Trial/ or Crossover Procedure/ or Double-Blind Procedure/ or Single-Blind Procedure/ (748472)
2 (cross-over or crossover or ((double or single or triple) adj blind*) or (phase adj (“3” or “III”)) or placebo or random*).tw,kw. (1866956)
3 1 or 2 (1979179)
4 exp Multiple Myeloma/ and (Relaps* or Refractory).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word] (15443)
5 ((relaps* or refractory) adj3 (myeloma* or Kahler* disease)).tw,kw. (6859)
6 4 or 5 (16033)
7 Bortezomib/ (32207)
8 bortezomib.tw,kw. (19713)
9 Carfilzomib/ (4642)
10 carfilzomib.tw,kw. (2991)
11 Daratumumab/ (3057)
12 daratumumab.tw,kw. (2186)
13 Elotuzumab/ (1186)
14 elotuzumab.tw,kw. (676)
15 Isatuximab/ (387)
16 isatuximab.tw,kw. (198)
17 Ixazomib/ (1508)
18 ixazomib.tw,kw. (951)
19 Lenalidomide/ (20351)
20 lenalidomide.tw,kw. (12796)
21 Panobinostat/ (4011)

22 panobinostat.tw,kw. (1691)
23 Pomalidomide/ (3635)
24 pomalidomid*.tw,kw. (2312)
25 Thalidomide/ (28149)
26 thalidomid*.tw,kw. (13408)
27 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 (69342)
28 3 and 6 and 27 (1874)
29 (exp Animal/ or exp Animal Experiment/) not exp Human/ (5079959)
30 (animal* or dog or dogs or "in vitro" or mouse or mice or rat or rats or rodent*).ti. (2066725)
31 29 or 30 (5536425)
32 28 not 31 (1841)
33 (2020083* or 202009* or 202010* or 202011* or 202012* or 202101* or 202102*).dc. (1351126)
34 32 and 33 (62)
35 limit 34 to embase status (26)

Search strategy for ongoing studies

Search date: June 2021
Search line: Multiple Myeloma AND (Relapse OR Refractory)

Update search for ongoing studies

Search period: June 2021 – May 2022
Search line: Multiple Myeloma AND (Relapse OR Refractory)
Search limitations in WHO ICTRP: None
Search limitations in Clinical Trials database: Under Recruitment, limited to Not yet recruiting, recruiting, enrolling by invitation, active not recruiting, unknown status. Under Study type: Interventional

Appendix 4

PRISMA NMA checklist

Modified PRISMA NMA checklist of items to include when reporting a systematic review involving a network meta-analysis

Section/Topic	Item #	Checklist Item	Where Reported
TITLE			
Title	1	Identify the report as a systematic review *incorporating a network meta-analysis (or related form of meta-analysis).*	*Not applicable to FHI HTAs*
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: **Background:** main objectives **Methods:** data sources; study eligibility criteria, participants, and interventions; study appraisal; and *synthesis methods, such as network meta-analysis.* **Results:** number of studies and participants identified; summary estimates with corresponding confidence/credible intervals; *treatment rankings may also be discussed. Authors may choose to summarize pairwise comparisons against a chosen treatment included in their analyses for brevity.* **Discussion/Conclusions:** limitations; conclusions and implications of findings. **Other:** primary source of funding; systematic review registration number with registry name.	*Executive summary*
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known*, including mention of why a network meta-analysis has been conducted.*	*Introduction chapter*
Objectives	4	Provide an explicit statement of questions being addressed, with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	*Clinical efficacy and safety chapter*
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists and if and where it can be accessed (e.g., Web address); and, if available, provide registration information, including registration number.	*Clinical efficacy and safety chapter*
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. *Clearly describe eligible treatments included in the treatment network, and note whether any have been clustered or merged into the same node (with justification).*	*Clinical efficacy and safety chapter*

Section/Topic	Item #	Checklist Item	Where Reported
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	*Clinical efficacy and safety chapter*
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	*Method chapter and Appendix 3*
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	*Clinical efficacy and safety chapter*
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	*Clinical efficacy and safety chapter*
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	*Clinical efficacy and safety chapter*
Geometry of the network	**S1**	Describe methods used to explore the geometry of the treatment network under study and potential biases related to it. This should include how the evidence base has been graphically summarized for presentation, and what characteristics were compiled and used to describe the evidence base to readers.	*Clinical efficacy and safety chapter*
Risk of bias within individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	*Clinical efficacy and safety chapter*
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means). *Also describe the use of additional summary measures assessed, such as treatment rankings and surface under the cumulative ranking curve (SUCRA) values, as well as modified approaches used to present summary findings from meta-analyses.*	*Clinical efficacy and safety chapter*
Planned methods of analysis	14	Describe the methods of handling data and combining results of studies for each network meta-analysis. This should include, but not be limited to: • *Handling of multi-arm trials;* • *Selection of variance structure;* • *Selection of prior distributions in Bayesian analyses; and* • *Assessment of model fit.*	*Clinical efficacy and safety chapter*
Assessment of Inconsistency	**S2**	Describe the statistical methods used to evaluate the agreement of direct and indirect evidence in the treatment network(s) studied. Describe efforts taken to address its presence when found.	*Clinical efficacy and safety chapter*
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	*Clinical efficacy and safety chapter*
Additional analyses	16	Describe methods of additional analyses if done, indicating which were pre-specified. This may include, but not be limited to, the following: • Sensitivity or subgroup analyses; • Meta-regression analyses; • *Alternative formulations of the treatment network; and* • *Use of alternative prior distributions for Bayesian analyses (if applicable).*	*Clinical efficacy and safety chapter*
RESULTS†			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	*Results chapter*

Section/Topic	Item #	Checklist Item	Where Reported
Presentation of network structure	**S3**	Provide a network graph of the included studies to enable visualization of the geometry of the treatment network.	*Results chapter*
Summary of network geometry	**S4**	Provide a brief overview of characteristics of the treatment network. This may include commentary on the abundance of trials and randomized patients for the different interventions and pairwise comparisons in the network, gaps of evidence in the treatment network, and potential biases reflected by the network structure.	*Results chapter*
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	*Results chapter*
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment.	*Results chapter*
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: 1) simple summary data for each intervention group, and 2) effect estimates and confidence intervals. *Modified approaches may be needed to deal with information from larger networks.*	*Not done — we judge it excessive to present all raw data in the main report*
Synthesis of results	21	Present results of each meta-analysis done, including confidence/credible intervals. *In larger networks, authors may focus on comparisons versus a particular comparator (e.g., placebo or standard care), with full findings presented in an appendix. League tables and forest plots may be considered to summarize pairwise comparisons.* If additional summary measures were explored (such as treatment rankings), these should also be presented.	*Results chapter*
Exploration for inconsistency	**S5**	Describe results from investigations of inconsistency. This may include such information as measures of model fit to compare consistency and inconsistency models, *P* values from statistical tests, or summary of inconsistency estimates from different parts of the treatment network.	*Results chapter*
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies for the evidence base being studied.	*Results chapter*
Results of additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression analyses*, alternative network geometries studied, alternative choice of prior distributions for Bayesian analyses,* and so forth).	*Additional analyses not performed*
DISCUSSION			
Summary of evidence	24	Summarize the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy-makers).	*Discussion chapter*
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias). *Comment on the validity of the assumptions, such as transitivity and consistency. Comment on any concerns regarding network geometry (e.g., avoidance of certain comparisons).*	*Methods, Results, and Discussion chapters*
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	*Discussion chapter*
FUNDING			

Section/Topic	Item #	Checklist Item	Where Reported
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. This should also include information regarding whether funding has been received from manufacturers of treatments in the network and/or whether some of the authors are content experts with professional conflicts of interest that could affect use of treatments in the network.	*No specific funding beyond salary cost from the NIPH budget*

PICOS = population, intervention, comparators, outcomes, study design.

** Text in italics indicates wording specific to reporting of network meta-analyses that has been added to guidance from the PRISMA statement.*

† Authors may wish to plan for use of appendices to present all relevant information in full detail for items in this section.

Appendix 5

Excluded studies

List of 29 references excluded through full text screening, with reason for exclusion.

Reference	Reason for exclusion
Bridoux, F., et al., Randomized Trial Comparing Double Versus Triple Bortezomib-Based Regimen in Patients With Multiple Myeloma and Acute Kidney Injury Due to Cast Nephropathy. Journal of Clinical Oncology, 2020. 38(23): p. 2647-2657.	Wrong population
Chanan-Khan, A.A., et al., Phase III randomised study of dexamethasone with or without oblimersen sodium for patients with advanced multiple myeloma. Leuk Lymphoma, 2009. 50(4): p. 559-65.	Wrong intervention
Chen, Z., J. Zhou, and H. Xu, Efficacy and Safety of Different Doses of Bortezomib Combined with Doxorubicin and Dexamethasone in the Treatment of Multiple Myeloma. Anti-Tumor Pharmacy, 2019. 9(4).	Language
Cohen, Y.C., et al., Daratumumab With Cetrelimab, an Anti-PD-1 Monoclonal Antibody, in Relapsed/Refractory Multiple Myeloma. Clinical lymphoma, myeloma & leukemia, 2020. 21(1): p. 46-54.e4.	Not randomised
Dimopoulos, M., et al., Response and progression-free survival according to planned treatment duration in patients with relapsed multiple myeloma treated with carfilzomib, lenalidomide, and dexamethasone (KRd) versus lenalidomide and dexamethasone (Rd) in the phase III ASPIRE study. Journal of hematology & oncology, 2018. 11(1): p. 49.	ASPIRE study: have included other publications from the same study with longer follow-up
Dimopoulos, M.A., et al., Oral ixazomib maintenance following autologous stem cell transplantation (TOURMALINE-MM3): a double-blind, randomised, placebo-controlled phase 3 trial. Lancet (London, England), 2019. 393(10168): p. 253-264.	Wrong population
Dimopoulos, M.A., et al., Carfilzomib or bortezomib in relapsed or refractory multiple myeloma (ENDEAVOR): an interim overall survival analysis of an open-label, randomised, phase 3 trial. The Lancet. Oncology, 2017. 18(10): p. 1327-1337.	ENDEAVOR study: have included other publications from the same study with longer follow-up
Dimopoulos, M.A., et al., Elotuzumab plus lenalidomide/dexamethasone for relapsed or refractory multiple myeloma: ELOQUENT-2 follow-up and post-hoc analyses on progression-free survival and tumour growth. British journal of haematology, 2017. 178(6): p. 896-905.	ELOQUENT-2 study: have included other publications from the same study with longer follow-up
Dimopoulos, M.A., et al., Daratumumab plus lenalidomide and dexamethasone versus lenalidomide and dexamethasone in relapsed or refractory multiple myeloma: updated analysis of POLLUX. Haematologica, 2018. 103(12): p. 2088-2096.	POLLUX study: have included other publications from the same study with longer follow-up

Dimopoulos, M.A., et al., Elotuzumab plus lenalidomide and dexamethasone in relapsed/refractory multiple myeloma: Extended 4-year follow-up and analysis of relative progression-free survival from the randomized ELOQUENT-2 trial. Cancer, 2018. 124(20): p. 4032-4043.	ELOQUENT-2 study: have included other publications from the same study with longer follow-up
Garderet, L., et al., Pomalidomide, cyclophosphamide, and dexamethasone for relapsed multiple myeloma. Blood, 2018. 132(24): p. 2555-2563.	Not randomized
Goldschmidt, H., et al., Carfilzomib-dexamethasone versus subcutaneous or intravenous bortezomib in relapsed or refractory multiple myeloma: secondary analysis of the phase 3 ENDEAVOR study. Leukemia & lymphoma, 2018. 59(6): p. 1364-1374.	Subgroup analysis
Gupta, N., et al., Dose and Schedule Selection of the Oral Proteasome Inhibitor Ixazomib in Relapsed/Refractory Multiple Myeloma: Clinical and Model-Based Analyses. Targeted oncology, 2017. 12(5): p. 643-654.	Wrong outcome
Isoda, A., et al., Intra-patient dose escalation of panobinostat in patients with relapsed/refractory multiple myeloma. Leuk Lymphoma, 2018. 59(5): p. 1277-1278.	Wrong study design
Kropff, M., et al., Thalidomide versus dexamethasone for the treatment of relapsed and/or refractory multiple myeloma: results from OPTIMUM, a randomized trial. Haematologica, 2012. 97(5): p. 784-91.	Wrong intervention
Lonial, S., et al., Elotuzumab Therapy for Relapsed or Refractory Multiple Myeloma. N Engl J Med, 2015. 373(7): p. 621-31.	ELOQUENT-2 study: have included other publications from the same study with longer follow-up
Mateos, M.V., et al., Effect of prior treatments on selinexor, bortezomib, and dexamethasone in previously treated multiple myeloma. Journal of hematology & oncology, 2021. 14(1): p. 59.	Subgroup analysis
Raab, M.S., et al., MOR202, a novel anti-CD38 monoclonal antibody, in patients with relapsed or refractory multiple myeloma: a first-in-human, multicentre, phase 1-2a trial. The Lancet Haematology, 2020. 7(5): p. e381-e394.	Wrong study design
Richardson, P., et al., A Phase III randomized, open-label study of isatuximab (SAR650984) plus pomalidomide and dexamethasone versus pom and dex in relapsed/refractory multiple myeloma. Haematologica - Volume 102, Issue 0, pp. 779 - published 2017-01-01	Conference abstract
San-Miguel, J.F., et al., Panobinostat plus bortezomib and dexamethasone: impact of dose intensity and administration frequency on safety in the PANORAMA 1 trial. British journal of haematology, 2017. 179(1): p. 66-74.	Subgroup analysis
Shah, J., et al., Oprozomib, pomalidomide, and Dexamethasone in Patients With Relapsed and/or Refractory Multiple Myeloma. Clinical Lymphoma, Myeloma and Leukemia, 2019. 19(9): p. 570.	Wrong study design
Soekojo, C.Y., et al., Pomalidomide and dexamethasone combination with additional cyclophosphamide in relapsed/refractory multiple myeloma (AMN001)-a trial by the Asian Myeloma Network. Blood Cancer Journal, 2019. 9(10): p. 83.	Wrong study design
Spencer, A., et al., Daratumumab plus bortezomib and dexamethasone versus bortezomib and dexamethasone in relapsed or refractory multiple myeloma: updated analysis of CASTOR. Haematologica, 2018. 103(12): p. 2079-2087.	CASTOR study: have included other publications from the same study with longer follow-up
Usmani, S.Z., et al., Greater treatment satisfaction in patients receiving daratumumab subcutaneous vs. intravenous for relapsed or refractory multiple myeloma: COLUMBA clinical trial results. Journal of Cancer Research & Clinical Oncology, 2020. 27: p. 27.	Wrong outcome

Voorhees, P.M., et al., A phase I/II study of ixazomib, pomalidomide, and dexamethasone for lenalidomide and proteasome inhibitor refractory multiple myeloma (Alliance A061202). American Journal of Hematology, 2021. 96(12): p. 1595-1603.	Not available data for the randomized phase 2 part of the study
Weisel, K., et al., Health-related quality of life of carfilzomib- and daratumumab-based therapies in patients with relapsed/refractory multiple myeloma, based on German benefit assessment data. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2020. 29(1): p. 69-79.	Wrong study design
Weisel, K., et al., Phase 3 ELOQUENT-2 study: Extended 4-year follow-up of Elotuzumab plus lenalidomide/dexamethasone vs lenalidomide/dexamethasone in relapsed/refractory multiple myeloma. Oncology Research and Treatment - Volume 40, Issue 0, pp. 35 - published 2017-01-01	Conference abstract
Wolf, J.L., et al., Phase II trial of the pan-deacetylase inhibitor panobinostat as a single agent in advanced relapsed/refractory multiple myeloma. Leuk Lymphoma, 2012. 53(9): p. 1820-3.	Wrong study design
Yakoub-Agha, I., et al., Low-dose vs. high-dose thalidomide for advanced multiple myeloma: a prospective trial from the Intergroupe Francophone du Myélome. Eur J Haematol, 2012. 88(3): p. 249-59.	Wrong intervention

Appendix 6

Characteristics of included studies

Detailed list of all included 50 studies. Shaded rows indicate studies that could not be included in the statistical analysis.

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
1703 ***Richardson 2015*** *(45)* NCT00742560 Phase 2, Open label Multicentre; International	**ERd, n=36** (E: 10 mg/kg i.v.; R: 25 mg p.o.; 40 p.o.)	Percent with >1 previous lines of treatment: 55% Percent refractory to bortezomib: 28%	1-3 previous lines of treatment. Disease progression since, or refractory to the most recent previous treatment.	None; see *Appendix 7*
	ERd, n=37 (E: 20 mg/kg i.v.; R: 25 mg p.o.; 40 p.o.)	Percent with >1 previous lines of treatment: 54% Number of people refractory to bortezomib: 19%		
AMBER ***White 2013*** *(46)* Phase 2; Blinded; Multicentre; North America	**BevV, n=49** (Bev: 15 mg/kg i.v.; V: 1,3 mg/m^2 i.v.)	Percent with >1 previous lines of treatment: 53%	Disease progression after 1-3 previous lines of treatment	OS, SAE, PFS, Disc.
	PboV, n=53 (Pbo: i.v.; V: 1,3 mg/m^2 i.v.)	Percent with >1 previous lines of treatment: 51%		
APEX ***Richardson 2005*** *(47),* ***Richardson 2007*** *(48),* ***Lee 2008*** *(49)*	**V, n=333** (V: 1,3 mg/m^2 i.v.)	Percent with >1 previous lines of treatment: 60%	Disease progression after 1-3 previous lines of treatment	OS, SAE, Disc

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
NCT00048230 Phase 3; Open label Multicentre; International	**d, n=336** (d: 40 mg p.o.)	Percent with >1 previous lines of treatment: 65%		
APOLLO ***Dimopoulus 2021*** *(50),* ***Terpos 2022*** *(51)* NCT03180736 Phase 3; Open label Multicentre; Europe	**DPd, n=151** (D: 1800 mg s.c.; P: 4 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 89% Percent refractory to PI: 47% Percent refractory to lenalidomide: 79%	≥1 previous lines of treatment with both lenalidomide and a PI, had at least a partial response to one or more previous lines of therapy, and were refractory to lenalidomide if they had received only one previous line of treatment	PFS, Disc.
	Pd, n=153 (P: 4 ,g p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 88% Percent refractory to PI: 49% Percent refractory to lenalidomide: 80%		
ARROW ***Moreau 2018*** *(52),* ***Moreau 2019*** *(53)* NCT02412878 Phase 3; Open label Multicentre; International	**Kd, n=240** (K: 70 mg/m^2 i.v. 1/3w/cycle; d: 40 mg)	Percent with >2 previous lines of treatment: 52% Percent refractory to bortezomib: 46% Percent refractory to lenalidomide: 86%	2-3 previous lines of treatments. Refractory to most recent therapy, including bortezomib or ixazomib. Previous exposure to PI (except carfilzomib) and IMiD	None; see *Appendix 7*
	Kd, n=238 (K: 27 mg/m^2 i.v. 2/3w/cycle; d: 40 mg)	Percent with >2 previous lines of treatment: 47% Percent refractory to bortezomib: 38% Percent refractory to lenalidomide: 82%		
ASPIRE ***Stewart 2015*** *(30),* ***Siegel 2018*** *(54)* NCT01080391 Phase 3; Open label Multicentre; International	**KRd, n=396** (K: 27 mg/m^2 i.v.; R: 25 mg p.o.; d: 40 mg)	Percent with >1 previous lines of treatment: 53% Percent refractory to lenalidomide: 7%	1-3 previous lines of treatment	OS, QoL, SAE, PFS, Disc.
	Rd, n=396 (R: 25 mg p.o.; d: 40 mg)	Percent with >1 previous lines of treatment: 60% Percent refractory to lenalidomide: 7%		
BELLINI ***Kumar 2020*** *(55)* NCT02755597 Phase 3; Double-blind Multicentre; International	**VenVd, n=194** (Ven: 800 mg p.o.; V: 1,3 mg/m^2 s.c./i.v.; d: 20 mg)	Percent with >1 previous lines of treatment: 53% Percent refractory to lenalidomide: 20%	1-3 previous lines of treatment. Sensitive or naïve to PI	OS, SAE, PFS, Disc.
	PboVd, n=97 (Pbo: p.o.; V: 1,3 mg/m^2 s.c./i.v.; d: 20 mg)	Percent with >1 previous lines of treatment: 55% Percent refractory to lenalidomide: 28%		
BOSTON ***Grosecki 2020*** *(56)*	**SeVd, n=195** (S: 100 mg p.o.; V: 1,3 mg/m2 s.c.; d: 20 mg)	Percent with >1 previous lines of treatment: 49%		OS, SAE, PFS, Disc.

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
NCT03110562 Phase 3; Open label Multicentre, International	**Vd, n=207** (V: 1,3 mg/m2 s.c.; d: 20 mg)	Percent with >1 previous line of treatment: 52%	1-3 previous lines of treatment. At least partial response to potential previous PI. Refractory to most recent treatment	
CANDOR ***Dimopoulos 2020*** *(57),* ***Siegel 2021*** *(59),* ***Usmani 2022*** *(58)* NCT03158688 Phase 3; Open label Multicentre; International	**DKd, n=312** (D: 16 mg/kg i.v.; K: 56 mg/m^2 i.v.; d: 40 mg p.o.)	Percent with >1 previous lines of treatment: 54% Percent refractory to bortezomib: 28% Percent refractory to lenalidomide: 32%	1-3 previous lines of treatment. Partial response or better to ≥1 treatment. Relapsed after last treatment.	OS, QoL, SAE, PFS, Disc.
	Kd, n=154 (K: 56 mg/m^2 i.v.; d: 40 mg p.o.)	Percent with >1 previous lines of treatment: 55% Percent refractory to bortezomib: 31% Percent refractory to lenalidomide: 36%		
CASTOR ***Mateos 2020*** *(40),* ***Hungria 2021*** *(61),* ***Palumbo 2016*** *(60)* NCT02136134 Phase 3; Open label Multicentre; International	**DVd** (D: 16 mg/kg i.v.; V: 1,3 mg/m^2 i.v.; d: 40 mg p.o./i.v.)	Percent with >1 previous lines of treatment: 51% Percent refractory to lenalidomide: 24%	≥1 previous line of treatment. At least partial response to ≥1 previous treatment.	OS, SAE, PFS, Disc.
	Vd (V: 1,3 mg/m^2 i.v.; d: 40 mg p.o./i.v.)	Percent with >1 previous lines of treatment: 54% Percent refractory to lenalidomide: 33%		
COLUMBA ***Mateos 2020*** *(62)* NCT03277105 Phase 3; Open label Multicentre; International	**D, n=263** (1800 mg s.c.)	Percent with >4 previous lines of treatment: 34%	≥3 previous lines of treatments, including a PI or an IMiD, or double refractory to both PI and IMiD. Response to ≥1 previous treatment regimen.	None; see *Appendix 7*
	D, n=259 (16 mg/kg i.v.)	Percent with >4 previous lines of treatment: 34%		
CREST ***Jagannath 2004*** *(63)* Phase 2; Open label Multicentre; USA	**Vd, n=28** (V: 1,0 mg/m^2 i.v.; d: 20 mg p.o.)	Median (range) number of previous lines of treatment: 3 (1-7)	Relapsed or refractory to front-line chemotherapy (first regimen in myeloma therapy).	None; see *Appendix 7*
	Vd, n=26 (V: 1,3 mg/m^2 i.v.; d: 20 mg p.o.)	Median (range) number of previous lines of treatment: 3 (1-7)		
DOXIL-MMY-3021 ***Orlowski 2007*** *(64),* ***Orlowski 2016*** *(65)* NCT00103506 Phase 3; Open label Multicentre; International	**V, n=322** (V: 1,3 mg/m^2 i,v.)	Percent with >2 previous lines of treatment: 66%	Progression after ≥1 previous line of treatment, or refractory to initial treatment. Bortezomib naïve.	OS, SAE, PFS, Disc.
	DoxV, n=324 (Dox: 30 mg/m^2 i,v.; V: 1,3 mg/m2 i.v.)	Percent with >2 previous lines of treatment: 66%		

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
ELOQUENT-2 ***Cella 2018*** *(66),* ***Dimopoulos 2020*** *(67)* NCT01239797 Phase 3; Open label Multicentre; International	**ERd, n=321** (E: 10 mg/kg i.v.; R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous lines of treatment: 53% Percent refractory to bortezomib: 22%	1-3 previous lines of treatment. Refractory to most recent therapy.	OS, QoL, SAE, PFS, Disc.
	Rd, n=325 (R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous lines of treatment: 51% Percent refractory to bortezomib: 21%		
ELOQUENT-3 ***Dimopoulos 2018*** *(68)* NCT02654132 Phase 2; Open label Multicentre; International	**EPd, n=60** (E: 20 mg/kg i.v.; P: 4 mg p.o.; d: 40 mg p.o.)	Percent with >3 previous lines of treatment: 40% Percent refractory to PI: 78% Percent refractory to lenalidomide: 90%	≥2 previous lines of treatments, including ≥2 consecutive cycles of lenalidomide and/or PI. Refractory to last treatment regime.	OS, SAE, PFS, Disc.
	Rd, n=57 (P: 4 mg p.o.; d: 40 mg p.o.)	Percent with >3 previous lines of treatment: 37% Percent refractory to PI: 82% Percent refractory to lenalidomide: 84%		
ENDEAVOR ***Dimopoulos 2016*** *(69),* ***Ludwig 2019*** *(70),* ***Orlowski 2019*** *(71)* NCT01568866 Phase 3; Open label Multicentre; International	**Kd, n=464** (K: 56 mg/m^2 i.v.; d: 20 mg p.o./i.v.)	Percent with >1 previous lines of treatment: 50%	1-3 previous lines of treatment. At least partial response to ≥1 previous treatment regime.	OS, QoL, SAE, PFS, Disc.
	Vd, n=465 (V: 1,3 mg/m^2 i.v./s.c.; d: 20 mg p.o./i.v.)	Percent with >1 previous lines of treatment: 50%		
FOCUS ***Hájek 2017*** *(72)* NCT01302392 Phase 3; Open label Multicentre; International	**K (± Cy), n=157** (K: 27 mg/m^2 i.v.)	Percent with >5 previous lines of treatment: 43% Percent refractory to bortezomib: 66% Percent refractory to lenalidomide:73%	≥3 previous lines of treatments, including bortezomib, alkylating agent, lenalidomide, thalidomide, corticosteroids, anthracycline. Refractory to the most recent therapy.	OS, SAE, PFS, Disc.
	cs (± Cy), n=158 (for d: 6 mg p.o./for p: 30 mg p.o.)	Percent with >5 previous lines of treatment: 42% Percent refractory to bortezomib: 68% Percent refractory to lenalidomide: 71%		
ICARIA-MM ***Attal 2019*** *(73)* NCT02990338	**Isd, n=154** (Is: 10 mg/kg i.v.; P: 4 mg p.o.; d: 40 mg p.o./i.v.)	Median (range) number of previous lines of treatment: 3 (2-4) Percent refractory to PI: 77% Percent refractory to lenalidomide: 94%	≥2 previous lines of treatments. No response to treatment with lenalidomide and/or a PI.	OS, QoL, SAE, PFS, Disc.

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
Phase 3; Open label Multicentre; International	**Pd, n=153** (P: 4 mg p.o.; d: 20 mg p.o./i.v.)	Median (range) number of previous lines of treatment: 3 (2-4) Percent refractory to PI: 75% Percent refractory to lenalidomide: 92%		
IFM 2009-02 ***Leleu 2013*** *(74)* NCT01053949 Phase 2; Open label Multicentre; France	**Pd, n=43** (P: 4 mg p.o. 21/28; d: 40 mg p.o.)	Percent with >6 previous line of treatment: 28% Percent refractory to bortezomib: 79%% Percent refractory to lenalidomide: 84%	≥1 previous line of treatment. No response to last line of treatment with lenalidomide	None; see *Appendix 7*
	Pd, n=41 (P: 4 mg p.o. 28/28; d; 40 mg p.o.)	Percent with >6 previous line of treatment: 17% Percent refractory to bortezomib: 83% Percent refractory to lenalidomide: 95%		
IKEMA ***Moreau 2021*** *(75)* NCT03275285 Phase 3; Open label Multicentre, International	**IsKd, n=179** (Is: 10 mg/kg i.v.; K: 56 mg/m^2 i.v., d: 20 mg i.v./p.o.)	Percent with >1 previous line of treatment: 56% Percent refractory to PI: 31% Percent refractory to lenalidomide: 32%	1-3 previous line of treatment. Not primary refractory	QoL, PFS, Disc.
	Kd, n=123 (K: 56 mg/m^2 i.v.; d: 20 mg i.v./p.o.)	Percent with >1 previous line of treatment: 55% Percent refractory to PI: 36% Percent refractory to lenalidomide: 34%		
KEYNOTE-183 ***Mateos 2019 (76)*** NCT02576977 Phase 3; Open label Multicentre; International	**PemPd, n=125** (Pem: 200 mg i.v.; P: 4 mg p.o.; d: 40 mg p.o.)	Median (range) number of previous lines of treatment: 3 (1-3) Percent refractory to lenalidomide: 86%	≥2 previous lines of treatments, including PIs or IMiDs. Failure of last line of therapy.	OS, SAE, PFS, Disc.
	Pd, n=124 (P: 4 mg p.o.; d: 40 mg p.o.)	Median (range) number of previous lines of treatment: 3 (1-3) Percent refractory to lenalidomide: 86%		
LEPUS ***Lu 2021*** *(39)* NCT03234972 Phase 3; Open label Multicentre; China	**DVd, n=141** (D: 16 mg/kg i.v.; V: 1,3 mg/m^2 s.c.; d: 20 mg i.v./p.o.)	Percent with >1 previous line of treatment: 71% Percent refractory to lenalidomide: 25%	≥1 previous line of treatment. At least partial response to ≥1 prior regimen.	OS, SAE, PFS, Disc.
	Vd, n=70 (V: 1,3 mg/m^2 i.v.; d: 20 mg i.v./p.o.)	Percent with >1 previous line of treatment: 72% Percent refractory to lenalidomide: 30%		
MM-002 ***Richardson 2014*** *(77)* NCT00833833 Phase 2; Open label Multicentre; North America	**Pd, n=113** (P: 4 mg p.o.; d: 40 mg p.o.)	Percent with >2 previous line of treatment: 95% Percent refractory to bortezomib: 71% Percent refractory to lenalidomide: 78%	≥2 previous lines of treatment, including ≥2 cycles of lenalidomide and ≥2 cycles of bortezomib, given separately or in combination.	OS, SAE, PFS, Disc.
	P, n=108 (P: 4 mg p.o.)	Percent with >2 previous line of treatment: 95% Percent refractory to bortezomib: 70% Percent refractory to lenalidomide: 80%		

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
MM-003 ***San Miguel 2013*** *(78),* ***Song 2015*** *(79),* ***Weisel 2015*** *(80)* NCT01311687 Phase 3; Open label Multicentre; International	**Pd, n=302** (P: 4 mg p.o.; d: 40 mg p.o. QW)	Percent with >2 previous line of treatment: 94% Percent refractory to bortezomib: 79% Percent refractory to lenalidomide: 95%	≥2 previous consecutive cycles of bortezomib and lenalidomide, alone or in combination. Failed treatment with bortezomib and lenalidomide.	OS, SAE, PFS, Disc.
	d, n=153 (d: 40 mg p.o. 4/3w/cycle)	Percent with >2 previous line of treatment: 95% Percent refractory to bortezomib: 79% Percent refractory to lenalidomide: 92%		
MM-009 ***Weber 2007*** *(81)* NCT00056160 Phase 3; Double blind Multicentre; North America	**Rd, n=177** (R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 62%	≥1 previous line of treatment. Not resistant to dexamethasone	OS, SAE, Disc.
	Pbo + d, n=176 (Pbo: p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 62%		
MM-010 ***Dimopoulos 2007*** *(82)* NCT00424047 Phase 3; Double blind Multicentre; International	**Rd, n=176** (R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 68%	≥1 previous line of treatment. Not resistant to dexamethasone	OS, SAE
	Pbo + d, n=175 (P: 4 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 67%		
MMVAR/IFM 2005-04 ***Garderet 2012*** *(83)* Phase 3; Open label Multicentre; Europe	**VTd, n=135** (V: 1,3 mg/m^2 i,v.; T: 200 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 47%	≥1 autologous stem cell transplantation. First relapse or progression.	PFS, Disc.
	Td, n=134 (T: 200 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 47%		
MMY-3033 ***Terpos 2018*** *(84)* NCT01910987 Phase 3; Open label Multicentre; Europe	**Vd, n=53** (V: 1,3 mg/m^2 s.c. 1/4w/cycle; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 42%	At least partial response to previous treatment with bortezomib	None; see *Appendix 7*
	Vd, n=27 (V: 1,3 mg/m^2 s.c. 2/2w/cycle; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 44%		
The Nordic Myeloma Study	**Td, n=67** (T: 200 mg p.o.; d: 40 mg p.o.)	Not reported	Refractory to melphalan	QoL, SAE

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
Hjorth 2012 *(25)* NCT00602511 Phase 3; Open label Multicentre; Scandinavia	**Vd, n=64** (V: 1,3 mg/m^2 i,v.; d: 20 mg p.o.)	Not reported		
OPTIMISMM ***Richardson 2019*** *(85)* ***Weisel 2020*** *(86)* NCT01734928 Phase 3; Open label Multicentre; International	**PVd, n=281** (P: 4 mg p.o.; V: 1,3 mg/m^2 i,v./s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 61% Percent refractory to bortezomib: 9% Percent refractory to lenalidomide: 71%	1-3 previous lines of treatment, including ≥2 cycles of lenalidomide treatment. Progressive disease during last treatment.	OS, SAE, PFS, Disc.
	Vd, n=278 (V: 1,3 mg/m^2 i,v./s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 58% Percent refractory to bortezomib: 12% Percent refractory to lenalidomide: 69%		
PANORAMA-1 ***San Miguel 2014*** *(87),* ***San Miguel 2016*** *(88),* ***Richardson 2018*** *(89)* NCT01023308 Phase 3; Double blind Multicentre; International	**FVd, n=387** (F: 20 mg p.o.; V: 1,3 mg/m^2 i.v.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 49%	1-3 previous lines of treatment. Bortezomib-sensitive. Not primary refractory.	OS, QoL, SAE, PFS, Disc.
	PboVd, n=381 (Pbo: p.o.; V: 1,3 mg/m^2 i.v.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 48%		
PANORAMA-3 ***Laubach 2021*** *(90)* NCT02654990 Phase 2; Open label Multicentre; International	**FVd, n=82** (F: 20 mg p.o. 3/2w/cycle; V: 1,3 mg/m^2 s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 59% Percent refractory to bortezomib: 1% Percent refractory to lenalidomide: 21%	1-4 previous lines of treatments, including an IMiD. Sensitive to bortezomib.	None; see *Appendix 7*
	FVd, n=83 (F: 20 mg p.o. 2/2w/cycle; V: 1,3 mg/m^2 s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 62% Percent refractory to bortezomib: 0% Percent refractory to lenalidomide: 12%		
	FVd, n=83 (F: 10 mg p.o. 3/2w/cycle; V: 1,3 mg/m^2 s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 62% Percent refractory to bortezomib: 0% Percent refractory to lenalidomide: 19%		
POLLUX ***Dimopoulos 2016*** *(91),* ***Bahlis 2020*** *(92),*	**DRd, n=286** (D: 16 mg/kg i.v.; R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 48% Percent refractory to PI: 20% Percent refractory to IMiD: 4%	≥1 previous line of treatment. Relapsed on/after last treatment	OS, SAE, PFS, Disc.

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
Plesner 2021 *(93)* NCT02076009 Phase 3; Open label Multicentre; International	**Rd, n=283** (R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 48% Percent refractory to PI: 16% Percent refractory to IMiD: 4%		
SIRIUS ***Lonial 2016*** *(94)* NCT01985126 Phase 2; Open label Multicentre; International	**D + mp, n=106** (D: 16 mg/kg i.v. QW/Q2W/Q4W; mp: 20 mg p.o.)	Percent with >3 previous line of treatment: 82% Percent refractory to bortezomib: 90% Percent refractory to lenalidomide: 88%	≥3 previous lines of treatments, including PI and IMiD. Responded to ≥1 previous treatment. Received alkylating agent alone or in combination with other myeloma treatment. Double refractory to most recent PI and IMiD treatment	None; see *Appendix 7*
	D + mp, n=18 (D: 16 mg/kg i.v. Q4W; mp: 20 mg p.o.)	Percent with >3 previous line of treatment: 67% Percent refractory to bortezomib: 89% Percent refractory to lenalidomide: 89%		
The China Continuation Study ***Hou 2017*** *(37)* NCT01564537 Phase 3; Double blind Multicentre; China	**IRd, n=57** (I: 4 mg p.o.; R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 56%	1-3 previous lines of treatment. Sensitive to lenalidomide and PI.	OS, SAE, PFS, Disc.
	PboRd, n=58 (Pbo: p.o.; R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 55%		
TOURMALINE-MM1 ***Moreau 2016*** *(95),* ***Leleu 2018*** *(26),* ***Richardson 2021*** *(38)* NCT01564537 Phase 3; Double blind Multicentre, International	**IRd, n=360** (I: 4 mg p.o.; R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 38% Percent refractory to PI: 16% Percent refractory to IMiD: 21%	1-3 previous lines of treatment. Sensitive to lenalidomide and PI.	OS, QoL, SAE, PFS, Disc.
	PboRd, n=362 (Pbo: p.o.; R: 25 mg p.o.; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 40% Percent refractory to PI: 2% Percent refractory to IMiD: 25%		
VANTAGE-088 ***Dimopoulos 2013*** *(96)* NCT00773747 Phase 3; Double blind Multicentre; International	**VorV, n=317** (Vor: 400 mg p.o.; V: 1,3 mg/m2 i.v.)	Percent with >1 previous line of treatment: 55%	1-3 previous lines of treatment. Achieved a response on previous treatment. Progressive disease after most recent treatment. Sensitive to bortezomib.	OS, SAE, PFS, Disc.
	PboV, n=320 (Pbo: p.o.; V: 1,3 mg/m^2 i.v.)	Percent with >1 previous line of treatment: 60%		
Ailawadhi 2020 (97) NCT01903811	**Kd, n=64** (K: 27 mg/m^2 i.v.; d: 20 mg i.v.)	Percent with 4-6 previous line of treatment: 22% Percent refractory to bortezomib: 50%	1-6 previous lines of treatments. Carfilzomib-naïve	None; see *Appendix 7*

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
Phase 2; Open label Multicentre; USA	**Kd, n=57** (K: 56 mg/m^2 i.v.; d: 20 mg i.v.)	Percent with 4-6 previous line of treatment: 26% Percent refractory to bortezomib: 49%		
Baz 2016 (98) NCT01432600 Phase 2; Open label Multicentre; USA	**Pd, n=36** (P: 4 mg p.o.; d: 40 mg)	Median (range) number of previous lines of treatment: 4 (2-12) Percent refractory to bortezomib: 78%	≥2 previous lines of treatments, including IMiD. Refractory to lenalidomide.	Disc.
	CyPd, n=34 (Cy: 400 mg p.o.; P: 4 mg p.o.; d: 40 mg)	Median (range) number of previous lines of treatment: 4 (2-9) Percent refractory to bortezomib: 71%		
Dimopoulus 2021 (99) NCT01084252 Phase 2; Open label Multicentre, International	**Is, n=109** (Is: 20 mg/kg i.v.)	Median (range) number of previous lines of treatment: 4 (2-10) Percent refractory to bortezomib: 65% Percent refractory to lenalidomide: 71%	≥3 previous lines of treatments, including a PI and an IMiD, or refractory to a PI and an IMiD. Achieved at least minimal response to a previous treatment.	OS, SAE, PFS, Disc.
	Is + d, n=55 (Is: 20 mg/kg i.v.; d: 40 mg i.v./p.o.)	Median (range) number of previous lines of treatment: 4 (2-10) Percent refractory to bortezomib: 67% Percent refractory to lenalidomide: 62%		
Iida 2016 (100) UMIN000003135 Phase 2; Open label Multicentre; Japan	**Vd, n=22** (V: 1,3 mg/m^2 i.v./s.c.; d: 20 mg p.o.)	Percent with ≥2 previous line of treatment: 22%	≥1 previous line of treatment. Treatment-naïve to bortezomib and thalidomide.	OS, SAE, PFS, Disc.
	Td, n=22 (T: 100-200 mg p.o.; d: 20 mg p.o.)	Percent with ≥2 previous line of treatment: 22%		
Jakubowiak 2016 (101) NCT01478048 Phase 2; Open label Multicentre; International	**EVd, n=77** (E: 10 mg/kg i.v.; V: 1,3 mg/m^2 i.v./s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 35%	Progression after 1-3 previous lines of treatment.	OS, SAE, PFS, Disc.
	Vd, n=75 (V: 1,3 mg/m^2 i.v./s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 32%		
Kropff 2017 (102) NCT00813150 Phase 3; Open label Multicentre; Germany	**Vd, n=46** (V: 1,3 mg/m^2 i.v.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 37%	1-3 previous lines of treatment. Primary refractory or relapsed disease.	OS, SAE, PFS, Disc.
	CyVd, n=47 (Cy: 50 mg p.o.; V: 1,3 mg/m^2 i.v.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 42%		
Mikhael 2020 (103) NCT01084252	**Is, n=23** (3 mg/kg, i.v. Q2W)	Median (range) number of previous lines of treatment: 5 (2-12) Percent refractory to bortezomib: 74% Percent refractory to lenalidomide: 83%	≥3 previous lines of treatments, including a PI and an IMiD. Received an alkylating	None; see *Appendix 7*

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
Phase 2; Open label Multicentre; International	**Is, n=25** (10 mg/kg, i.v. Q2W/Q4W)	Median (range) number of previous lines of treatment: 5 (3-14) Percent refractory to bortezomib: 88% Percent refractory to lenalidomide: 80%	agent. At least minimal response to ≥1 previous line of treatment.	
	Is, n=24 (10 mg/kg, i.v. Q2W)	Median (range) number of previous lines of treatment: 6 (2-13) Percent refractory to bortezomib: 67% Percent refractory to lenalidomide: 83%		
	Is, n=25 (20 mg/kg, i.v. QW/Q2W)	Median (range) number of previous lines of treatment: 5 (2-10) Percent refractory to bortezomib: 68% Percent refractory to lenalidomide: 88%		
Mina 2020 (104) NCT01913730 Phase 2; Open label Multicentre; Italy	**Vd, n=15** (V: 1,3 mg/m^2 s.c. Q2W; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 100%	1-3 previous lines of treatment. Bortezomib-based treatment as the last line of therapy (≥4 cycles) without progression.	None; see *Appendix 7*
	Std.care, n=20	Percent with >1 previous line of treatment: 80%		
	Vd, n=23 (V: 1,3 mg/m^2 s.c. QW; d: 40 mg p.o.)	Percent with >1 previous line of treatment: 100%		
Montefusco 2020 (105) EUDRACT 2010-021557-40 Phase 3; Open label Multicentre; Italy	**CyVd, n=76** (Cy: 500 mg/m^2 i.v.; V: 1,3 mg/m2 s.c.; d: 20 mg p.o./i.v.)	Not reported	First symptomatic relapse.	PFS
	CyRd, n=79 (Cy: 500 mg/m^2 i.v.; R: 25 mg p.o.; d: 20 mg p.o./i.v.)	Not reported		
Moreau 2011 (106) NCT00722566 Phase 3; Open Label Multicentre; Europe	**Vd, n=148** (V: 1,3 mg/m^2 s.c.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 38%	1-3 previous lines of treatment. Progression since last treatment.	None; see *Appendix 7*
	Vd, n=74 (V: 1,3 mg/m^2 i.v.; d: 20 mg p.o.)	Percent with >1 previous line of treatment: 35%		
Orlowski 2015 (107) NCT00401843 Phase 2; Double blind Multicentre; International	**SV, n=142** (S: 6 mg/kg i.v.; V: 1,3 mg/m^2 i.v.)	Percent with >1 previous line of treatment: 51%	1-3 previous lines of treatment. Bortezomib-naïve.	OS, PFS, Disc.
	PboV, n=144 (Pbo: i.v.; V: 1,3 mg/m^2 i.v.)	Percent with >1 previous line of treatment: 47%		

Study	Intervention and comparator	Previous lines of treatments and refractoriness	Inclusion criteria related to previous treatment	Outcomes in the analysis
Raje 2017 (108) NCT01602224 Phase 2; Double blind Multicentre; International	**TabVd, n=74** (Tab: 100 mg. i.v.; V: 1,3 mg/m^2 s.c.; d: 20 mg p.o.)	Median (range) number of previous lines of treatment: 1 (1-3)	1-3 previous lines of treatment.	OS, SAE, PFS, Disc.
	TabVd, n=74 (Tab: 300 mg. i.v.; V: 1,3 mg/m^2 s.c.; d: 20 mg p.o.)	Median (range) number of previous lines of treatment: 1 (0-3)		
	PboVd, n=72 (Pbo: i.v.; V: 1,3 mg/m^2 s.c.; d: 20 mg p.o.)	Median (range) number of previous lines of treatment: 1 (1-3)		
Sehgal 2015 (109) NCT01319422 Phase 2; Open label USA	**Pd, n=19** (P: 2 mg. p.o. 28/28; d: 40 mg)	Median number of previous lines of treatment: 4 Percent refractory to lenalidomide: 100%	≥2 previous lines of treatments. Refractory to lenalidomide.	None; see *Appendix 7*
	Pd, n=20 (P: 4 mg. p.o. 21/28; d: 40 mg)	Median number of previous lines of treatment: 4 Percent refractory to lenalidomide: 100%		

Bev: bevacizumab; cs.: corticosteroids; Cy: cyclophosphamide; d: dexamethasone; D: daratumumab; Disc: discontinuations due to adverse events; Dox: doxorubicin; E: elotuzumab; F: panobinostat (Farydak); I: ixazomib; Is: isatuximab; i.v.: intravenous; K: carfilzomib (Kyprolis); mp: methylprednisolone; OS: overall survival; P: pomalidomide; p: prednisone; Pbo: placebo; Pem: pembrolizumab; PFS: progression-free survival; p.o.: per oral; R: lenalidomide (Revlimid); S: siltuximab; SAE: severe adverse events; s.c.: subcutaneous; Se: Selinexor; SR: systematic review; T: thalidomide; Tab: tabalumab; V: bortezomib (Velcade); Vor: vorinostat

Appendix 7

Included articles not used in the analysis

List of included articles that were omitted from the statistical analysis, with reasons for why.

Author – year – [ref] Study name; study number	Study treatment	Reason for omission in the analysis
Ailawadhi 2020 (97) NCT01903811	[K + d] vs [K + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Hungria 2021 (61) *CASTOR;* NCT02136134	[DV + d] vs [V + d]	The study did not measure QoL on a scale that was compatible with the planned meta-analysis.
Jagannath 2004 (63) *CREST*	[V + d] vs [V + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Laubach 2021 (90) *PANORAMA-3;* NCT02654990	[FV + d] vs [FV + d] vs [FV + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Lee 2008 (49) *APEX;* NCT00048230	[V] vs [d]	The study did not measure QoL on a scale that was compatible with the planned meta-analysis.
Leleu 2013 (74) *IFM 2009-02;* NCT01053949	[P + d] vs [P + d] – different cycles	The study compared different regimens of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Leleu 2018 (26) *TOURMALINE-MM1;* NCT01564537	[IR + d] vs [Placebo + R + d]	Included for QoL, but could not be used in the analysis because required data were not available
Lonial 2016 (94) *SIRIUS;* NCT01985126	[D + mp] vs [D + mp] – different cycles	The study compared different regimens of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Ludwig 2019 (70) *ENDEAVOR;* NCT01568866	[K + d] vs [V + d]	Included for QoL, but could not be used in the analysis because required data were not available

Mateos 2020 (62) *COLUMBA;* NCT03277105	[D] vs [D] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Mikhael 2020 (103) NCT01913730	[Is] vs [Is] vs [Is] vs [Is] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Mina 2020 (104) NCT01913730	[V + d] vs [V + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Moreau 2011 (106) NCT00722566	[V + d] vs [V + d] – different administration methods	The study compared different administration methods of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Moreau 2018 (52) ARROW; NCT02412878	[K + d] vs [K + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Moreau 2019 (53) ARROW; NCT02412878	[K + d] vs [K + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Plesner 2021 (93) *POLLUX;* NCT02076009	[DR + d] vs [R + d]	The study did not measure QoL on a scale that was compatible with the planned meta-analysis.
Richardson 2015 (45) *1703;* NCT00742560	[ER + d] vs [ER + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Sehgal 2015 (109) NCT01319422	[P + d] vs [P + d] – different doses	The study compared different doses of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Song 2015 (79) *MM-003;* NCT01311687	[P + d] vs [d]	The study did not measure QoL on a scale that was compatible with the planned meta-analysis.
Terpos 2018 (84) *MMY-3033;* NCT01910987	[V + d] vs [V + d] – different cycles	The study compared different regimens of the same drug. Under our treatment definition (which does not distinguish between different posologies) this study reduces to a single arm trial without a comparison.
Terpos 2022 (51) *APOLLO;* NCT03180736	[DP + d] vs [P + d]	The study did not measure QoL on a scale that was compatible with the planned meta-analysis.
Weisel 2015 (80) *MM-003;* NCT01311687	[P + d] vs [d]	The study did not measure QoL on a scale that was compatible with the planned meta-analysis.
Weisel 2020 (86) *OPTIMISMM;* NCT01734928	[PV + d] vs [V + d]	The study did not measure QoL on a scale that was compatible with the planned meta-analysis.

Appendix 8

Ongoing studies

List of ongoing studies, listed by phase (phase 3, phase 2-3, phase 2, and phase 1-2).

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT03110562/Bortezomib, Selinexor, and Dexamethasone in Patients With Multiple Myeloma (BOSTON)	**BOSTON***	Active, not recruiting/ 2023	Arm 1: SeVd Arm 2: Vd	RCT phase 3 N=402	PFS
NCT02136134/Addition of Daratumumab to Combination of Bortezomib and Dexamethasone in Participants With Relapsed or Refractory Multiple Myeloma	**CASTOR***	Active, not recruiting/ 2024	Arm 1: DVd Arm 2: Vd	RCT phase 3 N=500	PFS
NCT03277105/A Study of Subcutaneous Versus (vs.) Intravenous Administration of Daratumumab in Participants With Relapsed or Refractory Multiple Myeloma	**COLUMBA***	Active, not recruiting/ 2023	Arm 1: D – i.v. Arm 2: D – s.c.	RCT phase 3 N=522	ORR, max trough concentration

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT02990338/Multinational Clinical Study Comparing Isatuximab, Pomalidomide, and Dexamethasone to Pomalidomide and Dexamethasone in Refractory or Relapsed and Refractory Multiple Myeloma Patients (ICARIA-MM)	**ICARIA-MM***	Active, not recruiting/ 2022	Arm 1: IsPd Arm 2: Pd	RCT phase 3 N=307	PFS
NCT03275285/Multinational Clinical Study Comparing Isatuximab, Carfilzomib And Dexamethasone To Carfilzomib And Dexamethasone In Relapse And/Or Refractory Multiple Myeloma Patients (IKEMA)	**IKEMA***	Active, not recruiting/ 2023	Arm 1: IsKd Arm 2: Kd	RCT phase 3 N=302	PFS
NCT02076009/A Study Comparing Daratumumab, Lenalidomide, and Dexamethasone With Lenalidomide and Dexamethasone in Relapsed or Refractory Multiple Myeloma	**POLLUX***	Active, not recruiting/ 2024	Arm 1: DRd Arm 2: Rd	RCT phase 3 N=570	PFS
NCT03562169/The Role of Ixazomib in Autologous Stem Cell Transplant in Relapsed Myeloma - Myeloma XII (ACCoRd)	ACCoRd	Recruiting/ 2027	Arm 1: conventional ASCT + ITd Arm 2: augmented ASCT + ITd	RCT phase 3 N=406	PFS
NCT03859427/A Study Comparing Once-weekly vs Twice-weekly Carfilzomib in Combination With Lenalidomide and Dexamethasone in Subjects With Relapsed or Refractory Multiple Myeloma (ARROW2)	ARROW2	Recruiting/ 2023	Arm 1: KRd – QW Arm 2: KRd – BIW	RCT phase 3 N=460	OR
NCT04939142/A Study of Evaluating the Safety and Efficacy of ATG-010, Bortezomib, and Dexamethasone (SVd) Versus Bortezomib and Dexamethasone (Vd) in Patients With Relapsed or Refractory Multiple Myeloma (RRMM) (BENCH)	BENCH	Recruiting/ 2024	Arm 1: SeVd Arm 2: Vd	RCT phase 3 N=150	PFS

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT04181827/A Study Comparing JNJ-68284528, a CAR-T Therapy Directed Against B-cell Maturation Antigen (BCMA), Versus Pomalidomide, Bortezomib and Dexamethasone (PVd) or Daratumumab, Pomalidomide and Dexamethasone (DPd) in Participants With Relapsed and Lenalidomide-Refractory Multiple Myeloma (CARTITUDE-4)	CARTITUDE-4	Active, not recruiting/ 2026	Arm 1: Ciltacabtagene Autoleucel Arm 2: PVd or DPd	RCT phase 3 N=419	PFS
NCT03836014/Study Comparing Continuous Versus Fixed Duration Therapy With Daratumumab, Lenalidomide, and Dexamethasone for Relapsed Multiple Myeloma (CONFIRM)	CONFIRM	Recruiting/ 2025	Arm 1: DRd - fixed Arm 2: DRd - continuous	RCT phase 3 N=434	OS
NCT04162210/Study of Single Agent Belantamab Mafodotin Versus Pomalidomide Plus Low-dose Dexamethasone (Pom/Dex) in Participants With Relapsed/Refractory Multiple Myeloma (RRMM)	DREAMM3	Recruiting/ 2025	Arm 1: Belantamab mafodotin Arm 2: Pd	RCT phase 3 N=380	PFS
NCT04246047/Evaluation of Efficacy and Safety of Belantamab Mafodotin, Bortezomib and Dexamethasone Versus Daratumumab, Bortezomib and Dexamethasone in Participants With Relapsed/Refractory Multiple Myeloma (DREAMM 7)	DREAMM7	Active, not recruiting/ 2026	Arm 1: Belantamab mafodotin + Vd Arm 2: DVd	RCT phase 3 N=575	PFS
NCT04484623/Belantamab Mafodotin Plus Pomalidomide and Dexamethasone (Pd) Versus Bortezomib Plus Pd in Relapsed/Refractory Multiple Myeloma (DREAMM 8)	DREAMM8	Recruiting/ 2028	Arm 1: Belantamab mafodotin + Pd Arm 2: PVd	RCT phase 3 N=450	PFS

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT03180736/Comparison of Pomalidomide and Dexamethasone With or Without Daratumumab in Subjects With Relapsed or Refractory Multiple Myeloma Previously Treated With Lenalidomide and a Proteasome Inhibitor Daratumumab/Pomalidomide/Dexamethasone vs Pomalidomide/Dexamethasone (EMN14)	EMN14	Active, not recruiting/ 2022	Arm 1: DPd Arm 2: Pd	RCT phase 3 N=304	PFS
NCT03651128/Efficacy and Safety Study of bb2121 Versus Standard Regimens in Subjects With Relapsed and Refractory Multiple Myeloma (RRMM) (KarMMa-3)	KarMMa-3	Recruiting/ 2025	Arm 1: BB2121 Arm 2: Std treatment (DPd, DVd, IRd, Kd, or EPd)	RCT phase 3 N=381	PFS
NCT05020236/MagnetisMM-5: Study of Elranatamab (PF-06863135) Monotherapy and Elranatamab + Daratumumab Versus Daratumumab + Pomalidomide + Dexamethasone in Participants With Relapsed/Refractory Multiple Myeloma (MAGNETISMM-5)	MAGNETISMM-5	Recruiting/ 2026	Arm 1: Elranatamab Arm 2: Elranatamab + D	RCT phase 3 N=476	Part 2: PFS
NCT05083169/A Study of Teclistamab in Combination With Daratumumab Subcutaneously (SC) (Tec-Dara) Versus Daratumumab SC, Pomalidomide, and Dexamethasone (DPd) or Daratumumab SC, Bortezomib, and Dexamethasone (DVd) in Participants With Relapsed or Refractory Multiple Myeloma (MajesTEC-3)	MajesTEC-3	Recruiting/ 2026	Arm 1: Teclistamab + D Arm 2: DPd or DVd	RCT phase 3 N=560	PFS
NCT03151811/A Study of Melphalan Flufenamide (Melflufen)-Dex or Pomalidomide-dex for RRMM Patients Refractory to Lenalidomide (OCEAN)	OCEAN	Active, not recruiting/ 2024	Arm 1: Melflufen + d Arm 2: Pd	RCT phase 3 N=495	PFS

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT03952091/TJ202, Lenalidomide and Dexamethasone vs. Lenalidomide and Dexamethasone in Subjects With Relapsed or Refractory Multiple Myeloma		Active, not recruiting/ 2022	Arm 1: TJ202 + Rd Arm 2: Rd	RCT phase 3 N=291	PFS
NCT03539744/A Study of Venetoclax and Dexamethasone Compared With Pomalidomide and Dexamethasone in Participants With Relapsed or Refractory Multiple Myeloma		Recruiting/ 2024	Arm 1: Venetoclax + d Arm 2: Pd	RCT phase 3 N=244	PFS
NCT04975997/Open-label Study Comparing Iberdomide, Daratumumab and Dexamethasone (IberDd) Versus Daratumumab, Bortezomib, and Dexamethasone (DVd) in Participants With Relapsed or Refractory Multiple Myeloma (RRMM)	EXCALIBER-RRMM	Not yet recruiting/ 2029	Arm 1: Iberdomide + Dd Arm 2: DVd	RCT phase 3 N=864	PFS
NCT03234972/A Study to Compare Daratumumab, Bortezomib, and Dexamethasone (DVd) vs Bortezomib and Dexamethasone (Vd) in Chinese Participants With Relapsed or Refractory Multiple Myeloma	**LEPUS***	Active, not recruiting/ 2022	Arm 1: DVd Arm 2: Vd	RCT phase 3 N=211	PFS
NCT05028348/A Study of Combination of Selinexor, Pomalidomide, and Dexamethasone (SPd) Versus Elotuzumab, Pomalidomide, and Dexamethasone (EloPd) in Subject With Previously Treated Multiple Myeloma		Recruiting/ 2028	Arm 1: SePd Arm 2: EPd	RCT phase 3 N=280	PFS
NCT03143049/Pomalidomide-Cyclophosphamide-Dexamethasone (PCD) Versus Pomalidomide-Dexamethasone (PD) in Relapse or Refractory Myeloma		Recruiting/ 2022	Arm 1: CyPd Arm 2: Pd	RCT phase 3 N=120	PFS
NCT05405166/SC Versus IV Isatuximab in Combination With Pomalidomide and Dexamethasone in RRMM		Recruiting/ 2026	Arm 1: IsPd (s.c.) Arm 2: IsPd (i.v.)	RCT phase 3 N=534	ORR

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT04513639/The Relapse From MRD Negativity as Indication for Treatment (REMNANT) Study	REMNANT	Recruiting/ 2031	Arm 1: DKd early treatment Arm 2: DKd standard treatment	RCT phase 2-3 N=176	PFS, OS
NCT02654990/Panobinostat/Bortezomib/Dexamethasone in Relapsed or Relapsed-and-refractory Multiple Myeloma (PANORAMA_3)	**PANORAMA-3***	Active, not recruiting/ 2024	Arm 1: FVd (F: 20 mg TIW) Arm 2: FVd (F: 20 mg BIW) Arm 3: FVd (F: 10 mg TIW)	RCT phase 2 N=249	ORR
NCT02406222/Pomalidomide in Relapsed and Refractory Multiple Myeloma (RRMM) (MUKseven)	MUKseven	Active, not recruiting/ 2023	Arm 1: CyPd Arm 2: Pd	RCT phase 2 N=124	PFS
NCT03697655/Pre-emptive Daratumumab Therapy of Minimal Residual Disease Reappearance or Biochemical Relapse in Multiple Myeloma (PREDATOR)	PREDATOR	Recruiting/ 2024	Arm 1: D Arm 2: Observation	RCT phase 2 N=274	Event-free survival
NCT04414475/A Study of Selinexor Plus Low-dose Dexamethasone in Participants With Penta-refractory Multiple Myeloma or Selinexor and Bortezomib Plus Low-dose Dexamethasone in Participants With Triple-class Refractory Multiple Myeloma		Recruiting/ 2023	Arm 1: SeVd Arm 2: Se (40 mg BIW) + d Arm 3: Se (80 mg BIW) + d Arm 4: Se (100 mg QW) + d	RCT phase 2 N=134	ORR
NCT01415882/Ixazomib Citrate in Treating Patients With Relapsed Multiple Myeloma That Is Not Refractory to Bortezomib		Active, not recruiting/ 2023	Arm 1: CyId Arm 2: CyID Arm 3-5: Id various doses	RCT phase 2 N=108	Proportion confirmed responses
NCT01745588/Autologous Stem Cell Transplant With Pomalidomide (CC-4047®) Maintenance Versus Continuous Clarithromycin/ Pomalidomide / Dexamethasone Salvage Therapy in Relapsed or Refractory Multiple Myeloma		Active, not recruiting/ 2023	Arm 1: Clarithromycin + Pd + ASCT Arm 2: Clarithromycin + Pd	RCT phase 2 N=23	ORR

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT02765854/Ixazomib and Dexamethasone Versus Ixazomib, Dexamethasone and Lenalidomide, Randomized With NFKB2 Rearrangement		Recruiting/ 2022	Arm 1: IRd – mutated NFκB Arm 2: Id – mutated NFκB Arm 3: Id – unmutated NFκB	RCT phase 2 N=90	VGPR rate
NCT03215524/A Study Of Daratumumab, Low-Dose Oral Dexamethasone and Cyclophosphamide With Or Without Pomalidomide (DCDP)		Active, not recruiting/ 2023	Arm 1: Cy + DPd Arm 2: Cy + Dd (+ P)	RCT phase 2 N=120	PFS
NCT01913730/Maintenance Therapy With Subcutaneous Bortezomib		Active, not recruiting/ 2021	Arm 1: Vd Arm 2: No prolonged therapy	RCT phase 2 N=63	Time to progression
NCT03336073/Carfilzomib, Cyclophosphamide, Dexamethasone in Multiple Myeloma		Active, not recruiting/ 2022	Arm 1: CyKd Arm 2: Kd	RCT phase 2 N=199	PFS
NCT03184194/Nivolumab Combined With Daratumumab With or Without Low-dose Cyclophosphamide		Active, not recruiting/ 2023	Arm 1: Nivolumab + D Arm 2: Nivolumab + CyDd	RCT phase 2 N=62	ORR
NCT03411031/Elotuzumab Plus Lenalidomide (Elo/Rev) for Serologic Relapse/Progression While on Lenalidomide		Active, not recruiting/ 2023	Arm 1: ERd (R: 25 mg) Arm 2: ERd (R: 10 mg)	RCT phase 2 N=18	PFS
NCT05064358/Study to Investigate Alternative Dosing Regimens of Belantamab Mafodotin in Participants With Relapsed or Refractory Multiple Myeloma (DREAMM 14)	DREAMM 14	Recruiting/ 2024	Arm 1-5: belantamab mafodotin – various posologies	RCT phase 2 N=180	Incidence rate
NCT03525678/A Study to Investigate the Efficacy and Safety of Two Doses of GSK2857916 in Participants With Multiple Myeloma Who Have Failed Prior Treatment With an Anti-CD38 Antibody		Active, not recruiting/ 2023	Arm 1-3: belantamab mafodotin – various posologies, frozen liquid or lyophilized powder	RCT phase 2 N=221	ORR
NCT03871829/Daratumumab Retreatment in Participants With Multiple Myeloma Who Have Been Previously Treated With Daratumumab		Recruiting/ 2026	Arm 1: DKd Arm 2: Kd	RCT phase 2 N=230	VGPR

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT05346809/Isatuximab During Stem Cell Collection and Transplant in Patients With Multiple Myeloma and Lymphoma		Recruiting/ 2024	Arm 1: Std ASCT + isatuximab Arm 2: Std ASCT	RCT phase 2 N=39	Total lymphocyte count change
NCT04126200/Platform Study of Belantamab Mafodotin as Monotherapy and in Combination With Anti-cancer Treatments in Participants With Relapsed/Refractory Multiple Myeloma (RRMM) (DREAMM 5)	DREAMM5	Recruiting/ 2028	Arm 1-5: Belantamab mafodotin + other treatment (GSK3174998, feladilimab, nirogacestat, dorstalimab, isatuximab)	RCT phase ½ N=464	Phase 2: ORR
NCT02343042/Selinexor and Backbone Treatments of Multiple Myeloma Patients (STOMP)	STOMP	Recruiting/ 2022	Arm 1-11: Selinexor + another treatment (Pd, Vd, Rd, PVd, Dd, Kd, Rd, Id, EPd, Belantamab + d, DPd)	RCT phase ½ N=518	Duration of response, clinical benefit rate
NCT04643002/Isatuximab in Combination With Novel Agents in RRMM - Master Protocol	UMBRELLA	Recruiting/ 2026	Arm 1: IsPd Arm 2: SAR439459 + Is + d Arm 3: Belantamab mafodotin + Is + d	RCT phase ½ N=117	VGPR rate
NCT04973605/A Phase 1b/2 Study of BGB-11417in Monotherapy and in Various Combinations With Dexamethasone and Carfilzomib in Multiple Myeloma		Recruiting/ 2025	Part 1: BGB-11417 + Kd dose escalation Part 2: BGB-11417 + Kd various posologies	RCT phase ½ N=167	Phase 2: ORR, VGPR, CR
NCT03314181/A Study of Combination Therapy With Venetoclax, Daratumumab and Dexamethasone (With and Without Bortezomib) in Participants With Relapsed or Refractory Multiple Myeloma		Active, not recruiting/ 2025	Arm 1: Venetoclax + DVd Arm 2: DVd	RCT phase ½ N=156	ORR

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT03194867/Isatuximab in Combination With Cemiplimab in Relapsed/Refractory Multiple Myeloma (RRMM) Patients		Active, not recruiting/ 2022	Arm 1: Cemiplimab + Is Arm 2: Cemiplimab + Is Arm 3: Is	RCT phase ½ N=109	ORR
NCT04150965/Immuno-Oncology Drugs Elotuzumab, Anti-LAG-3 and Anti-TIGIT		Recruiting/ 2024	Arm 1: EPd Arm 2: Anti LAG-3 + Pd Arm 3: Anti-TIGIT Arm 4: Anti-TIGIT + Pd	RCT phase ½ N=104	ORR, AE
NCT04892264/Belantamab Mafodotin, Lenalidomide, and Daratumumab for the Treatment of Relapsed, Refractory, or Previously Untreated Multiple Myeloma		Recruiting/ 2026	Arm 1: Belantamab mafodotin + DR Arm 2: Belantamab mafodotin + DRd	RCT phase ½ N=12	CR rate
NCT02004275/Pomalidomide and Dexamethasone With or Without Ixazomib in Treating Patients With Relapsed Multiple Myeloma		Active, not recruiting/ 2021	Arm 1: IPd Arm 2: Pd	RCT phase ½ N=117	Phase 2: PFS
NCT05427812/Phase ½ Study of ISB 1442 in Relapsed/Refractory Multiple Myeloma		Not yet recruiting/ 2027	Phase 1: ISB 1442 – dose escalation Phase 2: ISB 1442 in RRMM and RRMM post T-cell directed therapy participants	RCT phase ½ N=121	Phase 2: ORR
NCT01592370/An Investigational Immuno-Therapy Study to Determine the Safety and Effectiveness of Nivolumab and Daratumumab in Patients With Multiple Myeloma		Active, not recruiting/ 2022	Arm 1: Nivolumab + daratumumab Arm 2: daratumumab	RCT phase ½ N=316	AE, toxicities

Study ID/Study name Title	Studienavn	Status/ Estimated end	Treatments	Study design/ Enrollment (n)	Main outcome
NCT02773030/A Study to Determine Dose, Safety, Tolerability, Drug Levels, and Efficacy of CC-220 Monotherapy, and in Combination With Other Treatments in Participants With Multiple Myeloma		Recruiting/ 2028	Arm 1: Iberdomide + dexamethasone Arm 2: Iberdomide	RCT phase ½ N=532	ORR

**Ongoing study that has results and is included in our report*

AE: adverse events, ASCT: autologous stem cell transplant, BIW: twice weekly, Cy: cyclophosphamide, CR: complete response, D: daratumumab, d: dexamethasone, E: elotuzumab, I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NFκB: nuclear factor kappa B, LAG-3: lymphocyte-activation gene 3, ORR: overall respond rate, P: pomalidomide, PFS: progression-free survival, QW: once weekly, R: lenalidomide (Revlimid), RRMM: relapsed refractory multiple myeloma, s.c. subcutaneous, Se: selinexor, Std: standard, TIGIT: T cell immunoreceptor with Ig and ITIM domains, TIW: three times weekly, V: bortezomib (Velcade), VGPR: very good partial response

Appendix 9

Detailed GRADE: overall survival

Direct estimates

Table with detailed assessment of certainty of evidence (GRADE) for direct estimates of overall survival.

Treatment A	Treatment B	Study name	Starting point for GRADE	Reasons for downgrading	GRADE assessment	
BevV	V	AMBER	High	Imprecision	Low	⊕⊕
CyV + d	V + d	Kropff 2017	High	Imprecision	Low	⊕⊕
DK + d	K + d	CANDOR	High	Imprecision	Low	⊕⊕
DoxV	V	DOXIL-MMY-3021	High	Study limitations, imprecision	Low	⊕⊕
DR + d	R + d	POLLUX	High	Imprecision	Moderate	⊕⊕⊕
DV + d	V + d	LEPUS	High	Study limitations, indirectness	Very low	⊕
DV + d	V + d	CASTOR	High	Imprecision	Low	⊕⊕
EP + d	P + d	ELOQUENT-3	High	Imprecision	Low	⊕⊕
ER + d	R + d	ELOQUENT-2	High	Imprecision	Moderate	⊕⊕⊕
EV + d	V + d	Jakubowiak 2016	High	Study limitations, imprecision	Very low	⊕

Treatment A	Treatment B	Study name	Starting point for GRADE	Reasons for downgrading	GRADE assessment	
FV + d	V + d	PANORAMA-1	High	Imprecision	Moderate	⊕⊕⊕
IR + d	R + d	China Contin. study	High	Indirectness	Very low	⊕
IR + d	R + d	TOURMALINE-MM1	High	Imprecision	Very low	⊕
Is	Is + d	Dimopoulos 2021	High	Study limitations, imprecision	Very low	⊕
IsP + d	P + d	ICARIA-MM	High	Study limitations, imprecision	Low	⊕⊕
K	cs (± Cy)	FOCUS	High	Imprecision	Low	⊕⊕
K + d	V + d	ENDEAVOR	High	None	High	⊕⊕⊕⊕
KR + d	R + d	ASPIRE	High	None	High	⊕⊕⊕⊕
P	P + d	MM-002	High	Imprecision	Low	⊕⊕
P + d	d	MM-003	High	Imprecision	Moderate	⊕⊕⊕
PemP + d	P + d	KEYNOTE-183	High	Imprecision	Low	⊕⊕
PV + d	V + d	OPTIMISMM	High	Imprecision	Low	⊕⊕
R + d	d	MM-009	High	None	High	⊕⊕⊕⊕
R + d	d	MM-010	High	Imprecision	Moderate	⊕⊕⊕
SeV + d	V + d	BOSTON	High	Study limitations, imprecision	Very low	⊕
T + d	V + d	Iida 2018	High	Indirectness, imprecision	Very low	⊕
TabV + d 300 mg	V + d	Raje 2017	High	Study limitations, imprecision	Very low	⊕
TabV + d 100 mg	V + d	Raje 2017	High	Study limitations. imprecision	Very low	⊕
VeV + d	V + d	BELLINI	High	Imprecision	Moderate	⊕⊕⊕
VorV	V	VANTAGE-088	High	Imprecision	Low	⊕⊕

Bev: bevacizumab, Cy; cyclophosphamide, d: dexamethasone, D: daratumumab, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, Pem: pembrolizumab, R: lenalidomide (Revlimid), S: siltuximab, Se: selinexor, Tab: tabalumab, V: bortezomib, Ve: venetoclax, Vor: vorinostat.

The threshold of important effect was set at 0,8 and 1,25 and was used solely for the purpose as aid for assessing certainty of evidence.

NMA estimates – direct comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from direct comparisons within the matrix (*Table 7*), for overall survival.

Treatment A	Treatment B	Study name	Network	Starting point for GRADE	Reasons for downgrading	GRADE assessment
R + d	DR + d	POLLUX	NW1	High	Imprecision	Low ⊕⊕
EP + d	P + d	ELOQUENT-3	NW1	High	Imprecision	Low ⊕⊕
R + d	ER + d	ELOQUENT-2	NW1	High	Imprecision	Low ⊕⊕
R + d	IR + d	China Contin., TOURMALINE	NW1	High	Inconsistency, indirectness, imprecision	Very low ⊕
IsP + d	P + d	ICARIA-MM	NW1	High	Study limitations, imprecision	Very low ⊕
R + d	KR + d	ASPIRE	NW1	High	Imprecision	Low ⊕⊕
DK + d	K + d	CANDOR	NW2	High	Imprecision	Low ⊕⊕
DV + d	V + d	LEPUS, CASTOR	NW2	High	Study limitations, inconsistency, indirectness, imprecision	Very low ⊕
FV + d	V + d	PANORAMA-1	NW2	High	Imprecision	Low ⊕⊕
K + d	V + d	ENDEAVOR	NW2	High	Imprecision	Low ⊕⊕
PV + d	V + d	OPTIMISMM	NW2	High	Imprecision	Low ⊕⊕

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib
The threshold of important effect was set at 0,8 and 1,25 and was used solely for the purpose as aid for assessing certainty of evidence.

NMA estimates – indirect comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from indirect comparisons within the matrix, for overall survival.

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DoxV → DR + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → DR + d	POLLUX	NW1	Moderate		
DoxV → EP + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Intransitivity, imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
DoxV → ER + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → ER + d	ELOQUENT-2	NW1	High		
DoxV → IR + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → IR + d	China Continuation Study, TOURMALINE-MM1	NW1	Very low		
DoxV → IsP + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Intransitivity, imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → IsP + d	ICARIA-MM	NW1	Low		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DoxV → KR + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → KR + d	ASPIRE	NW1	High		
DoxV → P + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Intransitivity, imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → P + d	MM-003	NW1	High		
DoxV → R + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
DR + d → EP + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
DR + d → ER + d	DR + d → R + d	POLLUX	NW1	Moderate	Imprecision	Very low ⊕
	R + d → ER + d	ELOQUENT-2	NW1	High		
DR + d → IR + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → IR + d	China Continuation Study, TOURMALINE-MM1	NW1	Very low		
DR + d → IsP + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → IsP + d	ICARIA-MM	NW1	Low		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DR + d → P + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
EP + d → IR + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → IR + d	China Continuation Study, TOURMALINE-MM1	NW1	Very low		
EP + d → IsP + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Imprecision	Very low ⊕
	P + d → IsP + d	ICARIA-MM	NW1	Low		
EP + d → R + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
ER + d → EP + d	ER + d → R + d	ELOQUENT-2	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
ER + d → IR + d	ER + d → R + d	ELOQUENT-2	NW1	High	Imprecision	Very low ⊕
	R + d → IR + d	China Continuation Study, TOURMALINE-MM1	NW1	Very low		
ER + d → IsP + d	ER + d → R + d	ELOQUENT-2	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → IsP + d	ICARIA-MM	NW1	Low		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
ER + d → P + d	ER + d → R + d	ELOQUENT-2	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
IR + d → P + d	IR + d → R + d	China Continuation Study, TOURMALINE-MM1	NW1	Very low	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
IsP + d → IR + d	IsP + d → P + d	ICARIA-MM	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → IR + d	China Continuation Study, TOURMALINE-MM1	NW1	Very low		
IsP + d → R + d	IsP + d → P + d	ICARIA-MM	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	High		
	d → R + d	MM-009, MM-010	NW1	High		
KR + d → DR + d	KR + d → R + d	ASPIRE	NW1	High	Imprecision	Very low ⊕
	R + d → DR + d	POLLUX	NW1	Moderate		
KR + d → EP + d	KR + d → R + d	ASPIRE	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
KR + d → ER + d	KR + d → R + d	ASPIRE	NW1	High	Imprecision	Very low ⊕
	R + d → ER + d	ELOQUENT-2	NW1	High		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
KR + d →IR + d	KR + d → R + d	ASPIRE	NW1	High	Imprecision	Very low ⊕
	R + d → IR + d	China Continuation Study, TOURMALINE-MM1	NW1	Very low		
KR + d → IsP + d	KR + d → R + d	ASPIRE	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
	P + d → IsP + d	ICARIA-MM	NW1	Low		
KR + d → P + d	KR + d → R + d	ASPIRE	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d →	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	High		
R + d → P + d	R + d → d	MM-009, MM-010	NW1	High	Intransitivity, imprecision	Very low ⊕
	d → P +	MM-003	NW1	High		
DK + d → FV + d	DK + d → K + d	CANDOR	NW1	Low	Imprecision	Very low ⊕
	K + d → V + d	ENDEAVOR	NW2	High		
	V + d → FV + d	PANORAMA-1	NW2	Moderate		
DK + d → PV + d	DK + d → K + d	CANDOR	NW1	Low	Imprecision	Very low ⊕
	K + d → V + d	ENDEAVOR	NW2	High		
	V + d → PV + d	OPTIMISMM	NW2	Low		
DV + d → DK + d	DV + d → V + d	LEPUS, CASTOR	NW2	Very low	Imprecision	Very low ⊕
	V + d → K + d	ENDEAVOR	NW2	High		
	K + d → DK + d	CANDOR	NW1	Low		
DV + d → FV + d	DV + d → V + d	LEPUS, CASTOR	NW2	Very low	Imprecision	Very low ⊕
	V + d → FV + d	PANORAMA-1	NW2	Moderate		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DV + d → PV + d	DV + d → V + d	LEPUS, CASTOR	NW2	Very low	Imprecision	Very low ⊕
	V + d → PV + d	OPTIMISMM	NW2	Low		
FV + d → PV + d	FV + d → V + d	PANORAMA-1	NW2	Moderate	Imprecision	Very low ⊕
	V + d → PV + d	OPTIMISMM	NW2	Low		
K + d → DV + d	K + d → V + d	ENDEAVOR	NW2	High	Imprecision	Very low ⊕
	V + d → DV + d	LEPUS, CASTOR	NW2	Very low		
K + d → FV + d	K + d → V + d	ENDEAVOR	NW2	High	Imprecision	Very low ⊕
	V + d → FV + d	PANORAMA-1	NW2	Moderate		
K + d → PV + d	K + d → V + d	ENDEAVOR	NW2	High	Imprecision	Very low ⊕
	V + d → PV + d	OPTIMISMM	NW2	Low		
V + d → DK + d	V + d → K + d	ENDEAVOR	NW2	High	Imprecision	Very low ⊕
	K + d → DK + d	CANDOR	NW1	Low		

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib

The threshold of important effect was set at 0,8 and 1,25 and was used solely for the purpose as aid for assessing certainty of evidence.

Detailed GRADE: quality of life

NMA estimates – direct comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from direct comparisons within the matrix (*Table 9*), for quality of life.

Treatment A	Treatment B	Study name	Network	Starting point for GRADE	Reasons for downgrading	GRADE assessment	
R + d	ER + d	ELOQUENT-2	NW2	High	None	High	⊕⊕⊕⊕
R + d	IR + d	TOURMALINE-MM1	NW2	High	None	High	⊕⊕⊕⊕
IsP + d	P + d	ICARIA-MM	NW3	High	Study limitations, imprecision	Low	⊕⊕
DK + d	K + d	CANDOR	NW1	High	None	High	⊕⊕⊕⊕
IsK + d	K + d	IKEMA	NW1	High	Imprecision	Moderate	⊕⊕⊕
R + d	KR + d	ASPIRE	NW1	High	Imprecision	Moderate	⊕⊕⊕
FV + d	V + d	PANORAMA-1	NW1	High	None	High	⊕⊕⊕⊕
K + d	V + d	ENDEAVOR	NW1	High	None	High	⊕⊕⊕⊕

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib

The threshold of important effect was set at an absolute score change of 8 and was used solely for the purpose as aid for assessing certainty of evidence.

NMA estimates – indirect comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from indirect comparisons within the matrix (*Table 9*), for quality of life.

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DK + d → FV + d	DK + d → K + d	CANDOR	NV1	High	Imprecision	Moderate ⊕⊕⊕
	K + d → V + d	ENDEAVOR	NV1	High		
	V + d → FV + d	PANORAMA-1	NV1	High		
DK + d → IsK + d	DK + d → K + d	CANDOR	NV1	High	Imprecision	Low ⊕⊕
	K + d → IsK + d	IKEMA	NV1	Moderate		
ER + d → IR + d	ER + d → R + d	ELOQUENT-2	NV2	High	None	High ⊕⊕⊕⊕
	R + d → IR + d	TOURMALINE-MM1	NV2	High		
IsK + d → FV + d	IsK + d → K + d	IKEMA	NV1	Moderate	Imprecision	Low ⊕⊕
	K + d → V + d	ENDEAVOR	NV1	High		
	V + d → FV + d	PANORAMA-1	NV1	High		
K + d → FV + d	K + d → V + d	ENDEAVOR	NV1	High	None	High ⊕⊕⊕⊕
	V + d → FV + d	PANORAMA-1	NV1	High		
KR + d → ER + d	KR + d → R + d	ASPIRE	NV2	Moderate	Imprecision	Low ⊕⊕
	R + d → ER + d	ELOQUENT-2	NV2	High		
KR + d → IR + d	KR + d → R + d	ASPIRE	NV2	Moderate	Imprecision	Low ⊕⊕
	R + d → IR + d	TOURMALINE-MM1	NV2	High		
V + d → DK + d	V + d → K + d	ENDEAVOR	NV1	High	None	High ⊕⊕⊕⊕
	K + d → DK + d	CANDOR	NV1	High		
V + d → IsK + d	V + d → K + d	ENDEAVOR	NV1	High	None	Moderate ⊕⊕⊕
	K + d → IsK + d	IKEMA	NV1	Moderate		

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, R: lenalidomide (Revlimid), V: bortezomib (Velcade). The threshold of important effect was set at an absolute score change of 8 and was used solely for the purpose as aid for assessing certainty of evidence.

Detailed GRADE: severe adverse events

NMA estimates – direct comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from direct comparisons within the matrix (*Table 11*), for severe adverse events.

Treatment A	Treatment B	Study name	Network	Starting point for GRADE	Reasons for downgrading	GRADE assessment
R + d	DR + d	POLLUX	NW1	High	Imprecision	Moderate ⊕⊕⊕
EP + d	P + d	ELOQUENT-3	NW1	High	Imprecision	Low ⊕⊕
R + d	ER + d	ELOQUENT-2	NW1	High	Imprecision	Moderate ⊕⊕⊕
R + d	IR + d	China Contin., TOURMALINE-MM1	NW1	High	Indirectness	Low ⊕⊕
IsP + d	P + d	ICARIA-MM	NW1	High	Study limitations, imprecision	Very low ⊕
R + d	KR + d	ASPIRE	NW1	High	Imprecision	Moderate ⊕⊕⊕
DK + d	K + d	CANDOR	NW2	High	Imprecision	Moderate ⊕⊕⊕
DV + d	V + d	LUPUS, CASTOR	NW2	High	Study limitations, indirectness	Low ⊕⊕
FV + d	V + d	PANORAMA-1	NW2	High	None	High ⊕⊕⊕⊕
K + d	V + d	ENDEAVOR	NW2	High	None	High ⊕⊕⊕⊕
PV + d	V + d	OPTIMISMM	NW2	High	Imprecision	Moderate ⊕⊕⊕

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib
The threshold of important effect was set at 0,5 and 2 and was used solely for the purpose as an aid for assessing certainty of evidence.

NMA estimates – indirect comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from indirect comparisons within the matrix (*Table 11*), for severe adverse events.

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DoxV → DR + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → DR + d	POLLUX	NW1	Moderate		
DoxV → EP + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Intransitivity, imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → P + d	MM-003	NW1	Moderate		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
DoxV → ER + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → ER + d	ELOQUENT-2	NW1	Moderate		
DoxV → IR + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → IR + d	China Contin., TOURMALINE-MM1	NW1	Very low		
DoxV → IsP + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Intransitivity, imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → P + d	MM-003	NW1	Moderate		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	P + d → IsP + d	ICARIA-MM	NW1	Low		
DoxV → KR + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → KR + d	ASPIRE	NW1	Moderate		
DoxV → P + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Intransitivity, imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → P + d	MM-003	NW1	Moderate		
DoxV → R + d	DoxV → V	DOXIL-MMY-3021	NW1	Low	Imprecision	Very low ⊕
	V → d	APEX	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
DR + d → EP + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
DR + d → ER + d	DR + d → R + d	POLLUX	NW1	Moderate	Imprecision	Very low ⊕
	R + d → ER + d	ELOQUENT-2	NW1	Moderate		
DR + d → IR + d	DR + d → R + d	POLLUX	NW1	Moderate	Imprecision	Very low ⊕
	R + d → IR + d	China Contin. TOURMALINE-MM1	NW1	Very low		
DR + d → IsP + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → IsP + d	ICARIA-MM	NW1	Low		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DR + d → P + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity	Low ⊕⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
EP + d → IR + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → IR + d	China Contin. TOURMALINE-MM1	NW1	Very low		
EP + d → IsP + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Imprecision	Very low ⊕
	P + d → IsP + d	ICARIA-MM	NW1	Low		
EP + d → R + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
ER + d → EP + d	ER + d → R + d	ELOQUENT-2	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
ER + d → IR + d	ER + d → R + d	ELOQUENT-2	NW1	Moderate	Imprecision	Very low ⊕
	R + d → IR + d	China Contin. TOURMALINE	NW1	Very low		
ER + d → IsP + d	ER + d → R + d	ELOQUENT-2	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → IsP + d	ICARIA-MM	NW1	Low		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
ER + d → P + d	ER + d → R + d	ELOQUENT-2	NW1	Moderate	Intransitivity	Low ⊕⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
IR + d → P + d	IR + d → R + d	China Contin. TOURMALINE-MM1	NW1	Very low	Intransitivity	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
IsP + d → IR + d	IsP + d → P + d	ICARIA-MM	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
	R + d → IR + d	China Contin. TOURMALINE	NW1	Very low		
IsP + d → R + d	IsP + d → P + d	ICARIA-MM	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	Moderate		
	d → R + d	MM-009, MM-010	NW1	High		
KR + d → DR + d	KR + d → R + d	ASPIRE	NW1	Moderate	Imprecision	Very low ⊕
	R + d → DR + d	POLLUX	NW1	Moderate		
KR + d → EP + d	KR + d → R + d	ASPIRE	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
KR + d → ER + d	KR + d → R + d	ASPIRE	NW1	Moderate	Imprecision	Very low ⊕
	R + d → ER + d	ELOQUENT-2	NW1	Moderate		
KR + d → IR + d	KR + d → R + d	ASPIRE	NW1	Moderate	Imprecision	Very low ⊕

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	R + d → IR + d	China Contin. TOURMALINE-MM1	NW1	Very low		
KR + d → IsP + d	KR + d → R + d	ASPIRE	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → IsP + d	ICARIA-MM	NW1	Low		
KR + d → P + d	KR + d → R + d	ASPIRE	NW1	Moderate	Intransitivity	Low ⊕⊕
	R + d → d	MM-009, MM-010	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
R + d → P + d	R + d → d	MM-009, MM-010	NW1	High	Intransitivity, imprecision	Very low ⊕
	d → P + d	MM-003	NW1	Moderate		
DK + d → FV + d	DK + d → K + d	CANDOR	NW2	Moderate	Imprecision	Very low ⊕
	K + d → V + d	ENDEAVOR	NW2	Moderate		
	V + d → FV + d	PANORAMA-1	NW2	High		
DK + d → PV + d	DK + d → K + d	CANDOR	NW2	Moderate	Imprecision	Very low ⊕
	K + d → V + d	ENDEAVOR	NW2	Moderate		
	V + d → PV + d	OPTIMISMM	NW2	Moderate		
DV + d → DK + d	DV + d → V + d	LEPUS, CASTOR	NW2	Low	Imprecision	Very low ⊕
	V + d → K + d	ENDEAVOR	NW2	Moderate		
	K + d → DK + d	CANDOR	NW2	Moderate		
DV + d → FV + d	DV + d → V + d	LEPUS, CASTOR	NW2	Low	Imprecision	Very low ⊕
	V + d → FV + d	PANORAMA-1	NW2	High		
DV + d → PV + d	DV + d → V + d	LEPUS, CASTOR	NW2	Low	Imprecision	Very low ⊕

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	V + d → PV + d	OPTIMISMM	NW2	Moderate		
FV + d → PV + d	FV + d → V + d	PANORAMA-1	NW2	High	Imprecision	Very low ⊕
	V + d → PV + d	OPTIMISMM	NW2	Moderate		
K + d → DV + d	K + d → V + d	ENDEAVOR	NW2	Moderate	None	Low ⊕⊕
	V + d → DV + d	LEPUS, CASTOR	NW2	Low		
K + d → FV + d	K + d → V + d	ENDEAVOR	NW2	Moderate	Imprecision	Low ⊕⊕
	V + d → FV + d	PANORAMA-1	NW2	High		
K + d → PV + d	K + d → V + d	ENDEAVOR	NW2	Moderate	Imprecision	Low ⊕⊕
	V + d → PV + d	OPTIMISMM	NW2	Moderate		
V + d → DK + d	V + d → K + d	ENDEAVOR	NW2	Moderate	Imprecision	Low ⊕⊕
	K + d → DK + d	CANDOR	NW2	Moderate		

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib

The threshold of important effect was set at 0,5 and 2 and was used solely for the purpose as an aid for assessing certainty of evidence.

Detailed GRADE: progression-free survival

Direct estimates

Table with detailed assessment of certainty of evidence (GRADE) for direct estimates of progression-free survival.

Treatment A	Treatment B	Study name	Starting point for GRADE	Reasons for downgrading	GRADE assessment	
BevV	V	AMBER	High	Imprecision	Low	⊕⊕
CyV + d	V + d	Kropff 2017	High	Imprecision	Low	⊕⊕
CyV + d	CyR + d	Montefusco 2020	High	Imprecision	Low	⊕⊕
DK + d	K + d	CANDOR	High	None	High	⊕⊕⊕
DoxV	V	DOXIL-MMY-3021	High	Study limitations	Moderate	⊕⊕⊕
DP + d	P + d	APOLLO	High	None	High	⊕⊕⊕⊕
DR + d	R + d	POLLUX	High	None	High	⊕⊕⊕⊕
DV + d	V + d	CASTOR	High	None	High	⊕⊕⊕⊕
DV + d	V + d	LEPUS	High	Study limitations, indirectness	Very low	⊕
ER + d	R + d	ELOQUENT-2	High	None	High	⊕⊕⊕⊕
EP + d	P + d	ELOQUENT-3	High	None	High	⊕⊕⊕⊕
EV + d	V + d	Jakubowiak 2016	High	Study limitations, imprecision	Low	⊕⊕
FV + d	V + d	PANORAMA-1	High	None	High	⊕⊕⊕⊕
IR + d	R + d	China Contin. study	High	Study limitations, indirectness	Very low	⊕
IR + d	R + d	TOURMALINE-MM1	High	None	High	⊕⊕⊕⊕
Is	Is + d	Dimopoulos 2021	High	Imprecision	Low	⊕⊕
IsK + d	K + d	IKEMA	High	None	High	⊕⊕⊕⊕
IsP + d	P + d	ICARIA-MM	High	Study limitations	Moderate	⊕⊕⊕
K	cs (± Cy)	FOCUS	High	Imprecision	Low	⊕⊕

Treatment A	Treatment B	Study name	Starting point for GRADE	Reasons for downgrading	GRADE assessment	
K + d	V + d	ENDEAVOR	High	None	High	⊕⊕⊕⊕
KR + d	R + d	ASPIRE	High	None	High	⊕⊕⊕⊕
P	P + d	MM-002	High	None	High	⊕⊕⊕⊕
P + d	d	MM-003	High	None	High	⊕⊕⊕⊕
PemP + d	P + d	KEYNOTE-183	High	Imprecision	Moderate	⊕⊕⊕
PV + d	V + d	OPTIMISMM	High	None	High	⊕⊕⊕⊕
SeV + d	V + d	BOSTON	High	Study limitations	Moderate	⊕⊕⊕
TabV + d 300 mg	V + d	Raje 2017	High	Study limitations, imprecision	Very low	⊕
TabV + d 1 00 mg	V + d	Raje 2017	High	Study limitations, imprecision	Very low	⊕
T + d	V + d	Iida 2018	High	Indirectness, imprecision	Very low	⊕
VeV + d	V + d	BELLINI	High	None	High	⊕⊕⊕⊕
VorV	V	VANTAGE-088	High	None	High	⊕⊕⊕⊕
VT + d	T + d	MMVAR/IFM 2005-04	High	Study limitations	Moderate	⊕⊕⊕

Bev: bevacizumab, cs: corticosteroids, Cy; cyclophosphamide, d: dexamethasone, D: daratumumab, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, Pem: pembrolizumab, R: lenalidomide (Revlimid), Se: selinexor, T: thalidomide, Tab: tabalumab, V: bortezomib, Ve: venetoclax, Vor: vorinostat.
The threshold of important effect was set at 0,8 and 1,25 and was used solely for the purpose as an aid for assessing certainty of evidence.

NMA estimates – direct comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from direct comparisons within the matrix (*Table 13*), for progression-free survival.

Treatment A	Treatment B	Study name	Network	Starting point for GRADE	Reasons for downgrading	GRADE assessment	
R + d	DR + d	POLLUX	NW1	High	None	High	⊕⊕⊕⊕
R + d	ER + d	ELOQUENT-2	NW1	High	None	High	⊕⊕⊕⊕
R + d	IR + d	China Contin. TOURMALINE	NW1	High	Indirectness	Low	⊕⊕
R + d	KR + d	ASPIRE	NW1	High	None	High	⊕⊕⊕⊕
DK + d	K + d	CANDOR	NW2	High	None	High	⊕⊕⊕⊕
IsK + d	K + d	IKEMA	NW2	High	None	High	⊕⊕⊕⊕
DV + d	V + d	LEPUS, CASTOR	NW2	High	Study limitations, indirectness	Low	⊕⊕
FV + d	V + d	PANORAMA-1	NW2	High	None	High	⊕⊕⊕⊕
K + d	V + d	ENDEAVOR	NW2	High	None	High	⊕⊕⊕⊕
PV + d	V + d	OPTIMISMM	NW2	High	None	High	⊕⊕⊕⊕
EP + d	P + d	ELOQUENT-3	NW2	High	None	High	⊕⊕⊕⊕
IsP + d	P + d	ICARIA-MM	NW3	High	Study limitations	Moderate	⊕⊕⊕

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide; R: lenalidomide (Revlimid), V: bortezomib

The threshold of important effect was set at 0,8 and 1,25 and was used solely for the purpose as an aid for assessing certainty of evidence.

NMA estimates – indirect comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from indirect comparisons within the matrix (*Table 13*), for progression-free survival.

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DR + d ⟶ ER + d	DR + d ⟶ R + d	POLLUX	NW1	High	None	High ⊕⊕⊕⊕
	R + d ⟶ ER + d	ELOQUENT-2	NW1	High		
DR + d ⟶ IR + d	DR + d ⟶ R + d	POLLUX	NW1	High	None	High ⊕⊕⊕⊕
	R + d ⟶ IR + d	China Contin. TOURMALINE-MM1	NW1	Low		
ER + d ⟶ IR + d	ER + d ⟶ R + d	ELOQUENT-2	NW1	High	Imprecision	Very low ⊕
	R + d ⟶ IR + d	China Contin. TOURMALINE-MM1	NW1	Low		
KR + d ⟶ DR + d	KR + d ⟶ R + d	ASPIRE	NW1	High	None	High ⊕⊕⊕⊕
	R + d ⟶ DR + d	POLLUX	NW1	High		
KR + d ⟶ ER + d	KR + d ⟶ R + d	ASPIRE	NW1	High	Imprecision	Moderate ⊕⊕⊕
	R + d ⟶ ER + d	ELOQUENT-2	NW1	High		
KR + d ⟶ IR + d	KR + d ⟶ R + d	ASPIRE	NW1	High	Imprecision	Very low ⊕
	R + d ⟶ IR + d	China Contin. TOURMALINE-MM1	NW1	Low		
DK + d ⟶ FV + d	DK + d ⟶ K + d	CANDOR	NW2	High	None	High ⊕⊕⊕⊕
	K + d ⟶ V + d	ENDEAVOR	NW2	High		
	V + d ⟶ FV + d	PANORAMA-1	NW2	High		
DK + d ⟶ IsK + d	DK + d ⟶ K + d	CANDOR	NW2	High	Imprecision	Low ⊕⊕
	K + d ⟶ IsK + d	IKEMA	NW2	High		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DK + d ⟶ PV + d	DK + d ⟶ K + d	CANDOR	NW2	High	None	High ⊕⊕⊕⊕
	K + d ⟶ V + d	ENDEAVOR	NW2	High		
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
DV + d ⟶ DK + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Low	Imprecision	Very low ⊕
	V + d ⟶ K + d	ENDEAVOR	NW2	High		
	K + d ⟶ DK + d	CANDOR	NW2	High		
DV + d ⟶ FV + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Low	None	Low ⊕⊕
	V + d ⟶ FV + d	PANORAMA-1	NW2	High		
DV + d ⟶ IsK + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Low	Imprecision	Very low ⊕
	V + d ⟶ K + d	ENDEAVOR	NW2	High		
	K + d ⟶ IsK + d	IKEMA	NW2	High		
DV + d ⟶ PV + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Low	None	Low ⊕⊕
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
FV + d ⟶ PV + d	FV + d ⟶ V + d	PANORAMA-1	NW2	High	Imprecision	Low ⊕⊕
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
IsK + d ⟶ FV + d	IsK + d ⟶ K + d	IKEMA	NW2	High	None	High ⊕⊕⊕⊕
	K + d ⟶ V + d	ENDEAVOR	NW2	High		
	V + d ⟶ FV + d	PANORAMA-1	NW2	High		
IsK + d ⟶ PV + d	IsK + d ⟶ K + d	IKEMA	NW2	High	None	High ⊕⊕⊕⊕
	K + d ⟶ V + d	ENDEAVOR	NW2	High		
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
K + d ⟶ DV + d	K + d ⟶ V + d	ENDEAVOR	NW2	High	None	Low ⊕⊕
	V + d ⟶ DV + d	LEPUS, CASTOR	NW2	Low		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
K + d ⟶ FV + d	K + d ⟶ V + d	ENDEAVOR	NW2	High	Imprecision	Moderate ⊕⊕⊕
	V + d ⟶ FV + d	PANORAMA-1	NW2	High		
K + d ⟶ PV + d	K + d ⟶ V + d	ENDEAVOR	NW2	High	Imprecision	Moderate ⊕⊕⊕
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
V + d ⟶ DK + d	V + d ⟶ K + d	ENDEAVOR	NW2	High	None	High ⊕⊕⊕⊕
	K + d ⟶ DK + d	CANDOR	NW2	High		
EP + d ⟶ IsP + d	EP + d ⟶ P + d	ELOQUENT-3	NW3	High	Imprecision	Very low ⊕
	P + d ⟶ IsP + d	ICARIA-MM	NW3	Moderate		

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib

The threshold of important effect was set at 0,8 and 1,25 and was used solely for the purpose as an aid for assessing certainty of evidence.

Detailed GRADE: discontinuation due to adverse events

NMA estimates – direct comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from direct comparisons within the matrix (*Table 15*), for discontinuation due to adverse events.

Treatment A	Treatment B	Study name	Network	Starting point for GRADE	Reasons for downgrading	GRADE assessment	
R + d	DR + d	POLLUX	NW1	High	Imprecision	Low	⊕⊕
EP + d	P + d	ELOQUENT-3	NW1	High	Imprecision	Low	⊕⊕
R + d	ER + d	ELOQUENT-2	NW1	High	None	High	⊕⊕⊕⊕
R + d	IR + d	China Contin, TOURMALINE	NW1	High	Indirectness	Low	⊕⊕
IsP + d	P + d	ICARIA-MM	NW1	High	Study limitations, imprecision	Low	⊕⊕
R + d	KR + d	ASPIRE	NW1	High	None	High	⊕⊕⊕⊕
DK + d	K + d	CANDOR	NW2	High	None	High	⊕⊕⊕⊕
IsK + d	K + d	IKEMA	NW2	High	Imprecision	Moderate	⊕⊕⊕
DV + d	V + d	LEPUS, CASTOR	NW2	High	Study limitations, indirectness, imprecision	Very low	⊕
FV + d	V + d	PANORAMA-1	NW2	High	None	High	⊕⊕⊕⊕
K + d	V + d	ENDEAVOR	NW2	High	None	High	⊕⊕⊕⊕
PV + d	V + d	OPTIMISMM	NW2	High	None	High	⊕⊕⊕⊕

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib
The threshold of important effect was set at 0,5 and 2 and was used solely for the purpose as an aid for assessing certainty of evidence.

NMA estimates – indirect comparisons

Table with detailed assessment of certainty of evidence (GRADE) for NMA estimates from direct comparisons within the matrix (*Table 15*), for discontinuation due to adverse events.

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
DoxV → DR + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009	NW1	High		
	R + d → DR + d	POLLUX	NW1	Moderate		
DoxV → EP + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Intransitvity, imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
DoxV → ER + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Imprecision	Low ⊕⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009	NW1	High		
	R + d → ER + d	ELOQUENT-2	NW1	High		
DoxV → IR + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Imprecision	Very low ⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009	NW1	High		
	R + d → IR + d	China Contin., TOURMALINE-MM1	NW1	Low		
DoxV → IsP + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	V → d	APEX	NW1	High		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	d → P + d	MM-003	NW1	Moderate		
	P + d → IsP + d	ICARIA-MM	NW1	Low		
DoxV → KR + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Imprecision	Low ⊕⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009	NW1	High		
	R + d → KR + d	ASPIRE	NW1	High		
DoxV → P + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Intransitivity	Low ⊕⊕
	V → d	APEX	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
DoxV → R + d	DoxV → V	DOXIL-MMY-3021	NW1	Moderate	Imprecision	Low ⊕⊕
	V → d	APEX	NW1	High		
	d → R + d	MM-009	NW1	High		
DR + d → EP + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity	Very low ⊕
	R + d → d	MM-009	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
DR + d → ER + d	DR + d → R + d	POLLUX	NW1	Moderate	Imprecision	Very low ⊕
	R + d → ER + d	ELOQUENT-2	NW1	High		
DR + d → IR + d	DR + d → R + d	POLLUX	NW1	Moderate	Imprecision	Very low ⊕
	R + d → IR + d	China Contin., TOURMALINE-MM1	NW1	Low		
DR + d → IsP + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009	NW1	High		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	d → P + d	MM-003	NW1	Moderate		
	P + d → IsP + d	ICARIA-MM	NW1	Low		
DR + d → P + d	DR + d → R + d	POLLUX	NW1	Moderate	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
EP + d → IR + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	Moderate		
	d → R + d	MM-009	NW1	High		
	R + d → IR + d	China Contin., TOURMALINE-MM1	NW1	Low		
EP + d → IsP + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Imprecision	Very low ⊕
	P + d → IsP + d	ICARIA-MM	NW1	Low		
EP + d → R + d	EP + d → P + d	ELOQUENT-3	NW1	Low	Intransitivity, imprecision	Very low ⊕
	P + d → d	MM-003	NW1	Moderate		
	d → R + d	MM-009	NW1	High		
ER + d → EP + d	ER + d → R + d	ELOQUENT-2	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → EP + d	ELOQUENT-3	NW1	Low		
ER + d → IR + d	ER + d → R + d	ELOQUENT-2	NW1	High	Imprecision	Very low ⊕
	R + d → IR + d	China Contin., TOURMALINE-MM1	NW1	Low		
ER + d → IsP + d	ER + d → R + d	ELOQUENT-2	NW1	High		Very low ⊕

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	R + d ⟶ d	MM-009	NW1	High	Intransitivity, imprecision	
	d ⟶ P + d	MM-003	NW1	Moderate		
	P + d ⟶ IsP + d	ICARIA-MM	NW1	Low		
ER + d ⟶ P + d	ER + d ⟶ R + d	ELOQUENT-2	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d ⟶ d	MM-009	NW1	High		
	d ⟶ P + d	MM-003	NW1	Moderate		
IR + d ⟶ P + d	IR + d ⟶ R + d	China Contin., TOURMALINE-MM1	NW1	Low	Intransitivity, imprecision	Very low ⊕
	R + d ⟶ d	MM-009	NW1	High		
	d ⟶ P + d	MM-003	NW1	Moderate		
IsP + d ⟶ IR + d	IsP + d ⟶ P + d	ICARIA-MM	NW1	Low	Intransitivity	Very low ⊕
	P + d ⟶ d	MM-003	NW1	Moderate		
	d ⟶ R + d	MM-009	NW1	High		
	R + d ⟶ IR + d	China Contin., TOURMALINE-MM1	NW1	Low		
IsP + d ⟶ R + d	IsP + d ⟶ P + d	ICARIA-MM	NW1	Low	Intransitivity	Very low ⊕
	P + d ⟶ d	MM-003	NW1	Moderate		
	d ⟶ R + d	MM-009	NW1	High		
KR + d ⟶ DR + d	KR + d ⟶ R + d	ASPIRE	NW1	High	Imprecision	Very low ⊕
	R + d ⟶ DR + d	POLLUX	NW1	Moderate		
KR + d ⟶ EP + d	KR + d ⟶ R + d	ASPIRE	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d ⟶ d	MM-009	NW1	High		
	d ⟶ P + d	MM-003	NW1	Moderate		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	P + d → EP + d	ELOQUENT-3	NW1	Low		
KR + d → ER + d	KR + d → R + d	ASPIRE	NW1	High	Imprecision	Moderate ⊕⊕⊕
	R + d → ER + d	ELOQUENT-2	NW1	High		
KR + d → IR + d	KR + d → R + d	ASPIRE	NW1	High	Imprecision	Very low ⊕
	R + d → IR + d	China Contin., TOURMALINE-MM1	NW1	Low		
KR + d → IsP + d	KR + d → R + d	ASPIRE	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
	P + d → IsP + d	ICARIA-MM	NW1	Low		
KR + d → P + d	KR + d → R + d	ASPIRE	NW1	High	Intransitivity, imprecision	Very low ⊕
	R + d → d	MM-009	NW1	High		
	d → P + d	MM-003	NW1	Moderate		
R + d → P + d	R + d → d	MM-009	NW1	High	Intransitivity, imprecision	Very low ⊕
	d → P + d	MM-003	NW1	Moderate		
DK + d → FV + d	DK + d → K + d	CANDOR	NW2	High	Imprecision	Moderate ⊕⊕⊕
	K + d → V + d	ENDEAVOR	NW2	High		
	V + d → FV + d	PANORAMA-1	NW2	High		
DK + d → IsK + d	DK + d → K + d	CANDOR	NW2	High	Imprecision	Very low ⊕
	K + d → IsK + d	IKEMA	NW2	Low		
DK + d → PV + d	DK + d → K + d	CANDOR	NW2	High	Imprecision	Moderate ⊕⊕⊕
	K + d → V + d	ENDEAVOR	NW2	High		

Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
DV + d ⟶ DK + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Very low	Imprecision	Very low ⊕
	V + d ⟶ K + d	ENDEAVOR	NW2	High		
	K + d ⟶ DK + d	CANDOR	NW2	High		
DV + d ⟶ FV + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Very low	Imprecision	Very low ⊕
	V + d ⟶ FV + d	PANORAMA-1	NW2	High		
DV + d ⟶ IsK + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Very low	Imprecision	Very low ⊕
	V + d ⟶ K + d	ENDEAVOR	NW2	High		
	K + d ⟶ IsK + d	IKEMA	NW2	Low		
DV + d ⟶ PV + d	DV + d ⟶ V + d	LEPUS, CASTOR	NW2	Very low	Imprecision	Very low ⊕
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
FV + d ⟶ PV + d	FV + d ⟶ V + d	PANORAMA-1	NW2	High	None	High ⊕⊕⊕⊕
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
IsK + d ⟶ FV + d	IsK + d ⟶ K + d	IKEMA	NW2	Low	None	Low ⊕⊕
	K + d ⟶ V + d	ENDEAVOR	NW2	High		
	V + d ⟶ FV + d	PANORAMA-1	NW2	High		
IsK + d ⟶ PV + d	IsK + d ⟶ K + d	IKEMA	NW2	Low	Imprecision	Very low ⊕
	K + d ⟶ V + d	ENDEAVOR	NW2	High		
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
K + d ⟶ DV + d	K + d ⟶ V + d	ENDEAVOR	NW2	High	Imprecision	Very low ⊕
	V + d ⟶ DV + d	LEPUS, CASTOR	NW2	Very low		
K + d ⟶ FV + d	K + d ⟶ V + d	ENDEAVOR	NW2	High	None	High ⊕⊕⊕⊕

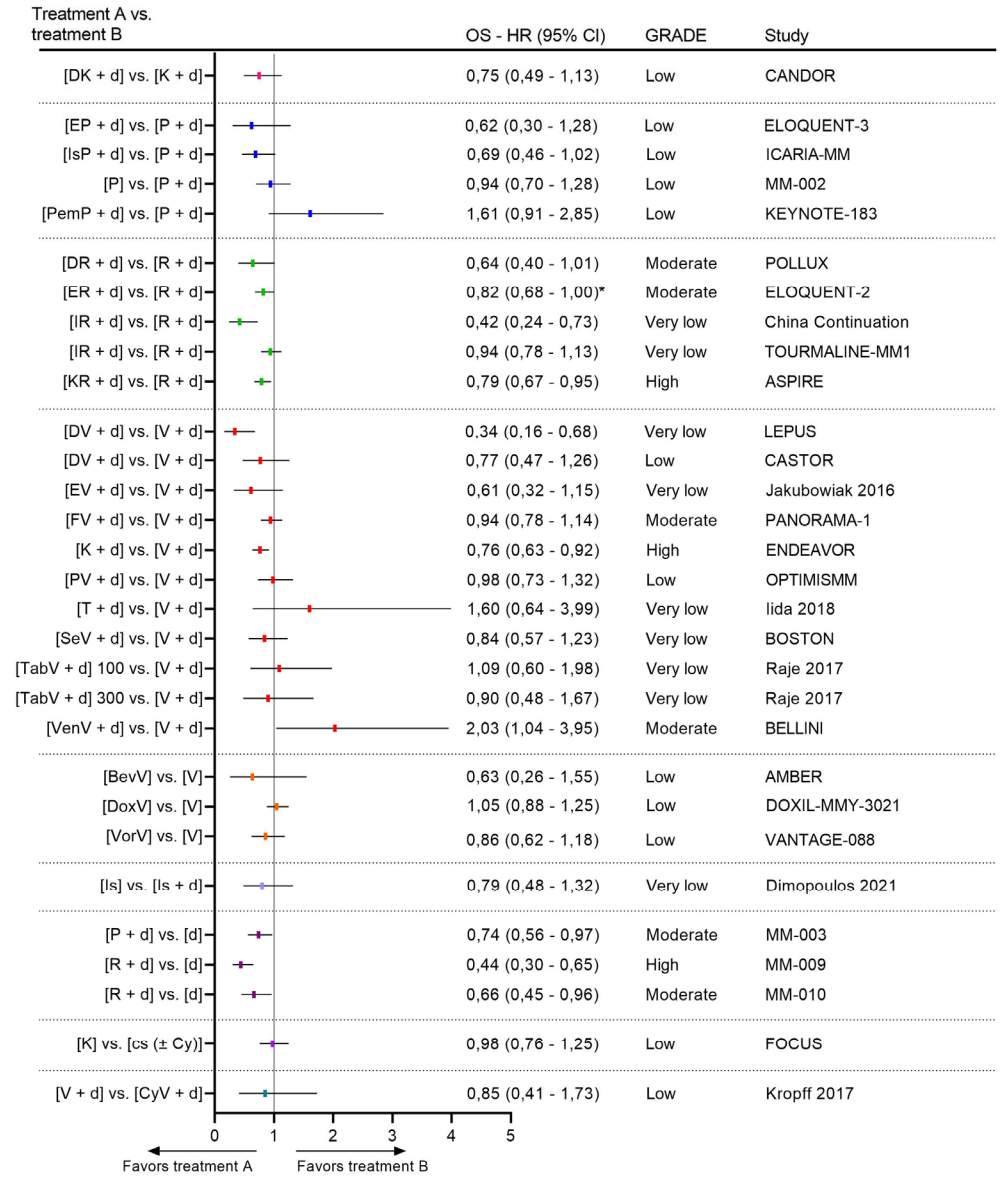
Indirect comparison (treatment A → treatment B)	Included direct treatments (treatment A → treatment B)	Study	Network	Direct GRADE assessment	Reasons for downgrading indirect comparison	Indirect GRADE assessment
	V + d ⟶ FV + d	PANORAMA-1	NW2	High		
K + d ⟶ PV + d	K + d ⟶ V + d	ENDEAVOR	NW2	High	Imprecision	Moderate ⊕⊕⊕
	V + d ⟶ PV + d	OPTIMISMM	NW2	High		
V + d ⟶ DK + d	V + d ⟶ K + d	ENDEAVOR	NW2	High	Imprecision	Moderate ⊕⊕⊕
	K + d ⟶ DK + d	CANDOR	NW2	High		
V + d ⟶ IsK + d	V + d ⟶ K + d	ENDEAVOR	NW2	High	Imprecision	Very low ⊕
	K + d ⟶ IsK + d	IKEMA	NW2	Low		

D: daratumumab, d: dexamethasone, Dox: doxorubicin, E: elotuzumab, F: panobinostat (Farydak), I: ixazomib, Is: isatuximab, K: carfilzomib (Kyprolis), NW: network, P: pomalidomide, R: lenalidomide (Revlimid), V: bortezomib

The threshold of important effect was set at 0,5 and 2 and was used solely for the purpose as an aid for assessing certainty of evidence.

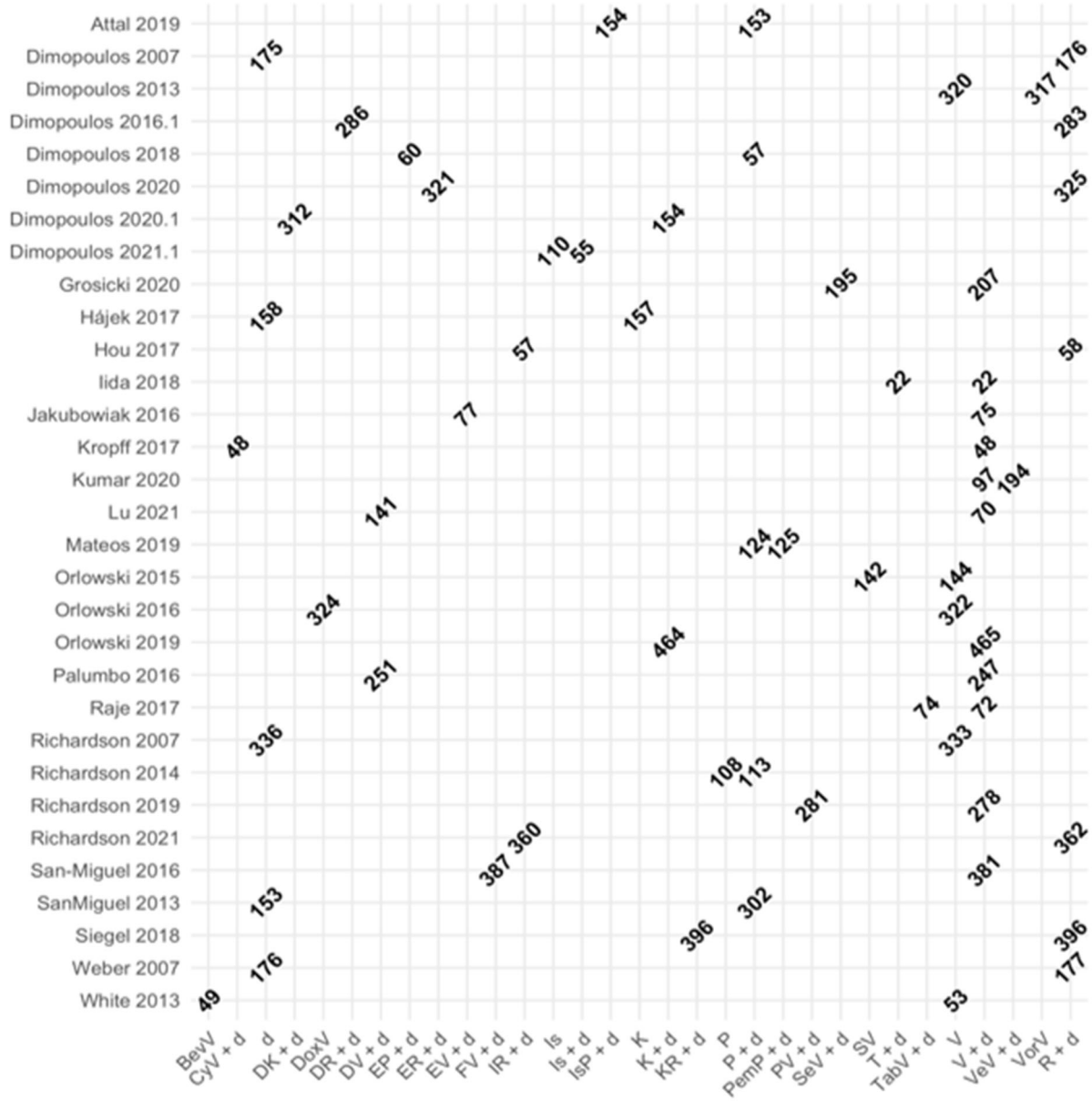
Appendix 10

Additional results: overall survival

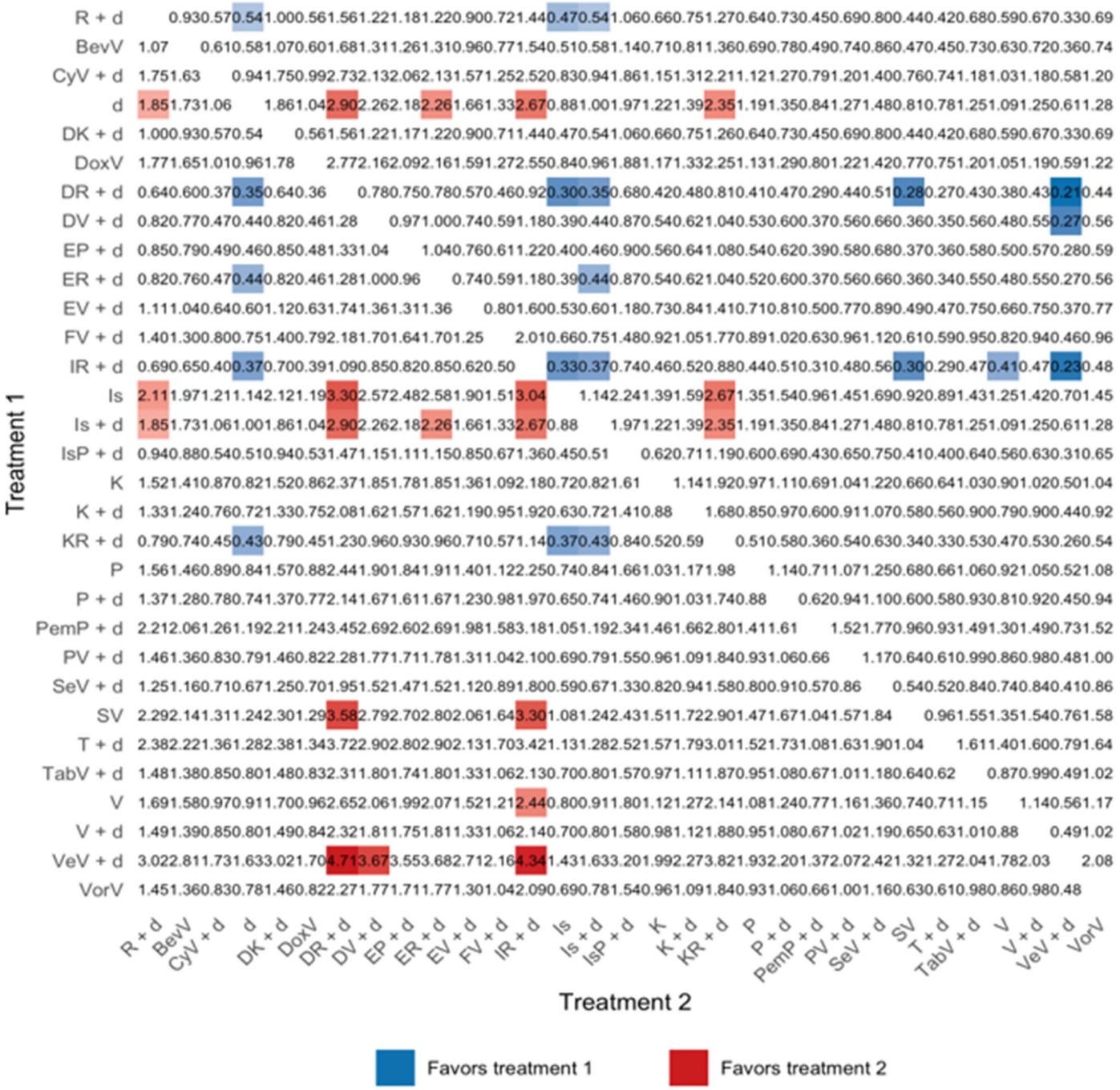


Forest plot of direct evidence - overall survival

** 95,4% confidence interval. CI: confidence interval, HR: hazard ratio, OS: overall survival. HR>1 favours treatment A, HR<1 favours treatment B. Note that almost all studies used dexamethasone (d), with the exception of one study (FOCUS) that used either dexamethasone (d) or methylprednisone. This abbreviated as "cs" in the forest plot.*

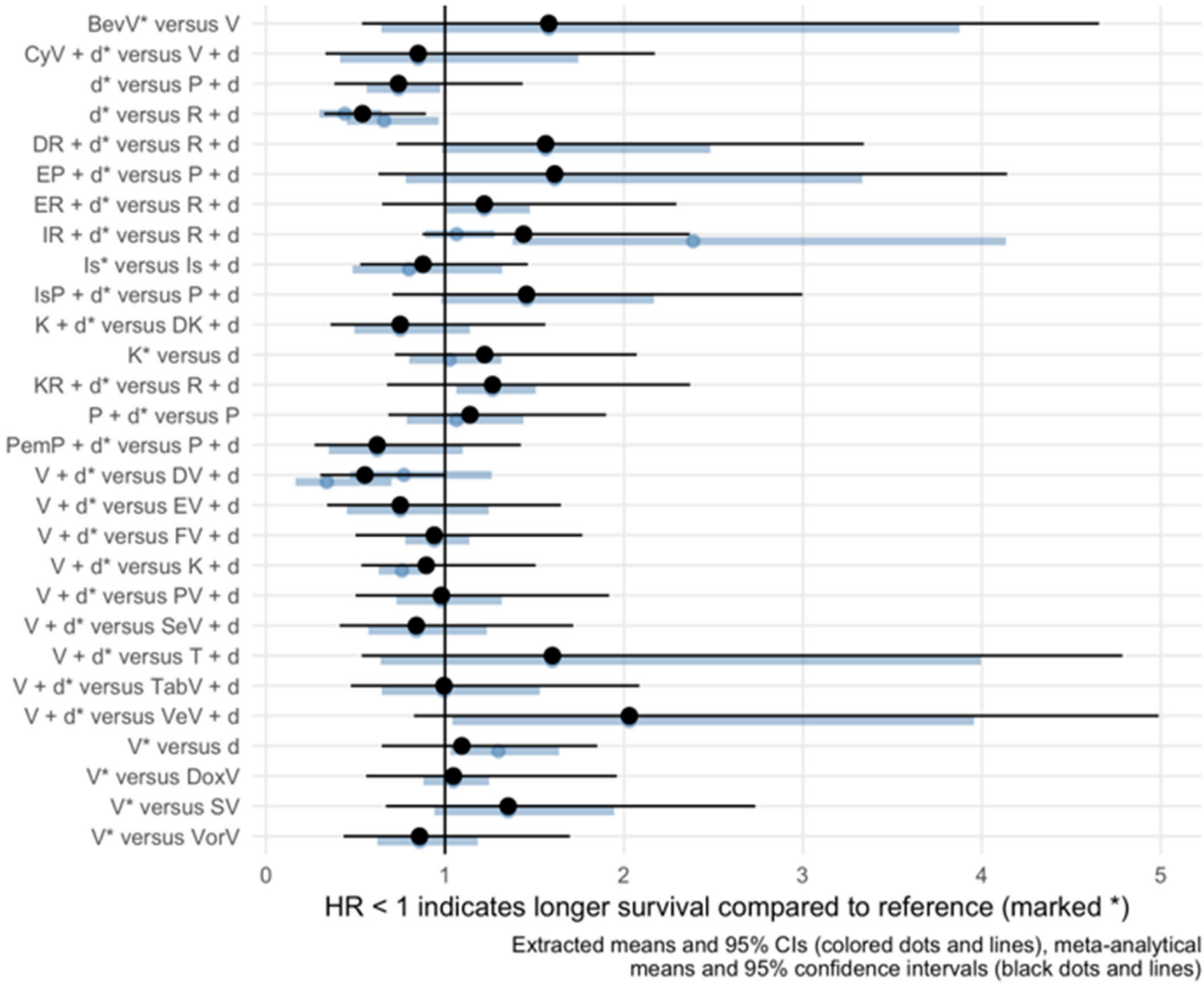


Study designs and sample sizes for overall survival



All meta-analytical point estimates of HR for overall survival

The matrix shows all possible comparisons and is symmetrical about its diagonal. Confidence intervals are not shown to improve readability. Relative treatment effect estimates whose 95% confidence intervals exclude the possibility of no difference in effect are shaded. Darker shading is used to indicate larger relative treatment effects.



Assessment of model fit for overall survival

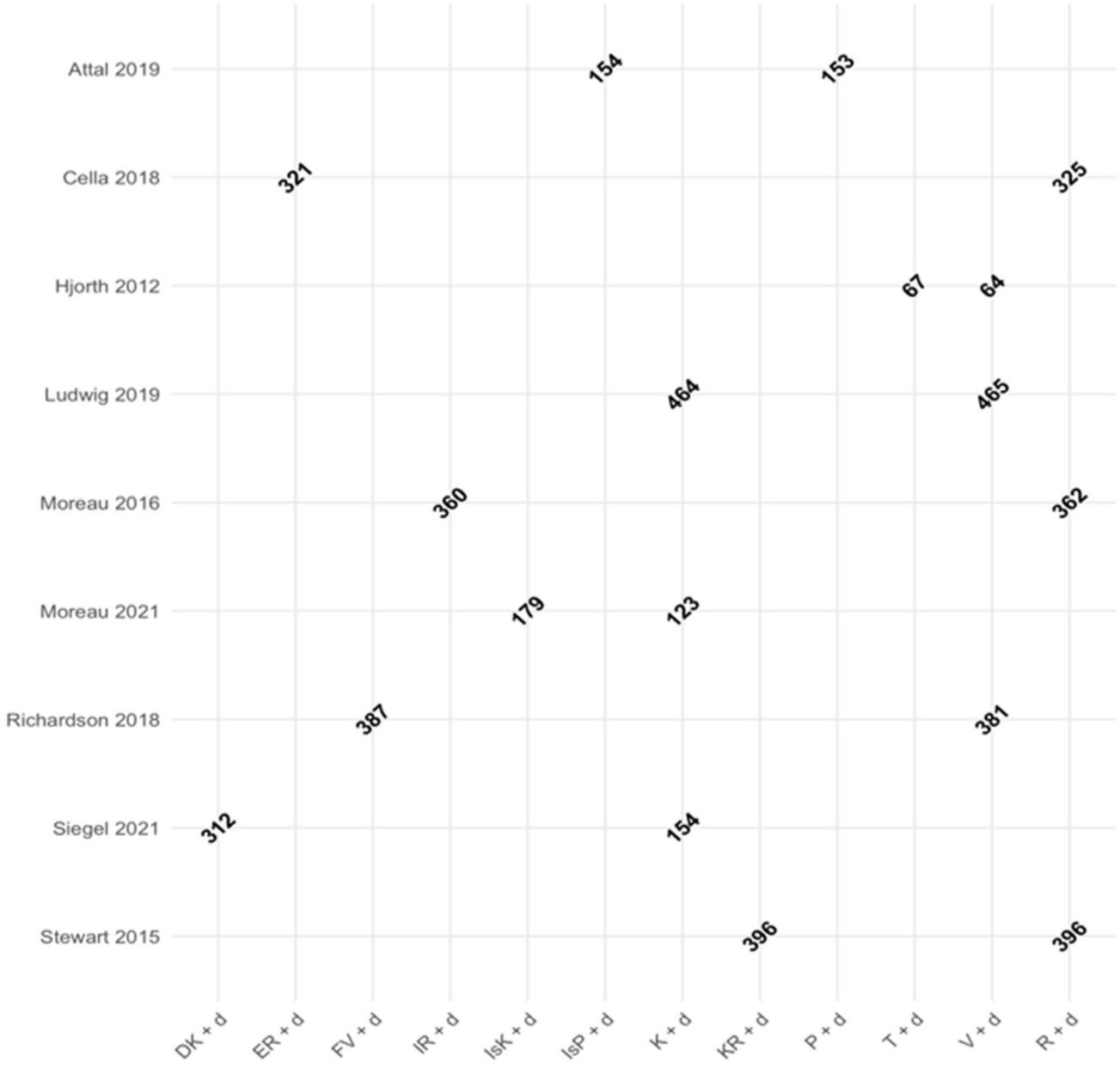
Extracted means and 95% confidence intervals (blue dots and lines) and component network meta-analytical means and 95% confidence intervals (black dots and lines). Confidence intervals from meta-analysis account for heterogeneity via random effects and may therefore be wider than those from the included studies.

Summary of findings for overall survival

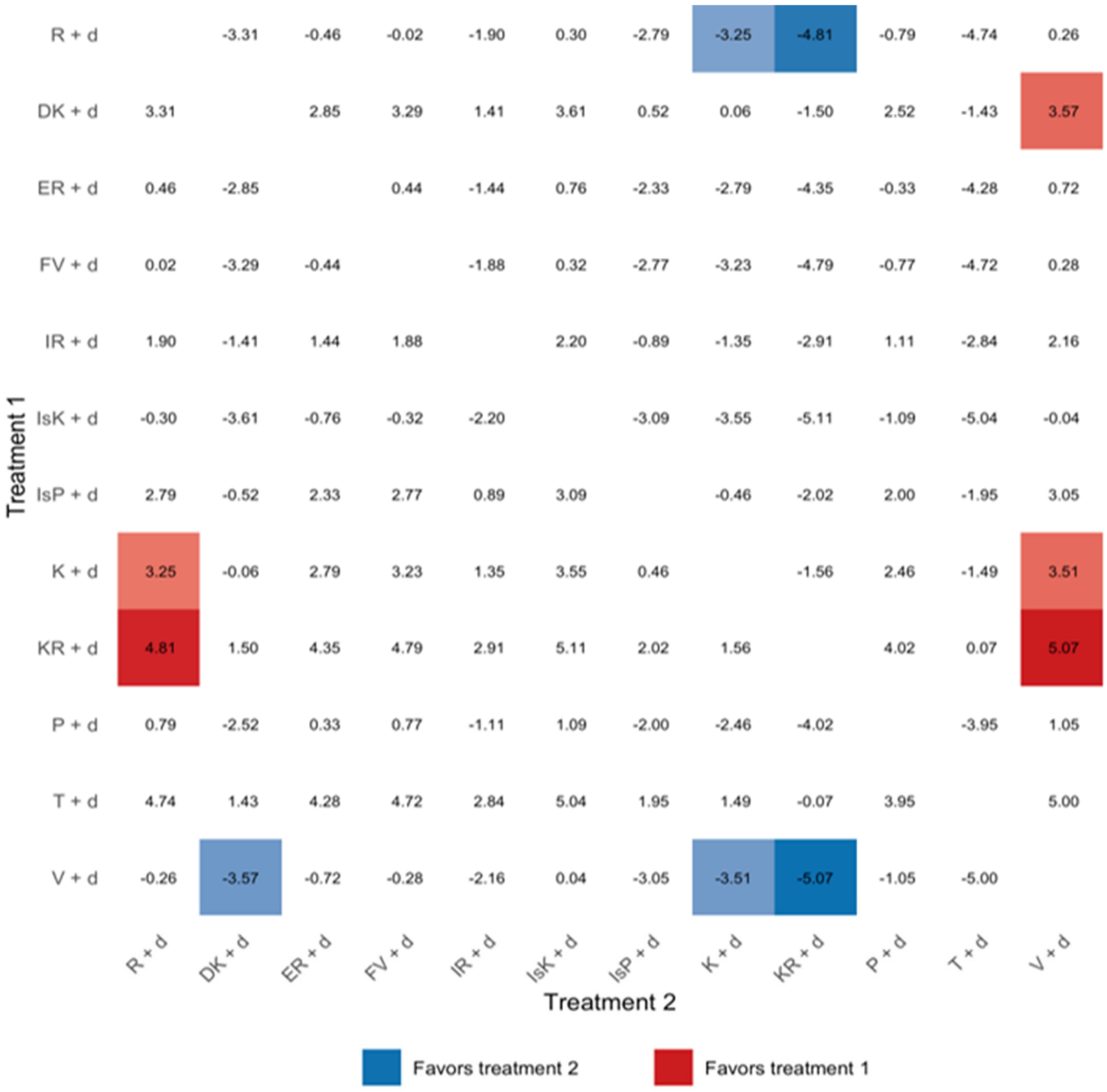
Treatment	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant for people who are:		
		Refractory to R	Refractory to V	Refractory to R + V
DR + d	1.00 (0.86)	NA	1.00 (0.84)	NA
IR + d	2.00 (0.85)	NA	2.00 (0.83)	NA
DV + d	3.00 (0.79)	1.00 (0.78)	NA	NA
KR + d	4.00 (0.78)	NA	3.00 (0.75)	NA
ER + d	5.00 (0.77)	NA	4.00 (0.73)	NA
EP + d	6.00 (0.73)	2.00 (0.75)	5.00 (0.67)	1.00 (0.77)
IsP + d	7.00 (0.70)	3.00 (0.72)	6.00 (0.63)	2.00 (0.74)
R + d	8.00 (0.68)	NA	7.00 (0.63)	NA
DK + d	9.00 (0.67)	4.00 (0.70)	8.00 (0.60)	3.00 (0.71)
BevV	10.00 (0.62)	5.00 (0.65)	NA	NA
EV + d	11.00 (0.62)	6.00 (0.64)	NA	NA
SeV + d	12.00 (0.56)	7.00 (0.59)	NA	NA
K + d	13.00 (0.52)	8.00 (0.57)	9.00 (0.45)	4.00 (0.58)
P + d	14.00 (0.50)	9.00 (0.56)	10.00 (0.44)	5.00 (0.57)
FV + d	15.00 (0.49)	10.00 (0.54)	NA	NA
VorV	16.00 (0.47)	11.00 (0.52)	NA	NA
PV + d	17.00 (0.46)	12.00 (0.52)	NA	NA
TabV + d	18.00 (0.45)	13.00 (0.51)	NA	NA
V + d	19.00 (0.44)	14.00 (0.51)	NA	NA
K	20.00 (0.43)	15.00 (0.50)	11.00 (0.38)	6.00 (0.51)
P	21.00 (0.42)	16.00 (0.48)	12.00 (0.37)	7.00 (0.49)
CyV + d	22.00 (0.37)	18.00 (0.43)	NA	NA
V	23.00 (0.35)	17.00 (0.43)	NA	NA
DoxV	24.00 (0.34)	19.00 (0.41)	NA	NA
d	25.50 (0.28)	20.50 (0.37)	13.50 (0.26)	8.50 (0.37)
Is + d	25.50 (0.28)	20.50 (0.37)	13.50 (0.26)	8.50 (0.37)
PemP + d	27.00 (0.25)	22.00 (0.32)	15.00 (0.23)	10.00 (0.31)
Is	28.00 (0.24)	23.00 (0.31)	17.00 (0.21)	12.00 (0.29)
T + d	29.00 (0.24)	24.00 (0.31)	16.00 (0.23)	11.00 (0.29)
SV	30.00 (0.21)	25.00 (0.29)	NA	NA
VeV + d	31.00 (0.13)	26.00 (0.20)	NA	NA

§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Ranking based on P-score for treatments relevant for non-refractory patients, patients who are refractory to lenalidomide (R), patients who are refractory to bortezomib (V), and patients who are refractory to both lenalidomide (R) and bortezomib (V). R: lenalidomide, V: bortezomib, NA: not applicable

Additional results: quality of life

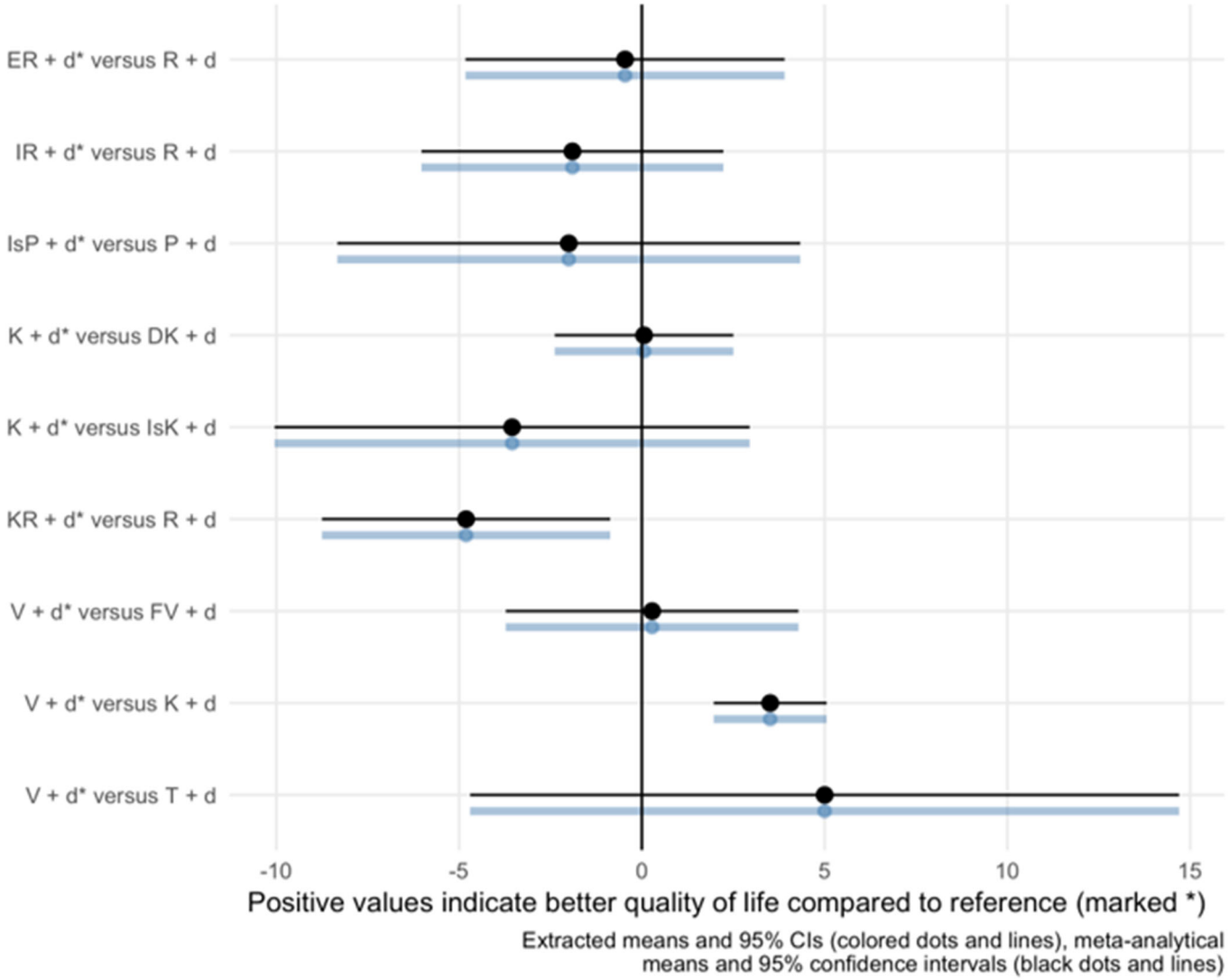


Study designs and sample sizes for overall survival



All meta-analytical point estimates of HR for quality of life

The matrix shows all possible comparisons and is symmetrical about its diagonal. Confidence intervals are not shown to improve readability. Relative treatment effect estimates whose 95% confidence intervals exclude the possibility of no difference in effect are shaded. Darker shading is used to indicate larger relative treatment effects.



Assessment of model fit for quality of life

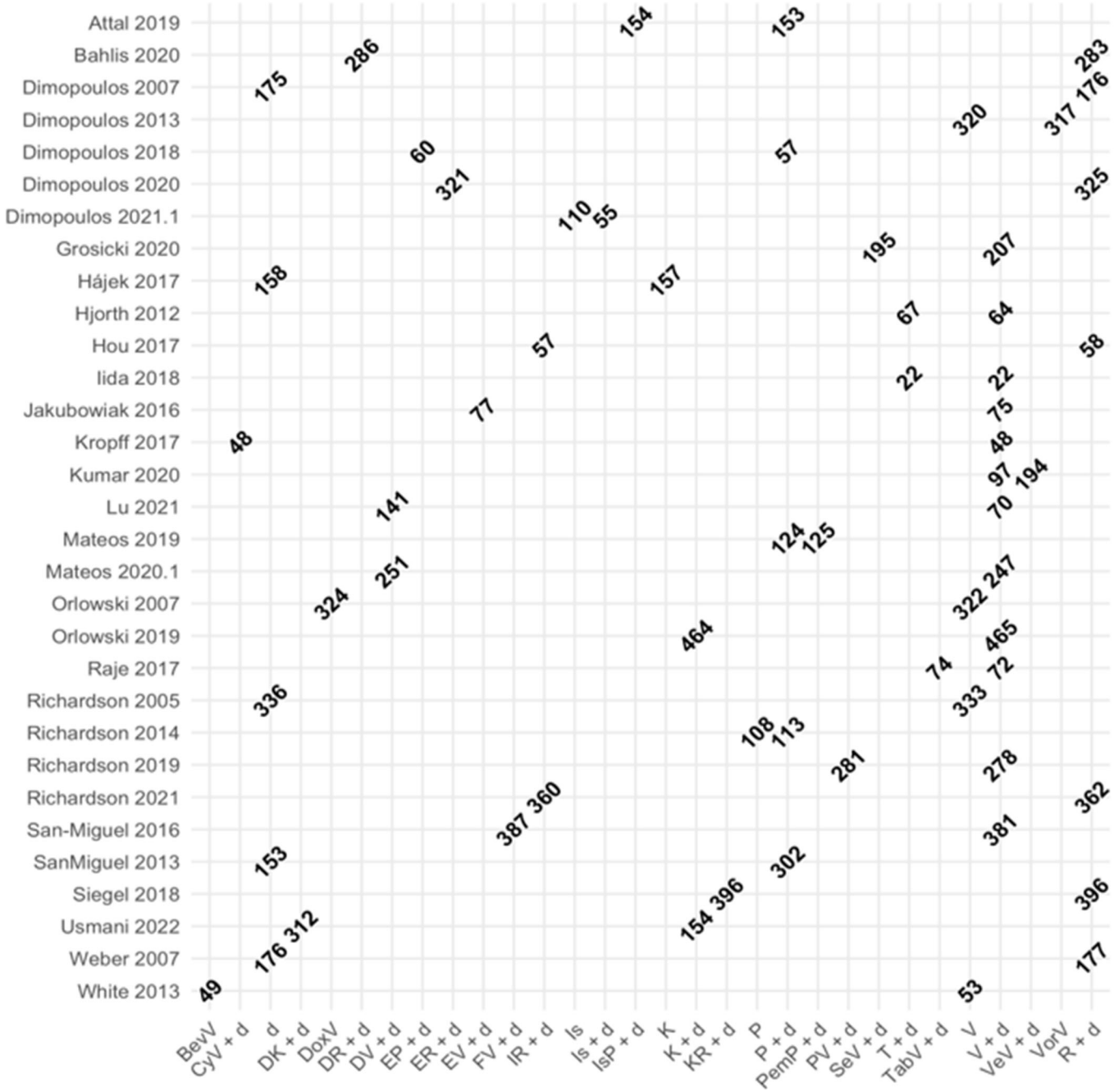
Extracted means and 95% confidence intervals (blue dots and lines) and component network meta-analytical means and 95% confidence intervals (black dots and lines). Confidence intervals from meta-analysis account for heterogeneity via random effects and may therefore be wider than those from the included studies.

Summary of findings for quality of life

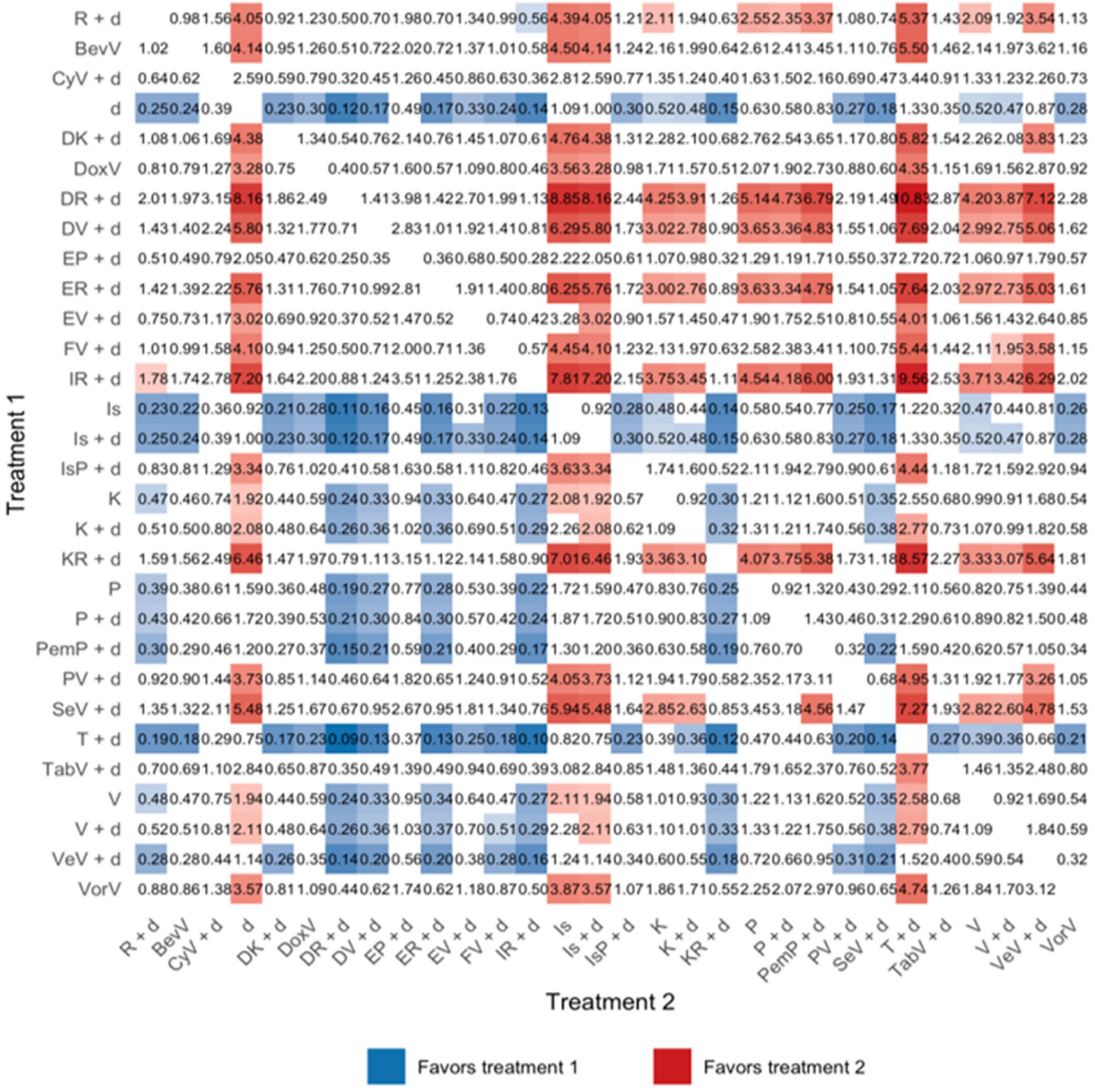
Treatment	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant for people who are:		
		Refractory to R	Refractory to V	Refractory to R + V
KR + d	1 (0.86)	NA	1 (0.81)	NA
T + d	2 (0.73)	1 (0.75)	2 (0.71)	1 (0.71)
K + d	3 (0.73)	2 (0.71)	3 (0.66)	3 (0.63)
DK + d	4 (0.72)	3 (0.71)	4 (0.66)	2 (0.63)
IsP + d	5 (0.65)	4 0.64 ()	5 (0.59)	4 (0.55)
IR + d	6 (0.53)	NA	6 (0.48)	NA
P + d	7 (0.38)	5 (0.38)	7 (0.33)	5 (0.27)
ER + d	8 (0.34)	NA	8 (0.30)	NA
IsK + d	9 (0.29)	7 (0.29)	9 (0.25)	6 (0.21)
FV + d	10 (0.29)	6 (0.29)	NA	NA
R + d	11 (0.25)	NA	10 (0.19)	NA
V + d	12 (0.22)	8 (0.23)	NA	NA

§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Ranking based on P-score for treatments relevant for non-refractory patients, patients who are refractory to lenalidomide (R), patients who are refractory to bortezomib (V), and patients who are refractory to both lenalidomide (R) and bortezomib (V). R: lenalidomide, V: bortezomib, NA: not applicable

Additional results: severe adverse events

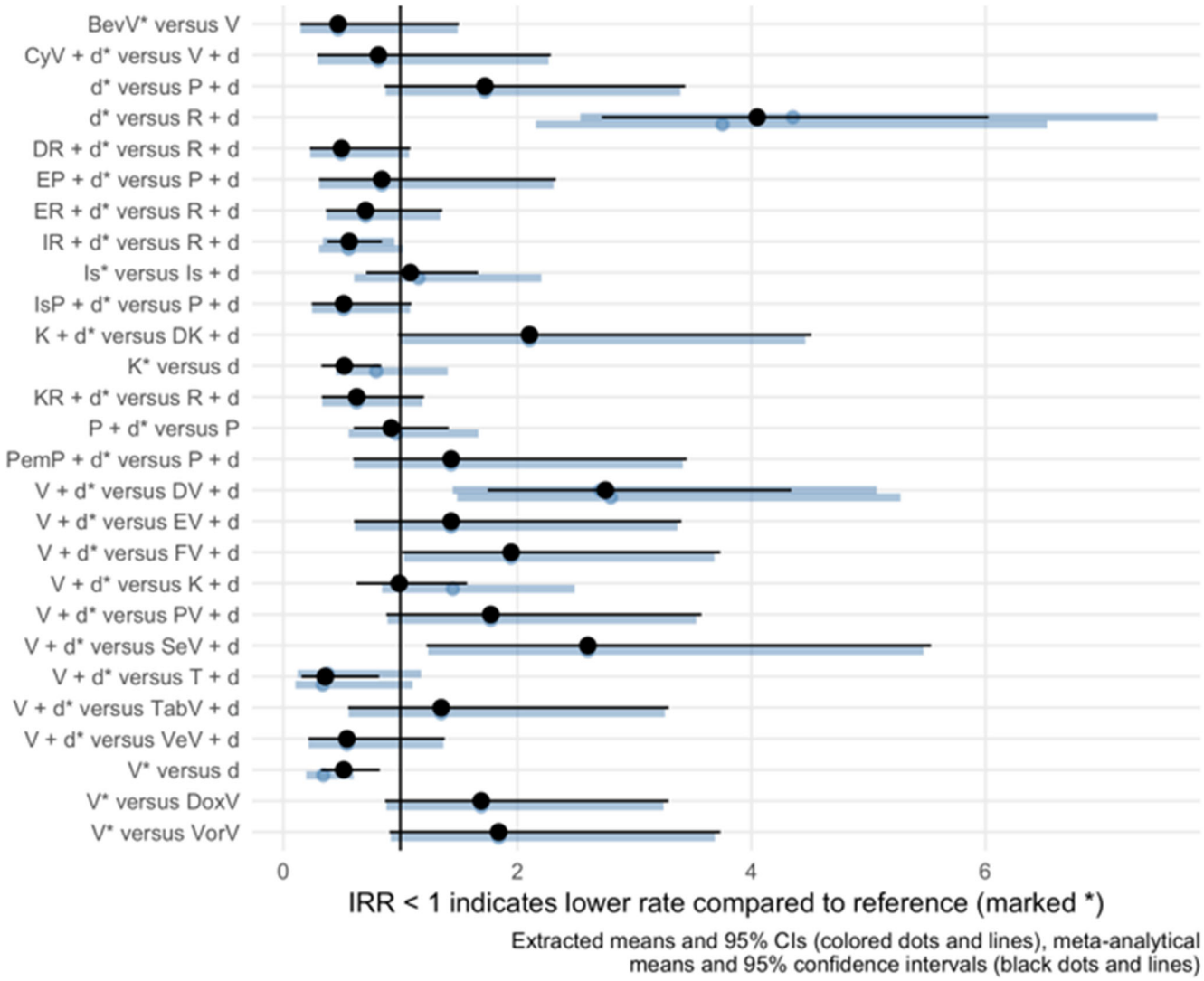


Study designs and sample sizes for overall survival



All meta-analytical point estimates of HR for severe adverse events

The matrix shows all possible comparisons and is symmetrical about its diagonal. Confidence intervals are not shown to improve readability. Relative treatment effect estimates whose 95% confidence intervals exclude the possibility of no difference in effect are shaded. Darker shading is used to indicate larger relative treatment effects.



Assessment of model fit for severe adverse events

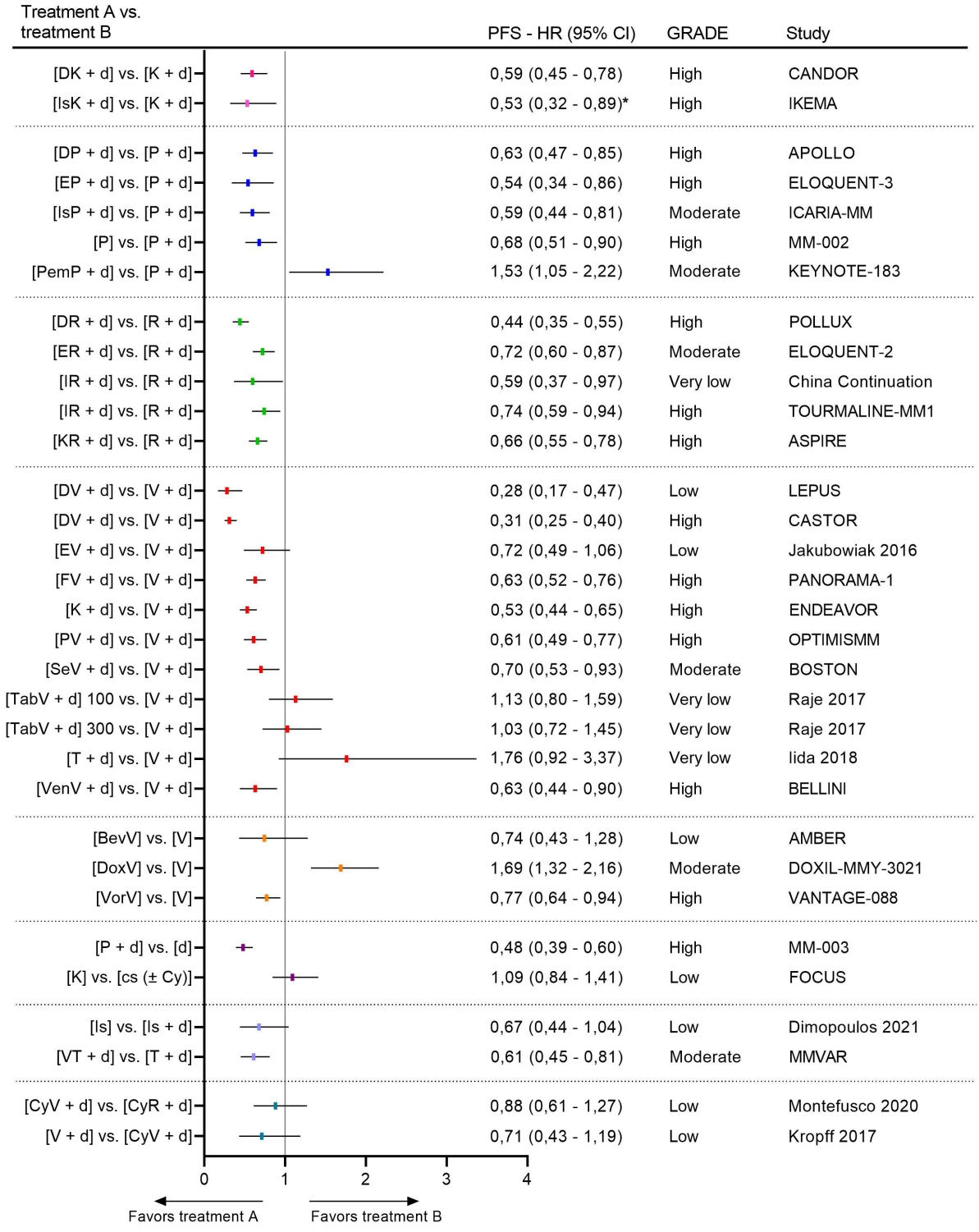
Extracted means and 95% confidence intervals (blue dots and lines) and component network meta-analytical means and 95% confidence intervals (black dots and lines). Confidence intervals from meta-analysis account for heterogeneity via random effects and may therefore be wider than those from the included studies.

Summary of findings for severe adverse events

Treatment	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant for people who are:		
		Refractory to R	Refractory to V	Refractory to R + V
T + d	1.00 (0.93)	1.00 (0.89)	1.00 (0.88)	1.00 (0.83)
Is	2.00 (0.91)	2.00 (0.88)	2.00 (0.86)	2.00 (0.79)
d	3.50 (0.91)	3.50 (0.86)	3.50 (0.83)	3.50 (0.75)
Is + d	3.50 (0.91)	3.50 (0.86)	3.50 (0.83)	3.50 (0.75)
VeV + d	5.00 (0.83)	5.00 (0.76)	NA	NA
PemP + d	6.00 (0.80)	6.00 (0.75)	5.00 (0.74)	5.00 (0.62)
P	7.00 (0.71)	7.00 (0.65)	6.00 (0.64)	6.00 (0.48)
P + d	8.00 (0.69)	8.00 (0.62)	7.00 (0.61)	7.00 (0.43)
V	9.00 (0.66)	10.00 (0.57)	NA	NA
K	10.00 (0.66)	9.00 (0.57)	8.00 (0.56)	9.00 (0.37)
V + d	11.00 (0.62)	13.00 (0.53)	NA	NA
K + d	12.00 (0.62)	12.00 (0.53)	10.00 (0.54)	10.00 (0.33)
EP + d	13.00 (0.60)	11.00 (0.54)	9.00 (0.55)	8.00 (0.37)
CyV + d	14.00 (0.52)	14.00 (0.44)	NA	NA
TabV + d	15.00 (0.48)	15.00 (0.40)	NA	NA
EV + d	16.00 (0.45)	16.00 (0.37)	NA	NA
DoxV	17.00 (0.41)	18.00 (0.33)	NA	NA
IsP + d	18.00 (0.40)	17.00 (0.33)	11.00 (0.37)	11.00 (0.17)
VorV	19.00 (0.37)	19.00 (0.30)	NA	NA
PV + d	20.00 (0.36)	20.00 (0.29)	NA	NA
BevV	21.00 (0.33)	21.00 (0.27)	NA	NA
R + d	22.00 (0.33)	NA	12.00 (0.31)	NA
FV + d	23.00 (0.32)	22.00 (0.25)	NA	NA
DK + d	24.00 (0.29)	23.00 (0.23)	13.00 (0.28)	12.00 (0.09)
SeV + d	25.00 (0.20)	24.00 (0.16)	NA	NA
ER + d	26.00 (0.19)	NA	14.00 (0.18)	NA
DV + d	27.00 (0.17)	25.00 (0.13)	NA	NA
KR + d	28.00 (0.15)	NA	15.00 (0.14)	NA
IR + d	29.00 (0.11)	NA	16.00 (0.10)	NA
DR + d	30.00 (0.09)	NA	17.00 (0.08)	NA

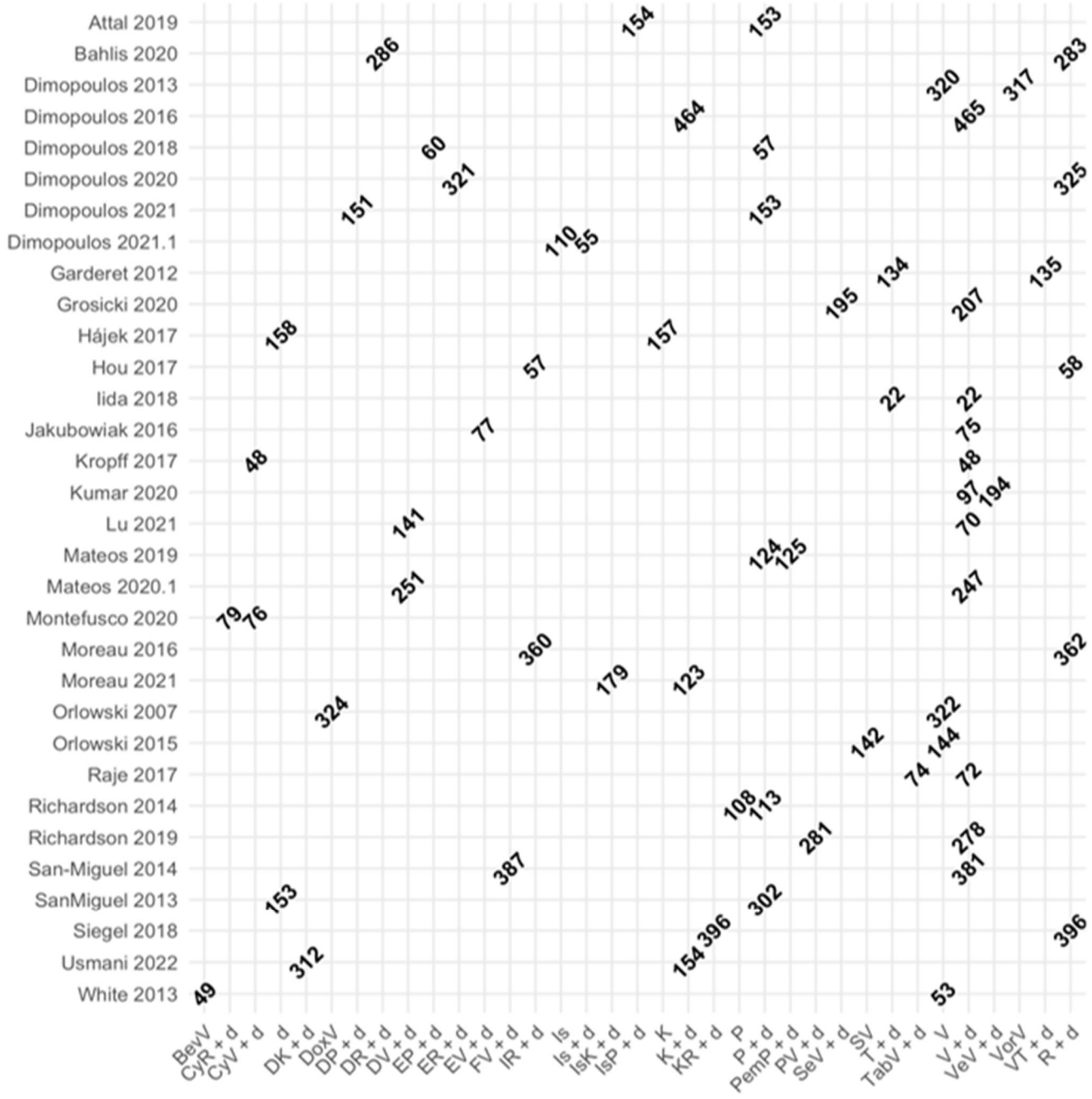
§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Ranking based on P-score for treatments relevant for non-refractory patients, patients who are refractory to lenalidomide (R), patients who are refractory to bortezomib (V), and patients who are refractory to both lenalidomide (R) and bortezomib (V).
R: lenalidomide, V: bortezomib, NA: not applicable

Additional results: progression-free survival

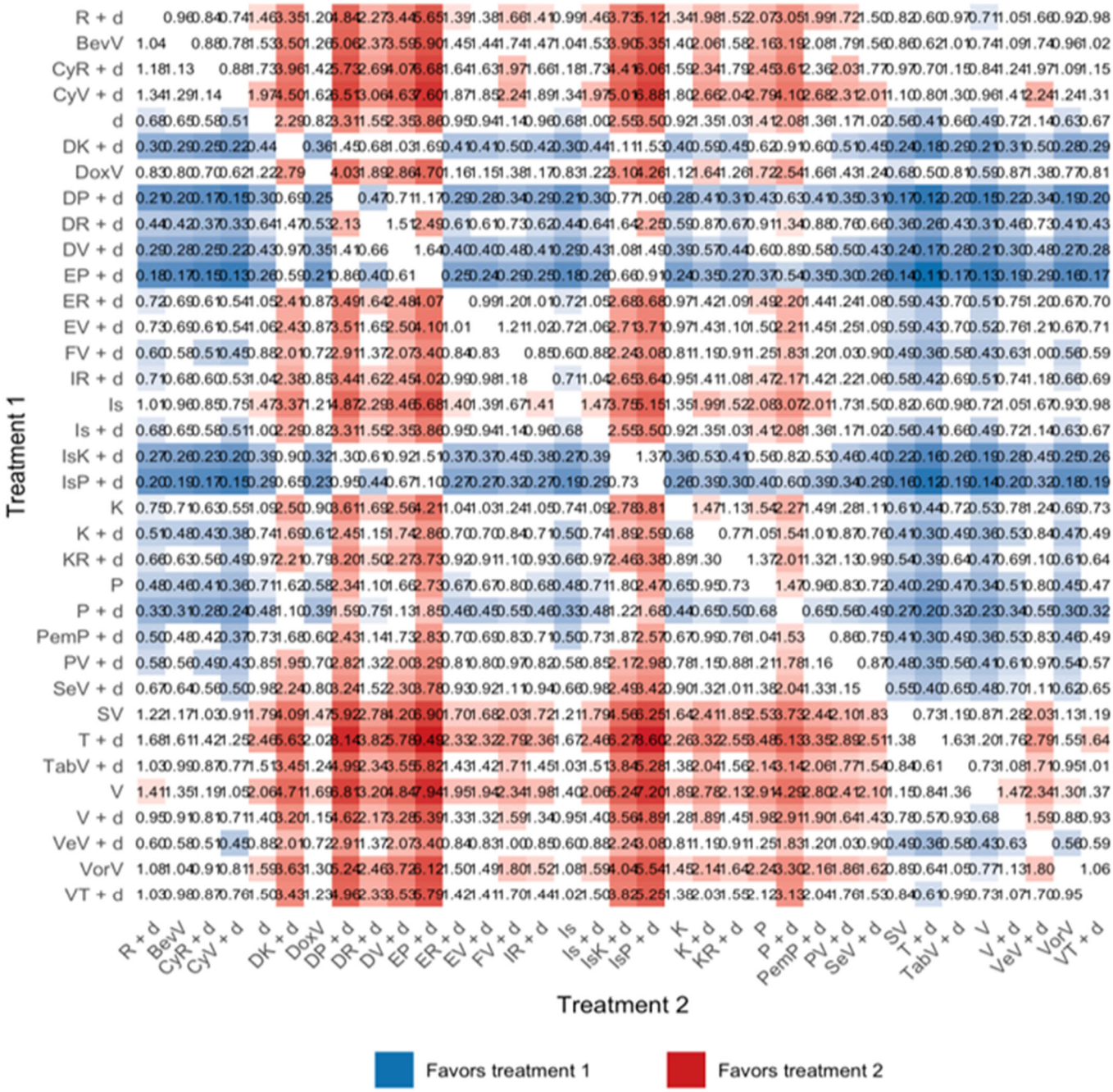


Forest plot of direct evidence – progression-free survival

** 99% confidence interval. CI: confidence interval, HR: hazard ratio, OS: overall survival. HR<1 favours treatment A, HR>1 favours treatment B. Note that almost all studies used dexamethasone (d), with the exception of one study (FOCUS) that used either dexamethasone (d) or methylprednisone. This abbreviated as "cs" in the forest plot.*

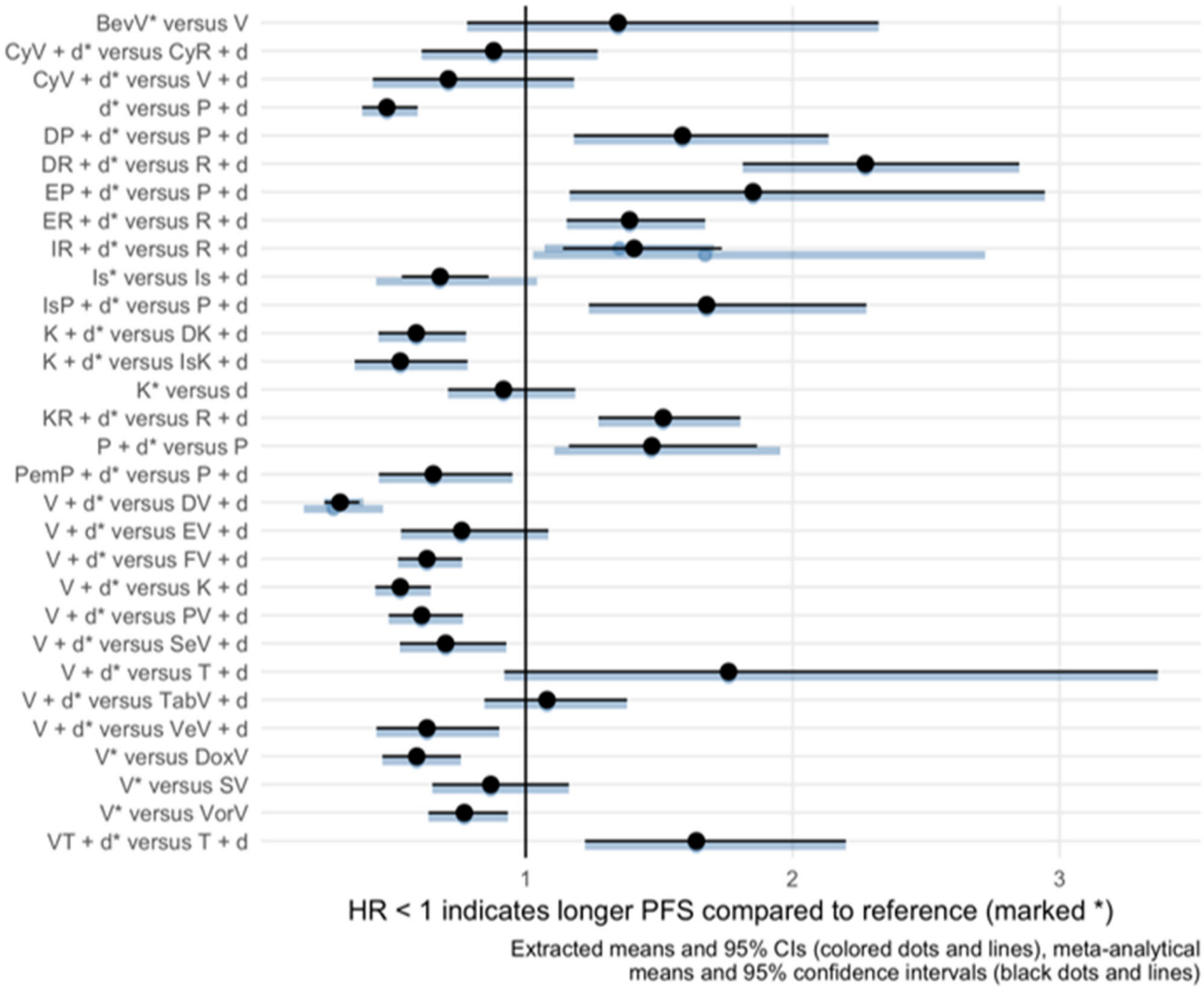


Study designs and sample sizes for progression-free survival



All meta-analytical point estimates of HR for progression-free survival

The matrix shows all possible comparisons and is symmetrical about its diagonal. Confidence intervals are not shown to improve readability. Relative treatment effect estimates whose 95% confidence intervals exclude the possibility of no difference in effect are shaded. Darker shading is used to indicate larger relative treatment effects.



Assessment of model fit for progression-free survival

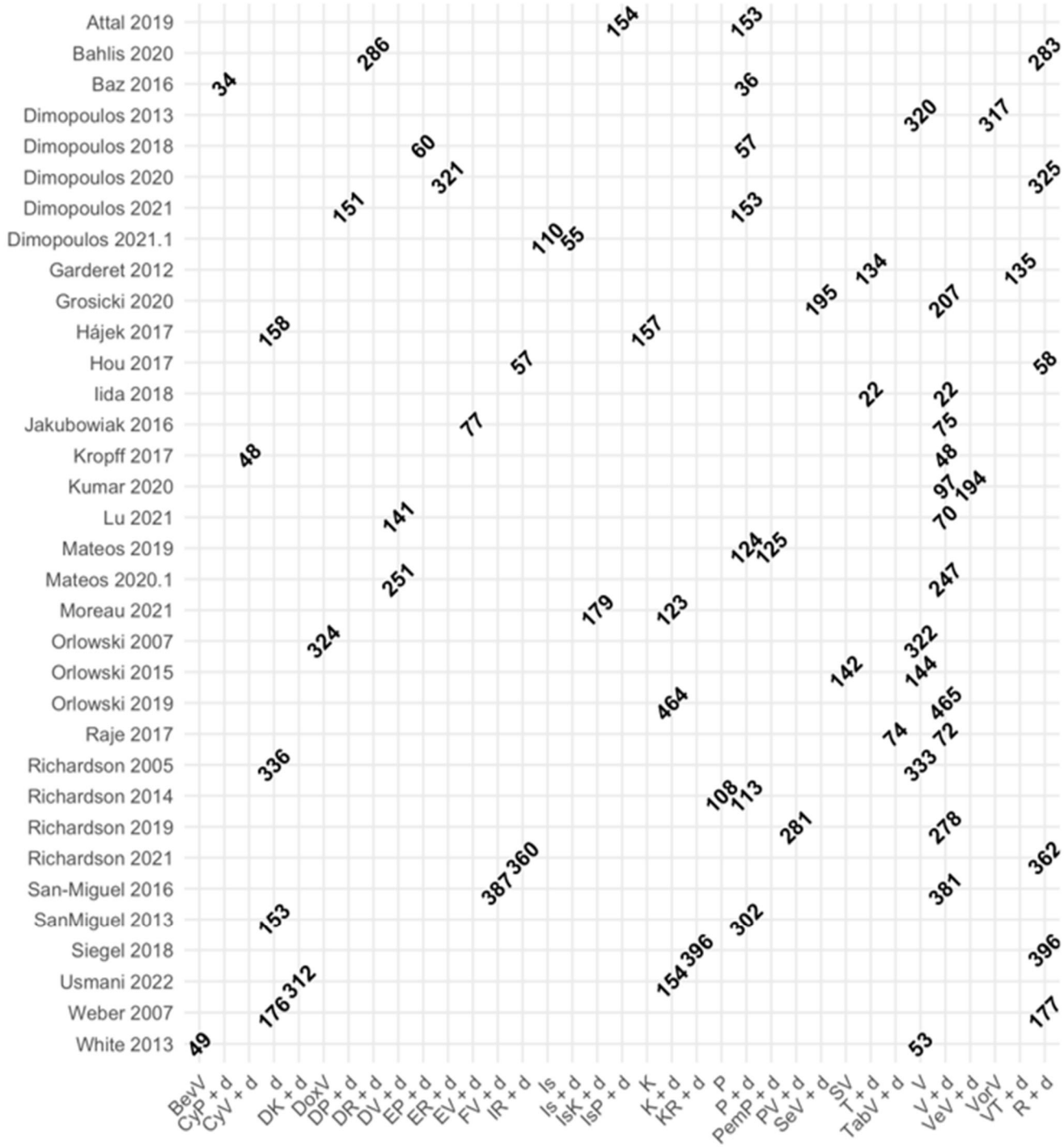
Extracted means and 95% confidence intervals (blue dots and lines) and component network meta-analytical means and 95% confidence intervals (black dots and lines). Confidence intervals from meta-analysis account for heterogeneity via random effects and may therefore be wider than those from the included studies.

Summary of findings for progression-free survival

Treatment	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant for people who are:		
		Refractory to R	Refractory to V	Refractory to R + V
EP + d	1.00 (0.97)	1.00 (0.96)	1.00 (0.95)	1.00 (0.93)
IsP + d	2.00 (0.96)	2.00 (0.95)	2.00 (0.93)	2.00 (0.90)
DP + d	3.00 (0.95)	3.00 (0.94)	3.00 (0.92)	3.00 (0.88)
IsK + d	4.00 (0.90)	4.00 (0.87)	4.00 (0.83)	4.00 (0.75)
DV + d	5.00 (0.87)	5.00 (0.85)	NA	NA
DK + d	6.00 (0.87)	6.00 (0.84)	5.00 (0.79)	5.00 (0.70)
P + d	7.00 (0.85)	7.00 (0.81)	6.00 (0.76)	6.00 (0.66)
DR + d	8.00 (0.76)	NA	7.00 (0.64)	NA
K + d	9.00 (0.71)	9.00 (0.66)	9.00 (0.57)	9.00 (0.45)
P	10.00 (0.71)	8.00 (0.68)	8.00 (0.60)	7.00 (0.47)
PemP + d	11.00 (0.69)	10.00 (0.66)	10.00 (0.57)	8.00 (0.45)
PV + d	12.00 (0.62)	11.00 (0.58)	NA	NA
FV + d	13.00 (0.60)	12.00 (0.56)	NA	NA
VeV + d	14.00 (0.59)	13.00 (0.56)	NA	NA
KR + d	15.00 (0.54)	NA	11.00 (0.40)	NA
SeV + d	16.00 (0.52)	14.00 (0.50)	NA	NA
d	17.50 (0.51)	15.50 (0.49)	12.50 (0.37)	10.50 (0.26)
Is + d	17.50 (0.51)	15.50 (0.49)	12.50 (0.37)	10.50 (0.26)
IR + d	19.00 (0.47)	NA	14.00 (0.33)	NA
EV + d	20.00 (0.46)	17.00 (0.44)	NA	NA
ER + d	21.00 (0.46)	NA	15.00 (0.31)	NA
K	22.00 (0.44)	18.00 (0.42)	16.00 (0.29)	12.00 (0.20)
DoxV	23.00 (0.36)	19.00 (0.36)	NA	NA
V + d	24.00 (0.27)	20.00 (0.27)	NA	NA
VT + d	25.00 (0.26)	21.00 (0.26)	NA	NA
BevV	26.00 (0.23)	22.00 (0.24)	NA	NA
Is	27.00 (0.23)	23.00 (0.24)	17.00 (0.11)	13.00 (0.07)
R + d	28.00 (0.23)	NA	18.00 (0.11)	NA
TabV + d	29.00 (0.22)	24.00 (0.23)	NA	NA
VorV	30.00 (0.19)	25.00 (0.20)	NA	NA
CyR + d	31.00 (0.18)	NA	19.00 (0.11)	NA
SV	32.00 (0.13)	26.00 (0.15)	NA	NA
CyV + d	33.00 (0.11)	27.00 (0.13)	NA	NA
V	34.00 (0.06)	28.00 (0.08)	NA	NA
T + d	35.00 (0.05)	29.00 (0.07)	20.00 (0.03)	14.00 (0.01)

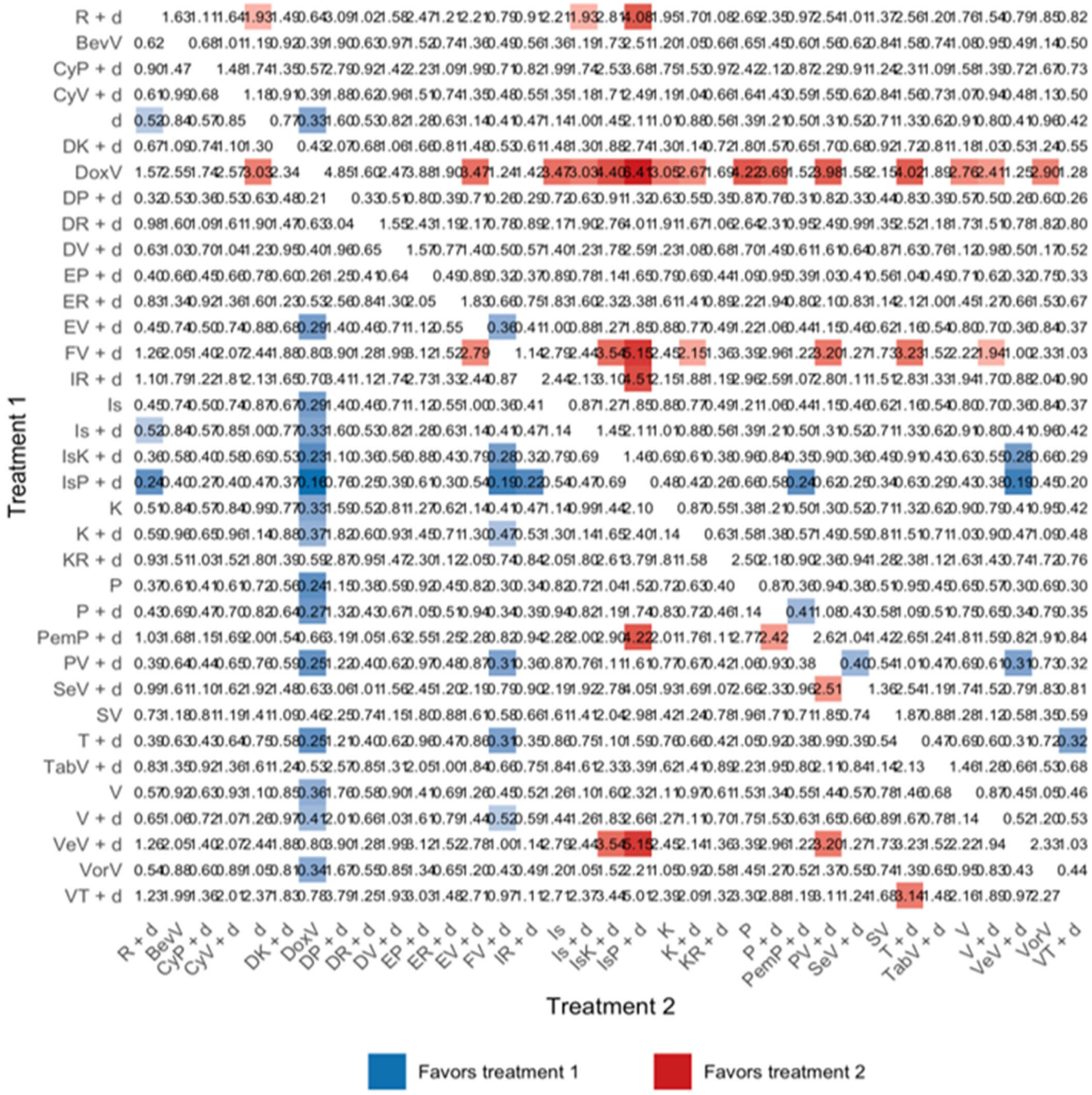
§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Ranking based on P-score for treatments relevant for non-refractory patients, patients who are refractory to lenalidomide (R), patients who are refractory to bortezomib (V), and patients who are refractory to both lenalidomide (R) and bortezomib (V).
R: lenalidomide, V: bortezomib, NA: not applicable

Additional results: discontinuation due to adverse events



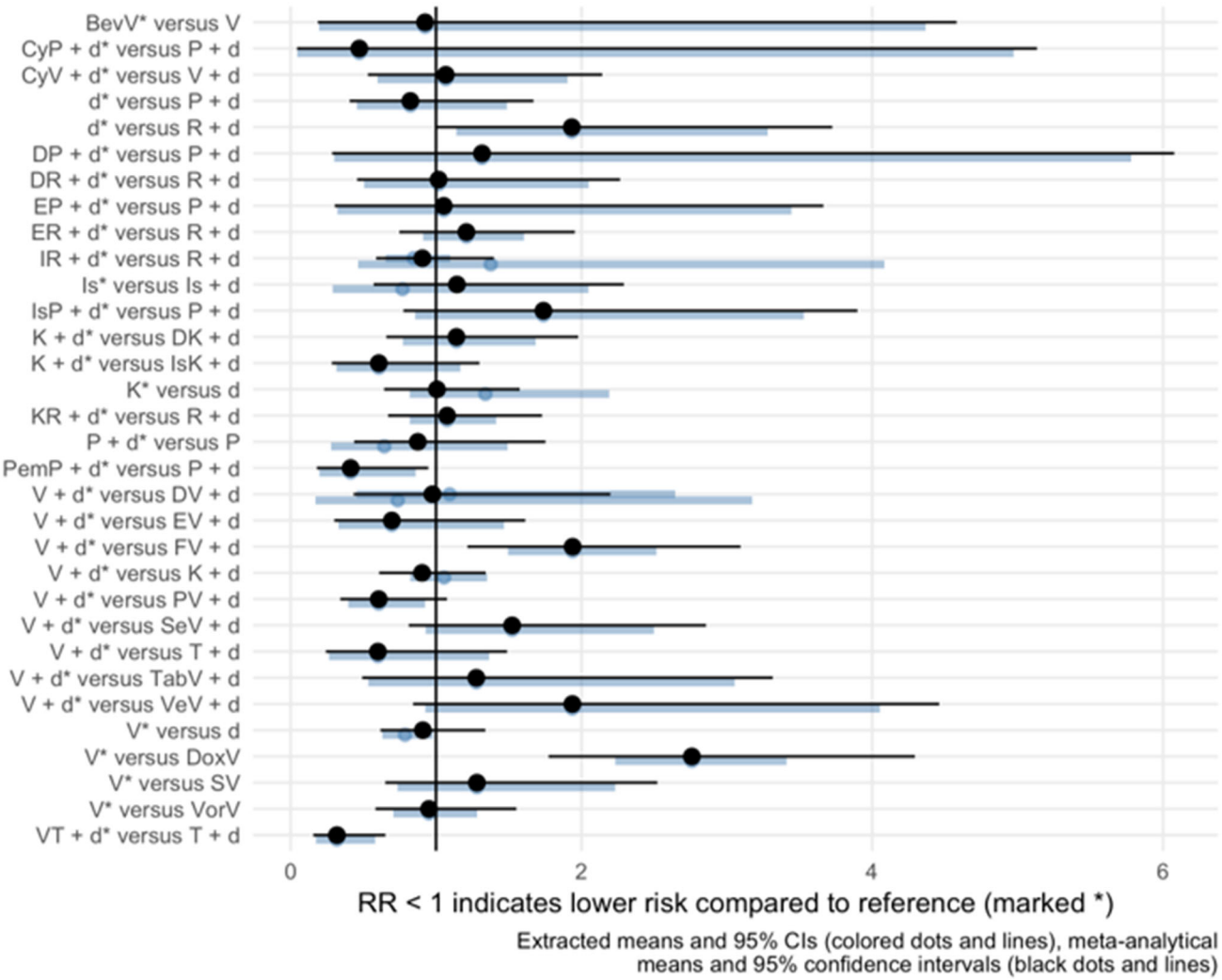
Summary of findings for discontinuation due to adverse events

Ranking based on P-score for treatments relevant for non-refractory patients, patients who are refractory to lenalidomide (R), patients who are refractory to bortezomib (V), and patients who are refractory to both lenalidomide (R) and bortezomib (V).



All meta-analytical point estimates of HR for discontinuation due to adverse events

The matrix shows all possible comparisons and is symmetrical about its diagonal. Confidence intervals are not shown to improve readability. Relative treatment effect estimates whose 95% confidence intervals exclude the possibility of no difference in effect are shaded. Darker shading is used to indicate larger relative treatment effects.



Assessment of model fit for discontinuation due to adverse events

Extracted means and 95% confidence intervals (blue dots and lines) and component network meta-analytical means and 95% confidence intervals (black dots and lines). Confidence intervals from meta-analysis account for heterogeneity via random effects and may therefore be wider than those from the included studies.

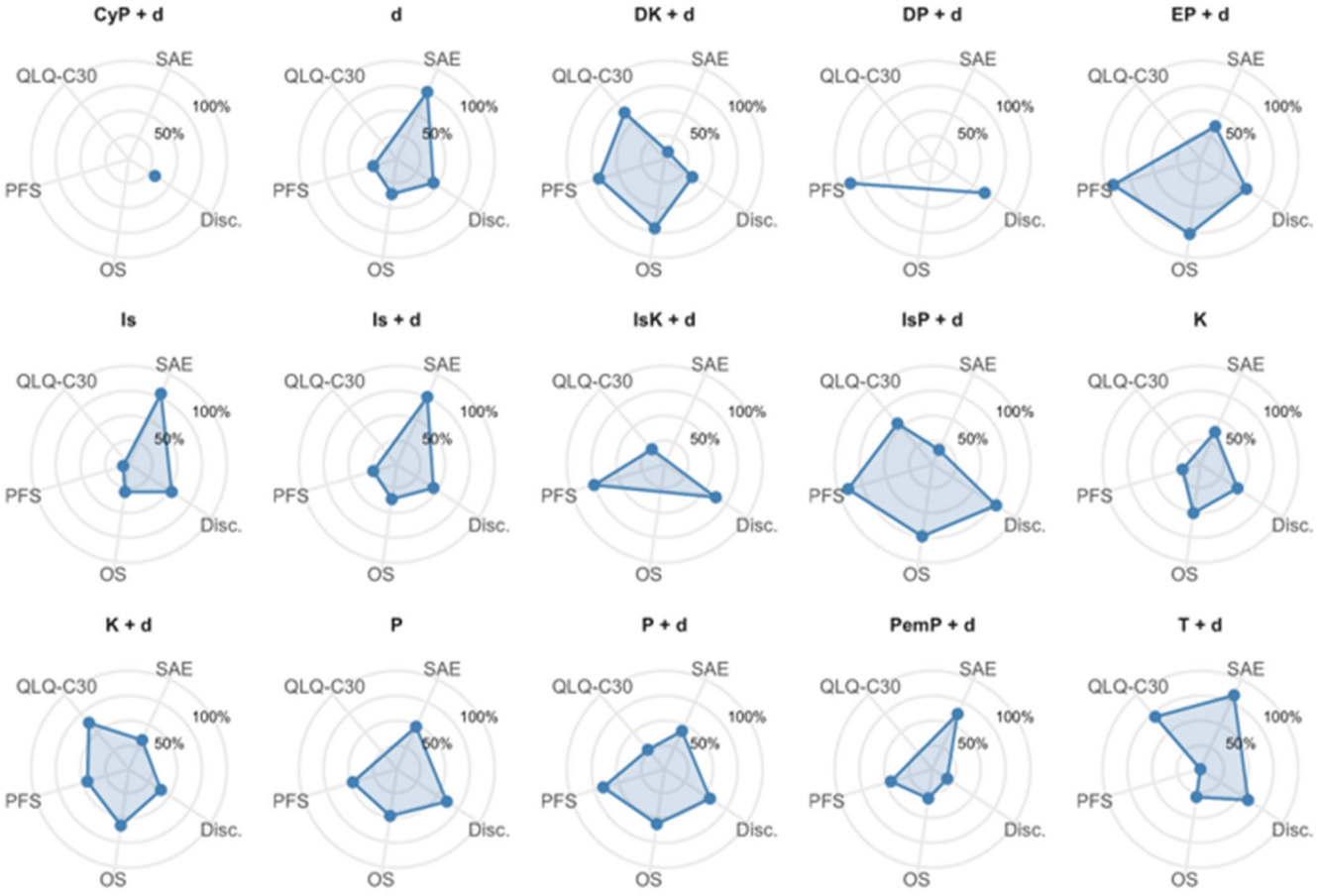
Summary of findings for discontinuation due to adverse events

Treatment	Rank (P-score)§ relevant for non-refractory patients	Rank (P-score)§ of treatments relevant for people who are:		
		Refractory to R	Refractory to V	Refractory to R + V
IsP + d	1.00 (0.89)	1.00 (0.83)	1.00 (0.83)	1.00 (0.78)
IsK + d	2.00 (0.78)	2.00 (0.70)	2.00 (0.70)	3.00 (0.63)
PV + d	3.00 (0.76)	5.00 (0.67)	NA	NA
DP + d	4.00 (0.74)	3.00 (0.70)	3.00 (0.70)	2.00 (0.64)
P	5.00 (0.74)	4.00 (0.69)	4.00 (0.69)	4.00 (0.61)
T + d	6.00 (0.74)	6.00 (0.67)	5.00 (0.66)	5.00 (0.58)
P + d	7.00 (0.71)	7.00 (0.65)	6.00 (0.65)	7.00 (0.56)
EP + d	8.00 (0.69)	8.00 (0.64)	7.00 (0.64)	6.00 (0.56)
EV + d	9.00 (0.68)	10.00 (0.61)	NA	NA
Is	10.00 (0.67)	9.00 (0.63)	8.00 (0.63)	8.00 (0.53)
d	11.50 (0.64)	12.50 (0.57)	9.50 (0.58)	10.50 (0.45)
Is + d	11.50 (0.64)	12.50 (0.57)	9.50 (0.58)	10.50 (0.45)
K	13.00 (0.64)	11.00 (0.57)	11.00 (0.57)	9.00 (0.46)
VorV	14.00 (0.60)	14.00 (0.54)	NA	NA
V	15.00 (0.57)	15.00 (0.52)	NA	NA
K + d	16.00 (0.56)	16.00 (0.51)	12.00 (0.50)	12.00 (0.40)
CyV + d	17.00 (0.53)	17.00 (0.49)	NA	NA
BevV	18.00 (0.52)	18.00 (0.49)	NA	NA
DV + d	19.00 (0.51)	19.00 (0.47)	NA	NA
V + d	20.00 (0.49)	20.00 (0.46)	NA	NA
DK + d	21.00 (0.48)	21.00 (0.45)	13.00 (0.44)	13.00 (0.34)
SV	22.00 (0.43)	22.00 (0.41)	NA	NA
CyP + d	23.00 (0.40)	23.00 (0.39)	14.00 (0.38)	14.00 (0.32)
TabV + d	24.00 (0.38)	24.00 (0.37)	NA	NA
ER + d	25.00 (0.38)	NA	15.00 (0.32)	NA
KR + d	26.00 (0.31)	NA	17.00 (0.26)	NA
DR + d	27.00 (0.30)	NA	18.00 (0.25)	NA
PemP + d	28.00 (0.28)	26.00 (0.29)	16.00 (0.26)	15.00 (0.19)
SeV + d	29.00 (0.28)	25.00 (0.30)	NA	NA
R + d	30.00 (0.27)	NA	19.00 (0.20)	NA
VT + d	31.00 (0.23)	27.00 (0.25)	NA	NA
IR + d	32.00 (0.23)	NA	20.00 (0.16)	NA
VeV + d	33.00 (0.20)	28.00 (0.22)	NA	NA
FV + d	34.00 (0.17)	29.00 (0.21)	NA	NA
DoxV	35.00 (0.10)	30.00 (0.13)	NA	NA

§ Treatments are ranked with respect to this outcome from best (rank 1) to worst according to P-score. A lower rank should not be interpreted to mean that a treatment is definitively worse than a higher-ranked treatment. Ranking based on P-score for treatments relevant for non-refractory patients, patients who are refractory to lenalidomide (R), patients who are refractory to bortezomib (V), and patients who are refractory to both lenalidomide (R) and bortezomib (V).
R: lenalidomide, V: bortezomib, NA: not applicable

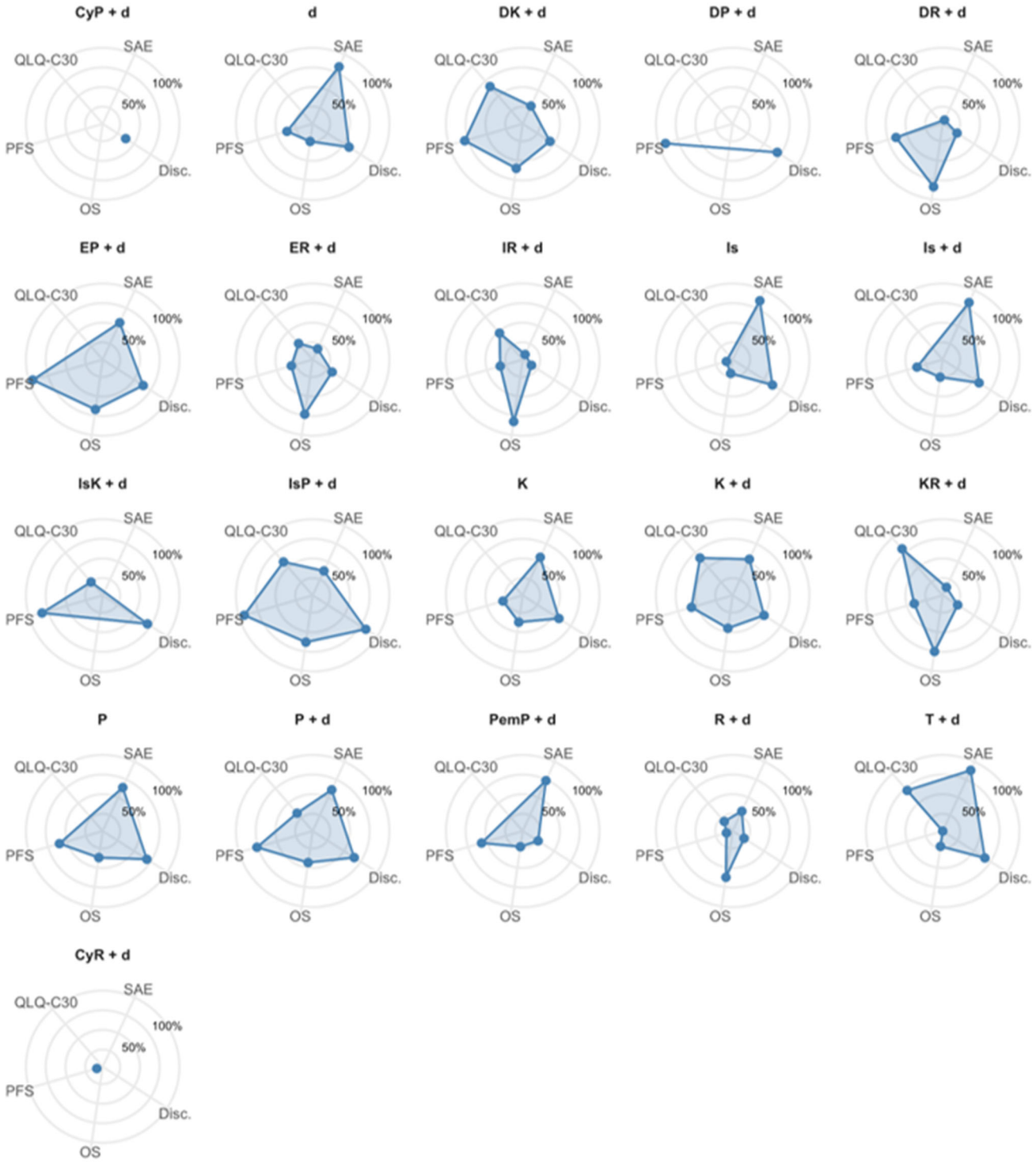
Appendix 11

Radarplots of all included treatments

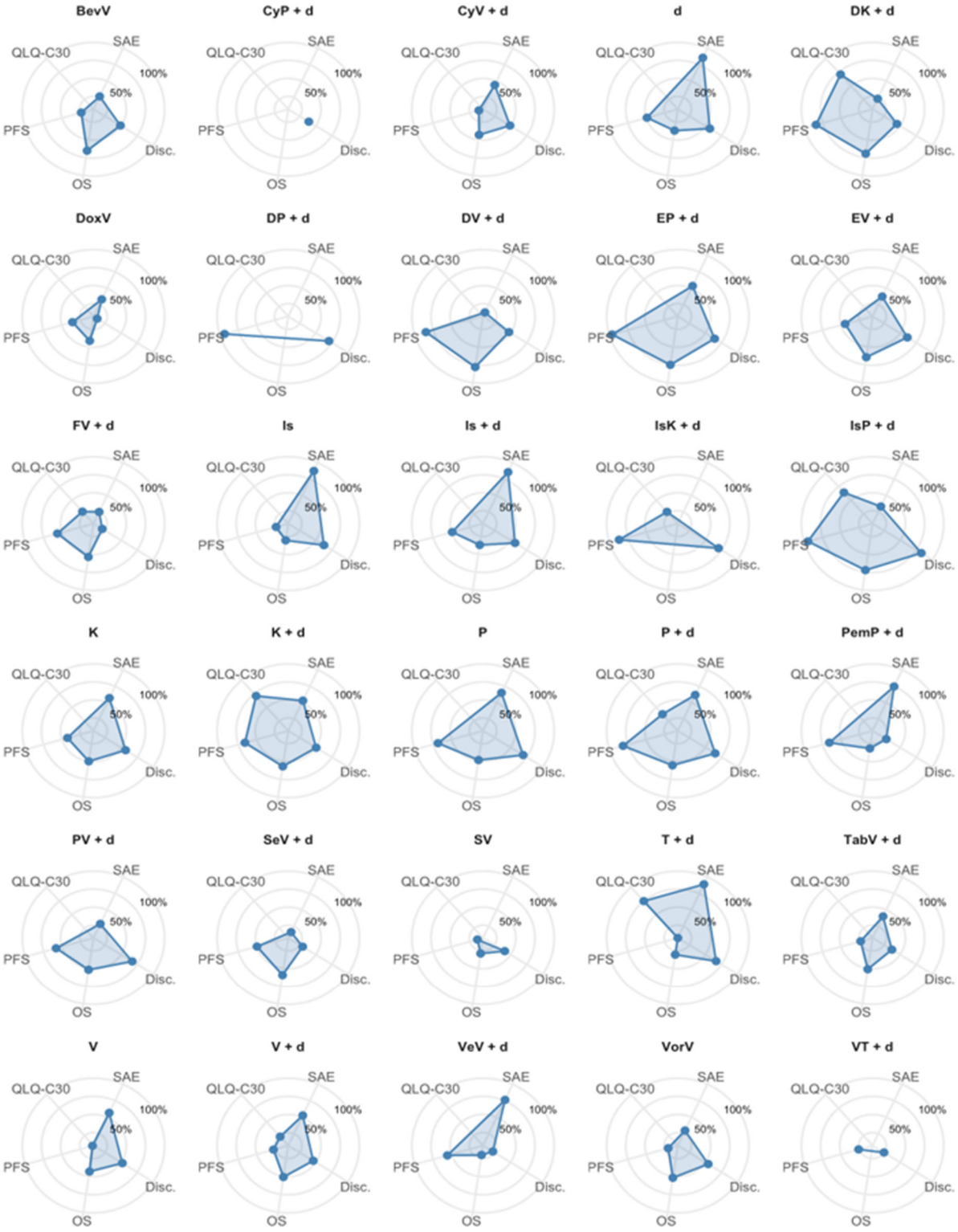


Radar plot of treatment regimens relevant for patients refractory to lenalidomide (R) and bortezomib (V)

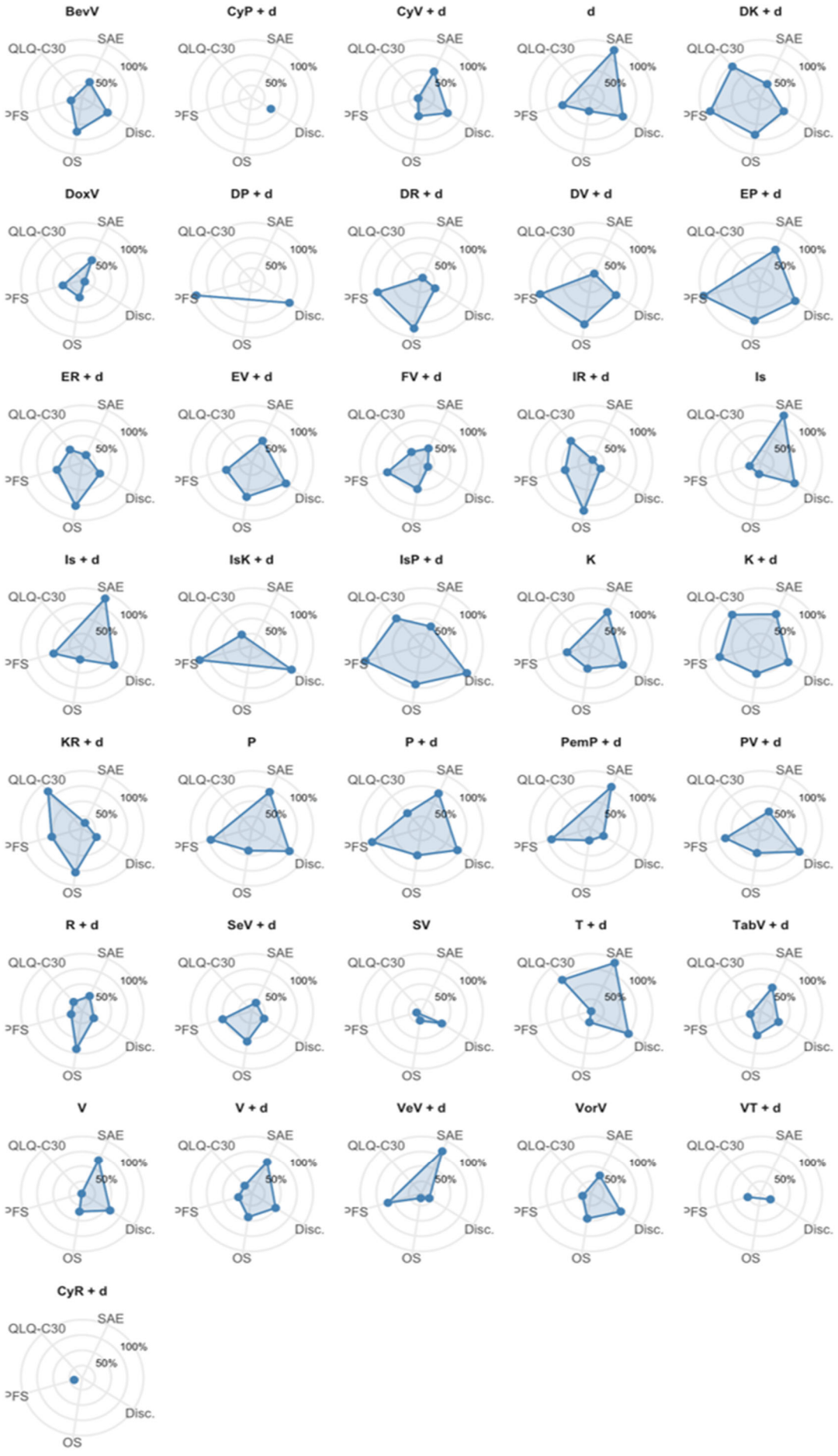
Cy: cyclophosphamide, D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), E: elotuzumab, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, Pem: pembrolizumab, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), SAE: severe adverse events (incidence rate ratio), T: thalidomide. The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.



Radar plot of treatment regimens relevant for patients refractory to bortezomib (V)
Cy: cyclophosphamide, D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), E: elotuzumab, I: ixazomib, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, Pem: pembrolizumab, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), R: lenalidomide, SAE: severe adverse events (incidence rate ratio), T: thalidomide. The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.



Radar plot of treatment regimens relevant for patients refractory to lenalidomide (R)
Bev: bevacizumab, Cy: cyclophosphamide, D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), Dox: doxorubicin, E: elotuzumab, F: panobinostat, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, Pem: pembrolizumab, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), S: siltuximab, SAE: severe adverse events (incidence rate ratio), Se: selinexor, T: thalidomide, Tab: tabalumab, V: bortezomib, Ve: venetoclax, Vor: vorinostat. The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.


BevV
CyP + d
CyV + d
d
DK + d
DoxV
DP + d
DR + d
DV + d
EP + d
ER + d
EV + d
FV + d
IR + d
Is
Is + d
IsK + d
IsP + d
K
K + d
KR + d
P
P + d
PemP + d
PV + d
R + d
SeV + d
SV
T + d
TabV + d
V
V + d
VeV + d
VorV
VT + d
CyR + d
SAE
QLQ-C30
PFS
Disc.
OS
100%
50%

Radar plot of treatment regimens relevant for non-refractory patients

Bev: bevacizumab, Cy: cyclophosphamide, D: daratumumab, d: dexamethasone, Disc.: discontinuation due to adverse events (risk ratio), Dox: doxorubicin, E: elotuzumab, F: panobinostat, I: ixazomib, Is: isatuximab, K: carfilzomib, OS: overall survival (hazard ratio), P: pomalidomide, Pem: pembrolizumab, PFS: progression-free survival (hazard ratio), QLQ-C30: quality of life (difference in mean score), R: lenalidomide, S: siltuximab, SAE: severe adverse events (incidence rate ratio), Se: selinexor, T: thalidomide, Tab: tabalumab, V: bortezomib, Ve: venetoclax, Vor: vorinostat. The radar plots summarize relative efficacy and safety but do not reflect assessments of the certainty of evidence or results of the health economic analysis.

Appendix 12

Relevant cost-effectiveness analysis studies

Table of relevant cost-effectiveness analyses of treatments for relapsed and/or refractory multiple myeloma (RRMM)

Author, year (reference)	Treatments	Type	Stage	(Model) info	Utility values	Conclusion
Borg, 2016 (113)	Pom d vs. HiDex	CUA. Societal perspective	RRMM	Discrete event simulation. Three states. model parameters that describe the disease course were populated from the MM-003 trial.	PomDex: 0.65 (pre-progression) PomDex: 0.62 (post-prog) HighDex: 0.61 (pre-progression) HighDex: 0.59 (post/prog)	HIDEX: LYs: 1.12, QALYs 0.65, costs SEK 179,976; Pomd Lys 2.23, QALYs 1.39, cost 767,064, ICER per QALY: Pomd vs. HIDEX: SEK 347072/0.7351 = 798,613 per QALY
Brown, 2013 (114)	Rd vs. dex	CUA	RRMM	Discrete event simulation.	Progressive disease (without treatment response): 0.64 Progressive disease (with response): 0.81	Ld vs. d: 3.22 LY, 2.2 QALYs, ICER: £66,483 LEN/DEX is cost effective compared to dex for MM patients with one prior treatment from NHS perspective.

Author, year (reference)	Treatments	Type	Stage	(Model) info	Utility values	Conclusion
Fragoulakis, 2013 (115)	Rd vs. V	CUA	RRMM	Discrete event simulation.	Non-responders (progressive disease): 0.64, Other response: 0.81	Rd vs. V: ICERs 29,415 euros/LYG; 38,268 euros/QALY. 95% likelihood of being cost-effective at the WHO recommended threshold of 3 times per capita income (= 60,000 euros in Greece).
Hornberger, 2010 (125)	(1) V vs. dex (2) V vs. Rd	CUA	RRMM	Lifetime horizon.	Prior to relapse: 0.81 After relapse: 0.645	Bortezomib (Velcade) vs. Dex: ICER 90,2874/QALY 2010 Swedish kroner. Bortezomib vs. Ldex: cost saving (dominant). Cost-effective from Swedish perspective.
Jakubowiak, 2016 (121)	KRd vs. Rd	CUA	RRMM	PartSA.	Pre-progression (at 1, 3, 6, 12, 18+ cycles): KRd 0.81, 0.818, 0.829, 0.840, 0.851; Rd 0.81, 0.798, 0.808, 0.818, 0.829; Post-progression: KRd 0.664; Rd 0.643	KRd vs. Rd: ICER $107,520/ QALY gained against Rd [RRMM (1–3 prior therapies]. Reimbursement of KRd for patients with RMM may represent an efficient allocation of health care resources in the health care budget.
Møller, 2011 (116)	Rd vs. V	CUA	RRMM	PartSA. States: (on or off	Progressive disease (0.64) was the same as for patients not responding to treatment. All levels of response had the same utility value (0.81). (Utilities from van Agthoven, et al 2004)	ICERs: Len-dex vs. Bortezomib NOK 247,978/QALY; NOK 198,714/LY
Aceituno, 2018 (117)	Rd vs. Vd	CUA	Second line treatment	Markov. Pre-progression on treatment; pre-progression off treatment, progression and death	Pre-progression: 0.81;	ICERs Rd vs. Vd: US$13,886/LY and US$23,565/QALY Lenalidomide in combination with dexamethasone represents a potentially cost-effective alternative for the second-line treatment of RRMM patients who are ineligible for transplantation. Chilean National Health Service perspective

Author, year (reference)	Treatments	Type	Stage	(Model) info	Utility values	Conclusion
Carlson, 2018 (126)	Rd; Vd; KRd; ERd; IRd; FVd; DRd; DVd	CUA	RRMM	PartSA, (based on a Network meta-analysis) with states: progression-free survival (PFS), progressed disease with subsequent treatments, and death.	Progression after 2 years: 0.77; Progression: 0.64 [Taken from Agthoven 2004, CUA intensive chemo vs. ASCT in NDMM stage II/III disease]	2nd line treatment: ICERs US$/QALY vs. Rd:
Jakubowiak, 2017 (122)	Kd vs. Vd	CUA	RRMM	Cycle length: one week	2nd line treatment:	Vd: 792,538, KRd: 211,458, ERd 430,009, IRd 454,684, DRd 187,728, DVd 50,704
Pelligra, 2017 (119)	Pom-d, Dara, Kar)	CUA	RRMM (heavily pre-treated patients - median 5 treatments)	PartSA. Lifetime time horizon.	PFS on treatment: 0.82 (0.78-0.88)	3rd line treatment: ICERs in $/QALY vs. Rd: Vd -853,800, KRd 252,293, ERd: 484,168, IRd: 508,021, PVd: Dominant, DRd: 216,360, DVd: 60.359
Zhang, 2018 (120)	DVd vs. Vd; DRd vs. Rd	CUA	RRMM	4-week cycles.	PFS off treatment: 0.84 (0.82-0.97)	ICER: Kd vs. Vd $121,828/QALY. Kd56 is cost-effective for patients with R/RMM at a willingness-to-pay threshold of$150,000/QALY. Trial data in the model may limit generalizability; however, SEER registry data mitigates this challenge. Kd56 provides additional value in key subgroups and remains cost-effective after steep comparator discounts.
Cai, 2019 (118)	IRd, Rd, VTd vs. Vd	CUA	RRMM	Markov model. time horizon 3 years. Health states: progression-free (PF), post-progression (PP), and death. 28-day cycle.	Progressed: 0.65 (0.62-0.74)	Results over 3 years: Pom-d: incr. LY vs. Dara (+ 0.02) & Car (+0.07) and incr. QALYS of (+0.01 & + 0.05); Cost savings $8,919 vs. Dara, $195 vs. Kar POM-d may be a cost-effective treatment option relative to DARA or KAR in heavily pretreated patients with RRMM in the US.

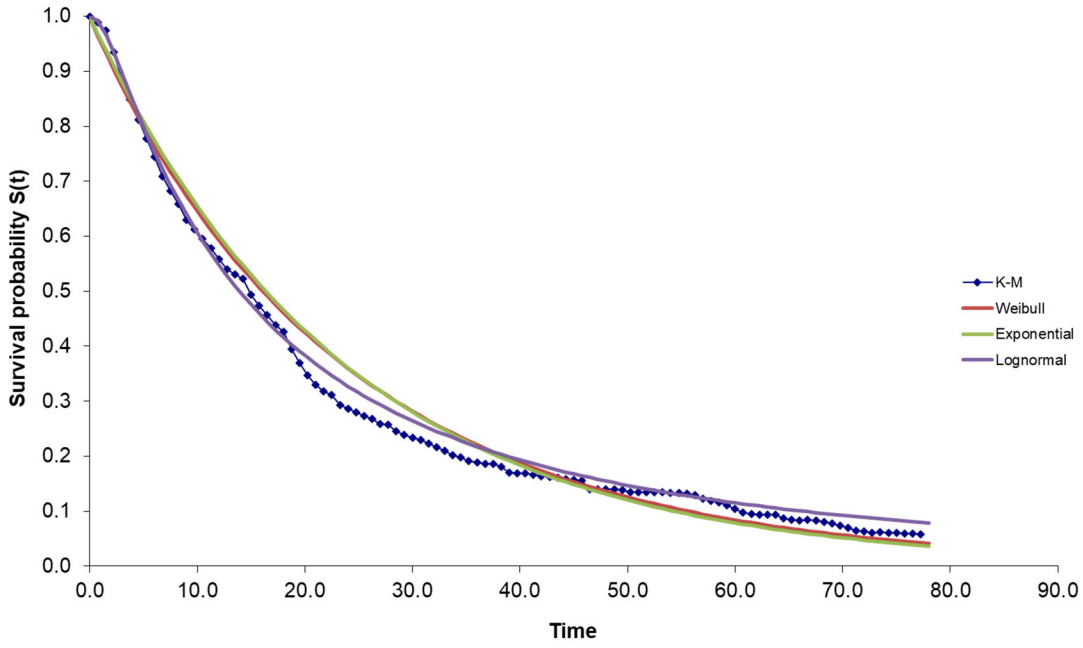
Author, year (reference)	Treatments	Type	Stage	(Model) info	Utility values	Conclusion
Campioni, 2020 (123)	KRd vs. Rd	CUA	RRMM	Semi-Markov. 10-year follow up. Stage length 4 weeks. States: PFS, progression and death.	3rd-line treatment:	ICERs: DVd vs. Vd US$284,180/QALY; DRd vs. Rd $1,369,062/QALY. 37% reduction in price of daratumumab required in combination with Vd to meet US WTP of $50,000. No amount of discount would make DRd cost-effect at US WTP
Wong, 2021 (124)	DRd vs. Rd	CUA	RRMM	Markov model. 10 yrs. States: PFS, progressed survival (PS), and death. [Chinese study]	PFS on treatment: 0.65 (0.52-0.78)	ICERS (vs. Vd) VTd: US$78,342/QALY; Rd: $52,713/QALY; IRD: $94,455/QALY; IRd vs. Rd: $228,030/QALY. Vd is considered cost-effective in China. Rd is an option for patients refractory to bortezomib.

D: daratumumab, d: dexamethasone, E: elotuzumab, F: panobinostat (Farydak), HighDex: high dose dexamethasone I: ixazomib, ICER: incremental cost effectiveness ratio, Is: isatuximab, K: carfilzomib (Kyprolis), P: pomalidomide, R: lenalidomide (Revlimid), T: thalidomide, V: bortezomib (Velcade). CUA: cost-utility analysis, PartSA: Partitioned Survival Analysis, PFS: progression-free survival, QALY: quality adjusted life years, RRMM: relapsen and/or refractory multiple myeloma, WTP: willingness-to-pay

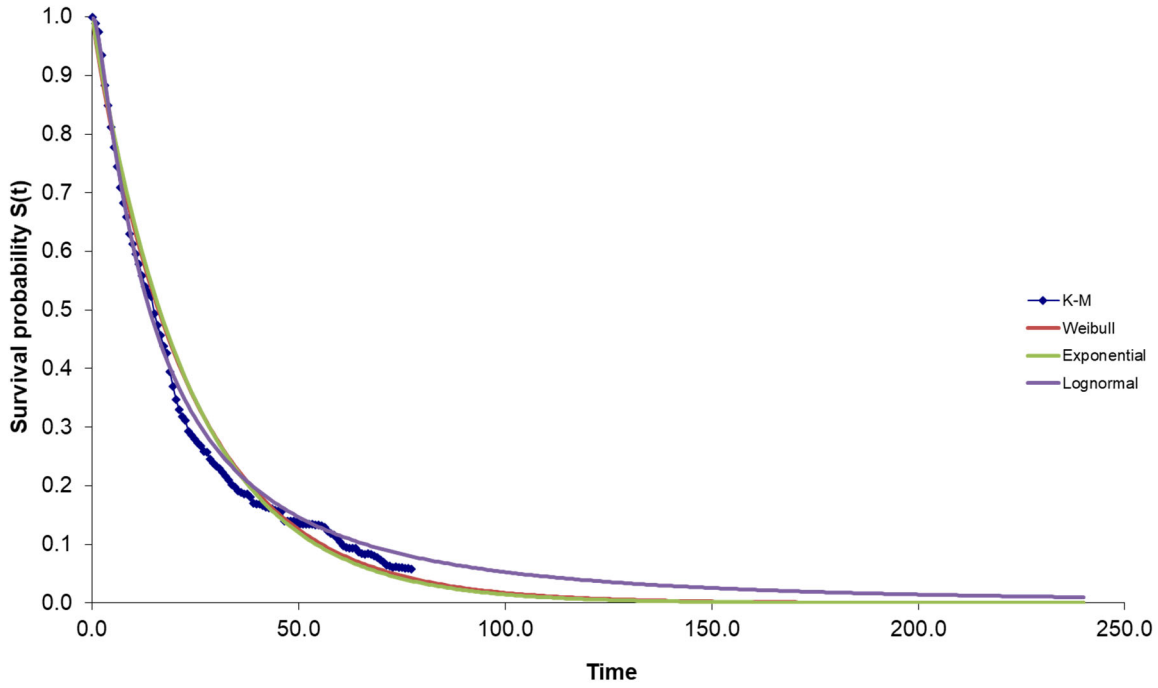
Appendix 13

Progression-free and overall survival curves for reference treatments

Survival curves for [R + d]

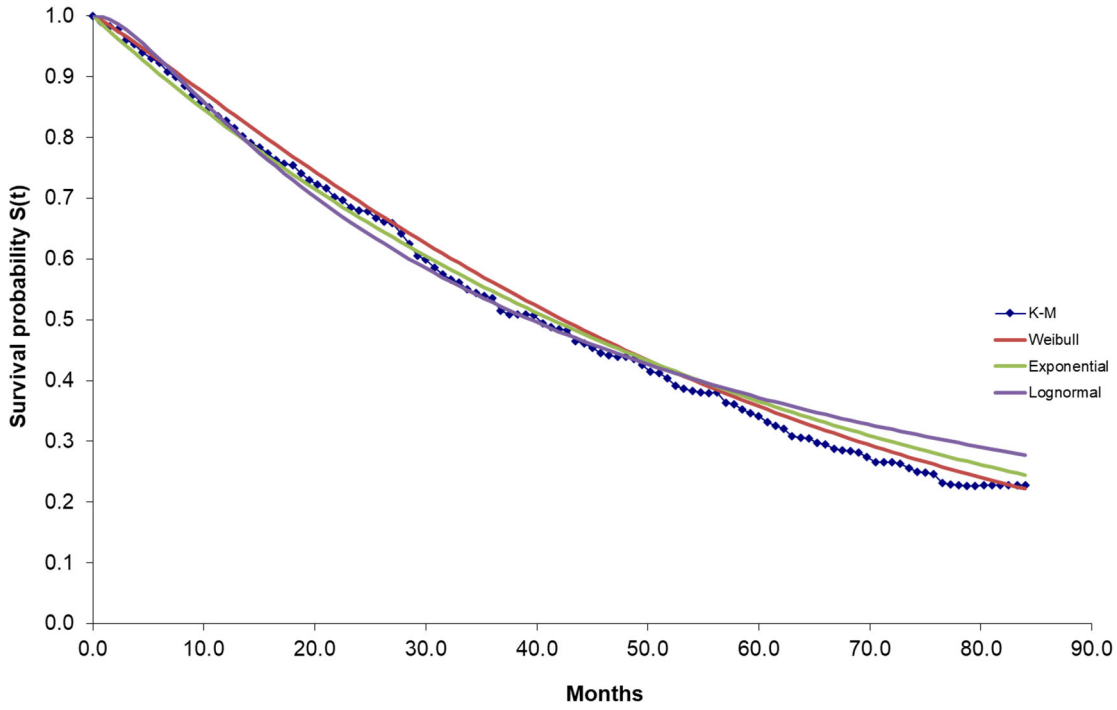


Progression free survival curve for [R + d] (months)
d: dexamethasone, R: lenalidomide (Revlimide)

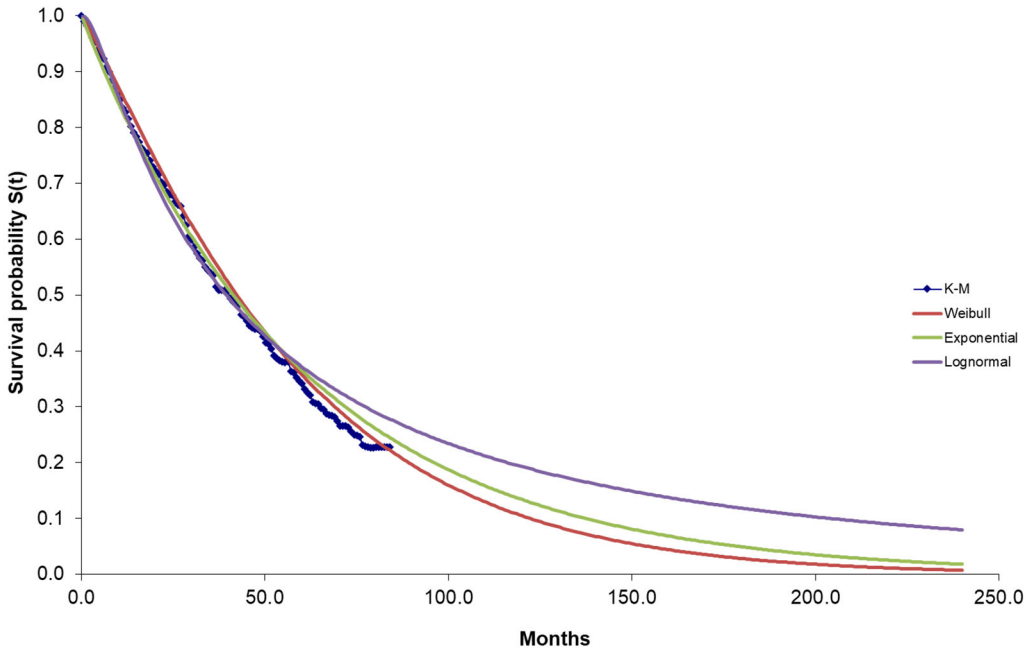


Progression free survival curve with extended time extrapolations for [R + d] (months)

d: dexamethasone, R: lenalidomide (Revlimide)



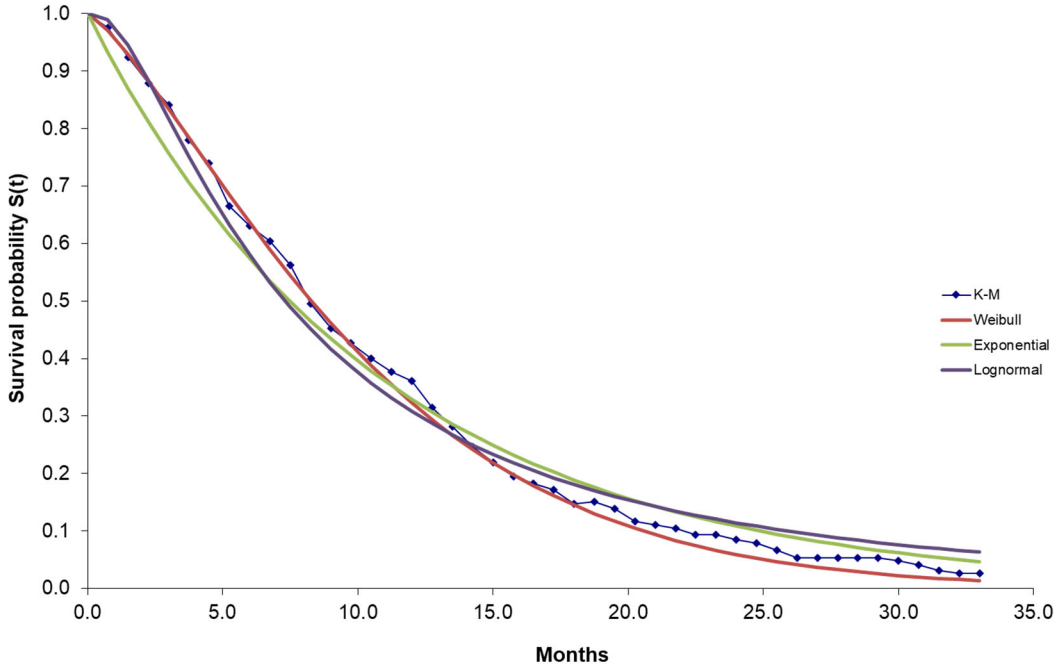
Overall survival curve for [R + d] (months)
d: dexamethasone, R: lenalidomide (Revlimide)



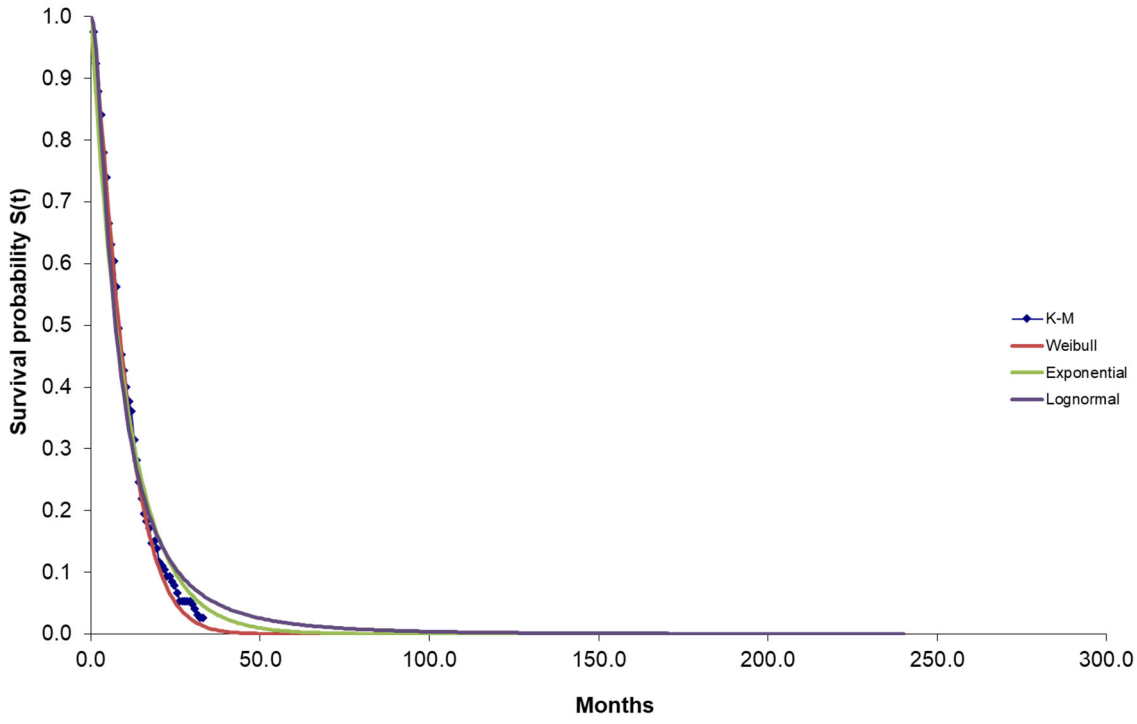
Overall survival curve with extended time extrapolations for [R + d] (months)[15]
d: dexamethasone, R: lenalidomide (Revlimide)

[15] The lognormal distribution overestimated the survival probabilities at the end of time horizon as the curve flattens. Pateints in the trial used for R+d were dead at the end of the trial period, therefore, based on the AIC and visual inspection the Weibull distribution was found to be a better fit for the reference.

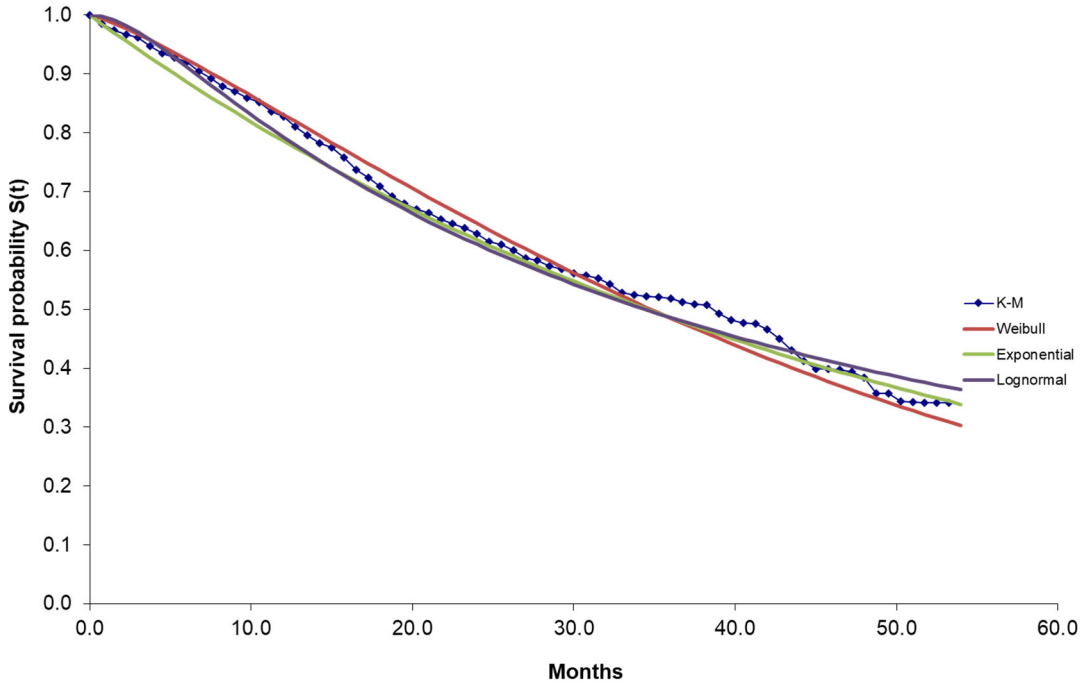
Survival curves for [V + d]



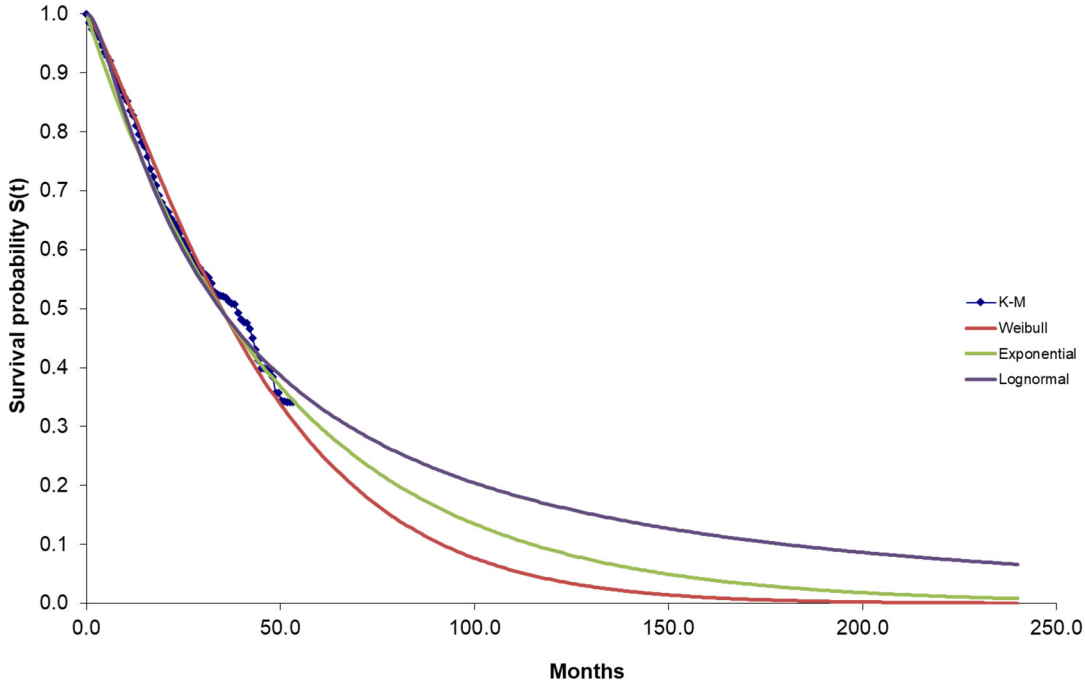
Progression free survival curve for [V + d] (months)
d: dexamethasone, V: bortezomib (Velcarde)



Progression free survival curve with extended time extrapolations for [V + d] (months)
d: dexamethasone, V: bortezomib (Velcarde)



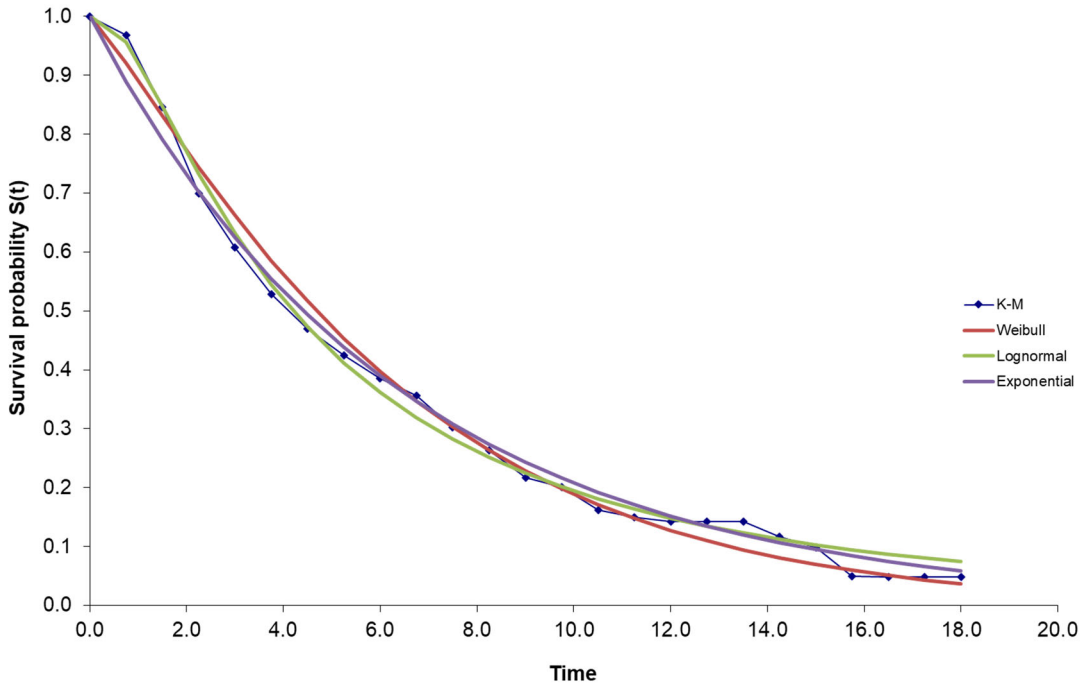
Overall survival curve for [V + d] (months)
d: dexamethasone, V: bortezomib (Velcarde)



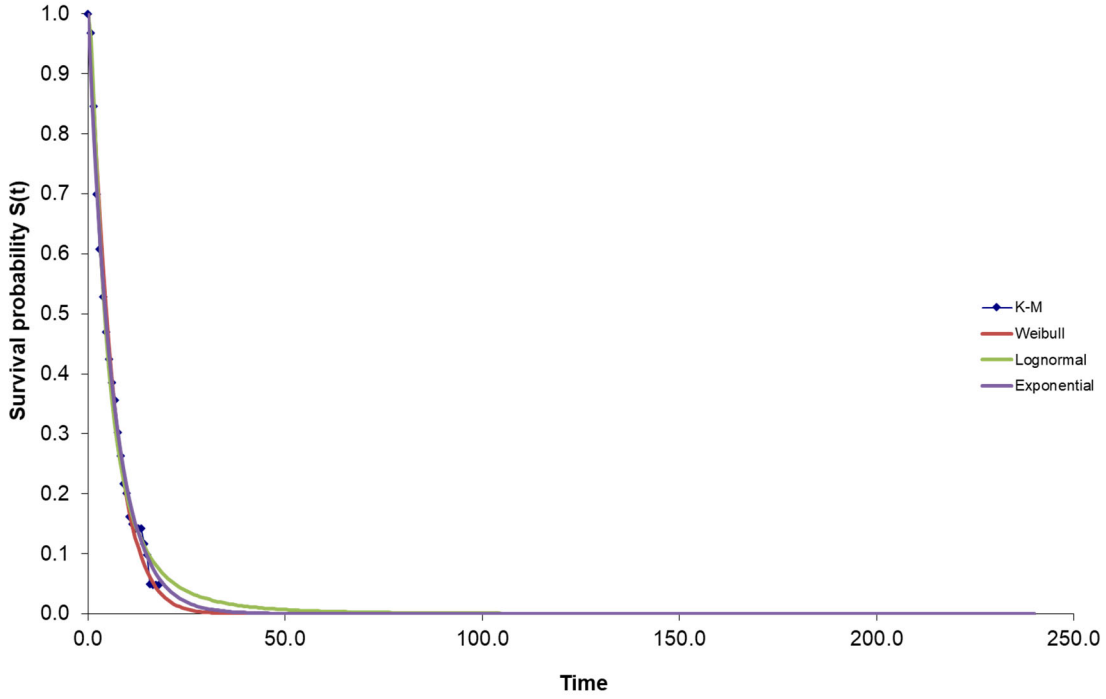
Overall survival curve with extended time extrapolations for [V + d] (months)[16]
d: dexamethasone, V: bortezomib (Velcarde)

[16] The lognormal distribution overestimated the survival probabilities at the end of time horizon as the curve flattens. Pateints in the trial used for V+d were dead at the end of the trial period, therefore, based on the AIC and visual inspection the Weibull distribution was found to be a better fit for the reference.

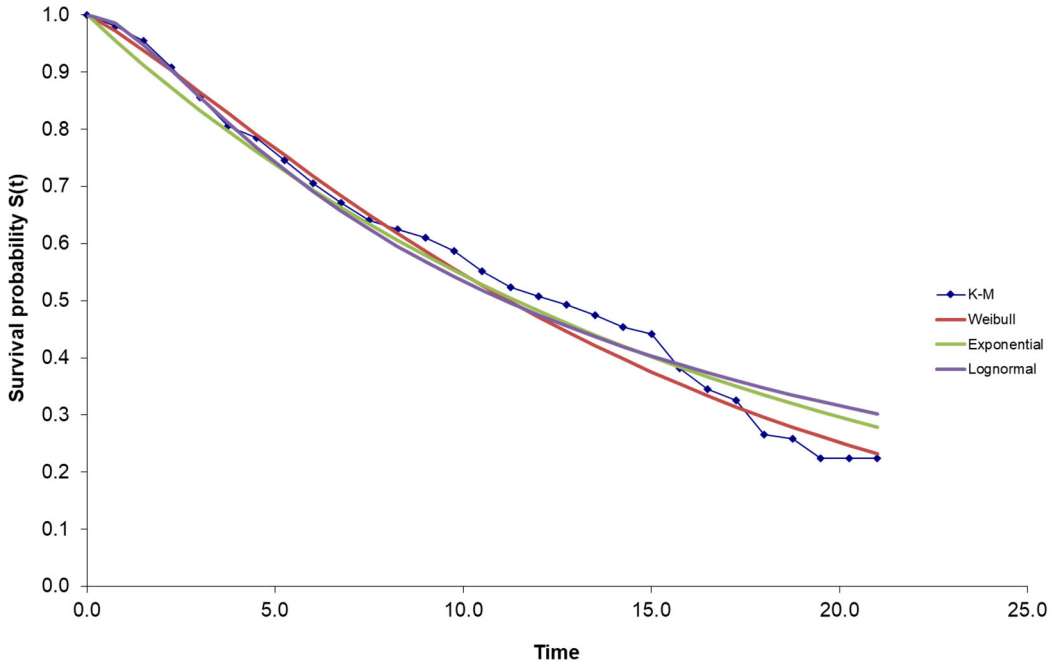
Survival curves for [P + d]



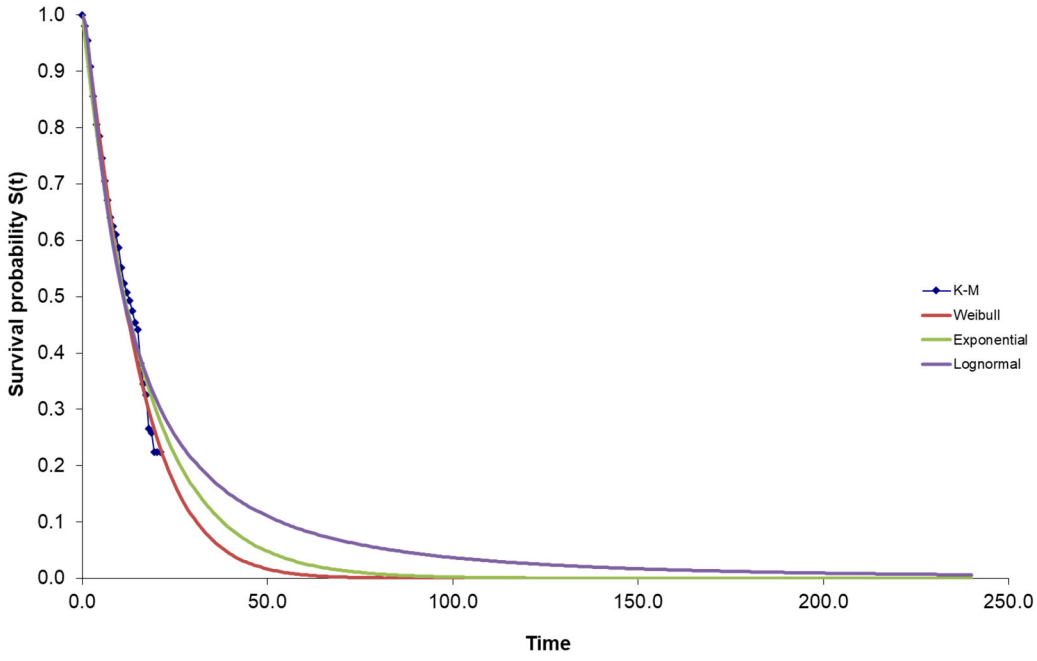
Progression free survival curve for [P + d] (months)
d: dexamethasone, P: pomalidomide



Progression free survival curve with extended time extrapolations for [P + d] (months)
d: dexamethasone, P: pomalidomide



Overall survival curve for [P + d] (months)
d: dexamethasone, P: pomalidomide



Overall survival curve with extended time extrapolations for [P + d] (months)[17]
d: dexamethasone, P: pomalidomide

[17] The lognormal distribution overestimated the survival probabilities after the the end trial time period for P+d. The patients were dead at the end of the trial period, therefore, based on the AIC and visual inspection the Weibull distribution was found to be a better fit for the reference.

Akaike Information Criteria for reference curves

Summary of Akaike Information criteria (AIC) for all reference treatments

Distribution	Rate (TreeAge)	Shape (TreeAge)	AIC Weibull	AIC Lognormal	AIC Exponential
	[R + d]				
Weibull (PFS)	0.04667	0.97	2218	2197	2217
Weibull (OS)	0.00996	1.1325	2408	2428	2412
	[V + d]				
Weibull (PFS)	0.042427	1.321	1909	1881	1909
Weibull (OS)	0.008471	1.2407	2878	2931	2891
	[P + d]				
Weibull (PFS)	0.11567	1.159	1478	1457	1483
Weibull (OS)	0.039298	1.1876	1200	1203	1204

The table provides the Akaike information criteria for all reference treatment curves according to the distribution of interest and the Weibull shape and scale parameter used to extrapolate survival functions. AIC: Akaike information criteria, d: dexamethasone, OS: overall survival, PFS: progression-free survival, R: lenalidomide (Revlimide), V: bortezomib (Velcade).

Point Calibration for reference curves

Summary of point calibration process on the Kaplan Meier plots for survival curves.

Time in Months	Number at risk = (Total – Censored patients – Events)	Events = At risk x (1 – HZ (t))	Censored patients	Event/At risk	Survival(t) (KM data) or S(t-1) x (1-(Event/Number at risk))	HZ (t)= S(t)/S(t-1)
0	325	0	(325-287) = 38	0	1	0
6	287	(287 x (1 – 0.923) = 22	287 – (255+22) = 10	0.077 or (1 – 0.923)	0.923	0.923/1 = 0.923
12	255	255 x (1 – 0.8969) = 26	255 – (26 + 228) = 1	0.103 or (1 – 0.8969)	0.828	0.828/0.923 = 0.8969
18	228	228 x (1 – 0.9109) = 20	228 – (20 + 208) = 0	0.0891 or (1 – 9109)	0.754	0.754/0.828 = 0.9109
X + 6 months	…	…	…	…	…	…

T= time, S= survival function of time, X= month, HZ= hazard as function of time

The table provides the steps undertaken to improve the accuracy of point estimates from the Kaplan Meier data extracted from the Web plot digitizer, the censored patients and events were calculated to ensure that the empirical survival probabilities were consistent with the points extracted from the web plot digitizer

Appendix 14

Cost Effectiveness Frontiers for all subgroups

Cost effectiveness frontier [R + d]

Cost effectiveness frontier [R + d]
The joined line represents all strategies that are not dominated by other treatments, dominated strategies and extendedly dominated strategies are presented separately from the frontier, see legend for further information. D: daratumumab, d: dexamethasone, E: elotuzumab, I: ixazomib, K: carfilzomib, R: lenalidomide

Cost effectiveness frontier [V + d]

Cost effectiveness frontier [V + d]
The joined line represents all strategies that are not dominated by other treatments, dominated strategies and extendedly dominated strategies are presented separately from the frontier, see legend for further information. D: daratumumab, d: dexamethasone, F: panobinostat, I: ixazomib, K: carfilzomib, P_ pomalidomide, V: bortezomib

Cost effectiveness frontier [P + d]

Cost effectiveness frontier [P + d]
The joined line represents all strategies that are not dominated by other treatments, dominated strategies and extendedly dominated strategies are presented separate from the frontier, see legend for further information. d: dexamethasone, E: elotuzumab, Is: isatuximab, P: pomalidomide

Cost Effectiveness Acceptability Frontiers for all subgroups

Cost effectiveness acceptability Frontier for [R + d]

Cost effectiveness acceptability Frontier for [R + d]
The graph presents strategies (i.e., treatments) that are cost effective at a certain willingness to pay threshold, as opposed to a Cost effectiveness acceptability curve that presents all strategies and their respective probability of cost effectiveness at a certain willingness to pay. D: daratumumab, d: dexamethasone, E: elotuzumab, I: ixazomib, K: carfilzomib, R: lenalidomide

Cost effectiveness acceptability Frontier for [V + d]

Cost effectiveness acceptability Frontier for [V + d]
The graph presents strategies (i.e., treatments) that are cost effective at a certain willingness to pay threshold, as opposed to a Cost effectiveness acceptability curve that presents all strategies and their respective probability of cost effectiveness at a certain willingness to pay. D: daratumumab, d: dexamethasone, F: panobinostat, K: carfilzomib, P: pomalidomide, V: bortezomib

Cost effectiveness acceptability Frontier for [P + d]

Cost effectiveness acceptability Frontier for [P + d]

The graph presents strategies (i.e., treatments) that are cost effective at a certain willingness to pay threshold, as opposed to a Cost effectiveness acceptability curve that presents all strategies and their respective probability of cost effectiveness at a certain willingness to pay. d: dexamethasone, E: elotuzumab, Is: isatuximab, P: pomalidomide

Deterministic results for all subgroups

Cost effective analysis for reference group [R + d]

Strategy	Cost (NOK)	Incr Cost (NOK)	Effect (QALYs)	Incr. effect	ICER (NOK)	ICER vs	Effect (Life-years)
R + d	[redacted]		2.90				4.00
IR + d	[redacted]	[redacted]	3.71	0.81	[redacted]	R + d	5.18
KR + d	[redacted]	[redacted]	3.45	- 0.26	[redacted]	IR + d	4.74
ER + d	[redacted]	[redacted]	3.34	- 0.37	[redacted]	IR + d	4.61
DR + d	[redacted]	[redacted]	4.08	0.37	[redacted]	IR + d	5.48

D: daratumumab, d: dexamethasone, E: elotuzumab, I: ixazomib, ICER: incremental cost-effectiveness ratios, Incr.: incremental, K: carfilzomib (Kyprolis), NOK: Norwegian krone, QALY: Quality adjusted life year. R: lenalidomide (Revlimide). The ICER values were rounded to the nearest thousand for easier visual comparisons, the strategies relevant to ICER calculation are also presented in the table.

Cost effective analysis for reference group [V + d]

Strategy	Cost (NOK)	Incr Cost (NOK)	Effect (QALYs)	Incr effect	ICER (NOK)	ICER vs	Effect (Life-years)
V + d	[redacted]		2.24				3.25
FV + d	[redacted]	[redacted]	2.38	0.14	[redacted]	V + d	3.39
PV + d	[redacted]	[redacted]	2.32	- 0.06	[redacted]	FV + d	3.28
K + d	[redacted]	[redacted]	2.49	0.11	[redacted]	FV + d	3.51
DV + d	[redacted]	[redacted]	3.48	1.24	[redacted]	V + d	4.91
DK + d	[redacted]	[redacted]	3.07	- 0.41	[redacted]	DV + d	4.29

D: daratumumab, d: dexamethasone, E: elotuzumab, F: pomalidomide, ICER: incremental cost-effectiveness ratios, Incr.: incremental, K: carfilzomib (Kyprolis), NOK: Norwegian krone, P: pomalidomide, QALY: Quality adjusted life year, V: bortezomib (Velcade). The ICER values were rounded to the nearest thousand for easier visual comparisons, the strategies relevant to ICER calculation are also presented in the table.

Cost effective analysis for reference group [P + d]

Strategy	Cost (NOK)	Incr Cost (NOK)	Effect (QALYs)	Incr effect	ICER (NOK)	ICER vs	Effect (Life-years)
P + d	[redacted]		0.81				1.15
EP + d	[redacted]	[redacted]	1.20	0.39	[redacted]	P + d	1.70
IsP + d	[redacted]	[redacted]	1.09	- 0.12	[redacted]	EP + d	1.56

d: dexamethasone, E: elotuzumab, ICER: incremental cost-effectiveness ratios, Incr.: incremental, Is: isatuximab, NOK: Norwegian krone, P: pomalidomide, QALY: Quality adjusted life year. The ICER values were rounded to the nearest thousand for easier visual comparisons, the strategies relevant to ICER calculation are also presented in the table.

Appendix 15

Progress log

Date	Action
27.05.2019	Commission given from the Regional Health Authorities to the Norwegian Institute of Public Health to perform a full HTA on the pharmacological treatment of multiple myeloma
18.11.2019	Start-up meeting with clinical experts and patient representative
16.12.2019	Sub-commission given from the Regional Health Authorities to the Norwegian Institute of Public Health to perform an evidence-mapping of different multiple myeloma populations
20.03.2020	Evidence-mapping submitted to the Regional Health Authorities
30.03.2020	Commission given from the Regional Health Authorities to the Norwegian Institute of Public Health to perform a full HTA on the pharmacological treatment of patients with relapsed and/or refractory multiple myeloma
10.09.2021	Report draft efficacy and safety - sent to internal reviewer
25.10.2021	Report draft efficacy and safety - sent to clinical experts for review
22.11.2021	The Norwegian Institute of Public Health informed the Regional Health Authorities about the negative review by the clinical expert, and the subsequent need to revise the set-up and analysis in the report
06.07.2022	Report draft efficacy and safety – sent to external reviewers and clinical expert
16.12.2022	Report draft health economic evaluation – sent to external reviewers and clinical expert.
24.02.2023	Report draft sent to department director for approval
01.03.2023	Report submitted to Commissioning Forum

NIPH

Published by the Norwegian Institute of Public Health
March 2023
P.O.B 4404 Nydalen
NO-0403 Oslo
Telefon: + 47-21 07 70 00
The report can be downloaded as pdf
at www.fhi.no/en/publ/