Longitudinal patient-reported outcomes

in patients with multiple myeloma

PhD dissertation

Lene Kongsgaard Nielsen

Faculty of Health Sciences
University of Southern Denmark

Quality of Life Research Center, Department of Haematology,
Odense University Hospital, Denmark

Supervisors

Niels Abildgaard, professor, chief physician, DMSc. (main supervisor)
Department of Haematology, Odense University Hospital, and Department of Clinical Research, University of Southern Denmark, Denmark

Mary Jarden, associate professor, PhD, MScN
Department of Haematology, Copenhagen University Hospital, and University of Copenhagen, Denmark

Christen Lykkegaard Andersen, PhD, associate professor, DMSc.
Department of Haematology, Copenhagen University Hospital, and University of Copenhagen, Denmark

Henrik Frederiksen, professor, chief physician, PhD
Department of Haematology, Odense University Hospital, and Department of Clinical Research, University of Southern Denmark, Denmark

International collaborator

Madeleine King, professor, director of the Quality of Life Office
Psycho-Oncology Co-operative Group, University of Sydney, Australia

Evaluation committee

Pernille Tine Jensen, professor, chief physician, PhD (Chair)
Department of Gynaecology and Obstetrics, Odense University Hospital, Denmark

Finn Wisløff, professor emeritus
Institute of Clinical Medicine, Oslo University, Norway

Fabio Efficace, MSc, PhD
Health Outcomes Research Unit, Italian Group for Adult Hematologic Diseases (GIMEMA) Foundation, Italy

Contact Information

Lene Kongsgaard Nielsen, MD, Specialist in Haematology
Quality of Life Research Center, Department of Haematology, Odense University Hospital, Denmark
Email: lene.kongsgaard.nielsen@rsyd.dk
Phone: +45 2284 6973

Preface

This PhD thesis is based on five studies, which were carried out in collaboration with my supervisors. Additional collaborators in study III and V were the Danish Myeloma Study Group (DMSG). Tobias Wirenfieldt Klausen, Department of Haematology, Herlev and Gentofte Hospital has contributed to study III and IV. In study II, the Nordic Myeloma Study Group (NMSG) and the Dutch-Belgium Cooperative Trial Group for Hematology Oncology (HOVON) were collaborators, especially MD, PhD student, Claudia Stege, professor, MD, Sonja Zweegman and statistician, PhD, Birgit Lissenberg-Witte, Cancer Center Amsterdam, Netherlands. Study V was carried out in collaboration with DMSG with contribution from professor, Madeleine King, Psycho-Oncology Co-operative Research Group, University of Sydney, Australia and Sören Möller, Odense Patient Data Explorative Network, Odense University Hospital.

I would like to express my sincere gratitude to all the collaborators and a special thank you to the people, who made this work possible:

First of all to my supervisors: Niels Abildgaard, who inspired me to do research. I am truly grateful for his persisting support and optimism over the years. I am sincerely grateful to Mary Jarden for introducing me to the field of patient-reported outcome measurement. I would also like to thank Christen Lykkegaard Andersen for his contribution and Henrik Frederiksen for his constructive feedback.

I am grateful to Henrik Gregersen for giving me the opportunity to collaborate in analysing and interpreting health-related quality of life data from the CLAIM study (study III). Also, I am grateful to professor Anders Waage, Norway and professor Sonja Zweegman, Netherlands for letting me be a part of the team on analysing and interpreting health-related quality of life data from the HOVON87/NMSG18 study (study II).

I wish to thank my colleagues in the DMSG, who have supported the idea of performing the population-based study "Quality of Life in multiple myeloma patients" (study V). A special thanks goes to the ten local investigators at the Danish departments of haematology who have implemented and are recruiting patients for the study. Also, a great thanks to the project nurses at the ten departments, who are including the patients and play an important role in this research project.

To my international collaborators in the field of quality of life research; Madeleine King, Sidney, Australia, thank you for your warm welcome during my study period abroad in the QOL Office, Sydney. Also thank you, professor, Carolyn Schwartz, Boston, USA for our meetings and correspondence in the area of response shift and quality of life appraisal.

I owe special thanks to the Head of Department of Haematology, Hanne Vestergaard, for your support. Also, to my colleagues at the Haematological Research Unit, Department of Haematology in Odense with a special thank you to PhD, project manager of the QoL Research Center OUH, Henrik Eshøj for professional sparring. Thank you to Sally Grant for advising me on legal and financial issues and Vickie Svane Kristensen for practical support and proofreading.

My warmest thanks go to my family – especially my husband, Brian Kongsgaard Nielsen, for your everyday support through ups and downs - to our three children Ida, Anton and Marie and my mother, Birthe, for your patience and understanding.

My final thoughts and appreciation go to all the patients, who have taken the time to complete the questionnaires to make this research possible.

December, 2018

Lene Kongsgaard Nielsen

List of publications

The dissertation is based on the following papers:

Nielsen LK, Jarden M, Andersen CL, Frederiksen F, Abildgaard N
A systematic review of health-related quality of life in longitudinal studies of myeloma patients
European Journal of Haematology, 2017;99(1):3-17

Nielsen LK[*], Stege C[*], Lissenberg-Witte B, van der Holt B, Mellqvist UH, Salomo M, Bos G, Levin MD, Visser-Wisselaar H, Hansson M, van der Velden A, Deenik W, Coenen J, Hinge M, Klein S, Tanis B, Szatkowski B, Brouwer R, Westerman M, Leys R, Sinnige H, Haukås E, van der Hem K, Durian K, Gimsing P, van de Donk N, Stevens-Kroef M, Sonneveld P, Waage A, Abildgaard N and Zweegman S
Health-related quality of life in transplant ineligible newly diagnosed multiple myeloma patients treated with either thalidomide or lenalidomide-based regimen until progression (HOVON87/NMSG18 study): a prospective, open-label, multicenter, randomized, phase 3 study [*]Equal contribution
Manuscript

Nielsen LK, Klausen TW, Jarden M, Frederiksen H, Vangsted AJ, Do T, Kristensen IB, Frølund UC, Andersen CL, Abildgaard N, Gregersen H
Clarithromycin added to Bortezomib-Cyclophosphamide-Dexamethasone impairs health-related quality of life in multiple myeloma patients
European Journal of Haematology, 2019;102(1):70-8.

Nielsen LK, Abildgaard N, Jarden M and Klausen TW
Methodological aspects of health-related quality of life measurement and analysis in patients with multiple myeloma
British Journal of Haematology 2018, accepted for publication

Nielsen LK, King M, Möller S, Jarden M, Andersen CL, Frederiksen F, Gregersen H, Klostergaard A, Steffensen MS, Pedersen PT, Hinge M, Frederiksen M, Jensen BA, Helleberg C, Mylin AK and Abildgaard N
Strategies to improve patient-reported outcome completion rate in longitudinal studies
Manuscript

List of abstract contributions at international conferences
25th Annual Conference of International Society of Quality of Life Research, Dublin, Ireland, October 2018
23rd Congress of the European Hematology Association, Stockholm, Sweden, June 2018
59th Annual Meeting of American Society of Hematology, Atlanta, USA, December 2017
24th Annual Conference of International Society of Quality of Life Research, Philadelphia, USA, October 2017
22nd Congress of the European Hematology Association, Madrid, Spain, June 2017
58th Annual Meeting of American Society of Hematology, San Diego, CA, December 2016
21st Congress of the European Haematology Association, Copenhagen, Denmark, June 2016

Abbreviations

BAI	Brief appraisal inventory
CRAB	Hypercalcaemia, renal failure, anaemia and bone lesions
DMSG	Danish Myeloma Study Group
EORTC	European Organisation for Research and Treatment of Cancer
EORTC QLQ-C30	European Organisation For Research And Treatment Of Cancer Quality of Life C30
EORTC QLQ-MY20	European Organisation For Research And Treatment Of Cancer multiple myeloma module
GRC	Global rating of change
HDT	High-dose therapy with autologous stem cell support
HOVON	Dutch-Belgium Cooperative Trial Group for Hematology Oncology
HRQoL	Health-related quality of life
IMID	Immunomodulatory drugs
IMWG	International Myeloma Working Group
MAR	Missing at random
MCAR	Missing completely at random
MGUS	Monoclonal Gammopathy of Undetermined Significance
MID	Minimal important difference
MM	Multiple myeloma
MNAR	Missing not at random
MPR	Melphalan-prednisolone-lenalidomide
MPR-R	Melphalan-prednisolone-lenalidomide and lenalidomide maintenance
MPT	Melphalan-prednisolone-thalidomide
MPT-T	Melphalan-prednisolone-thalidomide and thalidomide maintenance
MPV	Melphalan-prednisolone-bortezomib
NDMM	Newly diagnosed multiple myeloma
NMSG	Nordic Myeloma Study Group
NR	Non-responses
PD	Progressive disease
PFS	Progression free survival
PoCoG	Psycho-Oncology Co-operative Research Group
PRO	Patient-reported outcomes
QoL	Quality of Life
QoL-MM	Quality of Life in Danish multiple myeloma patients
r	Reliability
RMM	Relapsed multiple myeloma
Rd	Lenalidomide-dexamethasone
SD	Standard deviation
SEM	Standard error of measurement
SMM	Smoldering myeloma
VCD	Bortezomib-cyclophosphamide-dexamethasone
VRD	Bortezomib-lenalidomide-dexamethasone

Table of content

Danish summary (Dansk resume)

Myelomatose (MM) er en malign vækst af plasmaceller i knoglemarven. Sygdommen rammer især den ældre del af befolkningen, da medianalder ved diagnosen er ca. 70 år. På verdensplan konstateres årligt 86.000 nye tilfælde af sygdommen. MM er uhelbredelig, men følsom for behandling, og det typiske forløb er perioder med aktiv, behandlingskrævende sygdom samt perioder uden sygdomsaktivitet og behov for behandling. Prognosen er forbedret betydeligt gennem de seneste 20 år på grund af nye behandlingstilbud.

Livet som MM patient er forbundet med varierende symptomer fra sygdommen i form af knoglesmerter/-frakturer, lav blodprocent, nedsat nyrefunktion og/eller forhøjet kalk i blodet. Ligeledes må mange patienter leve med bivirkninger og senfølger efter sygdommen, såvel som behandling. Samlet benævnes dette patienternes *helbredsrelaterede livskvalitet* (HRQoL), som kan måles ved spørgeskemaundersøgelser, også kaldet patient-rapporteret outcomes (PRO).

Der er fundet metodemæssige udfordringer i tolkning af svarene fra spørgeskemaundersøgelser med gentagende HRQoL målinger. En af udfordringerne er at fastsætte den rigtige grænse for, hvad patienterne betragter som en klinisk meningsfuld effekt af behandlingen på HRQoL. En anden udfordring er, at patienter muligvis adapterer sig til ændringer i HRQoL over tid, hvilket kan påvirke, hvad disse patienter besvarer til spørgsmålene og dermed komplicere tolkningen af svarene. En tredje udfordring er, hvis patienterne ikke besvarer spørgeskemaerne, da der er risiko for skævvridning af resultaterne, hvis disse manglende besvarelser skyldes lav HRQoL.

Det overordnede formål med afhandlingen er at afgøre, om PRO-data kan belyse klinisk meningsfulde effekter af behandling på HRQoL set fra MM patienters perspektiv, på trods af metodemæssige udfordringer. Første delmål var at tolke offentliggjorte studier med gentagne HRQoL målinger, hvor spørgeskemaet EORTC QLQ-C30 blev anvendt, med afsæt i grænser for klinisk meningsfuld effekt af behandlingen. Andet delmål var at analysere den klinisk meningsfulde effekt på HRQoL af fire behandlinger ved brug af EORTC QLQ-C30 spørgeskemaet med tillæg af EORTC QLQ-MY20 og relatere disse til tidligere fund. Tredje delmål var at undersøge omfanget af manglende spørgeskemabesvarelser i HRQoL studier med MM patienter. Desuden var formålet at afgøre virkningen af de tiltag, som er indsat i et igangværende HRQoL studie med MM patienter, for at reducere manglende spørgeskemabesvarelser.

Til afdækning af delmål 1 identificeres, ved en systematisk litteratursøgning, 18 HRQoL studier med patienter med nykonstanteret MM og patienter med tilbagefald. Konklusionen blev, at patienter med nykonstanteret MM generelt rapporterer en klinisk meningsfuld forbedret global livskvalitet og fysisk funktion samt reduceret smerte og træthed. Dette er forskelligt fra patienter med MM, som behandles for tilbagefald, som rapporterer uændret eller ligefrem forringelse af HRQoL.

For at besvare delmål 2 analyseres spørgeskema besvarelser fra to kliniske studier med patienter med nykonstateret MM. Den generelle tendens kunne genfindes, da patienterne rapporterede en klinisk meningsfuld forbedret HRQoL. En nævneværdig undtagelse herfra var de patienter, der blev behandlet med clarithromycin i tillæg til bortezomib-cyclofosfamid-dexamethason-induktionsbehandling før højdosis kemoterapi med stamcellestøtte. Denne patientgruppe rapporterede forværring af træthed, søvnløshed og appetittab samt tiltagende score for bivirkninger under behandlingen. Dette resulterede i, at den forventede bedring af global livskvalitet og fysisk funktion efter primær behandlingen udeblev, samt at den social funktion og body image faldt under behandlingen. På trods heraf rapporterede patienterne

forbedring i den emotionelle funktion efter behandlingen, hvilket var forskellig fra kontrolgruppen, som rapporterede uændret emotionel funktion.

Til belysning af delmål 3 hentes oplysninger om andelen af manglende spørgeskemabesvarelser fra de 18 studier, identificeret til besvarelse af delmål 1, og de to studier, analyseret til besvarelse af delmål 2. Andelen af manglende besvarelser under opfølgningen var mellem 2 og 22%. Desuden var andelen af manglende spørgeskemabesvarelser grundet studieophør mellem 27 og 99%, sammenlignet med antallet patienter inkluderet ved studiestart. I et igangværende HRQoL studie med inklusion af den generelle population af patienter med MM uddannes sygeplejerskerne i at forebygge manglende besvarelser. I det pågældende studie modtager de patienter, der ikke har svaret, en påmindelse, og ved fortsat manglende besvarelse tages kontakt til patienten. Disse tiltag har gjort, at andelen af manglende besvarelser under opfølgningen kun er 5%.

Vi konkluderer, at PRO-data kan belyse den klinisk betydende effekt af behandling på HRQoL fra MM patienters perspektiv, hvis studiet er veldesignet, og metodemæssige udfordringer adresseres. Patienterne i valideringsstudiet for EORTC QLQ-C30 og EORTC QLQ-MY20 spørgeskemaerne var hovedsageligt patienter med nykonstanteret MM. Domænet bivirkninger, som er en del af EORTC QLQ-C30, er ikke udviklet til at måle patientoplevede bivirkninger til nyere behandlinger for MM.

De tre metodemæssige udfordringer ved gentagende HRQoL målinger bør undersøges nærmere ved patienter med MM. For det første mangles valide grænseværdier, der kan afgøre, hvilken ændring, der udgør en klinisk meningsfuld ændring over tid, set fra MM patienters perspektiv. For det andet bør tendensen til, at patienterne adopterer sig til ændring i HRQoL over tid undersøges nærmere og inddrages i grænseværdier for klinisk betydende ændringer over tid. For det tredje, selvom det nu er bevist, at andelen af manglende spørgeskemabesvarelser i opfølgning kan reduceres, mangles der stadig viden om statistiske metoder, der kan sikre, at uundgåelige manglende spørgeskemabesvarelser ikke skævvrider resultaterne.

English summary

Multiple myeloma (MM) is a malignancy of the plasma cells in the bone marrow. MM primarily affects the elderly with a median age at diagnosis of about 70 years. Worldwide, it is estimated that 86,000 patients are diagnosed with MM yearly. MM is incurable, but treatment-sensitive, and the typical course of the disease is periods with symptomatic, treatment-demanding disease alternating with periods of remission without need of treatment. Due to novel treatments, the prognosis has improved over the last two decades, and overall survival for patients below 70 years now exceeds 6-7 years.

As a result, myeloma patients might experience severe morbidity caused by bone destruction/fractures, renal dysfunction, bone marrow failure and high infection rates. In addition, they might be subject to adverse events and/or late sequelae to repeated lines of treatment. The impact of the disease and treatment on patients defines the *health-related quality of life* (HRQoL) and can be assessed by patient-reported outcome (PRO) questionnaires.

Methodological challenges in interpretation of longitudinal HRQoL results have been identified and we here present the three challenges relevant for this thesis. One of the challenges is to determine a correct threshold for clinical meaningful treatment effect on HRQoL from the patients´ perspective. Another challenge is that patients might adapt to changes in HRQoL, which could affect the patients´ answers to the questions in the questionnaires and complicate interpretation of results. A third challenge is, if the patients do not complete the questionnaires, since there is a risk of biased results, if the non-responses are related to the patients´ poor health.

The overall aim of the thesis is to determine if PRO data are valuable tools for assessing clinically meaningful effect on HRQoL from the MM patients´ perspective in spite of methodological challenges. The first aim was to interpret published longitudinal studies using the EORTC QLQ-C30 questionnaires for HRQoL measurement according to thresholds for clinically meaningful treatment effects on HRQoL. The second aim was to determine the clinically meaningful treatment effect on HRQoL of four first line treatment regimens using the EORTC QLQ-C30 questionnaires with the addition of EORTC QLQ-MY20 and relate the results to previous findings. The third aim was to analyse the magnitude of non-responses in longitudinal studies with MM and determine the effect of implemented strategies to reduce non-responses during follow-up in an ongoing longitudinal HRQoL study of patients with MM.

To elucidate the first aim, a systematic literature search was performed with identification of 18 longitudinal HRQoL studies of patients with newly diagnosed or relapsed MM, where the EORTC QLQ-C30 questionnaire was used. Clinically meaningful improvements in global QoL, physical functioning and reduction of pain and fatigue were found far more likely during primary treatment regimens, whereas relapsed patients reported no change or even deterioration in HRQoL.

To answer the second aim, HRQoL data from two clinical trials in newly diagnosed MM patients, where the EORTC QLQ-C30 and EORTC QLQ-MY20 instruments were used for HRQoL measurement were analysed. The general findings were confirmed, since the newly diagnosed patients with MM reported clinical meaningful improvement in HRQoL during treatment. However, a noteworthy exception was patients treated with clarithromycin added to the bortezomib-cyclophosphamide-dexamethasone induction before high dose therapy with stem cell support. These patients reported increased fatigue, insomnia and appetite loss and increasing score for side effects of treatment during treatment. This resulted in lack of clinically

meaningful improvement in global quality of life and physical functioning as well as decreased social functioning and body image during treatment. In spite of that, the patients reported increased emotional functioning after treatment, which differ from the patients in the placebo group, who reported unchanged emotional functioning.

To illuminate the third aim, information of the magnitude of non-responses from the 18 studies identified to elucidate the first aim, and the two studies analysed as part of answering the second aim were extracted. The non-responses rate during follow-up was between 2 and 22%. Compared to number of patients at baseline, the proportion of non-responses due to study discontinuation was between 27 and 99%. In the ongoing longitudinal HRQoL study, study nurses are being educated in prevention of non-responses. Patients, who have not completed the questionnaires receive reminders, and later, if the patients still have not completed the questionnaire, the study nurses contact the patients. The applied strategies resulted in a rate of non-responses in follow-up of 5%.

We concluded that PRO tools can assess clinically meaningful treatment effects on HRQoL from the MM patients´ perspective, if the study is well-designed and methodological challenges are addressed. The EORTC QLQ-C30 and EORTC QLQ-MY20 questionnaires are only validated in newly diagnosed MM patients, and the side effect of treatment domain of EORTC QLQ-MY20 is not developed to capture symptomatic toxicities of novel anti-myeloma drugs or drug combinations.

The three methodological challenges in longitudinal HRQoL measurement need further investigation in MM. Firstly, valid thresholds to assess clinically meaningful treatment effects from the MM patients´ perspective are needed. Secondly, possible adaption to change in HRQoL needs further investigation and should be integrated into thresholds for clinically meaningful change. Lastly, the high magnitude of non-responses to questionnaires found in HRQoL studies of patients with MM might have biased the existing HRQoL results. Now, we have practical tools to reduce NRs during follow-up, but statistical methods for handling of unavoidable NR in order to reduce selection bias need further development.

Introduction

Multiple myeloma

Multiple myeloma (MM) is a malignancy of the plasma cells in the bone marrow. MM is the second most common haematological malignancy (1). The annual incidence is about 4.3 per 100,000 with ethnical differences. MM is twice as common in Afro-Americans compared to Caucasians; moreover, the disease is slightly more common in males compared to females (1, 2). The cancer most often affects older people, with a median age of about 70 years at diagnosis (3). Worldwide, it is estimated that 86,000 patients are diagnosed with MM yearly, with the highest incidence in the industrialized countries (2). In Denmark, the yearly incidence of MM has increased and is now above 400 annually (4-7). In 2015, 1864 patients were living with MM in Denmark (4, 5).

Risk factors for developing MM are poorly understood (2). It is a very heterogeneous disease, both clinically and as defined by tumour genetics (8). MM is almost always preceded by an asymptomatic premalignant plasma cell disorder, monoclonal gammopathy of undetermined significance (MGUS) or smoldering myeloma (SMM) (9, 10). The overall prevalence of MGUS is 3.2 % in the Caucasian population at the age of 50 years or older (10, 11). The mean cumulative risk of progression from MGUS to SMM or MM is 10% at 10 years. The same figure for transformation from SMM to MM is about 10% per year for the first 5 years, hereafter the risk decreases to about 2 to 3% per year (11-13). The diagnosis of MM is based on the International Myeloma Working Group (IMWG) updated criteria (14). The IMWG diagnostic criteria are based on the findings of >10 % clonal plasma cells in the bone marrow or biopsy-proven plasmacytoma *and* one or more so-called CRAB features - hyperCalcaemia, Renal failure, Anaemia and Bone lesions (14, 15). Moreover, almost all patients have a monoclonal protein, called M-component, in serum or urine. In case of no measureable M-component in serum or urine, the free light kappa or lambda ratio in serum will almost always be abnormal. However, a few cases of true non-secretory MM exist (14).

MM patients experience variable morbidity caused by bone destruction/fractures, renal dysfunction, bone marrow failure, high infection rates and potential physical disability (14). The most frequent MM symptoms at diagnosis are bone pain, fatigue and loss of weight (16). The goal with initial therapy is to achieve significant tumour reduction, long progression free survival and overall survival, as well as relief of symptoms and prevention of disease complications with as few side effects and long term toxicities as possible. However, pre-treatment predictors of response are lacking, and it is difficult to foresee how long treatment response will last and the disease will progress (17, 18). This is one of the conditions that the patients with MM must try to cope with.

Eventually, MM will progress or relapse (RMM), and anti-myeloma therapy again becomes necessary to achieve disease control. IMWG defined criteria of progressive disease (PD) and RMM in 2006, which were updated in 2016 (19, 20). Criteria for PD are fulfilled if M-component in blood, urine and/or bone marrow plasma cells infiltration have increased by ≥ 25% from nadir. In case of non-secretory MM, an increase of ≥ 25% in the difference between the involved and not-involved serum free light chains (the increase must be at least 100 mg/l). Criteria for PD are also established, if the patient develops new skeleton lesions (≥ 50% increase from nadir of >1 lesion), progression of known lesions (≥ 50% increase in longest diameter and >1 cm in short axis) or ≥ 50% increase in circulating plasma cells (minimum 200 cells per μL) (20). Clinical

relapse is defined by progressive CRAB criteria as well as by tumoral growths of plasmacytomas or by hyperviscosity related to the M-component (20).

Typically, MM patients experience very individual courses of disease trajectories and often receive multiple lines of anti-myeloma therapies in different drug combinations (21). Therefore, myeloma patients are at risk of reversible side effects to administered anti-myeloma drugs, e.g. infusion-related reactions or hospitalization-demanding infections as well as possible irreversible side effects, e.g. peripheral neuropathy, fatigue, anxiety and depression (22-27). Documented by cross-sectional studies, the most prevalent patient-reported symptoms and challenges across the disease pattern from diagnosis to advanced disease stage are fatigue (59-99%), pain (50-73%), constipation (33-65%), insomnia (36%) and peripheral neuropathy (33-53%) as well as decreased physical (54-99%), cognitive (80%) and role (80%) functioning and financial difficulties (31-78%) (22, 28-32). Myeloma complications, and treatment side effects and late sequelae along with existing or adjacent comorbidity cause frailty and high risk of early death in some patients (23, 33, 34).

Until the middle of the 1990s, the median overall survival of patients diagnosed with MM treated in clinical trials was only three years (35). The treatment landscape for MM has evolved markedly hereafter. According to published data, the overall median survival has improved and is now above 6 years for patients diagnosed before the age of 65 years and almost 3 years for patient diagnosed past the age of 65 (21, 36-38). The improved prognosis is caused by the introduction of high dose chemotherapy with autologous stem cell support (HDT) in the 1990s, and new treatment options with thalidomide, bortezomib and lenalidomide (36, 38, 39). The prognosis is expected to improve even further due to the current introduction of new targeted agents, such as second and third generation proteasome inhibitors, monoclonal antibodies, new immunomodulatory drugs (IMIDs), deacetylase inhibitors and signalling pathways/kinase inhibitors (40).

Treatment of multiple myeloma

First line treatment of younger, newly diagnosed MM (NDMM) patients under the age of 70 years with good performance status and without severe comorbidity is induction therapy followed by HDT (39, 41, 42). Few patients above the age of 70 years may be eligible for this intensive treatment. In Denmark and internationally, the recommended induction treatment is a bortezomib-based regimen, and currently in Denmark, consists of four 21-day cycles of bortezomib-cyclophosphamide-dexamethasone (VCD) therapy (43, 44). For transplant ineligible and elderly NDMM patients, the recommended first line treatment is repeated cycles of anti-myeloma drug combinations for 6-9 months or longer (45). In Denmark, current recommended regimens are melphalan-prednisolone-bortezomib (MPV), lenalidomide-dexamethasone (Rd) or bortezomib-lenalidomide-dexamethasone (VRD) based on results from the VISTA, FIRST, and SWOG S0777 trials (15, 46-49).

For RMM, a variety of treatment regimens and combinations can be used including repeating the primary treatment regimen. If relapse occurs later than 18 months after primary HDT, re-induction therapy followed by salvage HDT is an option for eligible patients (50). At relapse, treatment decision-making is individualised, based on duration, observed complications and late effects of former therapies, current comorbidity and performance status, the expected response and toxicity profile of available drugs, as well

as the patients´ preferences concerning drug-delivery convenience, goals in life, and health-related quality of life (HRQoL) during and after treatment (51-54).

Approved drugs for RMM in Denmark by December 2018 are the IMIDs thalidomide, lenalidomide and pomalidomide (55-57), the proteasome inhibitors bortezomib, carfilzomib, and ixazomib (58-60), and the monoclonal antibodies elotuzumab and daratumumab (61, 62). The histone deacetylase inhibitor, panobinostat, which is approved by European Medicines Agency, is not approved as standard of care treatment in Denmark (63). Most frequently used traditional chemotherapy agents are melphalan, cyclophosphamide, adriamycin, and bendamustin (15, 64). High-dose or moderate dose glucocorticoids, prednisolone or dexamethasone are part of most combination treatments. Monoclonal antibodies in combination with an IMID or proteasome inhibitor and steroid have become a preferred treatment regimen in RMM because of the high efficacy, which has been reported in clinical trials (61, 65-67). Therefore, the Danish treatment recommendation anno 2018 at first relapse for non lenalidomide-refractory patients is daratumumab-lenalidomide-dexamethasone, and for lenalidomide-refractory patients it is daratumumab-bortezomib-dexamethasone, based on data from the POLLUX and CASTOR studies (61, 65). Alternative treatment recommendations for RMM are carfilzomib, ixazomib or elotuzumab in combination with Rd (15, 58, 59, 62).

Quality of life, health-related quality of life and patient-reported outcomes

Quality of life
Quality of life (QoL) is a key concern for patient-centred care (68). While there are many definitions of QoL, there is no standard definition that fits all purposes (68, 69). In 1984, Calman formulated a hypothesis of QoL in cancer patients, which has later been referred to as "Calman´s gap":

"Quality of life is a difficult concept to define and to measure. A hypothesis is proposed which suggests that the quality of life measures the difference, or the gap, at a particular period of time between the hopes and expectations of the individual and that individual´s present experiences. Quality of life can only be described by the individual and must take into account many aspects of life."(70)

The linkage between expectations, experiences and QoL has later been described in "A model of Quality of Life" (69, 71, 72). This leads to the fact that people with different expectations to their own QoL might report different QoL scores, even when they are in the same condition. Therefore, the concept of QoL should be considered complex, subjective and dynamic (69, 72, 73).

Health-related quality of life
HRQoL is considered a sub-element of the broader concept QoL (69). Again, there is no standard definition of HRQoL (74-76), however often explicitly focusing on the impact of disease or treatment. A widely accepted definition is by Osoba *et al*.:

"..a multidimensional construct encompassing perceptions of both positive and negative aspects of dimensions, such as physical, emotional, social, and cognitive functions, as well as the negative aspects of somatic discomfort and other symptoms produced by disease or its treatment." (77)

Multiple HRQoL models have been created, and the most commonly used model is "A Conceptual Model" of Wilson *et al.* in 1995 with a later revision (76, 78, 79). The Wilson Conceptual Model is presented in figure 1 and relies on a division of the outcomes measures into five levels of 1) biological and physiological variables, 2) symptoms status, 3) functional status, 4) general health perceptions and 5) overall QoL (78). Each level of outcomes measures is interrelated, and when moving from left toward right, the outcomes go from being distale to proximale measures and become increasingly difficult to define and measure, and increasingly affected by non-medical factors (78, 80).

| Biological and Physiological Variables | → | Symptom Status | → | Functional Status | → | General Health Perceptions | → | Overall Quality of Life |

Figure 1. The Conceptual Model adapted from Wilson *et al.* (78).

Patient-Reported Outcomes

The U.S. Department of Health and Human Services Food and Drug Administration (FDA) have defined patient-reported outcomes (PRO) as:

"A PRO is any report of the status of a patient´s health condition that comes directly from the patient, without interpretation of the patient´s response by a clinician or anyone else" (81)

PRO measures can be used to assess QoL or HRQoL, and the outcome can be presented in absolute terms or as a change from a previous measure. At group level, PROs are being used in clinical trials to measure the treatment effect and symptomatic toxicities from the patients´ perspective of a medical intervention or as part of treatment quality assessment in health care (81-83). At an individual level, PROs are being used in clinical practice for symptom and adverse event monitoring and have shown to improve communication with health care professionals, symptom control, patient satisfaction and overall survival (84-86). Also, it is evident that PRO measures can be used in prognostic modelling in MM, other haematological and solid cancers (87-90).

Patient-reported outcome instruments

PRO can be measured by validated questionnaires, which can be divided into generic, disease or domain specific instruments, or by interview based methods (68). A generic instrument is designed to measure QoL in the healthy population as well as in different patient cohorts (68). An example of a generic instrument is Short-form health survey version 2-4-week recall (91). Disease specific instruments are designed to measure HRQoL in a specific population, e.g. the European Organisation for Research and Treatment of Cancer Quality of Life C30 (EORTC QLQ-C30) in cancer patients (92). Domain specific instruments are designed to capture PRO issues of a specific symptom or condition (68). An example of a domain specific instrument is the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group Neurotoxicity subscale (93). An instrument consists of a varied number of items, which are calculated into domain scores. Domains based on an answer to one question are called "single-item domains", and domains calculated on the basis of two or more answers are called "multi-item domains".

Validation of a PRO instrument is the process of investigation of a instrument´s ability do measure what it is intended to measure and it is useful for its purpose (68). Aspects of validation of instruments relevant for the thesis will be mentioned here.

The content validity is an instrument's ability to capture a concept of interest by the items in the questionnaire (94). Thirteen disease or domain specific validated PRO instruments have been identified for use in MM patients (95). An investigation of the content validities of PRO instruments used in patients with MM, concluded that one single instrument does not capture all important HRQoL concepts (95). The EORTC QLQ-C30 with the addition of the Multiple Myeloma module EORTC QLQ-MY20 has been found to have the best coverage of domains of interest to patients with MM and is the most comprehensively psychometrically validated instrument for patients with MM (92, 95-98). This is supported by a Delphi consensus project, where the domains of EORTC QLQ-C30 reached the panellists' agreement for global standards for collecting PRO outcomes in NDMM with the addition of self-perception of body image, sexuality and pain (99).

Internal consistency refers to the extent to which the items of a multi-item domains are inter-related and is estimated by Cronbach´s α formula:

$$\alpha = N \cdot c / v + (N - 1) \cdot c$$
N; number of items, *c*; average covariance between item-pairs, *v;* average variance.

A Cronbach α level ≥0.7 is generally regarded as acceptable reliability for psychometrics scales, a level ≥0.8 is good, and ≥0.9 is excellent (68, 94).

EORTC QLQ-C30
The EORTC QLQ-C30 instrument consists of 30 items and 15 domains: one global QoL domain, five functional domains (physical, role, emotional, cognitive and social functioning), nine symptom domains (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea, and financial difficulties). The domain scores are calculated according to the EORTC manual and translated into a scale of 0-100 (100). A high score represents good HRQoL for functional domains, and a low score presents a low degree of symptoms for symptom domains. The EORTC QLQ-C30 has been validated in patients with MM by Wisloff *et al*, 1996 (98). The instrument has been found to be reliable for all multi-item scales, except for role functioning (Cronbach α of 0.54 - 0.59) and the mean scores for nausea and vomiting, dyspnoea, insomnia, appetite loss, constipation and diarrhoea were found to be substantially low. In terms of validity, a strong association was found between World Health Organization performance status and physical and role functioning, fatigue and pain. In addition, the instrument was found to have good responsiveness and be sensitive to changes in the patients´ clinical status over time (98).

EORTC QLQ-MY20
EORTC QLQ-MY20 consists of 20 items and four domains: two functional domains (future perspective and body image) and two symptom domains (disease symptoms and side effects of treatment) (96, 97). The module is administrated in addition to the EORTC QLQ-C30 questionnaire and all domains, except for body image, are multi-item domains. A high score represents good HRQoL for functional domains, and a low score represents a low degree of symptoms for symptom domains. The questionnaire is validated in

patients with MM and found with acceptable reliability for all three multi-item domains (Cronbach α of 0.7 - 0.82).

Other PRO instruments validated for MM patients is the EORTC Quality of Life Questionnaire high-dose chemotherapy and Functional Assessment of Cancer Therapy anaemia questionnaire (95). A detailed description of PRO instruments validated for MM patients can be found elsewhere (95, 101, 102).

Limitations in existing literature
Evidence-based knowledge of HRQoL during and after the variety of treatment regimens available for MM is an important part of clinical decision-making of myeloma patients. Therefore, longitudinal HRQoL evaluation has become an increasingly used endpoint in clinical trials with MM patients to assess risks and benefits of anti-myeloma therapies. A systematic review of published longitudinal HRQoL studies of patients with MM and an analysis of results for use in clinical decision-making has not been performed yet.

Measurement and interpretation of longitudinal HRQoL data
Measurement and interpretation aspects of longitudinal PRO data, relevant for the thesis will be presented here, including minimal important difference, missing PRO data and response shift.

Minimal important difference
For interpretation of PRO data results from clinical trials, a statistically significant change in HRQoL score is not necessarily important or even detectable to the patient and therefore not suitable for clinical decision-making. A minimal important difference (MID) is considered more relevant (103-105). Thirty years ago, in 1989, Jaeschke *et al.* defined minimal clinically important difference (106).

"The minimal clinically important difference can be defined as the smallest difference in score in the domain of interest which patients perceived as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patients´ management".

Several terminologies and definitions of MID have been used since clinically important difference was defined for the first time by Guyatt *et al.* in 1987 (104, 107). MID thresholds can be used to interpret clinically relevant difference *between* groups and clinically meaningful change *within* a group.

Clinically meaningful change
Approaches to determine MID can be categorized as anchor based, distribution based, a combination of anchor and distribution based (103, 108). For distribution based MIDs, the statistical distribution of the results obtained from a given cohort of patients is used, termed *minimal detectable change*, whereas in anchor based MIDs, an external criterion or "anchor" is used to compare the outcomes measured, termed *minimal important change* (104, 109). Since distribution based MIDs do not consider the patients´ or clinicians´ perspective, they have been recommended to be used as supportive evidence to anchor based MID (109, 110). Here, the thresholds for clinically meaningful change *within* group, which we use in this thesis, will be presented.

Distribution based MIDs

Cohen´s effect size for interpretation of magnitude of change in EORTC QLQ-C30 domain scores is based on the standard deviation (SD) of the mean score at baseline for the studied group (effect size x SD (baseline)) (111). Cohen suggested that an effect size of 0.2-0.5 represents a small group change, effect size of 0.5-0.8 represents a medium change, and an effect size of > 0.8 represents a large change (111). The validity of the medium effect size thresholds has been studied for different instruments, including for EORTC QLQ-C30 and has been found to be a suitable MID threshold for HRQoL in most circumstances, named *Norman´s rule of thumb* (112).

Standard error of measurement is adapted from the psychometric property of a PRO scale reliability (*r*). The SEM estimate of SEM=$SD\sqrt{1-r}$ can been used to calculate the MIDs for clinically meaningful change within groups (113, 114). The *r* is also a value for internal consistency and is estimated by Cronbach´s α formula. Cronbach α is depended on the number of items in the multi-item domain and the sample size. Therefore, the SEM based MIDs can only be calculated for multi-item domains (68, 94).

Anchor based MIDs
Kvam *et al.* used 2010 an anchor based method to define clinically meaningful score changes for four domains of the EORTC QLQ-C30 instrument in patients with MM: physical function, global QoL, fatigue and pain (115). Estimation of the MIDs was based on scores from 239 patients in different stages of MM. The patients in the study completed the questionnaire at baseline (T1) and after 3 months (T2). MIDs were assessed using three different approaches: An anchor based method with the addition of a global rating of change (GRC) question on patients´ perception of their change in QoL (the anchor MIDs), a method using receiver-operating characteristics, and a method based solely on the effect size of answers at baseline. The single GRC question used was whether the patient had noted improvement, no change or deterioration in QoL at T2. The MIDs were calculated for each of the four domains by applying the mean score changes from T1 to T2 among patients reporting improved or deteriorated QoL. The results suggested that the size of MID in MM patients may be affected by the direction of change and the domain. In conclusion, a general effect size of 0.3-0.5, which relates to a MID of 6-17 points, was found to be appropriate for the four domains investigated. In this thesis, these thresholds will be referred to as *Kvam's MID criteria*.

Cocks *et al.* established in 2012 guidelines to interpret changes in scores for the domains of the EORTC QLQ-C30 instrument. They estimated MIDs for the domains based on a literature review and data from patients with different cancer types (116). The authors used a meta-analytic technique combined with blinded expert opinions. The expert panel review consisted of health care professionals with experience in treatment of patients with cancer and the use of EORTC QLQ-C30 instrument (116, 117). They were asked to make a judgement on the relative size of the change over time, according to four levels of magnitude: large, medium, small and trivial deterioration or improvement. A large change was "an obvious and unequivocally clinically meaningful change", and medium changes was "likely to be clinically meaningful but to a lesser extent". Small changes indicated "a subtle, but nevertheless clinically meaningful change", and a trivial change was "unlikely to be clinically meaningful, or there was no difference". In this thesis, these thresholds will be referred to as *guideline of Cocks*.

Limitations of the existing literature
Anchor based and distribution based MID thresholds are available for the EORTC QLQ-C30 questionnaire, and only distribution based MID for the EORTC QLQ-MY20 instrument. Four anchor based MIDs established

by the GRC method of the MM patients´ perception of clinically meaningful change are available in the *Kvam´s MID criteria*. Limitations of the GRC method have been stated as; implicit theory of change, response shift, recall bias and that it is based on an invalidated single-item question (104, 118-123). Therefore, it is unknown, whether *Kvam´s MID criteria* or other MIDs are suitable for determination of clinical meaningful treatment effect on HRQoL in all patients with MM.

Missing PRO data

Missing data in clinical trials is an ongoing challenge in clinical research, and stated as one of the reasons for researchers not being able to draw definitive conclusions from the results (124-126). Missing PRO data can lead to a variety of problems, such as loss of study power and precision (127, 128). If the reason for missing PRO data is related to the patient's poor health status and not handled appropriately, missing PRO data may lead to biased results (129-132). Strategies to reduce missing PRO data should be integrated into the study design, protocol and data collection procedures as well as handled by appropriate statistical methods (127-129, 133-135).

Missing PRO data can be divided by causes and patterns; missing items, partial responses, complete non-responses (NR), intermittent NR and monotone NR (126). Partial responses are a partially completed questionnaire and a missing item is one missing answer to one question in a questionnaire. Missing items can be handled by the "Half-scale rule", if at least half of the items from a multi-item domain have been answered (100, 136). NR is a fully missing scheduled questionnaire and can be subdivided into; monotone (terminal), intermittent or mixed pattern. Monotone NR occurs when the scheduled questionnaires are completed until at a time, for example, when the patient drops out. Intermittent NR is when one or more NR are seen between completed questionnaires, and a mixed pattern is a combination of intermittent pattern until monotone NR occurs (137).

Three different missing data mechanisms have been described (138): "*Missing completely at random*" (MCAR) is, for example, if staff forgets to give the questionnaire to the patient. "*Missing at random*" (MAR) is, for example, if a specific subgroup of patients with similar outcomes e.g. poorer PRO scores has a higher proportion of NR. "*Missing not at random*" (MNAR) is, for example, if the PRO assessments are likely to be missed when patients are experiencing adverse events or complications, and are therefore termed "non-ignorable" or "informative" (128, 135).

Limitations of existing literature

The magnitude of NR in longitudinal HRQoL studies of patients with MM is unexplored. Therefore, the extent of this challenge in existing evidence-based knowledge of HRQoL in MM is unknown. Also, it has not previously been investigated which practical tools that are able to reduce NR to scheduled questionnaires in patients with MM.

Response shift

Deviated from the "Conception Models of QoL" stating that QoL is not a stable concept over time (71, 72, 139), Sprangers and Schwartz in 1999 described the phenomenon, "response shift"(140). The working definition refers to "*a change in the meaning of one´s self-evaluation of a target construct*". Response shift is a result of 1) a change in internal standards (recalibration), 2) a change in values (i.e. the importance of component domains constituting the target construct) or 3) a redefinition of the target construct

(reconceptualization) (120, 140). As a part of this model, response shift effect is initiated by a "catalyst", which is a relevant change in life or health e.g. being diagnosed with cancer (120).

In the concept of response shift it is explained why QoL might mean different things at different time points to the same person, since an accommodation and adaption to a change in life that might occur. Response shift has been recognized in the trajectory of several chronic illnesses including cancer and MM (141-147). Longitudinal PRO results might be incorrectly interpreted if response shift is not considered as implicated in the patients´ change in score over time (122, 148-152).

To be able to understand the phenomenon, adequate descriptions of inter- and intra-individual differences in QoL appraisal have been found to be fundamental (153). With linkage to this dynamic aspect of response shift, Rapkin and Schwartz in 2004 developed "The Theoretical model of QoL Appraisal" (153). The model includes four key parameters of appraisal;

- "*Frame of reference*" referring to the experiences individuals deem relevant to their response.
- *"Sample of experiences"* referring to inclusion of a sample of specific experiences within their frame of reference relevant for the response.
- *"Standards of comparison"* referring to the chosen sample of experiences to compare against when giving a response.
- *"Combinatory algorithm"* referring to the process summarizing the evaluation of relevant experiences and formulates a response.

Methods to detect response shift have been developed and include prospective interview methods or secondary statistical methods (121, 122, 154, 155). Barriers for not integrating response shift measurement into clinical trials are that existing methods are cumbersome to administer, score, analyse and with limitations (121-123, 156). Recently, a practical, low resource-intensive version of the QoL Appraisal Profile of 23 close-ended items to assess QoL appraisal has been developed, called the "Brief Appraisal Inventory" (BAI) (157-159). The use of BAI in future longitudinal HRQoL research will determine if BAI can detect response shifts over time (159).

Limitations of existing literature
Response shift in patients with MM were investigated as part of the same study where the *Kvam´s MID criteri*a were estimated (160). A clear indication of response shift existence was seen with impact the size of MID thresholds, especially for MID thresholds for deterioration. However, at present response shift adjusted MID thresholds are limited by the used method of the GRC method, which may also be more strongly associated with the appraisal process, than by a change in HRQoL (161). QoL appraisal processes in MM are unexplored, and a possible linkage between MID and QoL appraisal should be investigated further before integration into thresholds to determine clinically meaningful treatment effects on HRQoL.

Hypothesis of the thesis

The overall hypothesis was that PROs are valuable tools for assessing clinically meaningful treatment effects on HRQoL from the MM patients´ perspective in spite of methodological challenges in longitudinal PRO data measurement, analyses and interpretation.

Aims of the thesis

Part 1 and 2

The overall aim of part 1 and 2 was to determine if the EORTC QLQ-C30 and EORTC QLQ-MY20 questionnaires do capture clinically meaningful treatment effects on HRQoL from the MM patients´ perspective during first-line and relapse anti-myeloma regimens.

Part 1. Systematic review of longitudinal HRQoL studies in multiple myeloma patients (Paper I)

- Through previously published longitudinal HRQoL studies in MM to determine clinically meaningful treatment effects on global QoL, physical functioning, fatigue and pain during first line and relapse anti-myeloma regimens. (Study I)

Part 2. Clinical trials with multiple myeloma patients during first line therapies (Paper II and III)

- To determine clinically meaningful treatment effects on HRQoL during four first line treatment regimens from two clinical trials (Study II and III) and relate them to findings of study I.

Part 3

The aim of part 3 was to investigate the magnitude of NR and to propose and evaluate tools to minimise intermittent NR in longitudinal PRO studies of patients with MM.

Part 3. Non-responses in longitudinal PRO studies of multiple myeloma patients (Paper IV and V)

- To analyse the magnitude of intermittent and monotone NR in the studies identified in study I, II and III (Study IV).
- To determine the effect of implemented strategies in the study design, conduct and procedures to reduce intermittent NRs (Study V).

Part 1. Systematic review of longitudinal HRQoL studies in multiple myeloma patients (Study I)

Methods

A literature search was performed in May, 2016 to identify all previously published longitudinal HRQoL studies of patients with MM. A summary of the methods and the main findings is provided in this section of the thesis. Further description and full results can be found in Paper I.

A systematic search was performed in PubMed, Embase, PsycINFO and CINAHL, using the following search terms: *Multiple Myeloma* or *Myelomatosis* and *Quality of life* or *Life quality*. Qualified publications were studies of patients diagnosed with MM, using a longitudinal design and the EORTC QLQ-C30 instrument and which presented data for at least one of the following domains; physical function, global QoL, fatigue or pain. A prerequisite for inclusion was that baseline data and at least one follow-up evaluation were presented. The included studies were divided into first line and relapse treatment studies and others. For the four HRQoL domains (physical functioning, global health status, fatigue and/or pain), the mean change from baseline was calculated for every follow-up time point by subtracting the mean score at follow-up time point from the mean score at baseline. The size of every calculated mean change from baseline for each domain was interpreted according to *Kvam's MID criteria* (115).

Results

Twenty-three publications were eligible for the systematic review, corresponding to 11 first line treatment studies and seven relapse treatment studies and five others. The 18 first line and relapse studies are presented in table 1. An extended version with all identified publications can be found in paper I, with presentation of number of patients at baseline, mean age and mean baseline score for global QoL, physical functioning, fatigue and pain, number of follow-ups and time of last HRQoL assessment time point. The patient reported range of mean change of score from baseline for global QoL and physical functioning, fatigue and pain are presented in figure 2 for the 11 first line treatment and seven relapse studies included in the systematic review.

First-line treatment studies

The patients report clinically meaningful improvement in global QoL, physical functioning and pain reduction, except for the group of patients receiving melphalan-prednisolone-placebo and placebo maintenance, who did not report clinically meaningful pain reduction (162). Only some treatment regimens led to reporting of clinically meaningful improvement in fatigue, which was the case for patients treated with induction and HDT, the historical control group, pamidronate 30 or 90 mg and melphalan-prednisolone with or without the addition of interferon α2 or thalidomide (163-167).

Relapse treatment studies

Most relapse regimens led to unchanged global QoL, physical functioning, fatigue and pain. An exception from this was for pain during thalidomide monotherapy, which lead to reporting of clinically meaningful improvement. Clinically meaningful deterioration was reported during four regimens, which were in patients receiving dexamethasone monotherapy, bortezomib both with and without dexamethasone, who

reported clinically meaningful deterioration in global QoL, and the patients receiving high dose dexamethasone, who reported deterioration in fatigue (88, 168, 169).

Table 1. Eighteen of the 23 studies included in the systematic review were either primary or relapse studies. The design of the 18 studies was five phase II trials, 12 phase III trials and one evaluation study. Thirteen studies were randomized clinical trials of which four were double-blinded and three were placebo-controlled.

Author and year of publication	Study design	Treatment regime
First line treatment studies		
Delforge et al. 2015 (170) FIRST trial	Randomized phase III study	Lenalidomide-dexamethasone vs. Melphalan-prednisolone-thalidomide
Dimopoulos et al. 2013 (162) MM-015 study	Three-armed randomized double-blind placebo-controlled phase III study	Melphalan-prednisolone-lenalidomide and lenalidomide maintenance vs. Melphalan-prednisolone-lenalidomide and placebo maintenance vs. Melphalan-prednisolone-placebo and placebo maintenance
Ludwig et al. 2013 (171)	Randomized open-label phase II study	Bortezomib-thalidomide-dexamethasone + HDT vs. Bortezomib-thalidomide-dexamethasone-cyclophosphamide + HDT
Etto et al. 2011 (166)	Phase II study[1]	Induction therapy and HDT
Delforge et al. 2012 (172) VISTA trial	Randomized phase III study	Bortezomib-melphalan-prednisolone vs. Melphalan-prednisolone
Verelst et al. 2011 (173) HOVON 49	Randomized phase III study	Melphalan-prednisolone-thalidomide and thalidomide maintenance[2] vs. melphalan-prednisolone
Gimsing et al. 2010 (174)	Randomized double-blind phase III trial	Pamidronate 30 mg vs. pamidronate 90 mg
Waage et al. 2010 (163)	Randomized double-blind placebo-controlled phase III study	Melphalan-prednisolone-thalidomide and thalidomide maintenance[2] vs. melphalan-prednisolone placebo and placebo maintenance[2]
Gulbrandsen et al. 2001 (167)	Evaluation, phase II trial	VAD induction therapy and HDT vs. Historical control group
Wisloff et al. 1996 (175) NMSG 4/90	Evaluation study	Evaluation group
Wisloff et al. 1996 (176) NMSG 4/90	Randomized phase III study	Melphalan-prednisolone vs. Melphalan-prednisolone-Interferon-α2 and Interferon-α2 maintenance[2]
Relapse treatment studies		
Moreau et al. 2016 (58) TOURMALINE-MM1	Randomized, double-blind, placebo-controlled, phase III trial	Ixazomib-lenalidomide-dexamethasone vs. Placebo-lenalidomide-dexamethasone
Stewart et al. 2015 (59) ASPIRE trial	Randomized phase III study	Carfilzomib-lenalidomide-dexamethasone vs. Lenalidomide-dexamethasone
Song et al. 2015 (169) MM-003	Randomized phase III study	Pomalidomid-low dose dexamethasone vs. High dose-dexamethasone
Hjorth et al. 2012 (56) NMSG 17/07	Randomized phase III study[3]	Thalidomide-dexamethasone vs. Bortezomib-dexamethasone
Lee et al. 2008 (168) APEX study	Randomized phase III study	Bortezomib vs. Dexamethasone monotherapy
Dubois et al. 2006 (88) SUMMIT study	Open-label phase II study	Bortezomib
Waage et al. 2004 (55)	Phase II study	Thalidomide

HDT; High-dose therapy with autologous stem cell support, VAD; vincristine-doxorubicin-dexamethasone. [1]The patients at follow-up are not all of them the same as at diagnosis. [2]The patients continued maintenance/placebo or observation after plateau phase of the M component. [3]Cross-over at treatment failure.
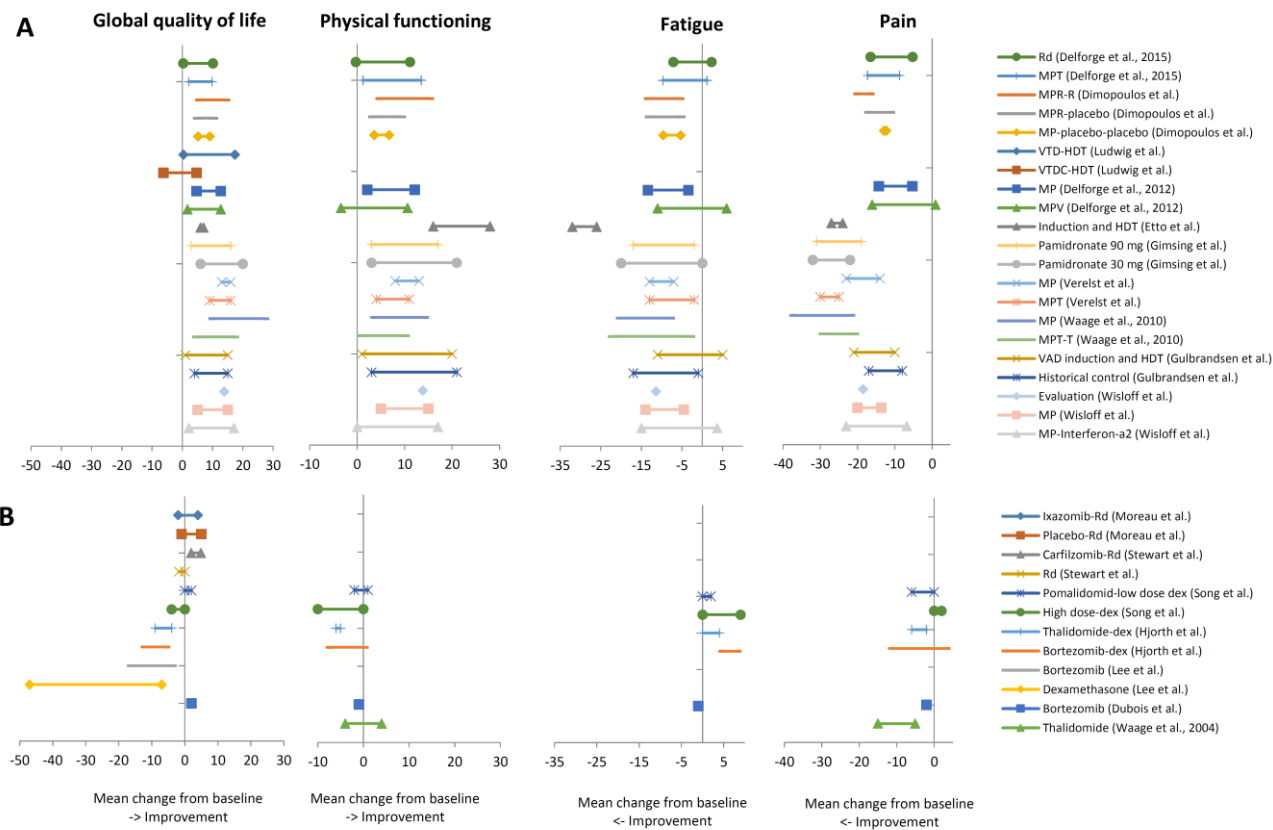
Figure 2. Patient-reported range of mean change of score from baseline for global quality of life, physical functioning, fatigue and pain. **A.** Represents the 11 included first-line treatment studies. **B.** Represents the seven relapse treatment studies.

Rd; lenalidomide-dexamethasone, MPT; melphalan-prednisolone-thalidomide, MPR-R; melphalan-prednisolone-lenalidomide followed by lenalidomide maintenance, MPR-placebo; MPR-placebo; melphalan-prednisolone-lenalidomide followed by placebo maintenance, MP-placebo-placebo; melphalan-prednisolone-placebo followed by placebo maintenance, VTD; bortezomib-thalidomide-dexamethasone, VTDC; bortezomib-thalidomide-dexamethasone-cyclophosphamide, HDT; high dose therapy with stem cell support; MP; melphalan-prednisolone, MPV; melphalan-bortezomib-prednisolone, VAD; vincristine-doxorubicin-dexamethasone, MPT-T; melphalan-prednisolone-thalidomide followed by thalidomide maintenance

Part 2. Clinical trials with multiple myeloma patients during first line therapies (Study II and III)

Methods

The aims of part 2 were to determine clinically meaningful treatment effects on HRQoL of four regimens from two clinical trials and relate the results to findings of the studies identified in study I.

In brief, study II, the HOVON87/NMSG18 study, was a multicentre, randomized phase III study with inclusion of HDT non-eligible NDMM patients with symptomatic disease. The patients were randomized between nine cycles of melphalan-prednisolone-thalidomide followed by thalidomide maintenance (MPT-T) or nine cycles of melphalan-prednisolone-lenalidomide followed by lenalidomide maintenance (MPR-R) (177). A detailed description of the HOVON87/NMSG18 HRQoL study and statistical analysis method can be found in paper II.

Study III, the CLAIM study, was a Danish randomized double-blind, placebo-controlled phase II study with inclusion of HDT eligible NDMM (178). The patients were randomized between addition of clarithromycin 500 mg p.o. or placebo twice daily for 63 days during bortezomib-cyclophosphamide-dexamethason (VCD) induction treatment with subsequent HDT (179). A detailed description of the CLAIM HRQoL study and statistical methods can be found in paper III.

HRQoL was a secondary explorative endpoint in study II and III protocols. The patients in both studies completed the EORTC QLQ-C30 with addition of the EORTC QLQ-MY20 questionnaire at baseline and prescheduled follow-up time points.

Interpretation of clinically meaningful treatment effects on HRQoL

Clinically meaningful treatment effects for global QoL, physical functioning, fatigue and pain were determined according to *Kvam's MID criteria*, since they are MM patient-derived ratings of change (115). For the remaining 11 EORTC QLQ-C30 domains without threshold defined by *Kvam's MID criteria*, the medium MID threshold defined by *guidelines of Cocks* were used, since the thresholds rely on cancer clinicians´ perspective of change. (116). The reason for choosing medium thresholds is when comparing the thresholds for a small change defined by *guidelines of Cocks* towards thresholds of *Kvam's MID criteria*, a small change was evaluated as no clinically meaningful change to patients with MM (180). There is no established anchor based MID thresholds for the domains of EORTC QLQ-MY20, therefore the three multi-item domains of EORTC QLQ-MY20 were interpreted according to SEM based MID thresholds (113, 114). For the single-item domain of body image, an effect size of 0.5 was used, which is adapted from the general finding of *Kvam's MID criteria* of a moderate effect size being clinically meaningful to patients with MM and *Normans´ rule of thumb* (112).

Kvam's MID criteria, *guidelines of Cocks* and the distribution based MID described above were used to relate the results from study II and III to the results for the studies identified in study I. Domains captured by EORTC QLQ-MY20 domains were published in two studies of Delforge *et al.* (170) and Dimopoulos *et al.* (162). The SEM based MID thresholds were calculated in the study of Dimopoulos *et al.* for disease symptoms (MID= -10) and side effects of treatment (MID= -6) and used in both. The medium MID threshold

of *guideline of Cocks* could not be evaluated for financial difficulties, and therefore the findings of this domain were not related to the findings from study I.

Results

The mean changes of score from baseline for the four first line regimens investigated in study II and III are presented in table 2. In Figure 3 A-E, the published domains of EORTC QLQ-C30 and EORTC QLQ-MY20 from the first line regimens of study I, II and III are gathered. Below is a description of differences and similarities in clinically meaningful treatment effects on HRQoL from study II and III compared to the findings of from study I. The distribution based MIDs used for the EORTC QLQ-MY20 domains of study II and III are presented in Table 3.

Global QoL, physical functioning, fatigue and pain

In global QoL, the patients in the MPT-T and MPR-R group reported clinically meaningful improvement, which is similar to the general reporting during the first line regimens in study I. However, this is in contrast to the patients treated with placebo or clarithromycin added to VCD induction and HDT, who reported unchanged global QoL, which was also reported during only one regimen in study I; the bortezomib-thalidomide-dexamethasone-cyclophosphamide induction therapy followed by HDT.

In physical functioning, the patients receiving clarithromycin added to VCD and HDT reported unchanged physical functioning, which is in contrast to the other patients treated with in first line. The patients in the MPT-T and MPR-R and the group of patients treated with placebo added to VCD induction and HDT reported clinically meaningful improvement, similar to the general reporting during the regimens of study I.

The patients treated with clarithromycin or placebo added to the VCD induction and HDT reported clinically meaningful increased fatigue, in contrast to the patients treated by other first line regimen, where either unchanged or reduced fatigue was reported. Clinically meaningful reduction in fatigue was reported by the patients treated with MPT-T. The patients from study II and III reported clinically meaningful reduction in pain in line with general findings in study I.

The remaining 11 EORTC QLQ-C30 domains

The patients treated with the drug combinations, investigated in study III, reported unchanged role functioning, which is similar to the patients treated with MPV as part of the VISTA study (172) and MPT-T as part of the study of Waage et al. (163). However, this is in contrast to the general reporting of clinically meaningful improvement in role functioning and also in contrast to the reportings of improved role functioning during the regimens investigated in study II. Of the patients treated in study II and III, only the patients treated with placebo added to VCD induction and HDT reported unchanged emotional functioning, which is similar to four out of 12 regimens in study I. The patients in study II and III reported unchanged cognitive functioning, which was similar to eight out of 12 regimens in study I. The patients treated with clarithromycin added to VCD induction and HDT reported decreased social functioning, which was in contrast to the patients treated with placebo added to VCD induction and HDT, where unchanged social functioning was reported. Clinically meaningful improvement in social functioning was the general finding in study I and also in study III.

Table 2. Mean change of scores from baseline and 95% confidence intervals for the domains of EORTC QLQ-C30 and EORTC QLQ-MY20 reported by the patients in study II and III. Green figures refers to a clinically meaningful, positive treatment effect and red figures refers to a clinically meaningful, negative treatment effect on HRQoL.

HRQoL domains	MPT-T				MPR-R				Clarithromycin added to VCD induction		Placebo added to VCD induction	
	After 3 MPT cycles	After 9 MPT cycles	After 6 months T	After 12 months T	After 3 MPR cycles	After 9 MPR cycles	After 6 months R	After 12 months R	Before priming	Two months after HDT	Before priming	Two months after HDT
EORTC QLQ-C30												
Global quality of life	5.6 (2.5;8.7)	7.9 (4.6;11.2)	13.0 (8.2;17.7)	13.4 (7.3;19.6)	4.1 (1.2;7.1)	8.5 (5.3;11.7)	14.8 (11.2;18.4)	14.8 (10.7;19.0)	-10.1 (-22.2;2.0)	-0.5 (-14.2;13.2)	6.1 (-3.9;16.0)	4.4 (-8.0;16.8)
Physical functioning	3.2 (0.31;6.0)	6.5 (3.5;9.6)	8.9 (4.6;13.3)	6.5 (0.9;12.2)	1.8 (-0.56;4.2)	6.1 (3.5;8.8)	9.4 (6.5;12.4)	12.4 (9.0;15.8)	-0.8 (-9.4;7.8)	-1.0 (-11.3;9.2)	7.0 (0.0;14.0)	5.1 (-4.1;14.2)
Role functioning	3.4 (-1.1;7.9)	10.7 (5.9;15.5)	14.7 (7.9;21.5)	12.4 (3.5;21.3)	3.1 (-0.81;7.0)	7.7 (3.5;12.0)	17.5 (12.6;22.3)	19.5 (13.9;25.0)	-10.1 (-24.7;4.6)	-6.4 (-21.3;8.5)	4.4 (-7.3;16.1)	9.1 (-4.1;-22.3)
Emotional functioning	3.1 (0.34;5.8)	7.6 (4.6;10.6)	10.6 (6.4;14.8)	9.1 (3.7;14.6)	7.7 (5.2;10.3)	7.8 (5.1;10.6)	13.1 (10.0;16.3)	12.4 (8.8;15.9)	-3.4 (-12.7;6.0)	13.0 (4.6;21.6)	3.9 (-3.5;11.4)	8.2 (0.7;15.7)
Cognitive functioning	-2.4 (-5.3;0.53)	-1.7 (-4.8;1.4)	0.9 (-3.5;5.3)	-3.2 (-9.0;2.6)	4.0 (1.4;6.6)	2.8 (0.0;5.6)	5.7 (2.5;8.9)	4.1 (0.4;7.7)	-2.2 (-12.9;8.6)	-0.1 (-12.6;12.3)	-0.1 (-8.7;8.4)	-4.3 (-15.2;6.7)
Social functioning	-0.8 (-4.7;3.1)	3.7 (-0.56;7.9)	12.9 (6.9;18.8)	14.6 (6.8;22.4)	0.1 (-0.34;3.6)	5.6 (1.8;9.4)	11.4 (7.1;15.7)	12.2 (7.2;17.1)	-14.3 (-25.7;-3.0)	-9.4 (-22.7;4.0)	-0.8 (-9.8;8.2)	-2.9 (-14.7;9.0)
Fatigue	-4.4 (-7.8;-1.0)	-7.7 (-11.4;-4.1)	-12.6 (-17.7;-7.4)	-13.8 (-20.5;-7.1)	0.6 (-2.6;3.8)	-2.1 (-5.6;1.4)	-8.9 (-12.8;-4.9)	-11.3 (-15.8;-6.7)	16.0 (3.5;28.4)	10.5 (-2.0;23.0)	9.4 (-1.1;19.8)	8.7 (-2.5;19.8)
Nausea and vomiting	-2.1 (-4.5;0.32)	-5.3 (-7.9;-2.8)	-5.9 (-9.5;-2.2)	-4.6 (-9.3;0.2)	-0.7 (-2.8;1.5)	-3.5 (-5.9;-1.2)	-4.6 (-7.2;-1.9)	-4.1 (-7.1;-1.0)	-5.4 (-12.1;1.4)	3.1 (-7.1;13.3)	-0.6 (-6.3;5.2)	1.1 (-8.0;10.1)
Pain	-22.7 (-26.8;-18.7)	-21.1 (-25.4;-16.7)	-23.0 (-29.1;-16.9)	-23.7 (-31.7;-15.7)	-13.2 (-16.8;-9.6)	-19.0 (-22.9;-15.1)	-23.9 (-28.4;-19.5)	-24.5 (-29.6;-19.4)	-10.8 (-25.2;3.7)	-19.7 (-34.7;-4.7)	-13.0 (-24.8;-1.2)	-21.5 (-35.0;-7.9)
Dyspnoea	0.22 (-3.9;4.3)	2.5 (-1.9;6.9)	-5.4 (-11.7;0.8)	-2.9 (-11.1;5.2)	4.5 (0.9;8.0)	3.0 (-0.8;6.9)	-4.7 (-9.1;0.3)	-3.8 (-8.9;1.2)	9.8 (-5.0;24.5)	5.0 (-9.8;19.6)	8.0 (-4.1;20.0)	9.1 (-4.2;22.2)
Insomnia	-16.0 (-19.9;-12.1)	-11.3 (-15.4;-7.1)	-14.2 (-20.1;-8.3)	-16.5 (-24.2;-8.8)	-6.0 (-9.4;-2.6)	-7.8 (-11.5;-4.1)	-11.1 (-15.2;-6.9)	-9.5 (-14.3;-4.7)	15.3 (-2.6;32.7)	-2.4 (-18.2;13.4)	4.7 (-9.2;18.6)	-15.7 (-29.8;-1.6)
Appetite loss	-5.2 (-9.5;-0.88)	-7.4 (-12.0;-2.7)	-17.0 (-23.6;-10.5)	-17.3 (-25.8;-8.8)	1.0 (-3.1;5.2)	-3.5 (-8.1;1.0)	-13.4 (-18.5;-8.3)	-10.0 (-15.9;-4.1)	6.3 (-5.9;18.4)	18.3 (-0.3;36.8)	2.6 (-6.9;12.1)	10.4 (-6.1;26.8)
Constipation	9.1 (4.6;13.6)	4.6 (-0.27;9.4)	3.8 (-3.0;10.6)	3.3 (-5.6;12.1)	0.3 (-3.7;4.3)	-3.4 (-7.8;1.0)	-8.5 (-13.5;-3.6)	-6.9 (-12.5;-1.2)	-2.9* (-20.6;14.8)	-25.0* (-36.5;-13.5)	2.7* (-11.7;17.1)	-15.6* (-26.3;-4.9)
Diarrhoea	-4.2 (-7.2;-1.3)	-5.0 (-8.2;-1.7)	-5.0 (-9.6;-0.5)	-3.2 (-9.1;2.7)	3.6 (0.0;7.2)	7.6 (3.7;11.5)	7.4 (3.0;11.7)	11.1 (6.0;16.1)	2.3 (-10.7;15.3)	18.5* (3.8;33.2)	8.9 (-1.2;19.0)	13.9* (0.9;26.9)
Financial difficulties	-0.0 (-2.2;2.1)	1.1 (-1.2;3.4)	0.30 (-3.0;3.6)	-2.6 (-6.8;1.7)	0.9 (-1.0;2.7)	1.0 (-1.1;3.0)	1.2 (-1.1;3.5)	0 (-2.7;2.6)	-1.4 (-8.9;6.2)	-5.4 (-11.5;0.8)	-3.4 (-9.3;2.6)	-0.1 (-5.4;5.3)
EORTC QLQ-MY20												
Disease Symptoms	-11.1 (-13.6;-8.6)	-8.9 (-11.6;-6.2)	-9.1 (-12.9;-5.4)	-9.1 (-14.0;-4.1)	-8.0 (-10.4;-5.6)	-8.7 (-11.3;-6.0)	-9.6 (-12.6;-6.7)	-10.0 (-13.4;-6.6)	-7.4 (-15.5;0.7)	-18.3 (-28.9;-7.6)	-14.5 (-21.2;-7.8)	-14.4 (-24.1;-4.7)
Side effects of treatment	4.5 (2.6;6.4)	3.8 (1.8;5.9)	1.8 (-1.1;4.6)	2.7 (-1.0;6.4)	1.1 (-0.6;2.8)	-0.6 (-2.4;1.3)	-3.7 (-5.8;-1.6)	-2.5 (-4.9;-0.1)	11.8 (5.5;18.1)	15.9 (7.9;23.8)	6.7 (1.7;11.6)	14.3 (7.1;21.4)
Future Perspective	8.1 (5.3;10.8)	13.1 (10.1;16.0)	15.8 (11.6;20.1)	13.6 (8.1;19.1)	11.1 (8.5;13.8)	14.0 (11.1;16.9)	19.3 (16.0;22.5)	19.1 (15.4;22.9)	15.9 (4.1;27.7)	18.2 (4.7;31.8)	16.5 (6.6;26.3)	10.6 (-1.4;22.6)
Body Image	1.2 (-2.6;5.1)	-2.4 (-6.6;1.7)	-1.4 (-7.3;4.5)	1.6 (-6.0;9.3)	-1.5 (-4.6;1.7)	1.3 (-2.2;4.8)	3.5 (-0.4;7.5)	2.5 (-2.0;7.1)	-30.5 (-45.3;-15.7)	-18.9 (-34.9;-3.0)	-4.3 (-15.9;7.3)	-16.9 (-31.0;-2.8)

*MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy, T; thalidomide maintenance, R; lenalidomide maintenance, VCD; bortezomib-cyclophosphamide-dexamethasone, HDT; high dose therapy with stem cell support, *Statistically significant time effect, further details can be found in paper III.*

Table 3. Calculated minimal important difference thresholds for clinically meaningful treatment effect for the four EORTC QLQ-MY20 domains of study II and III.
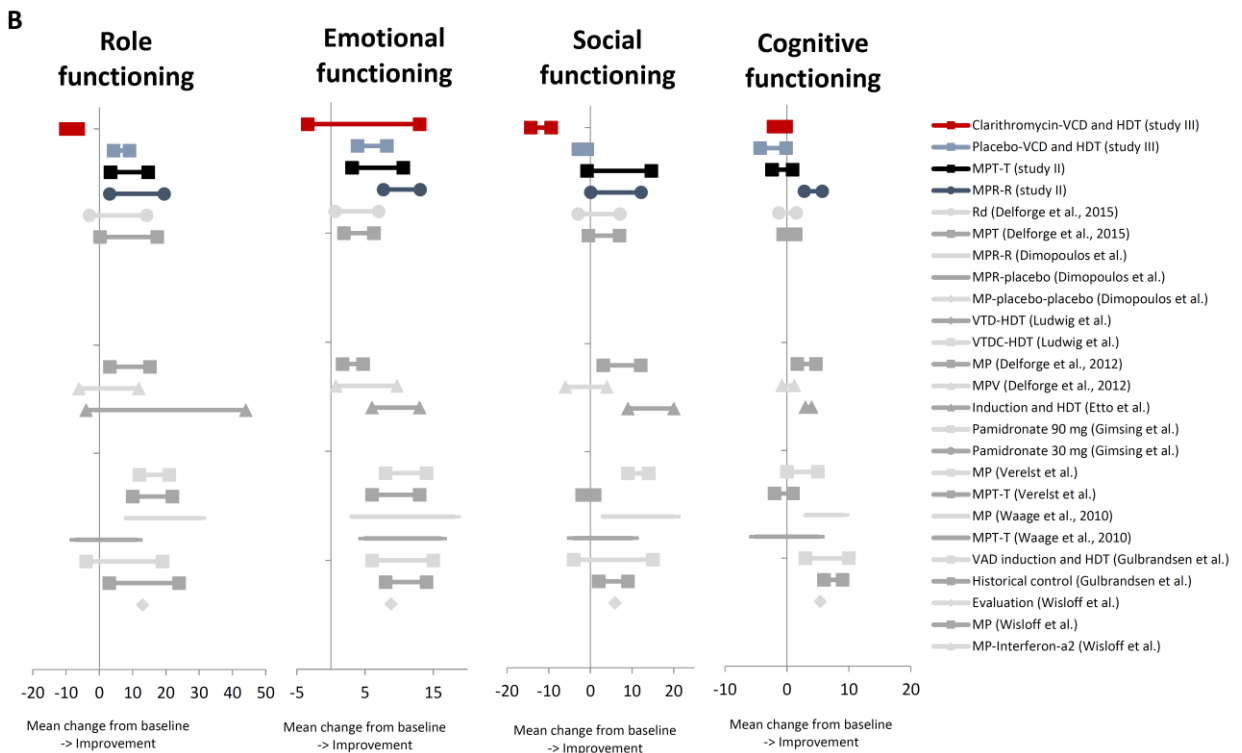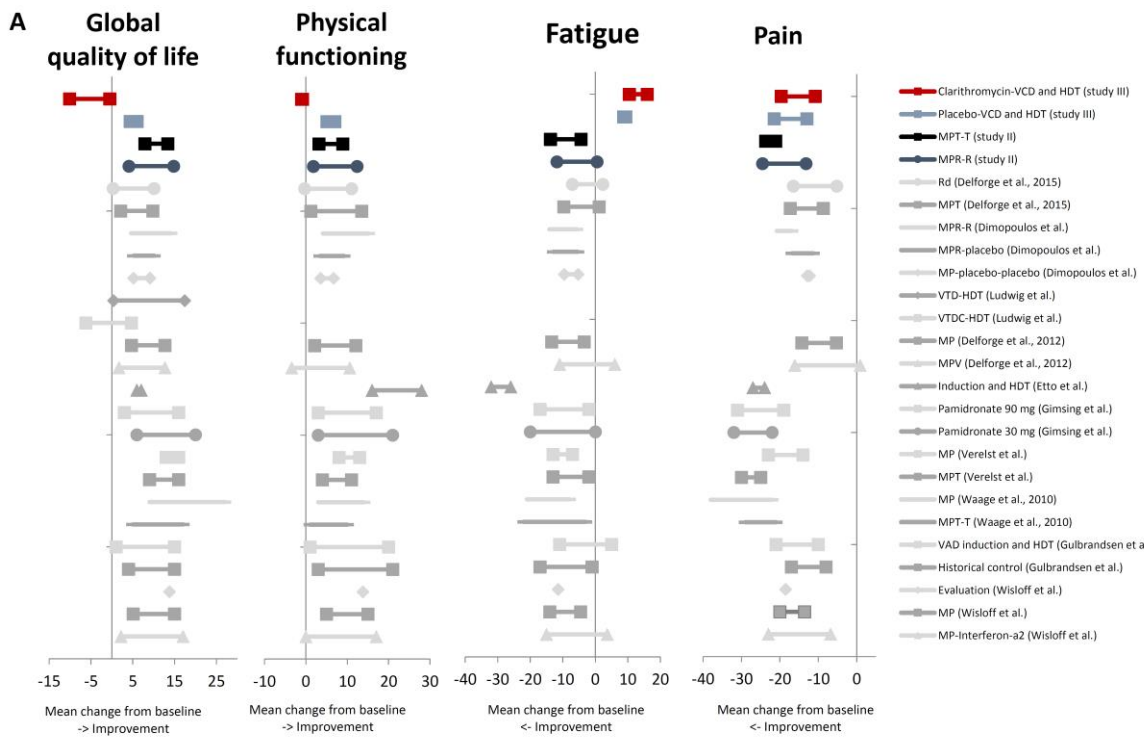
EORTC QLQ-MY20 domain	Study II	Study III
Disease Symptoms[1]	11.7	10.7
Side Effects of Treatment[1]	8.3	6.8
Future Perspective[1]	11.4	14
Body Image[2]	14.4	15.9

VCD; bortezomib-cyclophosphamide-dexamethasone, HDT; high dose melphalan, [1]multi-item domain, [2]single item-domain.

For the symptom domains, the patients with clarithromycin added to VCD induction reported clinically meaningful worsened insomnia and appetite loss, which is in contrast to the patients receiving other regimens, where either unchanged or reduced insomnia or appetite loss were reported. The patients treated in study II or III reported unchanged dyspnoea, nausea and vomiting, which is similar to the patients treated in other regimens. The patients treated with clarithromycin or placebo added to VCD induction and HDT reported reduction of constipation and increased diarrhoea. We found a statistically significant time effect for constipation and diarrhoea in study III. This means that the fact that patients received and completed the questionnaires earlier than two months after HDT had a significant impact on the mean score and the results from those two domains were inconclusive. The general reporting was unchanged constipation and diarrhoea during other regimens.

The EORTC QLQ-MY20 domains
Only two studies included in study I presented data from EORTC QLQ-MY20 (162, 170). The patients treated with all four regimens of study II and III reported clinically meaningful improvement in future perspectives, which is similar to the general reporting in future perspectives. The patients in study III reported clinically meaningful worsened side effects of treatment and body image, which is different compared to the other treatment regimens, where unchanged side effects of treatment and body image were reported. The patients in study III reported clinically meaningful reduction in disease symptoms, which was also the case for the patients treated with Rd (170) and MPR and planned lenalidomide maintenance (162).
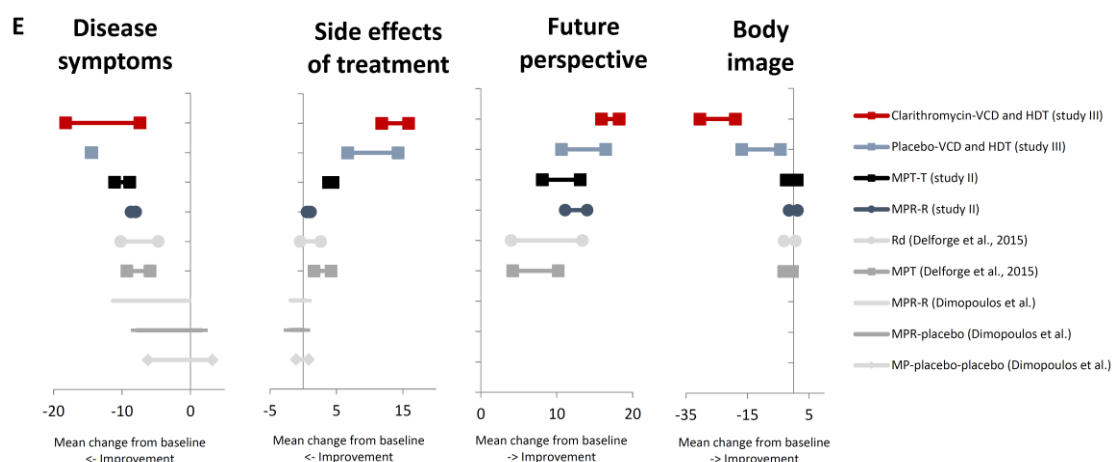
A

Global quality of life | Physical functioning | Fatigue | Pain

Mean change from baseline -> Improvement (Global quality of life, Physical functioning)
Mean change from baseline <- Improvement (Fatigue, Pain)

B

Role functioning | Emotional functioning | Social functioning | Cognitive functioning

Mean change from baseline -> Improvement

Legend:
- Clarithromycin-VCD and HDT (study III)
- Placebo-VCD and HDT (study III)
- MPT-T (study II)
- MPR-R (study II)
- Rd (Delforge et al., 2015)
- MPT (Delforge et al., 2015)
- MPR-R (Dimopoulos et al.)
- MPR-placebo (Dimopoulos et al.)
- MP-placebo-placebo (Dimopoulos et al.)
- VTD-HDT (Ludwig et al.)
- VTDC-HDT (Ludwig et al.)
- MP (Delforge et al., 2012)
- MPV (Delforge et al., 2012)
- Induction and HDT (Etto et al.)
- Pamidronate 90 mg (Gimsing et al.)
- Pamidronate 30 mg (Gimsing et al.)
- MP (Verelst et al.)
- MPT (Verelst et al.)
- MP (Waage et al., 2010)
- MPT-T (Waage et al., 2010)
- VAD induction and HDT (Gulbrandsen et al.)
- Historical control (Gulbrandsen et al.)
- Evaluation (Wisloff et al.)
- MP (Wisloff et al.)
- MP-Interferon-a2 (Wisloff et al.)

30

C

Nausea and vomiting | **Dyspnoea** | **Insomnia** | **Appetite loss**

Legend:
- Clarithromycin-VCD and HDT (study III)
- Placebo-VCD and HDT (study III)
- MPT-T (study II)
- MPR-R (study II)
- Rd (Delforge et al., 2015)
- MPT (Delforge et al., 2015)
- MPR-R (Dimopoulos et al.)
- MPR-placebo (Dimopoulos et al.)
- MP-placebo-placebo (Dimopoulos et al.)
- VTD-HDT (Ludwig et al.)
- VTDC-HDT (Ludwig et al.)
- MP (Delforge et al., 2012)
- MPV (Delforge et al., 2012)
- Induction and HDT (Etto et al.)
- Pamidronate 90 mg (Gimsing et al.)
- Pamidronate 30 mg (Gimsing et al.)
- MP (Verelst et al.)
- MPT-T (Verelst et al.)
- MP (Waage et al., 2010)
- MPT-T (Waage et al., 2010)
- VAD induction and HDT (Gulbrandsen et al.)
- Historical control (Gulbrandsen et al.)
- Evaluation (Wisloff et al.)
- MP (Wisloff et al.)
- MP-Interferon-a2 (Wisloff et al.)

Mean change from baseline
<- Improvement

D

Constipation | **Diarrhoea** | **Financial difficulties**

Mean change from baseline
<- Improvement

Figure 3. (A-E) The patient-reported range of mean change of score from baseline for the presented domains of the 19 EORTC QLQ-C30 and EORTC QLQ-MY20 domains for the 25 first-line treatment regimens included in study I-III. The published domains of primary treatment regimens identified in the systematic review of study I are presented in grey, and the treatment regimens of study II and III are presented in colour. The direction of improvement is indicated with an arrow for each domain.

VCD; bortezomib-cyclophosphamide-dexamethasone, HDT; high dose therapy with stem cell support; MPT-T; melphalan-prednisolone-thalidomide followed by thalidomide maintenance; MPR-R; melphalan-prednisolone-lenalidomide followed by lenalidomide maintenance, Rd; lenalidomide-dexamethasone, MPT; melphalan-prednisolone-thalidomide, MPR-placebo; melphalan-prednisolone-lenalidomide followed by placebo maintenance, MP-placebo-placebo; melphalan-prednisolone-placebo followed by placebo maintenance, VTD; bortezomib-thalidomide-dexamethasone, VTDC; bortezomib-thalidomide-dexamethasone-cyclophosphamide, MP; melphalan-prednisolone, MPV; melphalan-bortezomib-prednisolone, VAD; vincristine-doxorubicin-dexamethasone.

Part 3. Non-responses in longitudinal HRQoL studies of multiple myeloma patients (Study IV and V)

Non-responses to scheduled questionnaires in longitudinal HRQoL studies with multiple myeloma patients (Study IV)

Methods

The primary objective was to examine the magnitude of intermittent and monotone NR in the studies identified in study I, study II and III.

In brief, we reviewed the previous research from study I-III with MM patients treated in first line or for relapsed disease and extracted data to calculate intermittent and monotone NR. After the systematic literature search for study I was performed, separate publications with additional HRQoL results from the ASPIRE and TOURMALINE-MM1 studies were published (181, 182). Those two publications were included in the data extraction process.

The information extracted from the publications was 1) the number of participating patients from whom completed HRQoL assessments were expected at each scheduled HRQoL assessment time point, 2) the number of completed HRQoL assessments at each scheduled HRQoL assessment time point. In case the last published HRQoL follow-up time point was an end of study/treatment discontinuation assessment, the number of completed questionnaires at the former time point was used. Further details are to be found in paper IV.

The magnitude of intermittent NRs was estimated by calculating the proportion of patients, who did not complete scheduled HRQoL assessments of those from whom a completed HRQoL assessment was expected. This was done for all HRQoL assessments time points and added to the rate of intermittent NRs. The magnitude of monotone NRs was estimated by calculating the proportion of incomplete HRQoL assessments at last HRQoL assessment, compared to the number of patients participating in the study.

Results

In Table 4, all extracted information from the 20 primary and relapse studies identified in study I and study II and III are gathered. Details can be found in paper IV.

Magnitude of intermittent and monotone non-responses

In eight out of the 20 studies, the information for calculating the intermittent NR rate for each patient group was presented and was between 2% and 22%. In 17 out of 20 studies, the information for calculating the proportion of monotone NR for each group of patients was presented and was between 27% and 99%.

Table 4. Magnitude of intermittent and monotone non-responses

Study/author and year of publication	Intermittent non-responses			Monotone non-responses		
	Total number of patients expected to complete PRO assessment	Total number of completed PRO assessments	**Intermittent non-responses**	Number of patients participating in HRQoL reportings at baseline	Number of completed PRO assessments at last follow-up[1]	**Monotone non-responses**
First-line treatment studies						
Study II, paper II HOVON87/NMSG18 study	904 and 1509	751 and 904	17% and 15%	298 and 298[2]	38 and 89	87% and 70%
Study III, paper III CLAIM study	68 and 87	57 and 78	16% and 10%	30 and 25	22 and 17	27% and 32%
Delforge et al. 2015 (170) FIRST trial	5166 and 2492	4743 and 2179	8% and 13%	1076 and 547[3]	506 and 189	53% and 65%
Dimopoulos et al. 2013 (162) MM-015 study	NR	NR	-	152, 153 and 154[3]	65, 56 and 65[3]	57%, 63% and 58%
Ludwig et al. 2013 (171)	NR	NR	-	46 and 46	29 and 26	37% and 43%
Etto et al. 2011 (166)	NR	NR	-	29	14[5]	52%
Delforge et al. 2012 (172) VISTA trial	NR	NR	-	344 and 338	164 and 136[3]	52% and 60%
Verelst et al. 2011 (173) HOVON 49	NR	NR	-	149 and 135	NR	-
Gimsing et al. 2010 (174)	NR	NR	-	252 and 252	164 and 171[4] at 12 months	32% and 34%
Waage et al. 2010 (163)	NR	NR	-	NR	NR[6]	-
Gulbrandsen et al. 2001 (167)	1076 and 541	966 and 528	10% and 2%	274 and 120	72 and 38	74% and 68%
Wisloff et al. 1996 (175) NMSG 4/90, cohort study	2541	2055	19%	583	424	27%
Wisloff et al. 1996 (176) NMSG 4/90[3], randomized study	NR	NR	-	271 and 253	67 and 74	71% and 75%
Relapse treatment studies						
Leleu et al. 2018 (182) TOURMALINE-MM1	3242 and 3209	3007 and 2991	7% and 7%	360 and 360	7 and 2[7]	98% and 99%
Stewart et al. 2016 (181) ASPIRE trial	1706 and 1556	1543 and 1351	10% and 13%	348 and 348	227 and 148	35% and 57%
Song et al. 2015 (169) MM-003	NR	NR	-	289 and 144	51 and 6[3]	82% and 96%
Hjorth et al. 2012 (56) NMSG 17/07	NR	NR	-	67 and 61	NR[8]	-
Lee et al. 2008 (168) APEX study	NR	NR	-	333 and 336	65 and 81	80% and 76%
Dubois et al. 2006 (88) SUMMIT study	NR	NR	-	202	144[9]	29%
Waage et al. 2004 (55)	153	120	22%	62	20	78%

NR; not reported, [1]In case of HRQoL measurement at study discontinuation, the number of completed questionnaires at the former time point is presented, unless another time point is specified, [2]Participation in the HRQoL reportings was optional, [3]Based on mean score of physical functioning. [4]The number of questionnaires used for later follow-up time point evaluation is not reported, [5]The patients at follow-up are not all the same as at diagnosis, [6]50% of the patients in the melphalan-prednisone-thalidomide arm and 62% of the patients in the melphalan-prednisone arm. The exact numbers could not be extracted, [7]The results of the PRO data in the study was made at a specified cut-off date, and some patients were still in follow-up after, [8]29 vs. 29 patients were alive at the time of last follow-up. Study design with crossover at treatment failure, [9]The change in PRO over time was assessed by comparing the change in scores according to clinical response between baseline and best endpoint.

Strategies to reduce intermittent non-responses in a longitudinal study of patients with multiple myeloma (Study V)

Methods

The aim was to determine the effect of implemented strategies in the study design, conduct and procedures to reduce intermittent NRs.

The study design and data collection procedures for the ongoing population-based study of "Quality of life in Danish myeloma patients" is described in details in paper V. In brief, the study is a Danish multicentre, observational and primarily electronical survey. Treatment-demanding NDMM and RMM according to the IMWG criteria were eligible for inclusion (20). Patients with a mental disorder preventing the patient from completing a questionnaire or with the inability to understand the Danish language are ineligible for the study. The patients complete a set of baseline questionnaires at study entry and 12 times during 24 months follow-up. Schedule for completion of follow-up questionnaires is set to target dates every fourth week for the first six months and every 3 months thereafter and consists of 2-4 PRO instruments. The patients can choose between completing follow-up questionnaires electronically or on paper.

Strategies to minimize intermittent non-responses

All involved study nurses are educated in the importance of reducing NR and had access to a written manual of all study tasks. They also have access to support from the study office on weekdays. The patients are asked to complete the questionnaires on the pre-planned day (target date) and no later than day six after the target date. The study nurses are allowed to guide patients in completion of the questionnaires, if needed. If the patients using the electronic method have not completed the questionnaire at day four, they receive a reminder. If the patient have not completed the questionnaire at day seven, the local study nurse is notified by the study office as part of central real-time monitoring. In that case, the local study nurse contacts the patient, ascertain and document the reason for NR and invite the patient to complete the questionnaire. The effect of the implemented strategies to reduce NR was determined by calculating the rate of intermittent NRs.
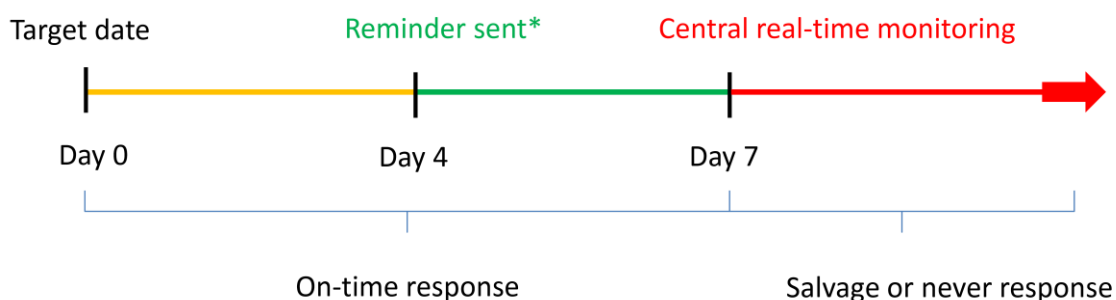
Figure 4. Study design with a seven-day time window for completion, reminders and central real-time monitoring of non-responses. *Reminders are only sent to patients completing questionnaires electronically.

Data analysis

This analysis included the patients participating in the study at August 16th 2018 and had reached the first follow-up HRQoL assessment at 4 weeks. Questionnaires completed before or within the seven-day window are defined as "on-time responses". In case the patient had completed the questionnaire at day seven after the target date or later, the response was defined as "salvage response", the remainders were categorised as a "never response". The study design is presented in Figure 4. NR to questionnaires was defined as an incomplete EORTC QLQ-C30 instrument, which was the first instrument in each set of questionnaires. The rate of intermittent NRs is calculated by the proportion of patients, who did not complete scheduled HRQoL assessments of those from whom a completed HRQoL assessment were expected. This proportion was calculated for each HRQoL assessments time points and added to the rate of intermittent NRs.

Results

The results are presented in paper V. In brief, 272 patients were included in the analysis and they had reached a total of 1441 scheduled questionnaires. Of the 1441, 1214 (84%) were completed on-time. Of the remaining 227 scheduled questionnaires, 153 (67%) were salvaged responses, and 74 (33%) were never responses, equivalent to an intermittent NRs rate of 5%.

Discussion

The overall aim of part 1 and 2 was to determine whether the EORTC QLQ-C30 and EORTC QLQ-MY20 questionnaires capture clinically meaningful treatment effects on HRQoL during first and relapse therapies from the MM patients´ perspective. We found that clinically meaningful improvement in global QoL, physical functioning, pain and fatigue are far more likely during first line compared to relapse treatment. We also found that when clarithromycin is added to VCD induction therapy in first line, the patients did not report improvement in HRQoL as expected. The patients treated with clarithromycin reported clinically meaningful worsened fatigue, insomnia, appetite loss and increasing in score for side effects of treatment. This resulted in a lack of improvement in global QoL and physical functioning, decreased social functioning and body image during treatment and increased emotional functioning two months after HDT.

Part 1 – Systematic review of longitudinal HRQoL studies in multiple myeloma patients

Methodological considerations

The identification of previously published longitudinal HRQoL studies of patients with MM has the strength in being based on a systematic literature search. For the selection strategy of publications, we chose only to focus on studies using EORTC QLQ-C30 for HRQoL measurement. This choice was made to meet the review aim of interpreting the HRQoL data by anchor based MIDs of *Kvam's MID criteria* and *guidelines of Cocks*, which are is based on the EORTC QLQ-C30 instrument. A possible limitation induced by our publication selection strategy could be that some studies using a different HRQoL instrument or studies published after we performed the literature search have found contradictive results.

As presented in the PRISMA flow diagram in paper I, 31 publications were excluded during the full-text publication review, since they used alternative instruments for HRQoL measurement. These 31 publications were secondarily reviewed to explore whether studies using a different HRQoL tool found contradictive results from our general findings in study I. Eighteen of the 31 publications were still not eligible, since they fell for one of the other exclusion criteria for the systematic review. 13 publications were remaining, and the HRQoL results of those were examined (183-195). Direct comparison of HRQoL results captured by another HRQoL instrument is not possible, and clinically meaningful treatment effects cannot be determined by the same MID thresholds, since MID is questionnaire specific. However, the HRQoL results from these 13 publications were corresponding to our general findings in study I. For example, the ECOG E1A06 study, where patients with NDMM, ineligible for HDT, reported improvement in FACT Functional and physical mean score during first line treatment with MPT and MPR (194).

Since May 2016, where the systematic literature search was performed, new HRQoL results from patients with MM using the EORTC QLQ-C30 instrument have been published. Apart from study II and III, we have identified two primary and five relapse studies, which have been published recently. The primary studies are the ALCYONE trial with comparison of Daratumumab-MPV versus MPV therapies and the Medical Research Council Myeloma XI trial (196-200). The general findings of clinically meaningful improvement in global QoL, physical functioning and pain were confirmed in both studies. Both patient groups in the ALCYONE trial reported unchanged fatigue, which corresponds to the previously investigated HRQoL reportings during MPV treatment in the VISTA study (172). The five relapse studies are the ENDEAVOR trial

(201), the PANORAMA-1 trial (202), the ELOQUENT-2 study (203), CASTOR (204) and POLLUX (205). Not in all publications, the results from global QoL, physical functioning, fatigue and pain were presented, but in general the RMM reporting unchanged global QoL during relapse treatment regimens in line with our findings in study I. In conclusion, the aim of part 1 of determine clinically meaningful treatment effects for the four domains was achieved by the performed literature review.

Part 2. Clinical trials with multiple myeloma patients during first line therapies

Methodological considerations

EORTC QLQ-C30 and EORTC QLQ-MY20 for HRQoL evaluation

The content validity of the EORTC QLQ-C30 questionnaire for HRQoL measurement from the NDMM patients´ perspective is well supported. This relies on the psychometric validation of the instrument and results from a Delphi consensus project for NDMM patients (95, 97-99). In the Delphi consensus project, it is recommended to measure body image, and since side effects of treatment are relevant in clinical decision-making in MM, the addition of the EORTC QLQ-MY20 questionnaires is a well-argued choice (24, 30, 99).

NDMM patients were only included in the validity study of EORTC QLQ-C30, and mainly (225 out of 240 patients) NDMM patients were included in the validity study of EORTC QLQ-MY20 (96, 98, 206). Therefore, the psychometric validity of the two questionnaires in patients with RMM is largely unknown, which is a limitation in the results of relapse treatment studies. The developer of the EORTC QLQ-MY20 questionnaire chose that the items for assessing the domain of side effects of treatment should capture expected adverse events of conventional chemotherapy and steroids (207). Melphalan is a conventional chemotherapeutic drug and used in HDT. HRQoL after HDT was investigation in study III, and the patients reported clinically meaningful worsening in the domain of side effects of treatment, which was not reported during other regimens in part 2.

In current years, conventional chemotherapy is used in drug combinations with IMIDs, proteasome inhibitors or monoclonal antibodies, resulting in a different toxicity profile than conventional chemotherapy alone (24). Therefore, the toxicities and impact of toxicities from the patients´ perspective to novel drugs or drug combinations are not necessarily elucidated by the side effects of treatment domain of EORTC QLQ-MY20. HRQoL, which includes toxicities to available drugs and drug combinations, is an important factor in clinical decision-making in MM. As demonstrated in study III, symptomatic toxicities from clinical trials are underreported by clinicians, which highlights the need for capturing symptomatic toxicities with PRO instruments (208). For this, we need flexible PRO tools with relevant items, which apply to the expected toxicity profile for the investigated drug. The EORTC item bank and the Common Terminologies Criteria for Adverse Events for self-reported toxicities, which are currently available, will probably make this possible (209, 210).

HRQoL study design

PRO data study design and timing of collection have an impact on the quality of HRQoL data and interpretation of the HRQoL results (124, 133, 211, 212). In study II, we found that the clinically meaningful treatment effect on HRQoL occurred 6 and 12 months after start of either thalidomide or lenalidomide maintenance. Since evidence concerning HRQoL during maintenance therapies in general are lacking, this would have been interesting to investigate as part of study II. The study was not designed with patient randomization after end of induction therapy to either lenalidomide or placebo maintenance *or* thalidomide or placebo maintenance. This would have been optimal for investigation of the HRQoL effect of maintenance versus placebo and interpretation of results during maintenance. It remains unsolved whether the improvement in HRQoL at 6 and 12 months, found in study II, is based on recovery after discontinuation of induction therapy, or can be assigned to maintenance therapy. Two aspects from the literature support that the measured improvement in mean score after start of maintenance could be assigned to recovery after end induction treatment. Firstly, cross-sectional studies of MM patients in different stages of disease have shown that the patients in the first treatment free interval report reduced symptom burden and higher function, compared to patients during primary treatments (31, 213). Secondly, HRQoL data during maintenance is also available from the MM-015 study, where the patients were randomized in three groups; MPR-R, MPR-placebo maintenance and MPplacebo-placebo maintenance (162, 214). Results from the MM-015 showed that going from induction to maintenance for patient treated with MPR-R compared to MPR-placebo, resulted in a further increased mean score from baseline for physical functioning and fatigue for both groups. A limitation in interpretation of results during maintenance in study II and MM-015 study is the high proportion of monotone NR. Moreover, the monotone NRs in the MM-015 study were handled by last observation carried forward, which might have overestimated the HRQoL results (162, 215, 216).

In the interpretation of HRQoL results from study III, we experienced that HRQoL results from a study designed with a placebo group do have advantages. When adding the knowledge gained from *between* group differences to the results from *within* group change to determine clinically meaningful toxicities to therapy provided a stronger conclusion, since we were able to compare to a "normal" trajectory in HRQoL during treatment. The results from the patients treated with placebo added to the VCD induction therapy acted like an "anchor" for the expected clinically meaningful change in HRQoL during and after the VCD induction treatment.

The comparisons of clinically meaningful treatment effects on HRQoL between studies performed in part 2 were based on the assumption that HRQoL assessment was performed at the most suitable time point. This might not always be the case. In study III, pre-planned seven day windows for completion were used as part of the PRO data collection design. Pre-planned time windows for completion are supposed to ensure that the PRO data collection is capturing the HRQoL at the clinically relevant time point (217). The Internet-based tool used in study III for web-based PRO data collection automatically delivered the questionnaires to the patients, at predefined time points, for completion in the pre-planned time window. No adjustment in timing of the automatic delivery of electronic questionnaires was made when treatment was rescheduled due to e.g. complications, which happened to be frequently in study III. Therefore, a number of patients completed the HRQoL questionnaires earlier than scheduled in the protocol, resulting in a statistically significant time effect for the results for constipation and diarrhoea. Due to this, we could not evaluate those two domains two months after HDT. We only investigated possible time effect for the eight

symptomatic toxicities, and whether this was also the case for the remaining domains is unknown. Possible heterogeneity in clinical treatment trajectories of the targeted population should be taken into account, when designing a study, to elucidate the HRQoL during treatments.

In study II, not all patients had completed the questionnaires as specified in the study protocol. This might be due to the fact that the pre-scheduled HRQoL time points were defined as approximately time points, for example "after cycle 3, approximately at 3 months after start of cycle 1". In order to include as many completed questionnaires as possible, we categorized the completed questionnaires according to the time frames, up to 3 months. This might have caused biased results, since patients with a temporary decline in HRQoL might have delayed completion until recovery. Therefore, it is important to make a clear description of all PRO data collection procedures as part of the study protocol and ensure that the staff is trained in how to collect the PRO data (129, 133).

Minimal important difference

In this thesis, we chose to follow recommendations of using available anchor based over distribution based MID thresholds, since anchor based MIDs are proposed to offer better estimates for minimal important changes than distribution based MIDs (109, 110, 119). *Kvam's MID criteria* are based on MM patients´ perception of change and were therefore preferred (115). Anchor based MID thresholds based on MM patients´ perception for the 11 remaining domains of EORTC QLQ-C30 and the four domains of EORTC QLQ-C30 MY20 are not available. To determine clinically meaningful treatment effect for those domains, we chose the thresholds from *guidelines of Cocks* for this thesis, which is based on cancer clinicians´ perception of change (116). The medium MID threshold of *guidelines of Cocks* was sensible to detect well known adverse events of e.g. improved insomnia and more constipation during thalidomide treatment. However, whether those findings reflect clinically meaningful change to patients with MM is unknown, which is a limitation in the findings.

For the four EORTC QLQ-MY20 domains, distribution based methods were the only way of addressing clinically meaningful treatment effects of those domains. We chose the SEM based method for the multi-item domains over effect size MIDs for this thesis, since SEM based MIDs rely not only on SD but also on the number of items and the sample size (113). Therefore, we considered the thresholds to be stronger, especially for interpretation of the mean change of scores from baseline in study III with low sample size. This choice is supported by the similarities of the estimated SEM based MID thresholds comparted to the estimated thresholds in the MM-015 study (162). For the single-item domain of body image, we used Cohen's effect size of 0.5, which is supported by the general findings of Kvam *et al.* of an effect size range of 0.3-0.5 being clinically meaningful to patients with MM and *Norman's rule of thumb* (112, 115). Even though the distribution based MIDs are able to detect e.g. increased side effects of treatment in patients treated with VCD induction and HDT, it is unknown, whether this is in accordance with clinically meaningful change for patients with MM, which is a limitation in the findings.

As argued here, the available MID thresholds for *within* group change of HRQoL scores of the EORTC QLQ-C30 and EORTC QLQ-MY20 questionnaires lack validity for being able to determine a clinically meaningful treatment effect on HRQoL for patients with MM. For interpretation of clinically meaningful change in

paper II and III, we chose distribution based MIDs. This choice was made in order to make interpretation simple and transparent to the readers, which might have been too simple (104). In this thesis, we chose to follow consensus recommendations of preferring available anchor based over distribution based MID thresholds in order to explore the gaps in existing MID thresholds in MM. Determination of anchor based MIDs based on patients´ perception of change are resource demanding with requirement of methodological considerations. The US Food and Drug Administration's PRO Guidance from 2009 has changed methods for determining clinically meaningful treatment effect in PROs from MIDs to "responder definition" (218, 219). Responder definition is based on the numbers of patients, who reported improvement, unchanged or worsening for each domain, during the study period (220). In order to evaluate HRQoL results from clinical trials using responder definition, strong anchors are important (221). Clearly, anchor based MID thresholds to define responders to treatments need further research and must be balanced against the practical and methodological aspects in their establishment.

We used the same thresholds for interpreting clinically meaningful change in patients treated during primary and relapse treatment. The differences in clinical context could be one of the explanations for our diversity in findings during first line, compared to relapse treatment. Often, the main goal of primary treatment is achieving a long, durable period of disease control, whereas in a relapse and/or refractory situation where long durable periods of disease control may not be possible, the treatment goal is more often to achieve disease control and prevent clinical relapse. The clinical context differs in the two settings and should be integrated in the MID thresholds. Moreover, the MIDs should reflect MM patients with different demographic and personality characteristics and disease trajectories (222, 223). Response shift does have an impact on some MID sizes in patients with MM (151, 160). In addition, in a study of response shift investigation, it has been proposed that amount of missing data and missing data mechanisms might have an impact on the response shift detections (224). Therefore, MID estimation should include a representative sample of patients with MM investigated in longitudinal study designs with high PRO completion rates and as few monotone NRs as possible. This needs further investigation as well as the share of response shift effect in the reported clinically meaningful interpretation of treatment effects. MIDs based on multiple clinical anchors are currently being developed and investigated and might include some of these aspects (225, 226).

Response shift

Our findings in clinically meaningful treatment effects on global QoL is supported by the "Conceptual Model" of Wilson *et al*. (78). Global QoL is a distale outcomes measure, placed to the left in the model with earlier described difficulties in its definition and measurement. In paper II, we found that 55% of the patients treated with MPT reported increased "*tingling hands and feet*" during treatment, but with no reflection on the global QoL mean score differences between the MPT and MPR treated patients. A possible explanation to this could be that the change score for the multidimensional domain of the global QoL was affected by e.g. response shift, and that patients may have adapted gradually to symptoms of peripheral neuropathy during thalidomide treatment. A limitation in this evaluation is that the "peripheral neuropathy" domain is not validated.

Another of our results could theoretically be explained by a possible response shift effect: In part 2, we found that the patients treated with clarithromycin added to VCD induction therapy and HDT reported clinically meaningful improvement in emotional functioning two months after HDT. This diverted from the patients treated with placebo added to VCD induction therapy and HDT, who reported unchanged emotional functioning. The reason for this could be assigned to a response shift effect and change in QoL appraisal over time. The domain of emotional functioning is calculated on the basis of answers to which extent the patients are feeling tense, irritable, depressed and worried. In theory, two months after HDT the patients might have compared their emotional functioning to the time after cyclophosphamide priming, where they reported decreased mean score for emotional functioning compared to baseline. This difference in samples of experiences and standards of comparison might have influenced the patients´ QoL appraisal and resulted in an improved emotional functioning two months after HDT compared to baseline.

As mentioned earlier, evidence for response shift effect in patients with MM has been presented previously (160). Here, consecutive patients with MM (NDMM, stable and RMM patients) were recruited in connection with admission to the hospital. Admission to the hospital was, in theory, the "catalyst" to initiate a response shift effect. To test the hypothesis of the existence of response shift in the trajectory of the MM disease, we aimed to include the BAI questionnaire in the ongoing study of "Quality of life in Danish multiple myeloma patients" presented in paper V. In collaboration with professor, Carolyn Schwartz, Boston, USA, we conducted a linguistic and conceptual validation study of the BAI translated into the Danish language. The translation procedure was based on EORTC Quality of life group's manual for forward-and-backward translation of questionnaires (227). The results for the pilot-testing interview were that patients had critical remarks to the introduction text as well as seven of the 23 items (data not shown). Concluded in collaboration with the two developers, Carolyn Schwartz and Bruce Rapkin, a qualitative study investigating the cognitive appraisal processes in the target cohort is found necessary to explore this further and develop a culture and disease adapted version of BAI suitable for MM patients in future studies (personal communication).

In conclusion, the aim of part 2 of determine clinically meaningful treatment effects for the four first line treatment regimens was achieved. However, the findings in relation of results to the findings in study I might be compromised by methodological difference in PRO study design and data collection. The overall aim of part 1 and 2 was to determine if the EORTC QLQ-C30 and EORTC QLQ-MY20 do capture clinical meaningful treatment effects on HRQoL from the MM patients´ perspective during first line and relapse anti-myeloma regimens. To achieve this overall aim, the validity of EORTC QLQ-C30 and EORTC QLQ-MY20 in RMM as well as during novel anti-myeloma treatments needs further investigation. Moreover, the applied MID threshold lack validity of been able to determine clinical meaningful treatment effects from all MM patients´ perspective.

Part 3. Non-responses in longitudinal PRO studies of multiple myeloma patients

The aim of part 3 was to investigate the magnitude of NR in longitudinal PRO studies of patients with MM, and to propose and evaluate tools to reduce intermittent NRs. Among the existing studies, the intermittent NRs rate was between 2% and 22% and the proportion of monotone NRs was between 27% and 99%. In a longitudinal HRQoL study in patients with MM strategies to minimize NR were integrated by staff education, use of reminders and real-time monitoring, which resulted in an intermittent NR rate of 5%.

Methodological considerations

Compliance or completion rate

It has been recognized that PRO compliance or PRO completion rates of clinical trials lack a standard definition and are currently been discussed (228). This means that in calculation of compliance rate there is no consensus on numerator and denominator. In paper V, we chose the definition of "PRO completion rate" as stated by Osoba *et al.* of "the number of completed PRO assessments over the number of scheduled PRO assessments expected to be completed" (220).

Impact and handling of non-responses

The high amount of NRs to scheduled questionnaires in HRQoL studies of patients with MM suggests that the evidence-based knowledge of HRQoL during and after anti-myeloma treatments is compromised. Implementing strategies to reduce NR in the study design and conduct has been recommended earlier (134, 135). However, a description of successfully strategies to reduce intermittent NRs, as demonstrated in study V, has not been available until now. In order to prevent monotone NR, data collection after discontinuation of the investigated drug must be a part of the protocol (128).

Intermittent and monotone NR to scheduled questionnaires in clinical and cohort studies of patients are practically unavoidable. When handling NRs statistically, ideally, the actual missing data mechanism should be ascertained and inform the choice of statistical method and ensure appropriated statistical handling of NR. This is important, also since the missing data mechanism often has a stronger impact on the results than the proportion of missing data (127, 134). However, currently there are no available methods for determining true missing data mechanisms in real datasets. Analytic methods that can distinguish between MCAR and MAR data are available, e.g. Little´s test (229). If data from clinical trials contain significant amounts of MNAR data, few methods can achieve unbiased estimates of change in PROs over time or level of PROs at a specific time (128, 137).

Sensitivity analysis performed by multiple imputation is considered one of the most reliable statistical methods for handling NR (230). Therefore, multiple imputation was used to test the robustness of the results in study II and III, but the results did not differ substantially from the analyses performed by linear mixed models of repeated measures. This was surprising, especially in study III, where grade 3 and 4 adverse events were found as a statistically significant predictor of NR. This information, which were considered to be a MNAR reason for NRs, were added to the multiple imputation model. Previous investigation of the multiple imputation method has shown that the bias reduction achieved by multiple imputation depends on the variables used in the model, and simulation studies have shown that information collected during follow-up should be included (230, 231).

NRs is an ongoing challenge in analysing PRO data (232). Demonstrated in study V, intermittent NR can be reduced substantially, but whether a salvage response is representative and unbiased in all cases has not yet been investigated. This, as well as statistical methods for handling NR in PRO studies, need further investigation. As part of study V, reasons for intermittent and monotone NR was collected. Future analyses

will reveal, if adding those reasons to the multiple imputation or another statistical method for handling longitudinal data with missing observations will reduce the bias of the results (231).

In conclusion, the aim of part 3 of investigation of the magnitude of NR in longitudinal studies of patient with MM was achieved, although not in all publications this was reported. Also, it is possible to reduce intermittent NRs by education of study nurses, use of reminders and real-time monitoring.

Limitations

Available HRQoL studies using EORTC QLQ-C30 and EORTC QLQ-MY20 do have limitations in applied methodologies, when assessing clinically meaningful treatment effects on HRQoL from the MM patients´ perspective.

The HRQoL studies, identified and analysed in part 1 and 2, were all clinical trials. NDMM patients included in clinical randomized phase III trials are not representative of the general population of MM patients (233). Patients in clinical trials are generally selected for better performance status, lower levels of comorbidity and younger age. This might limit the generalization of results and implication to daily clinical practice. Evidence of change in HRQoL over time, in unselected patients with MM is practically unknown, and whether HRQoL in the general population of patients with MM differs from MM patients in clinical trials is unexplored.

A limitation in our results is whether the applied MID thresholds are able to distinguish what change in mean score that is clinically meaningful for patients with MM. In addition, it is not clear, whether and how big a share of those thresholds that is due to a response shift effect, and how big a share is due to a treatment effect.

As examined in paper I and IV, the number and timing of HRQoL assessments were different among the studies. Some of the studies included measurement of long-term treatment effects on HRQoL, whereas other studies elucidated short term follow-up. The rationale for choosing the prescheduled HRQoL assessment time point for questionnaire completion in the investigated studies was, in general, not clear and might not have been the best suitable time point. We observed differences in mean age of the patients included and the sample sizes of each treatment group. The lack of statistical power is a limitation in the discrepancy from the general findings in HRQoL during the regimen of clarithromycin added to VCD induction therapy.

We cannot rule out that there is a limitation in the findings of side effect of treatment, since this is measured by the ten-item domain of side effect of treatment in EORTC QLQ-MY20, which is not based on the toxicity profiles of investigated novel drugs and drug combinations. Of notice, not in all studies, the EORTC QLQ-MY20 questionnaire was used, partly due to the fact that it was developed and validated in 2007. Few authors published results of all domains, and we were not able to include and evaluate results from the unpublished domains. In addition, the validity of the EORTC QLQ-C30 and EORTC QLQ-MY20 instruments in relapse patients is unknown.

We observed differences in applied statistical methods among the studies as well as a high magnitude of complete, intermittent and monotone NR. Some of the applied statistical methods treat NR as MCAR,

which is largely based on untested assumptions and is theoretically seldom the case in patients with MM. Furthermore, the included randomized trials were mainly designed to elucidate *between* treatment group differences and not *within* group change, which supposedly has guided the trial investigators in designing the studies. These mentioned methodological aspects might compromise comparison of results between trials carried out in part 1 and 2.

Main conclusions

Part 1 and 2

Concluded from the investigation of clinical meaningful treatment effect on HRQoL from 20 first line and relapse studies, NDMM reported clinically meaningful improvement in global QoL, physical functioning, pain and reduced fatigue during primary treatments, whereas RMM reported no change or even deterioration. The findings were unrelated to administrated drugs or drug combinations. A discrepancy from this general finding was seen in the reportings from the group of patients treated with clarithromycin added to VCD induction therapy and HDT. Those patients reported clinically meaningful increased fatigue, insomnia, appetite loss and side effects of treatment. This resulted in lack of expected improvement in global QoL and physical functioning, as well as decreasing social functioning and body image during treatment and increased emotional functioning, two months after HDT.

Methodological differences among the 20 investigated studies in the PRO data design, collection procedures, applied statistical analysis methods as well as high magnitude of NRs were seen. This might have compromise the comparison of results and caused bias. The choice of EORTC QLQ-C30 and EORTC QLQ-MY20 for HRQoL measurement is supported by the validation study of patients with for NDMM, but not in RMM, where no psychometric validation has ever been carried out. The longitudinal PRO data from first line as well as relapse studies were interpreted by the same MID thresholds for clinical meaningful treatment effect on HRQoL. In general, the available MID thresholds lack validation of being able to assess clinical meaningful treatment effect on HRQoL from the MM patients´ perspective and are not adjusted for a possible response shift effect. Moreover, the findings are based on patients included in clinical trials and might not be generalizable to the general population of patients with MM.

Distale outcomes measures of symptoms or symptomatic toxicities from EORTC QLQ-C30 and EORTC QLQ-MY20 are more sensible endpoints than proximale, multidimensional domains in patients with MM. With the use of IMIDs, proteasome inhibitors and monoclonal anti-bodies as treatment options in MM, rethinking toward flexible PRO tools to access symptomatic toxicities based on expected toxicity profile might be better to assess risks and benefits of treatments.

Part 3

Concluded from the investigation of NRs to questionnaires in longitudinal studies in MM, the finding of high magnitude of complete, intermittent and monotone NR suggest that the evidence-based knowledge of HRQoL in MM are compromised. Intermittent NR can be substantially reduced by implementing strategies of education of staff, using pre-defined time windows, reminders and real-time monitoring, which might increase the quality of HRQoL results.

Very little is known about the missing data mechanisms in patients with MM. We investigated this in study III and found prediction for NR from patients with severe adverse events. When adding this information to a multiple imputation analysis, we saw only minor indication of our results being biased. Statistical analysis methods for handling missing PRO data need further investigation in order to reduce bias of results.

Perspectives

PRO outcome results should be integrated in the information given to the patient at the time of treatment demanding disease to promote patient engagement and shared decision making in health care. Before this can be evidence based, methodological challenges in assessing clinically meaningful treatment effect on HRQoL must be addressed.

Well-designed studies for capturing of HRQoL in MM for clinical treatment decision making are needed and starts with a well-written protocol, where a description of study design, procedures related to PRO data collection and statistical analysis plan should be implemented. This should be done using available guidelines for PRO protocols to ensure that every step in the design of measuring, analysis and interpretation of results is thought through. A multidisciplinary team consisting of a representative sample of patients, clinicians with expertise within the disease and treatment, PRO researchers, statisticians and research staff should be involved in this process. Reporting of results in a way that is meaningful to clinicians and patients is important in order to integrate PRO results in clinical decision-making.

In this thesis, we have identified unresolved challenges in HRQoL measurement, analysis and interpretation in MM, which need further investigation. Firstly, strong anchor based MIDs for assessing clinically meaningful treatment effects from the MM patients´ perspective are needed. Secondly, response shift and change in QoL appraisal in patients with MM might have an impact on the size of the MID and needs further investigation. Lastly, NR to scheduled questionnaires have been pointed out as one of the major challenges in PRO data research. We have demonstrated practical tools for reducing intermittent NR, but statistical methods for handling unavoidable NR in order to reduce selection bias need further development.

The three methodological challenges are interrelated. Establishing of reliable MID thresholds for determining clinically meaningful effect of treatment on HRQoL from the MM patients´ perspective need integration of possible response shift effects. For this, data from a representative sample of patients with MM reflecting the clinical heterogeneity of the MM disease should be included.

The research performed in this thesis and discussed methodological challenges will not only have implications in clinical trials or in patients with MM. There is a growing demand for assessing real-world evidence of HRQoL in MM and other cancer patients and using PRO as a quality indicator for health care services. PROs are increasingly being added to national registries to monitor HRQoL and provide data for benchmarking, nationally and internationally. In order to do so the same challenges as identified and discussed here need to be addressed.

References

1. Kyle RA, Rajkumar SV. Multiple myeloma. Blood. 2008;111(6):2962-72.
2. Becker N. Epidemiology of multiple myeloma. Recent Results Cancer Research. 2011;183:25-35.
3. Ludwig H, Bolejack V, Crowley J, Blade J, Miguel JS, Kyle RA, et al. Survival and years of life lost in different age cohorts of patients with multiple myeloma. J Clin Oncol. 2010;28(9):1599-605.
4. Engholm G FJ, Christensen N, Hansen HL, Hertzum-Larsen R, Johannesen TB, Kejs AMT, Khan S, Ólafsdóttir E, Petersen T, Schmidt LKH, Virtanen A and Storm HH. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries http://www.ancr.nu2017 [updated Version 8.0 (20.12.2017).
5. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol. 2010;49(5):725-36.
6. Ocias LF, Larsen TS, Vestergaard H, Friis LS, Abildgaard N, Frederiksen H. Trends in hematological cancer in the elderly in Denmark, 1980-2012. Acta Oncol. 2016;55 Suppl 1:98-107.
7. Gimsing P, Holmstrom MO, Klausen TW, Andersen NF, Gregersen H, Pedersen RS, et al. The Danish National Multiple Myeloma Registry. Clinical epidemiology. 2016;8:583-7.
8. Fonseca R, Monge J, Dimopoulos MA. Staging and prognostication of multiple myeloma. Expert Rev Hematol. 2014;7(1):21-31.
9. Weiss BM, Abadie J, Verma P, Howard RS, Kuehl WM. A monoclonal gammopathy precedes multiple myeloma in most patients. Blood. 2009;113(22):5418-22.
10. Kyle RA, Rajkumar SV. Epidemiology of the plasma-cell disorders. Best practice & researchClinical haematology. 2007;20(4):637-64.
11. Landgren O, Kyle RA, Pfeiffer RM, Katzmann JA, Caporaso NE, Hayes RB, et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. Blood. 2009;113(22):5412-7.
12. Kyle RA, Larson DR, Therneau TM, Dispenzieri A, Kumar S, Cerhan JR, et al. Long-Term Follow-up of Monoclonal Gammopathy of Undetermined Significance. N Engl J Med. 2018;378(3):241-9.
13. Kyle RA, Durie BG, Rajkumar SV, Landgren O, Blade J, Merlini G, et al. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma: IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management. Leukemia. 2010;24(6):1121-7.
14. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. The Lancet Oncology. 2014;15(12):e538-48.
15. DMSG. Diagnostic and treatment of multiple myeloma. http://myeloma.hematology.dk/index.php/vejledninger-dmsg: Danish Myeloma Study Group; 2017.
16. Kyle RA, Gertz MA, Witzig TE, Lust JA, Lacy MQ, Dispenzieri A, et al. Review of 1027 patients with newly diagnosed multiple myeloma. Mayo Clin Proc. 2003;78(1):21-33.
17. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. J Clin Oncol. 2015;33(26):2863-9.
18. Lahuerta JJ, Paiva B, Vidriales MB, Cordon L, Cedena MT, Puig N, et al. Depth of Response in Multiple Myeloma: A Pooled Analysis of Three PETHEMA/GEM Clinical Trials. J Clin Oncol. 2017;35(25):2900-10.
19. Durie BGM, Harousseau J-L, Miguel JS, Bladé J, Barlogie B, Anderson K, et al. International uniform response criteria for multiple myeloma. Leukemia. 2006;20(9):1467.
20. Kumar S, Paiva B, Anderson KC, Durie B, Landgren O, Moreau P, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. The Lancet Oncology. 2016;17(8):e328-e46.

21. Mohty M, Terpos E, Mateos MV, Cavo M, Lejniece S, Beksac M, et al. Multiple Myeloma Treatment in Real-world Clinical Practice: Results of a Prospective, Multinational, Noninterventional Study. Clin Lymphoma Myeloma Leuk. 2018.

22. Boland E, Eiser C, Ezaydi Y, Greenfield DM, Ahmedzai SH, Snowden JA. Living with advanced but stable multiple myeloma: a study of the symptom burden and cumulative effects of disease and intensive (hematopoietic stem cell transplant-based) treatment on health-related quality of life. Journal of pain and symptom management. 2013;46(5):671-80.

23. Holmstrom MO, Gimsing P, Abildgaard N, Andersen NF, Helleberg C, Clausen NA, et al. Causes of early death in multiple myeloma patients who are ineligible for high-dose therapy with hematopoietic stem cell support: A study based on the nationwide Danish Myeloma Database. American Journal of Hematology. 2015;90(4):E73-4.

24. Mateos MV. Management of treatment-related adverse events in patients with multiple myeloma. Cancer Treatment Reviews. 2010;36(SUPPL. 2):S24-S32.

25. Richardson PG, Delforge M, Beksac M, Wen P, Jongen JL, Sezer O, et al. Management of treatment-emergent peripheral neuropathy in multiple myeloma. Leukemia. 2012;26(4):595-608.

26. Molassiotis A, Wilson B, Blair S, Howe T, Cavet J. Unmet supportive care needs, psychological well-being and quality of life in patients living with multiple myeloma and their partners. Psycho-oncology. 2011;20(1):88-97.

27. Lokhorst HM, Plesner T, Laubach JP, Nahi H, Gimsing P, Hansson M, et al. Targeting CD38 with Daratumumab Monotherapy in Multiple Myeloma. N Engl J Med. 2015;373(13):1207-19.

28. Ramsenthaler C, Kane P, Gao W, Siegert RJ, Edmonds PM, Schey SA, et al. Prevalence of symptoms in patients with multiple myeloma: a systematic review and meta-analysis. Eur J Haematol. 2016;97(5):416-29.

29. Johnsen AT, Tholstrup D, Petersen MA, Pedersen L, Groenvold M. Health related quality of life in a nationally representative sample of haematological patients. European journal of haematology. 2009;83(2):139-48.

30. Jordan K, Proskorovsky I, Lewis P, Ishak J, Payne K, Lordan N, et al. Effect of general symptom level, specific adverse events, treatment patterns, and patient characteristics on health-related quality of life in patients with multiple myeloma: results of a European, multicenter cohort study. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2014;22(2):417-26.

31. Ramsenthaler C, Osborne TR, Gao W, Siegert RJ, Edmonds PM, Schey SA, et al. The impact of disease-related symptoms and palliative care concerns on health-related quality of life in multiple myeloma: a multi-centre study. BMC Cancer. 2016;16:427.

32. Priscilla D, Hamidin A, Azhar MZ, Noorjan KON, Salmiah MS, Bahariah K. Quality of life among patients with hematological cancer in a Malaysian hospital. Medical Journal of Malaysia. 2011;66(2):117-20.

33. Gregersen H, Vangsted AJ, Abildgaard N, Andersen NF, Pedersen RS, Frolund UC, et al. The impact of comorbidity on mortality in multiple myeloma: a Danish nationwide population-based study. Cancer medicine. 2017.

34. Palumbo A, Bringhen S, Mateos MV, Larocca A, Facon T, Kumar SK, et al. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. Blood. 2015;125(13):2068-74.

35. Blade J, Rosinol L. Changing paradigms in the treatment of multiple myeloma. Haematologica. 2009;94(2):163-6.

36. Kumar SK, Rajkumar SV, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK, et al. Improved survival in multiple myeloma and the impact of novel therapies. Blood. 2008;111(5):2516-20.

37. Abildgaard N, Vangsted A, Gregersen H, Frost Andersen N, Schou Pedersen R, Plesner T, et al. Improvement of overall survival in danish multiple myeloma patients after 2008; a population-based study from the danish national multiple myeloma registry. Haematologica. 2016;101 (Supplement 1):537.

38. Kumar SK, Dispenzieri A, Lacy MQ, Gertz MA, Buadi FK, Pandey S, et al. Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. Leukemia. 2014;28(5):1122-8.

39. Lenhoff S, Hjorth M, Holmberg E, Turesson I, Westin J, Nielsen JL, et al. Impact on survival of high-dose therapy with autologous stem cell support in patients younger than 60 years with newly diagnosed multiple myeloma: a population-based study. Nordic Myeloma Study Group. Blood. 2000;95(1):7-11.

40. Ocio EM, Richardson PG, Rajkumar SV, Palumbo A, Mateos MV, Orlowski R, et al. New drugs and novel mechanisms of action in multiple myeloma in 2013: a report from the International Myeloma Working Group (IMWG). Leukemia. 2014;28(3):525-42.

41. Attal M, Harousseau JL, Stoppa AM, Sotto JJ, Fuzibet JG, Rossi JF, et al. A prospective, randomized trial of autologous bone marrow transplantation and chemotherapy in multiple myeloma. Intergroupe Francais du Myelome. N Engl J Med. 1996;335(2):91-7.

42. Mateos MV, San Miguel JF. Management of multiple myeloma in the newly diagnosed patient. Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program. 2017;2017(1):498-507.

43. Sonneveld P, Goldschmidt H, Rosinol L, Blade J, Lahuerta JJ, Cavo M, et al. Bortezomib-based versus nonbortezomib-based induction treatment before autologous stem-cell transplantation in patients with previously untreated multiple myeloma: a meta-analysis of phase III randomized, controlled trials. J Clin Oncol. 2013;31(26):3279-87.

44. Engelhardt M, Terpos E, Kleber M, Gay F, Wasch R, Morgan G, et al. European Myeloma Network recommendations on the evaluation and treatment of newly diagnosed patients with multiple myeloma. Haematologica. 2014;99(2):232-42.

45. Palumbo A, Sezer O, Kyle R, Miguel JS, Orlowski RZ, Moreau P, et al. International Myeloma Working Group guidelines for the management of multiple myeloma patients ineligible for standard high-dose chemotherapy with autologous stem cell transplantation. Leukemia. 2009;23(10):1716-30.

46. San Miguel JF, Schlag R, Khuageva NK, Dimopoulos MA, Shpilberg O, Kropff M, et al. Persistent overall survival benefit and no increased risk of second malignancies with bortezomib-melphalan-prednisone versus melphalan-prednisone in patients with previously untreated multiple myeloma. J Clin Oncol. 2013;31(4):448-55.

47. Benboubker L, Dimopoulos MA, Dispenzieri A, Catalano J, Belch AR, Cavo M, et al. Lenalidomide and dexamethasone in transplant-ineligible patients with myeloma. The New England journal of medicine. 2014;371(10):906-17.

48. Mateos MV, Richardson PG, Schlag R, Khuageva NK, Dimopoulos MA, Shpilberg O, et al. Bortezomib plus melphalan and prednisone compared with melphalan and prednisone in previously untreated multiple myeloma: updated follow-up and impact of subsequent therapy in the phase III VISTA trial. Journal of clinical oncology. 2010;28(13):2259-66.

49. Durie BG, Hoering A, Abidi MH, Rajkumar SV, Epstein J, Kahanic SP, et al. Bortezomib with lenalidomide and dexamethasone versus lenalidomide and dexamethasone alone in patients with newly diagnosed myeloma without intent for immediate autologous stem-cell transplant (SWOG S0777): a randomised, open-label, phase 3 trial. Lancet. 2017;389(10068):519-27.

50. Laubach J, Garderet L, Mahindra A, Gahrton G, Caers J, Sezer O, et al. Management of relapsed multiple myeloma: recommendations of the International Myeloma Working Group. Leukemia. 2016;30(5):1005-17.

51. Tariman JD, Doorenbos A, Schepp KG, Singhal S, Berry DL. Older adults newly diagnosed with symptomatic myeloma and treatment decision making. Oncology nursing forum. 2014;41(4):411-9.

52. Tariman JD, Doorenbos A, Schepp KG, Becker PS, Berry DL. Patient, Physician and Contextual Factors Are Influential in the Treatment Decision Making of Older Adults Newly Diagnosed with Symptomatic Myeloma. Cancer treatment communications. 2014;2(2-3):34-47.

53. Deber RB, Kraetschmer N, Urowitz S, Sharpe N. Do people want to be autonomous patients? Preferred roles in treatment decision-making in several patient populations. Health expectations : an international journal of public participation in health care and health policy. 2007;10(3):248-58.

54. Leleu X, Mateos M, Delforge M, Lewis P, Schindler T, Bibson C, et al. Assessment of Multiple Myeloma Patient Preferences on Treatment Choices: An International Discrete Choice Study. Blood. 2015;126(23):2086.

55. Waage A, Gimsing P, Juliusson G, Turesson I, Gulbrandsen N, Eriksson T, et al. Early response predicts thalidomide efficiency in patients with advanced multiple myeloma. British journal of haematology. 2004;125(2):149-55.

56. Hjorth M, Hjertner O, Knudsen LM, Gulbrandsen N, Holmberg E, Pedersen PT, et al. Thalidomide and dexamethasone vs. bortezomib and dexamethasone for melphalan refractory myeloma: a randomized study. European journal of haematology. 2012;88(6):485-96.

57. Miguel JS, 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. The Lancet Oncology. 2013;14(11):1055-66.

58. Moreau P, Masszi T, Grzasko N, Bahlis NJ, Hansson M, Pour L, et al. Oral Ixazomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. The New England journal of medicine. 2016;374(17):1621-34.

59. Stewart AK, Rajkumar SV, Dimopoulos MA, Masszi T, Spicka I, Oriol A, et al. Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. The New England journal of medicine. 2015;372(2):142-52.

60. Richardson PG, Barlogie B, Berenson J, Singhal S, Jagannath S, Irwin D, et al. A phase 2 study of Bortezomib in relapsed, refractory myeloma. New England Journal of Medicine. 2003;348(26):2609-17.

61. 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.

62. Lonial S, Dimopoulos M, Palumbo A, White D, Grosicki S, Spicka I, et al. Elotuzumab therapy for relapsed or refractory multiple myeloma. New England Journal of Medicine. 2015;373(7):621-31.

63. San-Miguel J, Hungria VT, Yoon SS, Beksac M, Dimopoulos MA, Elghandour A, et al. Randomized, double-blind, placebo-controlled phase 3 study of panobinostat or placebo plus bortezomib and dexamethasone in relapsed or relapsed and refractory multiple myeloma (panorama 1). Haematologica. 2014;99:219-20.

64. Stohr E, Schmeel FC, Schmeel LC, Hanel M, Schmidt-Wolf IG. Bendamustine in heavily pre-treated patients with relapsed or refractory multiple myeloma. J Cancer Res Clin Oncol. 2015;141(12):2205-12.

65. 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.

66. Suzuki K, Dimopoulos MA, Takezako N, Okamoto S, Shinagawa A, Matsumoto M, et al. Daratumumab, lenalidomide, and dexamethasone in East Asian patients with relapsed or refractory multiple myeloma: subgroup analyses of the phase 3 POLLUX study. Blood cancer journal. 2018;8(4):41.

67. Varga C, Maglio M, Ghobrial IM, Richardson PG. Current use of monoclonal antibodies in the treatment of multiple myeloma. Br J Haematol. 2018;181(4):447-59.

68. Fayers PM, Machin D. Quality of life : the assessment, analysis and interpretation of patient-reported outcomes. Chichester: John Wiley; 2007. xxi, 544 s., illustreret p.

69. Carr AJ, Gibson B, Robinson PG. Measuring quality of life: Is quality of life determined by expectations or experience? BMJ (Clinical research ed). 2001;322(7296):1240-3.

70. Calman KC. Quality of life in cancer patients--an hypothesis. Journal of medical ethics. 1984;10(3):124-7.

71. Allison PJ, Locker D, Feine JS. Quality of life: a dynamic construct. Soc Sci Med. 1997;45(2):221-30.

72. Albrecht GL, Devlieger PJ. The disability paradox: high quality of life against all odds. Social science & medicine (1982). 1999;48(8):977-88.

73. Moons P, Budts W, De Geest S. Critique on the conceptualisation of quality of life: a review and evaluation of different conceptual approaches. International journal of nursing studies. 2006;43(7):891-901.

74. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res. 2000;9(8):887-900.

75. Cella DF. Quality of life: concepts and definition. Journal of pain and symptom management. 1994;9(3):186-92.

76. Bakas T, McLennon SM, Carpenter JS, Buelow JM, Otte JL, Hanna KM, et al. Systematic review of health-related quality of life models. Health Qual Life Outcomes. 2012;10:134.

77. Osoba D. Lessons learned from measuring health-related quality of life in oncology. J Clin Oncol. 1994;12(3):608-16.

78. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. Jama. 1995;273(1):59-65.

79. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual model of health-related quality of life. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing. 2005;37(4):336-42.

80. Luckett T, King MT. Choosing patient-reported outcome measures for cancer clinical research--practical principles and an algorithm to assist non-specialist researchers. Eur J Cancer. 2010;46(18):3149-57.

81. US Department of Health and Human Services Food and Drug Administration. US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: Use in medical product development to support labeling claims. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf [updated December 2009.

82. Ibfelt EH, Jensen DV, Hetland ML. The Danish nationwide clinical register for patients with rheumatoid arthritis: DANBIO. Clinical epidemiology. 2016;8:737-42.

83. Basch E, Dueck AC, Rogak LJ, Minasian LM, Kelly WK, O'Mara AM, et al. Feasibility Assessment of Patient Reporting of Symptomatic Adverse Events in Multicenter Cancer Clinical Trials. JAMA oncology. 2017;3(8):1043-50.

84. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol. 2014;32(14):1480-501.

85. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. N Engl J Med. 2017;376(2):105-8.

86. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. Jama. 2017;318(2):197-8.

87. Viala M, Bhakar AL, de la Loge C, van de Velde H, Esseltine D, Chang M, et al. Patient-reported outcomes helped predict survival in multiple myeloma using partial least squares analysis. Journal of clinical epidemiology. 2007;60(7):670-9.

88. Dubois D, Dhawan R, van de Velde H, Esseltine D, Gupta S, Viala M, et al. Descriptive and prognostic value of patient-reported outcomes: the bortezomib experience in relapsed and refractory multiple myeloma. Journal of Clinical Oncology. 2006;24(6):976-82.

89. Efficace F, Cottone F, Abel G, Niscola P, Gaidano G, Bonnetain F, et al. Patient-reported outcomes enhance the survival prediction of traditional disease risk classifications: An international study in patients with myelodysplastic syndromes. Cancer. 2018;124(6):1251-9.

90. Jerkeman M, Kaasa S, Hjermstad M, Kvaloy S, Cavallin-Stahl E. Health-related quality of life and its potential prognostic implications in patients with aggressive lymphoma: a Nordic Lymphoma Group Trial. Med Oncol. 2001;18(1):85-94.

91. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Medical care. 1996;34(3):220-33.

92. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85(5):365-76.

93. Huang HQ, Brady MF, Cella D, Fleming G. Validation and reduction of FACT/GOG-Ntx subscale for platinum/paclitaxel-induced neurologic symptoms: a gynecologic oncology group study. International journal of gynecological cancer : official journal of the International Gynecological Cancer Society. 2007;17(2):387-93.

94. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.

95. Osborne TR, Ramsenthaler C, Siegert RJ, Edmonds PM, Schey SA, Higginson IJ. What issues matter most to people with multiple myeloma and how well are we measuring them? A systematic review of quality of life tools. European journal of haematology. 2012;89(6):437-57.

96. Cocks K, Cohen D, Wisloff F, Sezer O, Lee S, Hippe E, et al. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. European journal of cancer (Oxford, England : 1990). 2007;43(11):1670-8.

97. Stead ML, Brown JM, Velikova G, Kaasa S, Wisloff F, Child JA, et al. Development of an EORTC questionnaire module to be used in health-related quality-of-life assessment for patients with multiple myeloma. European Organization for Research and Treatment of Cancer Study Group on Quality of Life. Br J Haematol. 1999;104(3):605-11.

98. Wisloff F, Eika S, Hippe E, Hjorth M, Holmberg E, Kaasa S, et al. Measurement of health-related quality of life in multiple myeloma. British Journal of Haematology. 1996;92(3):604-13.

99. Blade J, Calleja MA, Lahuerta JJ, Poveda JL, de Paz HD, Lizan L. Defining a set of standardised outcome measures for newly diagnosed patients with multiple myeloma using the Delphi consensus method: the IMPORTA project. The BMJ Open. 2018;8(2):e018850.

100. Fayers P, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). European Organisation for Research and Treatment of Canceer, Brussels 2001.

101. Osborne T, Ramsenthaler C, Siegert R, Edmonds P, Schey S, Higginson I. Health related quality of life instruments for use in people with multiple myeloma: A systematic literature review. Haematologica. 2011;96:366.

102. King TA, King MT, White KJ. Patient Reported Outcomes in Optimizing Myeloma Patients' Health-Related Quality of Life. Semin Oncol Nurs. 2017;33(3):299-315.

103. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clin Proc. 2002;77(4):371-83.

104. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert review of pharmacoeconomics & outcomes research. 2011;11(2):171-84.

105. King MT, Fayers PM. Making quality-of-life results more meaningful for clinicians. Lancet. 2008;371 North American Edition(9614):709-10 2p.

106. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-15.

107. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171-8.

108. Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Life Res. 1993;2(3):221-6.

109. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006;4:54.

110. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102-9.

111. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale N.J.: Lawrence Erlbaum Associates; 1988.

112. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical care. 2003;41(5):582-92.

113. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol. 1999;52(9):861-73.

114. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? Journal of biopharmaceutical statistics. 2004;14(1):97-110.

115. Kvam AK, Fayers P, Wisloff F. What changes in health-related quality of life matter to multiple myeloma patients? A prospective study. European journal of haematology. 2010;84(4):345-53.

116. Cocks K, King MT, Velikova G, de Castro G, Jr., Martyn St-James M, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Eur J Cancer. 2012;48(11):1713-21.

117. Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol. 2011;29(1):89-96.

118. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. Qual Life Res. 2003;12(3):239-49.

119. Fayers PM, Hays RD. Don't middle your MIDs: regression to the mean shrinks estimates of minimally important differences. Qual Life Res. 2014;23(1):1-4.

120. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Social science & medicine (1982). 1999;48(11):1507-15.

121. Taminiau-Bloem EF, Schwartz CE, van Zuuren FJ, Koeneman MA, Visser MR, Tishelman C, et al. Using a retrospective pretest instead of a conventional pretest is replacing biases: a qualitative study of cognitive processes underlying responses to thentest items. Qual Life Res. 2016;25(6):1327-37.

122. Hamidou Z, Dabakuyo TS, Bonnetain F. Impact of response shift on longitudinal quality-of-life assessment in cancer clinical trials. Expert review of pharmacoeconomics & outcomes research. 2011;11(5):549-59.

123. Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal data. Psychology & Health. 2004;19(1):51-69.

124. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. The New England journal of medicine. 2012;367(14):1355-60.

125. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. BMC Medical Research Methodology. 2014;14:118.

126. National Research Council Panel on Handling Missing Data in Clinical T. The Prevention and Treatment of Missing Data in Clinical Trials. Washington (DC): National Academies Press (US)

Copyright 2010 by the National Academy of Sciences. All rights reserved.; 2010.

127. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Statistical methods in medical research. 2014;23(5):440-59.

128. Fairclough DL. Design and analysis of quality of life studies in clinical trials. 2. ed. ed. Boca Raton: Chapman & Hall/CRC; 2010. 404 p.

129. Mercieca-Bebber R, Calvert M, Kyte D, Stockler M, King MT. The administration of patient-reported outcome questionnaires in cancer trials: Interviews with trial coordinators regarding their roles, experiences, challenges and training. Contemporary Clinical Trials Communications. 2018;9:23-32.

130. Mercieca-Bebber R, Kyte DG, Calvert MJ, Stockler M, King MT. Challenges and training needs of study coordinators responsible for administration of patient-reported outcome questionnaires in cancer trials. Quality of Life Research. 2017;26 (1 Supplement 1):102.

131. Fairclough DL, Peterson HF, Cella D, Bonomi P. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. Stat Med. 1998;17(5-7):781-96.

132. Bernhard J, Cella DF, Coates AS, Fallowfield L, Ganz PA, Moinpour CM, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. Stat Med. 1998;17(5-7):517-32.

133. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. The Journal of the American Medical Association. 2018;319(5):483-94.

134. Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. The BMJ Open. 2016;6(6):e010938.

135. Palmer MJ, Mercieca-Bebber R, King M, Calvert M, Richardson H, Brundage M. A systematic review and development of a classification framework for factors associated with missing patient-reported outcome data. Clinical trials. 2018;15(1):95-106.

136. Maruish ME. User´s manual for the SF-12v2 Health Survey 3rd ed: Lincoln, RI: QualityMetric Incorporated.; 2012.

137. Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. Health and quality of life outcomes. 2009;7:57.

138. Rubin DB. Inference and Missing Data. Biometrika. 1976;63(3):581-92.

139. Schwartz CE, Feinberg RG, Jilinskaia E, Applegate JC. An evaluation of a psychosocial intervention for survivors of childhood cancer: paradoxical effects of response shift over time. Psychooncology. 1999;8(4):344-54.

140. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med. 1999;48(11):1531-48.

141. Schwartz C, Ahmed S, Sawatzky R, Sajobi T, Mayo N, Finkelstein J, et al. Guidelines for secondary analysis in search of response shift.

142. Visser MR, Smets EM, Sprangers MA, de Haes HJ. How response shift may affect the measurement of change in fatigue. Journal of pain and symptom management. 2000;20(1):12-8.

143. Liu JJ, Davis GE. The significance of response shift in sinus surgery outcomes. International forum of allergy & rhinology. 2015;5(1):55-9.

144. Yang J, Hanna-Pladdy B, Gruber-Baldini AL, Barr E, von Coelln R, Armstrong MJ, et al. Response shift - The experience of disease progression in Parkinson disease. Parkinsonism & related disorders. 2017;36:52-6.

145. Korfage IJ, de Koning HJ, Essink-Bot ML. Response shift due to diagnosis and primary treatment of localized prostate cancer: a then-test and a vignette study. Qual Life Res. 2007;16(10):1627-34.

146. Schwartz CE, Sprangers MA, Oort FJ, Ahmed S, Bode R, Li Y, et al. Response shift in patients with multiple sclerosis: an application of three statistical techniques. Qual Life Res. 2011;20(10):1561-72.

147. 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. European journal of haematology. 2011;87(4):330-7.

148. Schwartz CE, Bode R, Repucci N, Becker J, Sprangers MA, Fayers PM. The clinical significance of adaptation to changing health: a meta-analysis of response shift. Qual Life Res. 2006;15(9):1533-50.

149. Hamidou Z, Dabakuyo-Yonli TS, Guillemin F, Conroy T, Velten M, Jolly D, et al. Impact of response shift on time to deterioration in quality of life scores in breast cancer patients. PloS one. 2014;9(5):e96848.

150. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. Health Qual Life Outcomes. 2004;2:16.

151. Schwartz CE, Zhang J, Rapkin BD, Finkelstein JA. Reconsidering the minimally important difference: evidence of instability over time and across groups. The spine journal : official journal of the North American Spine Society. 2018.

152. Ousmen A, Conroy T, Guillemin F, Velten M, Jolly D, Mercier M, et al. Impact of the occurrence of a response shift on the determination of the minimal important difference in a health-related quality of life score over time. Health Qual Life Outcomes. 2016;14(1):167.

153. Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. Health and quality of life outcomes. 2004;2:14.

154. Schwartz CE, Ahmed S, Sawatzky R, Sajobi T, Mayo N, Finkelstein J, et al. Guidelines for secondary analysis in search of response shift. Qual Life Res. 2013;22(10):2663-73.

155. Schwartz CE, Sprangers MA. Guidelines for improving the stringency of response shift research using the thentest. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2010;19(4):455-64.

156. Schwartz CE, Li J, Rapkin BD. Refining a Web-based goal assessment interview: item reduction based on reliability and predictive validity. Qual Life Res. 2016;25(9):2201-12.

157. Rapkin BD, Schwartz CE. Distilling the essence of appraisal: a mixed methods study of people with multiple sclerosis. Qual Life Res. 2016;25(4):793-805.

158. Schwartz CE, Finkelstein JA, Rapkin BD. Appraisal assessment in patient-reported outcome research: methods for uncovering the personal context and meaning of quality of life. Qual Life Res. 2017;26(3):545-54.

159. Rapkin BD, Garcia I, Michael W, Zhang J, Schwartz CE. Development of a practical outcome measure to account for individual differences in quality-of-life appraisal: the Brief Appraisal Inventory. Qual Life Res. 2018;27(3):823-33.

160. Kvam AK, Wisloff F, Fayers PM. Minimal important differences and response shift in health-related quality of life; a longitudinal study in patients with multiple myeloma. Health and quality of life outcomes. 2010;8:79-7525-8-79.

161. Schwartz CE, Powell VE, Rapkin BD. When global rating of change contradicts observed change: Examining appraisal processes underlying paradoxical responses over time. Qual Life Res. 2017;26(4):847-57.

162. Dimopoulos MA, Delforge M, Hajek R, Kropff M, Petrucci MT, Lewis P, et al. Lenalidomide, melphalan, and prednisone, followed by lenalidomide maintenance, improves health-related quality of life in newly diagnosed multiple myeloma patients aged 65 years or older: results of a randomized phase III trial. Haematologica. 2013;98(5):784-8.

163. Waage A, Gimsing P, Fayers P, Abildgaard N, Ahlberg L, Bjorkstrand B, et al. Melphalan and prednisone plus thalidomide or placebo in elderly patients with multiple myeloma. Blood. 2010;116(9):1405-12.

164. Wisloff F, Hjorth M, Kaasa S, Westin J. Effect of interferon on the health-related quality of life of multiple myeloma patients: results of a Nordic randomized trial comparing melphalan-prednisone to melphalan-prednisone + alpha-interferon. The Nordic Myeloma Study Group. British journal of haematology. 1996;94(2):324-32.

165. Gimsing P, Carlson K, Turesson I, Fayers P, Waage A, Vangsted A, et al. Effect of pamidronate 30 mg versus 90 mg on physical function in patients with newly diagnosed multiple myeloma (Nordic Myeloma Study Group): A double-blind, randomised controlled trial. The Lancet Oncology. 2010;11(10):973-82.

166. Etto LY, Morelli VM, Silva VC, Hungria VT, Ciconelli RM, Almeida MS, et al. Autologous stem cell transplantation improves quality of life in economically challenged, Brazilian multiple myeloma patients. Clinics (Sao Paulo, Brazil). 2011;66(11):1855-9.

167. Gulbrandsen N, Wisloff F, Brinch L, Carlson K, Dahl IM, Gimsing P, et al. Health-related quality of life in multiple myeloma patients receiving high-dose chemotherapy with autologous blood stem-cell support. Medical oncology (Northwood, London, England). 2001;18(1):65-77.

168. 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. British journal of haematology. 2008;143(4):511-9.

169. 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.

170. Delforge M, Minuk L, Eisenmann JC, Arnulf B, Canepa L, Fragasso A, et al. Health-related quality-of-life in patients with newly diagnosed multiple myeloma in the FIRST trial: lenalidomide plus low-dose dexamethasone versus melphalan, prednisone, thalidomide. Haematologica. 2015;100(6):826-33.

171. Ludwig H, Viterbo L, Greil R, Masszi T, Spicka I, Shpilberg O, et al. Randomized phase II study of bortezomib, thalidomide, and dexamethasone with or without cyclophosphamide as induction therapy in previously untreated multiple myeloma. Journal of clinical oncology. 2013;31(2):247-55.

172. Delforge M, Dhawan R, Robinson D, Jr., Meunier J, Regnault A, Esseltine DL, et al. Health-related quality of life in elderly, newly diagnosed multiple myeloma patients treated with VMP vs. MP: results from the VISTA trial. European journal of haematology. 2012;89(1):16-27.

173. Verelst SG, Termorshuizen F, Uyl-de Groot CA, Schaafsma MR, Ammerlaan AH, Wittebol S, et al. Effect of thalidomide with melphalan and prednisone on health-related quality of life (HRQoL) in elderly patients with newly diagnosed multiple myeloma: a prospective analysis in a randomized trial. Annals of hematology. 2011;90(12):1427-39.

174. Gimsing P, Carlson K, Turesson I, Fayers P, Waage A, Vangsted A, et al. Effect of pamidronate 30 mg versus 90 mg on physical function in patients with newly diagnosed multiple myeloma (Nordic Myeloma Study Group): a double-blind, randomised controlled trial. The LancetOncology. 2010;11(10):973-82.

175. Wisloff F, Eika S, Hippe E, Hjorth M, Holmberg E, Kaasa S, et al. Measurement of health-related quality of life in multiple myeloma. Nordic Myeloma Study Group. Br J Haematol. 1996;92(3):604-13.

176. Wisloff F, Hjorth M, Kaasa S, Westin J. Effect of interferon on the health-related quality of life of multiple myeloma patients: Results of a Nordic randomized trial comparing melphalan-prednisone to melphalan-prednisone+alpha-interferon. British Journal of Haematology. 1996;94(2):324-32.

177. Zweegman S, van der Holt B, Mellqvist UH, Salomo M, Bos GM, Levin MD, et al. Melphalan, prednisone, and lenalidomide versus melphalan, prednisone, and thalidomide in untreated multiple myeloma. Blood. 2016;127(9):1109-16.

178. Gregersen H, Do T, Kristensen IB, Frølund UC, Andersen NF, Nielsen LK, et al. A randomized placebo-controlled phase II study of clarithromycin or placebo combined with VCD induction therapy prior to high-dose melphalan with stem cell support in patients with newly diagnosed multiple myeloma. Experimental Hematology & Oncology. 2018;7(1).

179. Gregersen H, Abildgaard N, Do T, Kristensen I, Frølund U, Anderssen N, et al. A randomized placebo-controlled phase II study of clarithromycin or placebo combined with VCD induction therapy prior to high-dose melphalan with stem cell support in patients with newly diagnosed multiple myeloma. 59th Annual Meeting of the American Society of Hematology, 2017, Abstract 3129.

180. Nielsen LK, Jarden M, Andersen CL, Frederiksen H, Abildgaard N. A systematic review of health-related quality of life in longitudinal studies of myeloma patients. European journal of haematology. 2017;99(1):3-17.

181. Stewart AK, Dimopoulos MA, Masszi T, Spicka I, Oriol A, Hajek R, et al. Health-Related Quality of Life Results From the Open-Label, Randomized, Phase III ASPIRE Trial Evaluating Carfilzomib, Lenalidomide, and Dexamethasone Versus Lenalidomide and Dexamethasone in Patients With Relapsed Multiple Myeloma. Journal of clinical oncology. 2016.

182. 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.

183. Anderson KO, Giralt SA, Mendoza TR, Brown JO, Neumann JL, Mobley GM, et al. Symptom burden in patients undergoing autologous stem-cell transplantation. Bone Marrow Transplant. 2007;39(12):759-66.

184. Astolfi S, Scaramuzzo L, Logroscino CA. A minimally invasive surgical treatment possibility of osteolytic vertebral collapse in multiple myeloma. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2009;18 Suppl 1:115-21.

185. Chen LH, Hsieh MK, Niu CC, Fu TS, Lai PL, Chen WJ. Percutaneous vertebroplasty for pathological vertebral compression fractures secondary to multiple myeloma. Archives of orthopaedic and trauma surgery. 2012;132(6):759-64.

186. Coluzzi F, Raffa RB, Pergolizzi J, Rocco A, Locarini P, Cenfra N, et al. Tapentadol prolonged release for patients with multiple myeloma suffering from moderate-to-severe cancer pain due to bone disease. Journal of Pain Research. 2015;8:229-38.

187. Groeneveldt L, Mein G, Garrod R, Jewell AP, Someren KV, Stephens R, et al. A mixed exercise training programme is feasible and safe and may improve quality of life and muscle strength in multiple myeloma survivors. BMC Cancer. 2013;13(31).

188. Henon P, Donatini B, Eisenmann JC, Becker M, Beck-Wirth G. Comparative survival, quality of life and cost-effectiveness of intensive therapy with autologous blood cell transplantation or conventional chemotherapy in multiple myeloma. Bone Marrow Transplantation. 1995;16(1):19-25.

189. Khanna AJ, Reinhardt MK, Togawa D, Lieberman IH. Functional outcomes of kyphoplasty for the treatment of osteoporotic and osteolytic vertebral compression fractures. Osteoporosis International. 2006;17(6):817-26.

190. Saccilotto IC, Bittencourt RI, Fischer CC, Quevedo A, Hirakata VN, Picon PD. Efficacy of a referral center for patient-centered care in multiple myeloma: a cohort study. BMC health services research. 2015;15(1):455.

191. Schaar CG, Kluin-Nelemans HC, Te Marvelde C, le Cessie S, Breed WP, Fibbe WE, et al. Interferon-alpha as maintenance therapy in patients with multiple myeloma. Annals of oncology : official journal of the European Society for Medical Oncology. 2005;16(4):634-9.

192. Schmitt T, Goldschmidt H, Neben K, Freiberger A, Husing J, Gronkowski M, et al. Aprepitant, granisetron, and dexamethasone for prevention of chemotherapy-induced nausea and vomiting after high-dose melphalan in autologous transplantation for multiple myeloma: results of a randomized, placebo-controlled phase III trial. J Clin Oncol. 2014;32(30):3413-20.

193. Sherman AC, Simonton S, Latif U, Plante TG, Anaissie EJ. Changes in quality-of-life and psychosocial adjustment among multiple myeloma patients treated with high-dose melphalan and autologous stem cell transplantation. Biol Blood Marrow Transplant. 2009;15(1):12-20.

194. Stewart AK, Jacobus S, Fonseca R, Weiss M, Callander NS, Chanan-Khan AA, et al. Melphalan, prednisone, and thalidomide vs melphalan, prednisone, and lenalidomide (ECOG E1A06) in untreated multiple myeloma. Blood. 2015;126(11):1294-301.

195. Zou J, Mei X, Gan M, Yang H. Kyphoplasty for spinal fractures from multiple myeloma. Journal of Surgical Oncology. 2010;102(1):43-7.

196. Mateos MV, Dimopoulos MA, Cavo M, Suzuki K, Jakubowiak A, Knop S, et al. Daratumumab plus Bortezomib, Melphalan, and Prednisone for Untreated Myeloma. N Engl J Med. 2018;378(6):518-28.

197. Gries K, Fastenau J, Chen Y, Wang J, Ho KF, Wroblewski S, et al. Health-related quality of life in patients with newly diagnosed multiple myeloma who are ineligible for stem cell transplantation: Results from the ALCYONE trial. Journal of Clinical Oncology. 2018;36(15_suppl):8042-.

198. Royle KL, Gregory WM, Cairns DA, Bell SE, Cook G, Owen RG, et al. Quality of life during and following sequential treatment of previously untreated patients with multiple myeloma: findings of the Medical Research Council Myeloma IX randomised study. Br J Haematol. 2018.

199. Morgan GJ, Davies FE, Gregory WM, Bell SE, Szubert AJ, Navarro Coy N, et al. Cyclophosphamide, thalidomide, and dexamethasone as induction therapy for newly diagnosed multiple myeloma patients destined for autologous stem-cell transplantation: MRC Myeloma IX randomized trial results. Haematologica. 2012;97(3):442-50.

200. Morgan GJ, Davies FE, Gregory WM, Russell NH, Bell SE, Szubert AJ, et al. Cyclophosphamide, thalidomide, and dexamethasone (CTD) as initial therapy for patients with multiple myeloma unsuitable for autologous transplantation. Blood. 2011;118(5):1231-8.

201. Ludwig H, Moreau P, Dimopoulos M, Mateos M, Kaiser M, Feng S, et al. Health related quality of life results from the open-label, randomized, phase III endeavor trial evaluating carfilzomib and dexamethasone versus bortezomib and dexamethasone in patients with relapsed or refractory multiple myeloma. Blood Conference: 58th Annual Meeting of the American Society of Hematology, ASH 2016 United States. 2016.

202. 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. Br J Haematol. 2018;181(5):628-36.

203. 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. Ann Hematol. 2018;97(12):2455-63.

204. Spencer A, Lentzsch S, Weisel K, Avet-Loiseau H, Mark TM, Spicka I, et al. Daratumumab plus bortezomib and dexamethasone versus bortezomib and dexamethasone in relapsed or refractory multiple myeloma: updated analysis of CASTOR. Haematologica. 2018.

205. Dimopoulos MA, San-Miguel J, Belch A, White D, Benboubker L, Cook G, et al. Daratumumab plus lenalidomide and dexamethasone versus lenalidomide and dexamethasone in relapsed or refractory multiple myeloma: updated analysis of POLLUX. Haematologica. 2018.

206. Wisloff F, Eika S, Hippe E, Hjorth M, Holmberg E, Kaasa S, et al. Measurement of health-related quality of life in multiple myeloma. Nordic Myeloma Study Group. British journal of haematology. 1996;92(3):604-13.

207. Stead ML, Brown JM, Velikova G, Kaasa S, Wisloff F, Child JA, et al. Development of an EORTC questionnaire module to be used in health-related quality-of-life assessment for patients with multiple myeloma. European Organization for Research and Treatment of Cancer Study Group on Quality of Life. British journal of haematology. 1999;104(3):605-11.

208. Thanarajasingam G, Minasian LM, Baron F, Cavalli F, De Claro RA, Dueck AC, et al. Beyond maximum grade: modernising the assessment and reporting of adverse events in haematological malignancies. The Lancet Haematology. 2018.

209. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst. 2014;106(9).

210. Petersen MA, Aaronson NK, Arraras JI, Chie WC, Conroy T, Costantini A, et al. The EORTC CAT Core-The computer adaptive version of the EORTC QLQ-C30 questionnaire. Eur J Cancer. 2018;100:8-16.

211. Efficace F, Gaidano G, Lo-Coco F. Patient-reported outcomes in hematology: is it time to focus more on them in clinical trials and hematology practice? Blood. 2017;130(7):859-66.

212. Sonneveld P, Verelst SG, Lewis P, Gray-Schopfer V, Hutchings A, Nixon A, et al. Review of health-related quality of life data in multiple myeloma patients treated with novel agents. Leukemia. 2013;27(10):1959-69.

213. Acaster S, Gaugris S, Velikova G, Yong K, Lloyd AJ. Impact of the treatment-free interval on health-related quality of life in patients with multiple myeloma: a UK cross-sectional survey. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2013;21(2):599-607.

214. Palumbo A, Hajek R, Delforge M, Kropff M, Petrucci MT, Catalano J, et al. Continuous lenalidomide treatment for newly diagnosed multiple myeloma. The New England journal of medicine. 2012;366(19):1759.

215. Lavori PW, Brown CH, Duan N, Gibbons RD, Greenhouse J. Missing Data in Longitudinal Clinical Trials Part A: Design and Conceptual Issues. Psychiatric Annnals. 2008;38(12):784-92.

216. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. The BMJ. 2013;346:e8668.

217. Klee MC, King MT, Machin D, Hansen HH. A clinical model for quality of life assessment in cancer patients receiving chemotherapy. Annals of oncology : official journal of the European Society for Medical Oncology. 2000;11(1):23-30.

218. Coon CD, Cappelleri JC. Interpreting Change in Scores on Patient-Reported Outcome Instruments. Therapeutic innovation & regulatory science. 2016;50(1):22-9.

219. Speight J, Barendse SM. FDA guidance on patient reported outcomes. The BMJ. 2010;340:c2921.

220. Osoba D, Bezjak A, Brundage M, Zee B, Tu D, Pater J. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. Eur J Cancer. 2005;41(2):280-7.

221. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res. 2018;27(1):33-40.

222. Mols F, Thong MS, van de Poll-Franse LV, Roukema JA, Denollet J. Type D (distressed) personality is associated with poor quality of life and mental health among 3080 cancer survivors. Journal of affective disorders. 2012;136(1-2):26-34.

223. Robinson D, Jr., Esseltine DL, Regnault A, Meunier J, Liu K, van de Velde H. The influence of baseline characteristics and disease stage on health-related quality of life in multiple myeloma: findings from six randomized controlled trials. Br J Haematol. 2016;174(3):368-81.

224. Schwartz CE, Sajobi TT, Verdam MG, Sebille V, Lix LM, Guilleux A, et al. Method variation in the impact of missing data on response shift detection. Qual Life Res. 2015;24(3):521-8.

225. Musoro ZJ, Hamel JF, Ediebah DE, Cocks K, King MT, Groenvold M, et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. BMJ Open. 2018;8(1):e019117.

226. Musoro JZ, Bottomley A, Coens C, Eggermont AM, King MT, Cocks K, et al. Interpreting European Organisation for Research and Treatment for Cancer Quality of life Questionnaire core 30 scores as minimally importantly different for patients with malignant melanoma. Eur J Cancer. 2018;104:169-81.

227. EORTC. EORTC Quality of life group translation procedure: European Organisation for Research and Treatment of Cancer; 2016 [Fourth:[

228. Bottomley A, Pe M, Sloan J, Basch E, Bonnetain F, Calvert M, et al. Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. Clinical trials. 2018:1740774518795637.

229. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association. 1988;83(404):1198-202.

230. Rombach I, Gray AM, Jenkinson C, Murray DW, Rivero-Arias O. Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. BMC Med Res Methodol. 2018;18(1):87.

231. Cornish RP, Macleod J, Carpenter JR, Tilling K. Multiple imputation using linked proxy outcome data resulted in important bias reduction and efficiency gains: a simulation study. Emerging themes in epidemiology. 2017;14:14.

232. Pe M, Dorme L, Coens C, Basch E, Calvert M, Campbell A, et al. Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. The Lancet Oncology. 2018;19(9):e459-e69.

233. Klausen TW, Gregersen H, Abildgaard N, Andersen NF, Frolund UC, Gimsing P, et al. The majority of newly diagnosed myeloma patients do not fulfill the inclusion criteria in clinical phase III trials. Leukemia. 2018.

Appendices

Appendix I

Paper I and supplementary material

REVIEW ARTICLE

WILEY | Haematology

A systematic review of health-related quality of life in longitudinal studies of myeloma patients

Lene Kongsgaard Nielsen[1,2] ORCID | Mary Jarden[3] | Christen Lykkegaard Andersen[4] | Henrik Frederiksen[1,2] | Niels Abildgaard[1,2]

[1]Quality of Life Research Centre, Department of Haematology, Odense University Hospital, Odense, Denmark

[2]The Academy of Geriatric Cancer Research (AgeCare), Odense University Hospital, Odense, Denmark

[3]University Hospitals Centre for Health Research, Copenhagen University Hospital, Copenhagen, Denmark

[4]Department of Haematology, Roskilde University Hospital, Roskilde, Denmark

Correspondence
Lene Kongsgaard Nielsen, Quality of Life Research Centre, Department of Haematology, Odense University Hospital, Odense, Denmark.
Email: lene.kongsgaard.nielsen@rsyd.dk

Abstract

Objectives: Multiple myeloma (MM) patients report high symptom burden and reduced health-related quality of life (HRQoL) compared to patients with other haematological malignancies. The aim of this review was to analyse published longitudinal studies including MM patients according to a change in HRQoL scores, which is perceived as beneficial to the patient according to two published guidelines.

Methods: A literature search was performed May 2016. Publications with longitudinal follow-up using the EORTC QLQ-C30 instrument for HRQoL measurement of physical functioning, global quality of life, fatigue and/or pain were included. An analysis of mean change from baseline was carried out according to minimal important difference (MID).

Results: Large and medium HRQoL improvements were reported during first-line treatments. No clinically beneficial change or deteriorations in scores of global QoL or fatigue were reported during relapse treatment. HRQoL data during maintenance therapy are sparse and inconclusive.

Conclusions: Guidelines for interpreting changes in HRQoL including definitions of MID have been developed; however, consensus is missing. Improvements in HRQoL are far more likely to occur during first-line compared to relapsed treatment regimens. The background of these findings should be in focus in future studies, and HRQoL measurements should be integrated in maintenance studies.

KEYWORDS
Health-related quality of life, multiple myeloma, longitudinal studies

1 | INTRODUCTION

The prognosis of multiple myeloma (MM) has improved markedly over the past 20 years, and the median survival of MM patients under the age of 70 at the time of diagnosis now exceeds 6-7 years.[1,2] The prognosis is expected to improve further in the coming years due to new generations of proteasome inhibitors, third-generation immunomodulatory drugs, monoclonal antibodies and epigenetic therapies.[3]

Patients with MM report high symptom burden and low health-related quality of life (HRQoL) compared to patients with other haematological malignancies.[4-7] Along with improvements in survival, HRQoL is an increasingly important dimension in treatment and care

of myeloma patients. HRQoL has been an endpoint in several clinical trials as Wisloff et al.[8] more than 20 years ago published the first results of patient-reported HRQoL during primary treatment of MM patients.

Several validated cancer-specific instruments are available for HRQoL measurement. The European Organisation for Research and Treatment of Cancer Quality of life questionnaire, EORTC QLQ-C30 (QLQ-C30),[9] is the most frequently used instrument for cancer-specific HRQoL assessment (scale range 0-100) in myeloma patients. In addition, the myeloma cancer module, EORTC QLQ-MY20,[10] is often used. However, consensus of interpretation of change in HRQoL scores over time is lacking. As King et al.[11] in 1996 published the first review on

how to interpret HRQoL scores obtained by QLQ-C30, the method of interpretation of meaningful changes in many longitudinal studies has moved from focusing not only on statistical significance, but also on minimal important difference (MID).[12]

MID was defined in 1989 by Jaeschke et al. as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which will mandate, in the absence of troublesome side effects and excessive costs, a change in a patients' management".[13] As a statistically significant change in HRQoL score is not necessarily important to the patient, MID is a more useful way of interpreting changes in HRQoL scores for clinical decision-making.[14,15]

Kvam et al. defined clinically meaningful score changes in the QLQ-C30 instrument in patients with MM for the four domains: physical function, global quality of life (global QoL), fatigue and pain.[16] The calculations of MID are based on QLQ-C30 scores from 239 patients in different stages of the myeloma disease. An anchor-derived method based on the addition of a single question on patient's perception of their change in QoL was used to calculate MID for the four domains in the situation of deterioration and improvement. The result of the mean change by global rating of change suggests that the MID may be affected by the direction of change and that the calculated MID is domain specific.

Cocks et al. has published an Evidence-Based Interpretation Guidelines[17] with the aim to improve current guidelines for sample size calculation and interpretation by utilising published study results of QLQ-C30 outcomes. The method of estimating the clinically relevant change was a meta-analytical technique in combination with blinded expert opinions. The experts were asked to make a judgement on the relative size of the change over time according to four levels of magnitude: large, medium, small and trivial. Large changes were with obvious and unequivocal clinical relevance, and medium changes were likely to be clinically relevant but to a lesser extent. Small changes indicated that a subtle, but nevertheless clinically relevant and a trivial change was unlikely to have any clinical relevance.

The main objective of this systematic review was to analyse the published longitudinal studies of MM patients according to a change in HRQoL scores, which are perceived as clinically beneficial to the patient for the four domains of physical function, global QoL, fatigue and/or pain according to the guidelines by Kvam[16] and Cocks.[17]

2 | MATERIAL AND METHODS

2.1 | Search strategy

A literature search was performed on 24 May 2016 in PubMed, EMBASE, PsycINFO and CINAHL. The PubMed searches were carried out using the MeSH terms "Multiple Myeloma" or title/abstract "Multiple Myeloma" or MeSH term "Myelomatosis" or title/abstract "Myelomatosis" AND MeSH term "Quality of life" or title/abstract "Quality of life" or MeSH term "Life quality" or title/abstract "life quality." The same search was repeated in the three other databases and collected in EndNote X7. The publications were collected in EndNote X7 and duplicates were removed automatically, as

well as manually. Titles and abstracts of the remaining publications were reviewed first by title/abstracts and next by full-text reading to identify eligible publications.

2.2 | Publication selection

A publication was qualified for the review, if the following criteria for inclusion were met; the patients in the study were diagnosed with MM. The study was designed as a longitudinal follow-up using the QLQ-C30 instrument for HRQoL measurement of physical function, global QoL, fatigue and/or pain, and mean score at baseline and minimum one follow-up time point are presented in the text, a table or a figure, or the change in mean score from baseline was specified. Baseline is defined as start of treatment, consolidation, maintenance or observation. Publications concerning patients with mixed haematological diagnoses were excluded, if the results were not presented for MM patients separately. Articles in languages other than English were also excluded. There was no time limit set for the literature search. The literature search and publication selection are illustrated in Figure 1 by the PRISMA diagram flow.[18]

2.3 | Data analysis

The included studies were divided into six categories; 1) first-line treatment studies including induction therapy and autologous haematopoietic stem cell transplantation (ASCT), 2) first-line treatment studies without ASCT, 3) consolidation treatment studies, hereunder reporting from the ASCT regimens only, 4) maintenance treatment studies, 5) relapse treatment studies and 6) non-interventional studies.

For the four HRQoL domains (physical function, global QoL, fatigue and/or pain), the mean change from baseline was calculated for every follow-up time point by subtracting the mean score at follow-up time point from the mean score at baseline. For studies specifying the number of patients at follow-up time points, the data evaluation was terminated if the patient dropout rate increased to 75% from baseline. The size of every calculated mean change from baseline for each domain was interpreted according to Kvams' MID criteria[16] and the guideline of Cocks.[17]

3 | RESULTS

3.1 | Studies included for data analysis and interpretation

A total of 23 longitudinal studies were included in this systematic review, corresponding to three first-line treatment studies including induction therapy and ASCT, eight first-line treatment studies without ASCT, three consolidation treatment studies, two maintenance treatment studies and seven relapse treatment studies and one non-interventional study. In one of the studies, HRQoL data for both "first-line treatment not including ASCT" and "maintenance treatment" are presented separately, and the study was included in both categories. The studies are listed according to categories in Table 1.
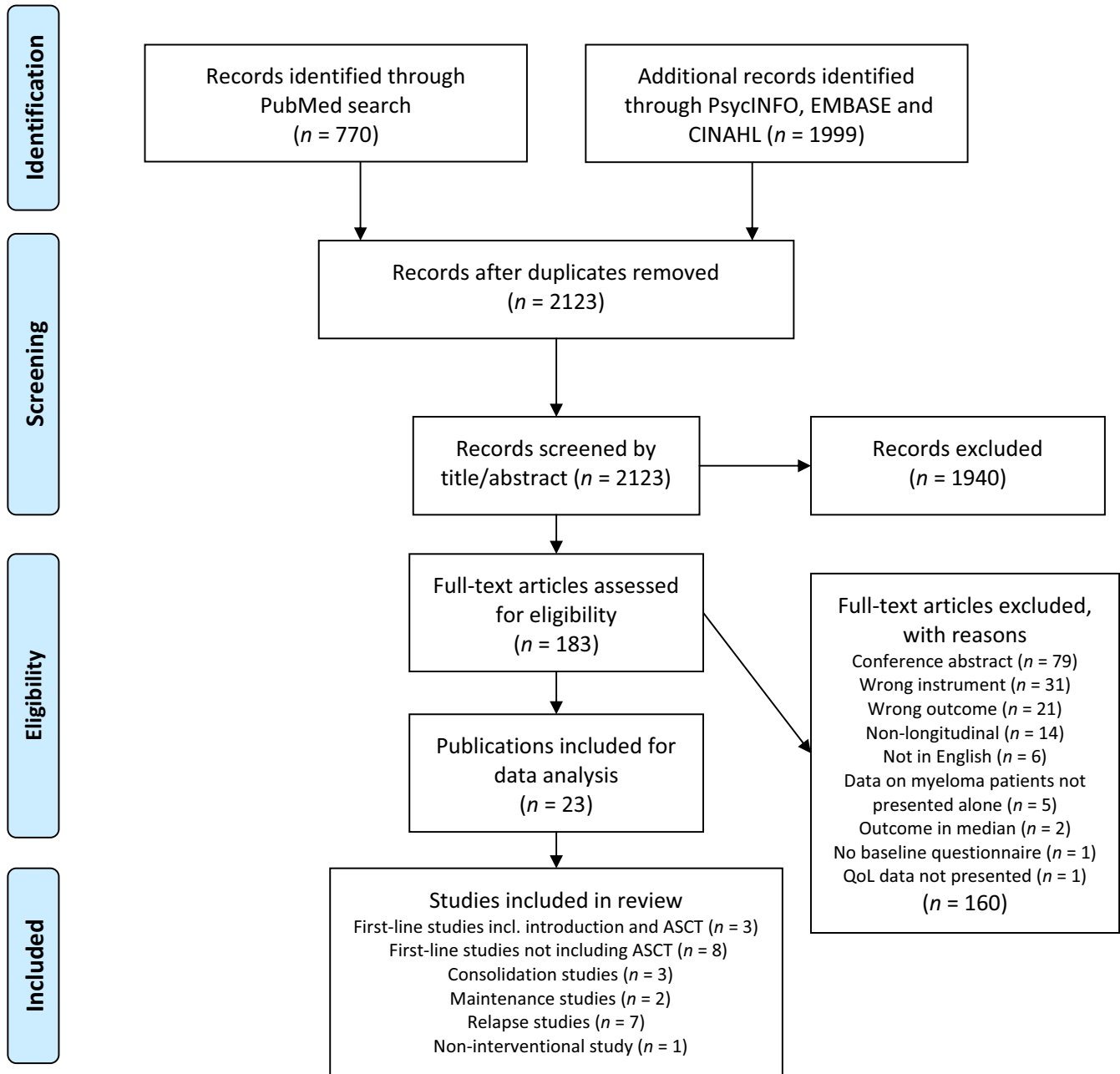
FIGURE 1 PRISMA flow diagram of the literature search of PubMed. ASCT, autologous haematopoietic stem cell transplantation; QoL, Quality of life

3.2 | The interpretation of the mean change over time by Kvam vs Cocks

The main difference between the two guidelines are that a small change in mean score from baseline defined as a subtle, but nevertheless clinically relevant using guideline of Cocks, is evaluated as no clinically meaningful change using Kvams' MID criteria. Also, improvement in fatigue and pain or deterioration in pain, which is perceived as beneficial to MM patients, requires a considerably larger change in score according to Kvams' MID criteria than defined by guideline of Cocks. The changes in scores are listed in Table 2.

3.3 | First-line studies including induction therapy and ASCT

Three studies were included in this category[19-21] and the patient-reported range of mean change in score from baseline for global QoL and physical functioning is illustrated in Figure 2, and for fatigue and pain in Figure 3. Patients reported deterioration in global QoL of up to 6.2 points from baseline to completion of induction treatment with bortezomib-thalidomide-cyclophosphamide-dexamethasone (VTD-Cyclo). After ASCT, the largest reported improvement was for fatigue with a mean change in score of 32 points.

TABLE 1 The publications included in the systematic review divided into six categories

Author, year of publication	Study design	Treatment regime	Baseline No. of patients	Mean age	Mean baseline score				No. follow-up time points presented	Last follow-up HRQoL presented
					Physical function	Global QoL	Fatigue	Pain		
First-line treatment studies including induction therapy and ASCT										
Ludwig et al. 2013 [19]	Randomized open-label phase II study	VTD + ASCT	46	57	-	53,6	-	-	7	Pre-PD follow-up visit[a]
		VTD-Cyclo + ASCT	46	58	-	56,2	-	-		
Etto et al. 2011 [20]	Phase II study[b]	ASCT	29	55.4	43	64	53	58	3	Day 100 after ASCT[a]
Gulbrandsen et al. 2001 [21]	Evaluation, phase II trial	ASCT	274	52	54	49	43	48		36 mo
		Historical control group	113	55	51	48	49	48	6	
First-line treatment studies without ASCT										
Delforge et al. 2015 [22] FIRST trial	Randomized phase III study	Len-dex	1076	73	61,5	51,7	44,8	44,7	6 and SD	18 mo and SD[a]
		MPT	547	73	59,9	50,8	46	45,3		
Dimopoulos et al. 2013 [23] MM-015 study	Three-armed randomized double-blind placebo-controlled phase III study	MPR(-R maintenance)	152	71	57,5	49,6	49,3	48,8		End of induction (day 1 of cycle 10)[a]
		MPR(-placebo maintenance)	153	71	64	53,2	44,2	44,9	3	
		MP placebo (-placebo maintenance)	154	72	65,5	52,8	42,6	43,5		
Delforge et al. 2012 [24] VISTA trial	Randomized phase III study	VMP	331	71	59,4	50,3	46	45,1	10	Day 1 of cycle 9 and EOT[a]
		MP	318	71	60,9	50,3	45,4	45,1		
Verelst et al. 2011 [25] HOVON 49	Randomized phase III study	MPT and Thalidomide maintenance[c]	135	71	58	54	43	49	4 and off protocol	18 mo and off protocol
		MP	149	72	55	49	47	50		
Gimsing et al. 2010 [26]	Randomized double-blind phase III trial	Pamidronate 30 mg	252	62	50	44	55	60	13	36 mo
		Pamidronate 90 mg	250	63	54	47	55	57		
Waage et al. 2010 [27]	Randomized double-blind placebo-controlled phase III study	MPT and Thalidomide maintenance[c]	182	74.6	53	48	54	54	13	36 mo
		MP placebo and placebo maintenance[c]	175	74.1	57	47	55	58		
Wisloff et al. 1996 [28] NMSG 4/90	Evaluation study	Evaluation group	524	66	47,9	45,4	51,9	50,9	2	6 mo
Wisloff et al. 1996 [29] NMSG 4/90[d]	Randomized phase III study	MP	271	-	48	47	51	50	6	36 mo
		MP-IFN	253	-	48	44	53	52		
Consolidation treatment studies										
Mellqvist et al. 2013 [30]	Randomized phase II trial after ASCT	Bortezomib	187	59.1	-	-	31	-	12	42 mo
		Control	183	58.7	-	-	32	-		
Frödin et al. 2011 [31]	Phase II study	ASCT	56	-	73	61	36	28	4	3 y
Khalafallah et al. 2011 [32]	Phase II study	Tandem ASCT	18	57.5	57,2	41,7	-	-	5	2. and subsequent follow-up[a]

(Continues)

TABLE 1 (Continued)

Author, year of publication	Study design	Treatment regime	Baseline No. of patients	Mean age	Mean baseline score				No. follow-up time points presented	Last follow-up HRQoL presented
					Physical function	Global QoL	Fatigue	Pain		
Maintenance treatment studies										
Dimopoulos et al. 2013 [23] MM-015 study	Randomized double-blind placebo-controlled phase III study.	Lenalidomid maintenance after MPR	90[e]	-	68,6	63,7	39,4	31,7	2	Day 1 of maintenance cycle 6[a]
		Placebo maintenance after MPR	91[e]	-	71	61,4	36,3	28,1		
		Placebo maintenance after MP placebo	100[e]	-	72,2	61,5	33,3	31,2		
Sirohi et al. 2007 [33]	Randomized phase II study	P-IFN	30	50	-	-	-	-	2	3 mo
		IFN	30	52	-	-	-	-		
Relapse treatment studies										
Moreau et al. 2016 [34] TOURMALINE-MM1	Randomized, double-blind, placebo-controlled, phase III trial	Ixazomib Len-dex	360	66	-	58	-	-	18 and EOT	Day 1 of 34 cycle and EOT[a]
		Placebo Len-dex	362	66	-	56	-	-		
Stewart et al. 2015 [35] ASPIRE trial	Randomized phase III study	Carfilzomib Len-dex	396	64	-	58,5	-	-	5	18 mo
		Len-dex	396	65	-	58,5	-	-		
Song et al. 2015 [36] MM-003	Randomized phase III study	Pomalidomid-Lodex	271[f]	63.5	-	-	-	-	11	Day 1 of cycle 10[a]
		HiDEX	134[f]	64	-	-	-	-		
Hjorth et al. 2012 [37] NMSG 17/07	Randomized phase III study[g]	Thalidomide-dex	67	71	68	59	42	40	5	52 wks
		Bortezomib-dex	64	71	63	58	47	43		
Lee et al. 2008 [38] APEX study	Randomized phase III study	Bortezomib	296	62	-	61	-	-	8	42 wks
		Dexamethasone monotherapy	302	61	-	59	-	-		
Dubois et al. 2006 [39] SUMMIT study	Open-label phase II study	Bortezomib	143	60	-	-	-	-	2	End of study[a]
Waage et al. 2004 [40]	Phase II study	Thalidomide monotherapy	62	63	54	-	-	50	3	24 wks
Non-interventional study										
Mols et al. 2012 [41] PROFILES registry	Cohort study	Cohort	80	66	69	68	33	33	2	1 y

Twenty-three studies are included in the systematic review of which 22 were clinical trials and one is a cohort study. The design of the clinical trials was nine phase II trials, 12 phase III trials and one evaluation study. Fifteen studies are randomized clinical trials of which 4 are double-blinded and 3 are placebo-controlled (the MM-015 trial is only counted once). The number of patients and the mean age for the patients included in the studies/treatment arms at baseline range from 18 - 1,076 patients and 50–74.6 y, respectively. The number of HRQoL assessments range from 2–18 times and the last follow-up HRQoL measurements are 42 months.

-, Data not available; ASCT, Autologous haematopoietic stem cell transplantation; Bortezomib-dex, Bortezomib-Dexamethasone; EOT, End of treatment; HiDEX, High-dose dexamethasone; INF, Interferon-a2; Len-dex, Lenalidomid-dexamethasone; MP, Melphalan-Prednisolone; MP-IFN, Melphalan-Prednisolone-Interferon alfa-2b; MPplacebo, Melphalan-Prednisolone-placebo; MPR, Melphalan-Prednisolone-Lenalidomid; MPT, Melphalan-Prednisolone-Thalidomide; P-IFN, Pegylated Interferon-alfa2b; Pomalidomid-Lodex, Pomalidomide low-dose dexamethasone; Pre-PD, Pre-Progressive disease; SD, Study discontinuation; Thalidomide-dex, Thalidomide-dexamethasone; VMP, Bortezomib-Melphalan-Prednisone; VTD, Bortezomib-Thalidomide-Dexamethason; VTD-Cyclo, Bortezomib-Thalidomide-Cyclophosphamide-Dexamethasone.

[a]The precise time point is not specified in the publication.

[b]The patients at follow-up are not all of them the same as at diagnosis.

[c]The patients continued maintenance/placebo or observation after plateau phase of the M component, therefor the start of maintenance was at different time points and cannot be evaluated separately.

[d]The maintenance phase of the study with IFN cannot be evaluated separately since the baseline score for start of maintenance are not specified in the publication.

[e]Number of patients at baseline for Global QoL score.

[f]The number of patients at baseline is for physical functioning, fatigue and pain.

[g]Cross-over at treatment failure.

TABLE 2 The minimal important difference and guideline for interpreting mean change over time

EORTC QLQ-C30 domain	Kvam et al (16) Minimal important difference	Cocks et al. (17) Small difference[a]	Medium difference[b]	Large difference[c]
Physical functioning				
Improved	+6.2	2-7	>7	NE
Deteriorated	−12.8	−10 to −5	−17 to −10	<−17
Global quality of life				
Improved	+7.6	5-8	>8	NE
Deteriorated	−12.1	−10 to −5	−16 to −10	<−16
Fatigue[d]				
Improved	−13.5	4-9	>9	NE
Deteriorated	+8.6	−10 to −5	−15 to −10	<-15
Pain[d]				
Improved	−14.7	5-9	9-14	>14
Deteriorated	+17.3	−11 to −3	−20 to −11	<−20

The clinically meaningful changes in score according to Kvams' MID criteria for the four included domains. The clinically relevant change in score estimated by Cocks guideline for the included four domains divided into a small, medium and large difference.

NE, Not evaluable (a guideline for that size class was unobtainable)

[a]Small difference; indicated a subtle, but nevertheless clinically relevant change.

[b]Medium difference: a change likely to be clinically relevant but to a lesser extent. Upper limits for medium improvements could not generally be estimated.

[c]Large difference; a change with obvious and unequivocal clinical relevance.

[d]Kvam et al. indicates an improvement in fatigue and pain with positive sign, which is the opposite by Cocks et al., who indicates an improvement in fatigue and pain with a negative sign.

According to Kvams' MID criteria, patients report clinically meaningful improvement in all four domains after induction therapy and ASCT. The deterioration of global QoL of up to 6.2 points reported during induction therapy with VTD-Cyclo is not perceived as clinically meaningful to the patients.

In accordance with guideline of Cocks, patients report large improvement in pain and medium improvement in physical functioning and fatigue, which represent a likely clinically relevant change. The patients receiving induction with VTD-Cyclo reported a small deterioration in global QoL at the end of induction treatment and no difference at preprogressive disease follow-up after ASCT. The mean change in scores from baseline is presented in the Appendix S1, figures F1-F3.

3.4 | First-line treatment studies without ASCT

Eight studies were included in this category,[22-29] and the patient-reported range of mean change in score from baseline for the four domains is illustrated in Figure 2 for global QoL and physical functioning, and Figure 3 for fatigue and pain. During three regimens, the patients report deterioration in mean score from baseline and the largest deterioration was 6 points for fatigue reported by patients receiving bortezomib—melphalan-prednisolone (VMP) at day 1 at cycle 4, which is the last cycles of twice-weekly bortezomib administration. For all regimens, the patients report improvement in mean score from baseline at follow-up with the largest improvement in pain at 38 points at the

FIGURE 2 (A,B) Patient-reported range of mean change of score from baseline for global quality of life and physical functioning. (A) The eleven first-line treatment regimens. (B) The seven relapse treatment regimens. In general the patients report improvements in mean score from baseline during first-line treatment for global quality of life and physical functioning, but during relapse treatment the patients report either stabilization or deterioration. *Reports of study discontinuation, off protocol measurements and measure time points with patient drop-out rate over 75% are not maintained in the range.* Len-dex: Lenalidomide–dexamethasone, MPT: melphalan–prednisolone–thalidomide, MPR(-R): melphalan–prednisolone–lenalidomide introduction and planned lenalidomide maintenance, MPR(-pl): melphalan–prednisolone–lenalidomide induction and planned placebo maintenance, MPpl(-pl): melphalan–prednisolone-placebo induction and planned placebo maintenance, VTD: bortezomib–thalidomide–dexamethasone, VTD-Cyclo: bortezomib–thalidomide–cyclophosphamide–dexamethasone, MP: melphalan–prednisolone, VMP: bortezomib–melphalan–prednisone, ASCT: autologous haematopoietic stem cell transplantation, P-90; pamidronate 90 mg, P-30: pamidronate 30 mg, MP-IFN: melphalan–prednisone–interferon alfa-2b, Ixa-Len-dex: ixazomib–lenalidomide–dexamethasone, placebo-Len-dex: placebo–lenalidomide–dexamethasone, Car-Len-dex: carfilzomib–lenalidomide–dexamethasone, Pom-Lodex: pomalidomide–low-dose dexamethasone, HiDEX: high-dose dexamethasone, Thal-dex: thalidomide–dexamethasone, Bort-dex: bortezomib–dexamethasone. [1]*Data for global quality not shown in the publication.* [2] *Data for physical functioning not shown in the publication. The number in parenthesis refers to the publication the data is extracted from*

(A)

Global quality of life

Physical functioning
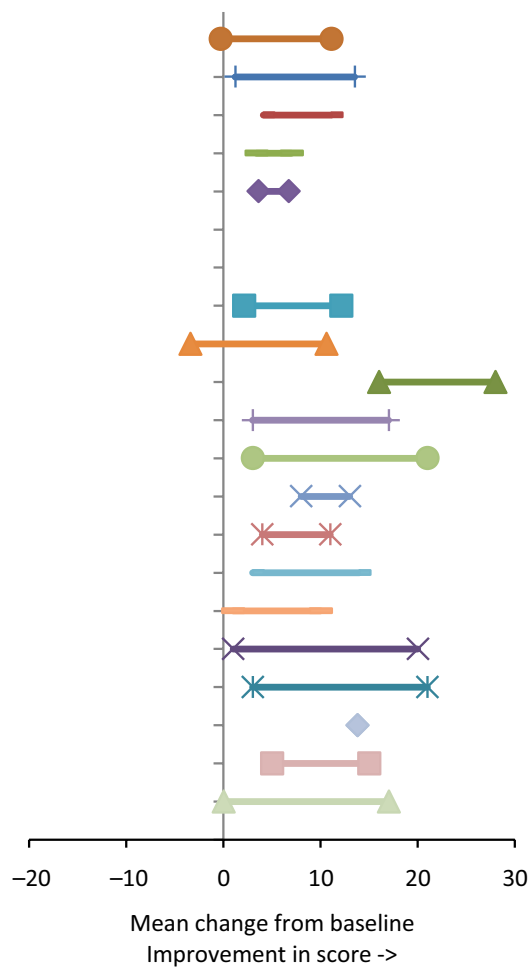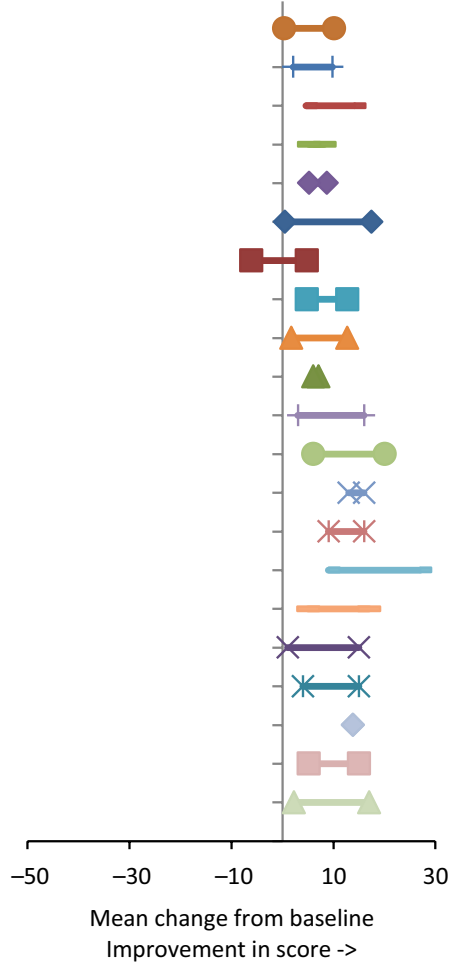
Len-Dex (22)
MPT (22)
MPR(-R) (23)
MPR(-pl) (23)
MPpl(-pl) (23)
VTD² (19)
VTD-Cyclo² (19)
MP (24)
VMP (24)
ASCT (20)
P-90 mg (26)
P-30 mg (26)
MP (25)
MPT (25)
MP (27)
MPT (27)
ASCT (21)
Control (21)
Evaluation (28)
MP (29)
MP-IFN (29)

Mean change from baseline
Improvement in score ->

Mean change from baseline
Improvement in score ->

(B)

Global quality of life

Physical functioning
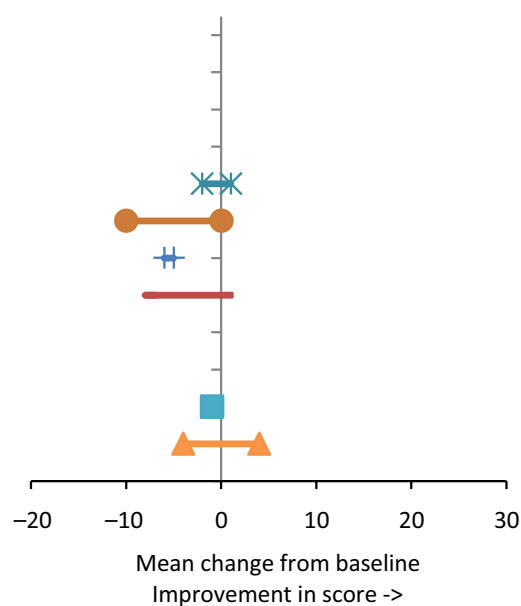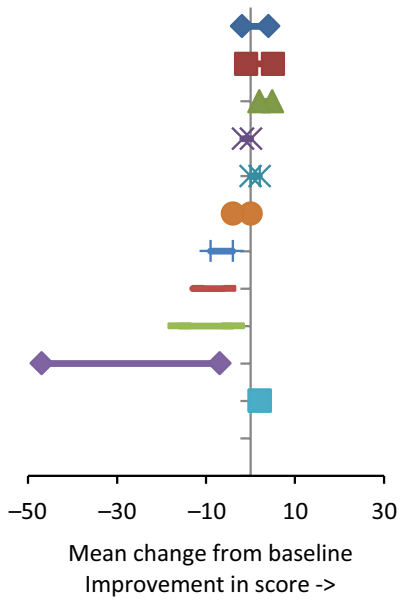
Ixa-Len-dex² (34)
Placebo-Len-dex² (34)
Car-Len-dex² (35)
Len-dex² (35)
Pom-Lodex (36)
HiDEX (36)
Thal-dex (37)
Bort-dex (37)
Bortezomib² (38)
Dexamethasone² (38)
Bortezomib (39)
Thalidomide¹ (40)

Mean change from baseline
Improvement in score ->

Mean change from baseline
Improvement in score ->

(C)

Fatigue

Pain

Mean change from baseline
<- Improvement in score

Mean change from baseline
<- Improvement in score

- Len-Dex (22)
- MPT (22)
- MPR(-R) (23)
- MPR(-pl) (23)
- MPpl(-pl) (23)
- VTD³ (19)
- VTD-Cyclo³ (19)
- MP (24)
- VMP (24)
- ASCT (20)
- P-90 mg (26)
- P-30 mg (26)
- MP (25)
- MPT (25)
- MP (27)
- MPT (27)
- ASCT (21)
- Control (21)
- Evaluation (28)
- MP (29)
- MP-IFN (29)

(D)

Fatigue

Pain

Mean change from baseline
<- Improvement in score

Mean change from baseline
<- Improvement in score

- Ixa-Len-dex³ (34)
- Placebo-Len-dex³ (34)
- Car-Len-dex³ (35)
- Len-dex³ (35)
- Pom-Lodex (36)
- HiDEX (36)
- Thal-dex (37)
- Bort-dex (37)
- Bortezomib³ (38)
- Dexamethasone³ (38)
- Bortezomib (39)
- Thalidomide⁴ (40)

36-month follow-up for patients treated with melphalan-Prednisone in the study by Waage et al.[27]

According to Kvams' MID criteria, all patients report clinically meaningful improvements in global QoL, physical functioning and pain with exception for pain in the group receiving melphalan-prednisolone-placebo induction phase with planned placebo maintenance (MPpl(-pl maintenance)). Clinically meaningful reduction in fatigue was reported in both treatment groups of three included studies.[26,27,29]

Using the guideline of Cocks, the reported deterioration in fatigue during the VMP regime is categorised as a subtle, but nevertheless clinically relevant change to the patients. During all treatment regimens, the patients report a likely clinically relevant improvement for physical functioning and an unequivocal clinically relevant improvement for pain with the same exception as earlier described: the MPpl(-pl maintenance) group. For all treatment arms, patients report a medium improvement in global QoL and fatigue, which is likely

FIGURE 3 (C,D) Patient-reported range of mean change of score from baseline for fatigue and pain. (C) The eleven first-line treatment regimens. (D) The seven relapse treatment regimens. In general, the patients report improvements in mean score from baseline during first-line treatment in fatigue and pain, but during relapse treatment the patients report stabilization or deterioration in fatigue and stabilization or improvement in pain. *Reports of study discontinuation, off-protocol measurements and time points with patient dropout rate over 75% are not maintained in the range.* Len-dex: Lenalidomide–dexamethasone, MPT: melphalan–prednisolone–thalidomide, MPR(-R): melphalan–prednisolone–lenalidomide introduction and planned lenalidomide maintenance, MPR(-pl): melphalan–prednisolone–lenalidomide induction and planned placebo maintenance, MPpl(-pl): melphalan–prednisolone–placebo induction and planned placebo maintenance, VTD: bortezomib–thalidomide–dexamethasone, VTD-Cyclo: bortezomib–thalidomide–cyclophosphamide–dexamethasone, MP: melphalan–prednisolone, VMP: bortezomib–melphalan–prednisone, ASCT: autologous haematopoietic stem cell transplantation, P-90: pamidronate 90 mg, P-30: pamidronate 30 mg, MP-IFN: melphalan–prednisone–interferon alfa-2b, Ixa-Len-dex: ixazomib–lenalidomide–dexamethasone, placebo-Len-dex: placebo–lenalidomide–dexamethasone, Car-Len-dex: carfilzomib–lenalidomide–dexamethasone, Pom-Lodex: pomalidomide–low-dose dexamethasone, HiDEX: high-dose dexamethasone, Thal-dex: thalidomide–dexamethasone, Bort-dex: bortezomib–dexamethasone. [3]*Data for fatigue and pain not shown in the publication.* [4]*Data for fatigue not shown in the publication. The number in parenthesis refers to the publication the data is extracted from*

clinically relevant, with exception of fatigue in the Len-Dex group. Here, the patients report a small improvement for fatigue, which is categorised as a subtle, but nevertheless clinically relevant improvement. The mean change from baseline is presented in the Appendix S1, figures P1-P8.

3.5 | Consolidation treatment studies

Three studies were included in this category.[30-32] Deteriorations in all four domains during single ASCT were reported by the patients with the largest deterioration in mean score from baseline at 33 points for fatigue 2 weeks after ASCT. The largest improvement after consolidation therapy was after tandem ASCT for global QoL, where the patients report a mean change in score of 29.4 points at the first follow-up visit.

According to Kvams' MID criteria, the patients report clinically meaningful deterioration in all four domains during single ASCT with recovery in all domains at follow-up. During tandem ASCT, clinically meaningful improvement was reported at first, second and subsequent follow-ups for physical functioning and global QoL.

Interpreted by guidelines of Cocks, temporary obvious and unequivocal clinically relevant deterioration for physical functioning, global QoL and fatigue are reported 2 weeks after single ASCT with reports of recovered HRQoL for all four domains 2 months after ASCT. After the second ASCT, the patients reported a smaller but likely clinically relevant improvement in physical functioning and global QoL. Patients receiving bortezomib consolidation therapy report initial small increased fatigue, followed by a small decrease and finally a small increase in fatigue compared to baseline. The mean change from baseline is presented in the Appendix S1, figures C1-C3.

3.6 | Maintenance treatment studies

Two studies were included in this category,[23,33] and the patient-reported range of mean change in score from baseline for the four domains is illustrated in Figure 4. The largest improvement in mean score from baseline during maintenance treatment was 9.1 points reported

for pain by the patients receiving pegylated interferon-alfa2b maintenance after induction treatment. The largest deterioration in mean score from baseline was reported during the placebo maintenance treatment after MPR.

According to Kvams' MID criteria, none of the included maintenance regimens resulted in a clinically meaningful mean change in score.

When the reported mean changes in score are interpreted using guidelines of Cocks, the patients report subtle, but nevertheless clinically relevant improvements in physical functioning and fatigue during the lenalidomide as well as placebo maintenance treatment after melphalan-prednisolone-lenalidomide (MPR) induction phase. During placebo maintenance after MPR, a subtle, but nevertheless clinically relevant deterioration was reported for pain. Patients receiving maintenance treatment with Pegylated interferon-alfa2b report a medium improvement in pain and small improvement in physical functioning, whereas patients treated with Interferon-a2 report a small deterioration in global QoL and fatigue. The mean change from baseline is presented in the Appendix S1, figures M1-M2.

3.7 | Relapse studies

Seven relapse studies are included in the review[34-40] and the patient-reported range of mean changes in score from baseline for the four domains are illustrated in Figure 2 for global QoL and physical functioning, and Figure 3 for fatigue and pain. The largest improvement in mean score from baseline was 15 points reported for pain at 24-week follow-up of treatment with thalidomide monotherapy. The largest deterioration was reported by patients receiving dexamethasone monotherapy in global QoL with a mean change from baseline at follow-up of 47 points at 36-week follow-up.

Using Kvams' MID criteria, only one study showed clinically meaningful improvement of HRQoL parameters and that was for pain in patients receiving thalidomide monotherapy. The patients in four treatment arms report a clinically meaningful deterioration for fatigue during high-dose dexamethasone treatment and for global QoL during bortezomib-dexamethasone therapy, bortezomib monotherapy and dexamethasone monotherapy.
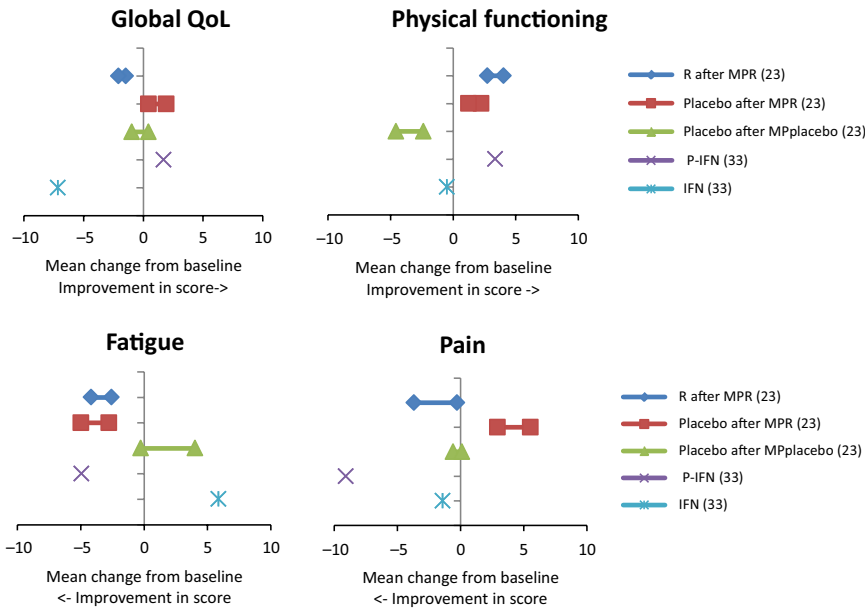
FIGURE 4 Patient-reported range of mean change in score from baseline for the two included maintenance treatment studies. The number in parenthesis refers to the publication the data are extracted from. Based on this systematic review, data on patient reported health-related quality of life during maintenance treatment are sparse and inconclusive. R after MPR: lenalidomide maintenance after melphalan–prednisolone–lenalidomide induction therapy, placebo after MPR: placebo maintenance after melphalan–prednisolone–lenalidomide induction therapy, placebo after MPplacebo: placebo maintenance after melphalan–prednisolone–placebo induction therapy, P-IFN: pegylated interferon-alfa2b, INF: interferon-a2

Interpreted by guidelines of Cocks, the patients reported a small improvement in pain during two treatment regimens: pomalidomide-low-dose dexamethasone and thalidomide-dexamethasone treatment. The patients also reported a likely clinically relevant improvement in pain during bortezomib-dexamethasone treatment and an obvious and unequivocal clinically relevant improvement in pain during thalidomide monotherapy. Mostly, the patients reported a subtle, but nevertheless clinically relevant deterioration during several relapse treatment regimens. A large deterioration in global QoL was reported during dexamethasone monotherapy and a medium deterioration was reported by the groups receiving Bort-dex and bortezomib monotherapy. The mean change from baseline is presented in the Appendix S1, figures R1-R7.

3.8 | Non-interventional study

A single non-interventional study was included in the systematic review,[41] which is a cohort study of MM patients in different disease stages. No improvements were reported, and the largest deterioration was reported for global QoL with a mean change from baseline at 13 points at one-year follow-up.

According to Kvams' MID criteria, the patients reported a clinically meaningful deterioration in global QoL at one-year follow-up.

According to guidelines of Cocks, the patients reported a medium deterioration, which is likely to be clinically relevant in global QoL and a small deterioration, which is subtle, but nevertheless clinically relevant in fatigue and pain after one year. The mean change from baseline is presented in the Appendix S1, figure B1.

4 | DISCUSSION

The main objective of this systematic review was to analyse published longitudinal studies of MM patients according to a changes in HRQoL scores, which are perceived as clinically beneficial to the patient for

the four domains of physical function, global QoL, fatigue and/or pain according to the Kvams' MID criteria and guideline of Cocks.

Based on this systematic review, clinically beneficial improvements in HRQoL are far more likely during primary treatments compared to relapse treatment regimens. During ASCT, temporary obvious and unequivocal clinically relevant deterioration for physical functioning, global QoL and fatigue are found but this is followed by recovered HRQoL for all four domains 2 months after ASCT.

Using Kvams' MID criteria at therapy demanding relapse, a clinically meaningful improvement was only reported for pain in one relapse study.

Even though several included studies are designed with a maintenance phase, HRQoL data are only reported and available for interpretation from two studies. Thus, data on patient-reported HRQoL during maintenance treatment are sparse and inconclusive.

In the only non-interventional study eligible for this systematic review, the patients in different disease stages reported a clinically meaningful deterioration in global QoL one year later.

4.1 | Discussion of our results from a clinical standpoint

There is a growing interest in assessing and understanding HRQoL in patients with MM and to use patient-reported outcomes in clinical decision-making regarding choice of treatment in combination with data on response, and expected progression-free survival and overall survival.[12] Therefore, it is essential, that HRQoL results are communicated precisely and in a clinical context.[42,43]

The results of our systematic review suggest that a clinically realistic HRQoL outcome for patients undergoing first-line treatment is different from that expected for patients treated for relapse. In this regard, patients should be informed that HRQoL is expected to improve during first-line treatment with different degrees of possible deterioration during the first cycles. The same degree of improvement

in HRQoL cannot be expected during conventionally studied relapse treatments, where the patients should be prepared for only a stabilisation of HRQoL. In addition, in the relapse situation it might be a relevant time point for facilitation of interdisciplinary rehabilitation aimed to improve HRQoL.

HRQoL outcome results should be integrated in the information given to the patient at the time of treatment-demanding disease to promote patient engagement and shared decision-making in health care. Before this is possible in an evidence-based matter, the interpretation of the HRQoL outcomes data using MID should be validated in a clinical setting including MM patients in different stages of the disease and with different socio-demographic characteristics.

The underlying causes of the varying HRQoL results in patients treated for newly diagnosed disease versus for relapsed disease have to our knowledge never been explored and are therefore unknown. Several reasons are hypothetically possible:

4.2 | Results of HRQoL during first-line versus relapse treatment regimens

In general, the newly diagnosed MM patients are often more symptomatic than relapse patients, as treatment of relapsed disease often occurs before the patient experiences a symptomatic relapse.[44] A comparison of baseline scores and clinical data from six randomised clinical trials with HRQoL measurement has been published.[45] The findings of this comparison support that symptoms are better controlled in patients treated at relapse than in newly diagnosed patients and this finding can be confirmed here. For the studies included in this review, the average mean baseline scores for fatigue and pain were 47.7 and 48.4 points for first-line treated patients and 44.4 and 44.2 points for relapsed patients, respectively. An interesting finding in the published comparison study is that the mean score at baseline for global QoL score is similar over the course of the myeloma disease. For the included studies in this review, the average mean scores in global QoL at baseline for the first-line versus relapse treatment regimens were 49.7 versus 58.4 points, respectively, indicating that relapsed patients report better global QoL than previously untreated patients at the entrance of clinical trials.

In terms of internal validity in comparison of baseline scores, the mean age of the patients in the first-line versus relapse treatment categories was 68.8 versus 64.2 years, respectively. The fact that the relapse patients are younger than the previously untreated patients indicates that patients included in relapse studies are more selected than patients included in first-line treatment studies.

The treatment regimens used in first-line compared to relapse situations are different and this may in part explain the differences in HRQoL outcomes as demonstrated in this systematic review for first-line versus relapse treatment. Of the studies included, HRQoL is reported by Len-dex-treated patients in one-first-line study, the FIRST trial, and two relapse studies, the ASPIRE and TOURMALINE-MM1 trials, of which the last mentioned was placebo-controlled. With respect to the differences of the cohorts included in the three studies, the patients reported medium improvement in global QoL during the

first-line study and no MID in the relapse studies. Concluded by this observation, the same treatment regimens are not as effective on HRQoL outcomes in relapse MM patients as in previously untreated MM patients.

4.3 | Biology of the plasma cell clone

One explanation for the different findings of mean change in HRQoL from baseline in first-line versus relapse treated patients could also be the biology of the clone of the plasma cells, the clonal dynamics and the trajectory of the MM disease.[46] Whether increasing dominance of a partly resistant clone at relapse leaves the MM disease and symptoms relative resistant to anti-myeloma treatment is unexplored.

4.4 | Response shift in quality of life reportings

Response shift in serial patient-reported outcome measurements and QoL has been known since Calman et al. in 1984 described the concept quality of life as "the differences, or the gap, at a particular period of time between the hopes and expectations of the individual and that individual's present experiences".[47] Schwartz and Sprangers published a working definition of response shift in 1999, which is a psychological phenomenon of change in internal standards, values and/or in the conceptualisation of QoL catalysed by health state changes. Response shift refers to three considerations: a change in the meaning of one's self-evaluation of a target construct as a result of a change in the respondents' internal standards of measurement (recalibration), a change in the respondents' values (reprioritisation) and redefinition of the target construct (reconceptualisation).[48,49]

Response shift has been recognised in several chronic illnesses including cancer[50] and has also been explored in MM patients calculating MID by Kvam using the anchor-based method.[51] This is a further work of Kvams' MID criteria where the magnitude and direction of response shift are calculated based on the pretest—then-test data and Cohen's criteria for effect size.[52] The findings indicate response shift among MM patients, mainly in those patients who deteriorated over a 3-month observation period. The anchor-based method, which is often used to detect response shift, is likely to introduce another bias, such as recall bias, and may not be a suitable method in controlling for inconsistencies in the respondents' cognitive processes over time.[53]

Response shift is described as a "meta-construct" and the findings of an exploring study of response shift in longitudinal data provided evidence for limitations and validity of the then-test approach to measuring recalibration response shift.[54] Rapkin and Schwartz in 2004 developed "The Theoretical model of QoL Appraisal," which links to the aspects of response shift; change in frame of reference, sample of experiences, standards of comparison and combinatory algorithm. In the process of development of a practical, low resource-intensive version of the QoL Appraisal Profile, the Cohen's criteria for effect size was tested and could only explain a moderate amount of variance in the domains tested.[55] This has led to a fundamental recommendation; whenever one measures QoL, one should measure appraisal.[55,56]

Before calculated thresholds for response shift based on the anchor-based methods can be integrated in the clinically meaningful interpretation of HRQoL outcomes data, these thresholds should to be tested in longitudinal prospective studies with MM patients and validated upon a test of response shift detection in parallel.

Response shift could be a possible explanation of the difference in HRQoL result during first-line versus relapse treatment; the relapsed MM patients might have gone through a psychological adaption to the consequences of the myeloma disease and reduced expectations for the future, which might have resulted in a change in the patients' internal standard of measurement. This could make HRQoL changes during relapse treatment conservative, and therefore improvements more difficult to demonstrate later than after first-line treatment. Also, this could in part explain the findings of different mean change in scores in subgroups of patients according to response to treatment.

4.5 | How to interpret changes in HRQoL in longitudinal studies

Guidelines for interpretation of changes in score over time perceived as beneficial to the patient have been calculated and defined. These tools may be useful for interpretation of patient-reported outcome data, but of note these methods are not definite and confirming studies and validation of the guidelines are necessary and should be performed.

The change in score, which is perceived beneficial to the patient, may vary by population and context with possible differences in terms of the patients' demographic characteristics. In addition, the baseline score might have an impact on the sizes of change in score, which is perceived as beneficial to the patient.[15] Rapid and sustainable symptom relief is valuable for the patient. However, time to MID and time duration of MID are not integrated in the MID definition. A standardised duration of time between baseline and the follow-up time point for interpretation of mean change in relation to MID is needed.

HRQoL results are often presented by the *mean* score for the cohort. During the publication selection for this review, two studies were excluded because the results of HRQoL were presented as *median* scores.[57,58] A reported normal distributed score evaluated as a mean change in score over time for a given HRQoL domain may include subgroups that experiences improvement, deterioration and/or stable conditions. A way to address the magnitude of this is to present percentages of the patients, who report improved, stable or deteriorated status during follow-up. These data are presented in three of the included studies.[23,39,59] After the literature search for this systematic review was completed, an additional analysis of the HRQoL data from the ASPIRE trial is published including these aspects.[60]

Confidence intervals (95%) or standard deviations are reported in 15 of the 23 publications and the standard deviations ranged up to 34 on a 0-100 point scale. This clearly indicates that groups of patients in the treatment arms report very different scores, more than illustrated by the mean score. Individualised treatment for MM patients based on pretreatment prognostic markers is a research field of increasing interest. Except from cytogenetic prognostic markers,[61] scores of

comorbidity, geriatric assessments and/or frailty scores could make the pretreatment assessment more powerful.[62]

One of the challenges in interpretation of QoL data from longitudinal studies are low compliance, high dropout rates and missing data.[63] Missing data in QoL studies are often not missing at random, as the patients staying in the study are often selected for the best outcome and therefore continue to follow up and this is a source of bias. A thoughtful design to reduce missing data and the applied statistical approach for dealing with missing data should be stated in the publication.[43,64] Sixteen of the included studies have published data on compliance in the HRQoL follow-up measurement. In the FIRST trial, compliance at 18 months was 85-65%.[22] A much lower compliance has been demonstrated in the relapse studies, for example 28-23% in the APEX study[38] and 4-19% in the MM-003 study.[23] Conclusions of mean change in scores drawn on such low compliance rates are not of value in clinical decision-making. We chose to end the interpretation of the follow-up for this review, when patient dropout rates exceeded 75%. In future longitudinal HRQoL studies methodologies assess whether the impact of this bias should be applied.

4.6 | HRQoL in the general population of multiple myeloma patients

Of the studies included in this review, only one was population based. It is a fact that patients included in clinical trials are not representative of the general MM population, as clinical trial participants are often younger and have better performance status and less comorbidity. The results of our review may therefore not be transferable to the general population of MM patients. Population-based QoL studies including elderly, frail and comorbid myeloma patients, the so-called real-time data collection, are needed to be able to provide "patient-like-me" information to the variety of MM patients in expected HRQoL outcome.

4.7 | HRQoL during maintenance therapy and treatment continuation until progressive disease

Six of the included studies in this review included maintenance therapy after effective anti-myeloma treatment, but separate HRQoL data reporting on the maintenance phase was only presented in two studies.[23,33] Often the reports of HRQoL are stopped before the maintenance phase starts or there is no clear baseline score for the maintenance phase. Maintenance therapy was found to improve progression-free survival and overall survival in MM patients,[65] but was associated with increased risk of grade 3-4 adverse events.[66] Based on this systematic review, data on HRQoL during maintenance treatment are sparse and inconclusive and HRQoL measurement should be a part of future protocols including maintenance therapy.

Four trial protocols in this systematic review included studies with treatment continuation until progressive disease.[22,34-36] In the FIRST and ASPIRE trials, the HRQoL measurement stopped at 18 months, which is the end of Len-dex treatment for the fixed 18-month duration (Rd18) in the FIRST trial and at the end of administration of carfilzomib

in the ASPIRE trial. For the TOURMALINE-MM1 and MM-003, the duration of published HRQoL follow-up is day 1 of treatment cycles 34 and 10, respectively, but with very low compliance rates. Results of progression-free survival and overall survival from these studies are based on treatment continuation until progressive disease, but leave HRQoL during the late phases of treatment continuation unexplored.

4.8 | HRQoL after study discontinuation and off protocol

In two-first-line studies, the FIRST and the HOVON 49 trials,[22,25] data on HRQoL measurement at study discontinuation and after going off protocol are presented. In both studies, the patients report that improvement reported during first-line treatment disappeared at study discontinuation and at the off-protocol measurement, except for pain in the HOVON 49 trial. This could be due to the patients being in a relapse situation at that time point. In general, there has been very little focus on the patients' HRQoL during periods without cancer treatment.[67]

This is the first systematic review that includes published longitudinal HRQoL studies with the aim of interpreting HRQoL data obtained by QLQ-C30 reported by MM patients according to available guidelines for changes in score, which is perceived beneficial to the patients. Consensus of analysis and presenting data from HRQoL studies are essential in order to include HRQoL data in clinical decision-making. A precise clinically relevant method of interpreting HRQoL data is important to give the patient evidence-based information on what to expect during anti-myeloma treatment in aspects of improved patient engagement and shared decision-making. We recommend future prospective population-based longitudinal studies to validate MID thresholds in MM patients at different stages of the disease and with different socio-demographic characteristics and also to address the challenges of taking into account the response shift issue and the handling of missing data in longitudinal QoL studies.

REFERENCES

1. Kumar SK, Dispenzieri A, Lacy MQ, et al. Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia*. 2014;28(5):1122-1128.
2. Kumar SK, Rajkumar SV, Dispenzieri A, et al. Improved survival in multiple myeloma and the impact of novel therapies. *Blood*. 2008;111(5):2516-2520.
3. Ocio EM, Richardson PG, Rajkumar SV, et al. New drugs and novel mechanisms of action in multiple myeloma in 2013: a report from the International Myeloma Working Group (IMWG). *Leukemia*. 2014;28(3):525-542.
4. Jordan K, Proskorovsky I, Lewis P, et al. Effect of general symptom level, specific adverse events, treatment patterns, and patient characteristics on health-related quality of life in patients with multiple myeloma: results of a European, multicenter cohort study. *Support Care Cancer*. 2014;22(2):417-426.
5. Johnsen AT, Tholstrup D, Petersen MA, Pedersen L, Groenvold M. Health related quality of life in a nationally representative sample of haematological patients. *Eur J Haematol*. 2009;83(2):139-148.
6. Boland E, Eiser C, Ezaydi Y, Greenfield DM, Ahmedzai SH, Snowden JA. Living with advanced but stable multiple myeloma: a study of the symptom burden and cumulative effects of disease and intensive (hematopoietic stem cell transplant-based) treatment on health-related quality of life. *J Pain Symptom Manage*. 2013;46(5):671-680.
7. Priscilla D, Hamidin A, Azhar MZ, Noorjan KON, Salmiah MS, Bahariah K. Quality of life among patients with hematological cancer in a Malaysian hospital. *Med J Malaysia*. 2011;66(2):117.
8. Wisloff F, Eika S, Hippe E, et al. Measurement of health-related quality of life in multiple myeloma. *Br J Haematol*. 1996;92(3):604-613.
9. Aaronson NK, Ahmedzai S, Bergman B, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365-376.
10. Cocks K, Cohen D, Wisloff F, et al. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. *Eur J Cancer*. 2007;43(11):1670-1678.
11. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res*. 1996;5(6):555-567.
12. Kvam AK, Waage A. Health-related quality of life in patients with multiple myeloma–does it matter? *Haematologica*. 2015;100(6):704-705.
13. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415.
14. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77(4):371-383.
15. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171-184.
16. Kvam AK, Fayers P, Wisloff F. What changes in health-related quality of life matter to multiple myeloma patients? A prospective study. *Eur J Haematol*. 2010;84(4):345-353.
17. Cocks K, King MT, Velikova G, et al. Evidence-based guidelines for interpreting change scores for the European organisation for the research and treatment of cancer quality of life questionnaire core 30. *Eur J Cancer*. 2012;48(11):1713-1721.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
19. Ludwig H, Viterbo L, Greil R, et al. Randomized phase II study of bortezomib, thalidomide, and dexamethasone with or without cyclophosphamide as induction therapy in previously untreated multiple myeloma. *J Clin Oncol*. 2013;31(2):247-255.
20. Etto LY, Morelli VM, Silva VC, Hungria VTM, Ciconelli RM, Almeida MSS, et al. Autologous stem cell transplantation improves quality of life in economically challenged, Brazilian multiple myeloma patients. *Clinics*. 2011;66(11):1855-1859.
21. Gulbrandsen N, Wisloff F, Brinch L, et al. Health-related quality of life in multiple myeloma patients receiving high-dose chemotherapy with autologous blood stem-cell support. *Med Oncol*. 2001;18(1):65-77.
22. Delforge M, Minuk L, Eisenmann JC, et al. Health-related quality-of-life in patients with newly diagnosed multiple myeloma in the FIRST trial: lenalidomide plus low-dose dexamethasone versus melphalan, prednisone, thalidomide. *Haematologica*. 2015;100(6):826-833.
23. Dimopoulos MA, Delforge M, Hajek R, et al. Lenalidomide, melphalan, and prednisone, followed by lenalidomide maintenance, improves health-related quality of life in newly diagnosed multiple myeloma patients aged 65 years or older: results of a randomized phase III trial. *Haematologica*. 2013;98(5):784-788.
24. Delforge M, Dhawan R, Robinson D Jr, et al. Health-related quality of life in elderly, newly diagnosed multiple myeloma patients

treated with VMP vs. MP: results from the VISTA trial. *Eur J Haematol*. 2012;89(1):16-27.

25. Verelst SG, Termorshuizen F, Uyl-de Groot CA, et al. Effect of thalidomide with melphalan and prednisone on health-related quality of life (HRQoL) in elderly patients with newly diagnosed multiple myeloma: a prospective analysis in a randomized trial. *Ann Hematol*. 2011;90(12):1427-1439.

26. Gimsing P, Carlson K, Turesson I, et al. Effect of pamidronate 30 mg versus 90 mg on physical function in patients with newly diagnosed multiple myeloma (Nordic Myeloma Study Group): a double-blind, randomised controlled trial. *Lancet Oncol*. 2010;11(10):973-982.

27. Waage A, Gimsing P, Fayers P, et al. Melphalan and prednisone plus thalidomide or placebo in elderly patients with multiple myeloma. *Blood*. 2010;116(9):1405-1412.

28. Wisloff F, Eika S, Hippe E, et al. Measurement of health-related quality of life in multiple myeloma. Nordic Myeloma Study Group. *Br J Haematol*. 1996;92(3):604-613.

29. Wisloff F, Hjorth M, Kaasa S, Westin J. Effect of interferon on the health-related quality of life of multiple myeloma patients: results of a Nordic randomized trial comparing melphalan-prednisone to melphalan-prednisone + alpha-interferon. The Nordic Myeloma Study Group. *Br J Haematol*. 1996;94(2):324-332.

30. Mellqvist UH, Gimsing P, Hjertner O, et al. Bortezomib consolidation after autologous stem cell transplantation in multiple myeloma: a Nordic Myeloma Study Group randomized phase 3 trial. *Blood*. 2013;121(23):4647-4654.

31. Frodin U, Borjeson S, Lyth J, Lotfi K. A prospective evaluation of patients' health-related quality of life during auto-SCT: a 3-year follow-up. *Bone Marrow Transplant*. 2011;46(10):1345-1352.

32. Khalafallah A, McDonnell K, Dawar HU, Robertson I, Woods D. Quality of life assessment in multiple myeloma patients undergoing dose-reduced tandem autologous stem cell transplantation. *Mediterr J Hematol Infect Dis*. 2011;3(1):e2011057.

33. Sirohi B, Powles R, Lawrence D, et al. An open, randomized, controlled, phase II, single centre, two-period cross-over study to compare the quality of life and toxicity experienced on PEG interferon with interferon-alpha2b in patients with multiple myeloma maintained on a steady dose of interferon-alpha2b. *Ann Oncol*. 2007;18(8):1388-1394.

34. Moreau P, Masszi T, Grzasko N, et al. Oral Ixazomib, lenalidomide, and dexamethasone for multiple myeloma. *N Engl J Med*. 2016;374(17):1621-1634.

35. Stewart AK, Rajkumar SV, Dimopoulos MA, et al. Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. *N Engl J Med*. 2015;372(2):142-152.

36. Song KW, Dimopoulos MA, Weisel KC, 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-e67.

37. Hjorth M, Hjertner O, Knudsen LM, et al. Thalidomide and dexamethasone vs. bortezomib and dexamethasone for melphalan refractory myeloma: a randomized study. *Eur J Haematol*. 2012;88(6):485-496.

38. Lee SJ, Richardson PG, Sonneveld P, 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):511-519.

39. Dubois D, Dhawan R, van de Velde H, et al. Descriptive and prognostic value of patient-reported outcomes: the bortezomib experience in relapsed and refractory multiple myeloma. *J Clin Oncol*. 2006;24(6):976-982.

40. Waage A, Gimsing P, Juliusson G, et al. Early response predicts thalidomide efficiency in patients with advanced multiple myeloma. *Br J Haematol*. 2004;125(2):149-155.

41. Mols F, Oerlemans S, Vos AH, et al. Health-related quality of life and disease-specific complaints among multiple myeloma patients up to 10 yr after diagnosis: results from a population-based study using the PROFILES registry. *Eur J Haematol*. 2012;89(4):311-319.

42. King MT, Fayers PM. Making quality-of-life results more meaningful for clinicians. *Lancet*. 2008;371(9614):709-710.

43. Calvert M, Brundage M, Jacobsen PB, Schunemann HJ, Efficace F. The CONSORT Patient-Reported Outcome (PRO) extension: implications for clinical trials and practice. *Health Qual Life Outcomes*. 2013;11:184.

44. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 2014;15(12):e538-e548.

45. Robinson D Jr, Esseltine DL, Regnault A, Meunier J, Liu K, van de Velde H. The influence of baseline characteristics and disease stage on health-related quality of life in multiple myeloma: findings from six randomized controlled trials. *Br J Haematol*. 2016;174(3):368-381.

46. Keats JJ, Chesi M, Egan JB, et al. Clonal competition with alternating dominance in multiple myeloma. *Blood*. 2012;120(5):1067-1076.

47. Calman KC. Quality of life in cancer patients–an hypothesis. *J Med Ethics*. 1984;10(3):124-127.

48. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Social science & medicine (1982)*. 1999;48(11):1507-1515.

49. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med*. 1999;48(11):1531-1548.

50. Schwartz C, Ahmed S, Sawatzky R, et al. Guidelines for secondary analysis in search of response shift. *Qual Life Res*. 2013;22(10):2663-2673.

51. Kvam AK, Wisloff F, Fayers PM. Minimal important differences and response shift in health-related quality of life; a longitudinal study in patients with multiple myeloma. *Health Qual Life Outcomes*. 2010;8:79-7525.

52. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988. xxi, 567 sider p.

53. Taminiau-Bloem EF, Schwartz CE, van Zuuren FJ, et al. Using a retrospective pretest instead of a conventional pretest is replacing biases: a qualitative study of cognitive processes underlying responses to thentest items. *Qual Life Res*. 2016;25(6):1327-1337.

54. Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal data. *Psychol Health*. 2004;19(1):51-69.

55. Rapkin BD, Schwartz CE. Distilling the essence of appraisal: a mixed methods study of people with multiple sclerosis. *Qual Life Res*. 2016;25(4):793-805.

56. Schwartz CE, Li J, Rapkin BD. Refining a web-based goal assessment interview: item reduction based on reliability and predictive validity. *Qual Life Res*. 2016;25(9):2201-2212.

57. Alegre A, Oriol-Rocafiguera A, Garcia-Larana J, et al. Efficacy, safety and quality-of-life associated with lenalidomide plus dexamethasone for the treatment of relapsed or refractory multiple myeloma: the Spanish experience. *Leuk Lymphoma*. 2012;53(9):1714-1721.

58. Iversen PO, Wisloff F, Gulbrandsen N. Reduced nutritional status among multiple myeloma patients during treatment with high-dose chemotherapy and autologous stem cell support. *Clin Nutr*. 2010;29(4):488-491.

59. Weisel K, Dimopoulos M, Song KW, 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):519-530.

60. Stewart AK, Dimopoulos MA, Masszi T, et al. Health-Related Quality of Life Results From the Open-Label, Randomized, Phase III ASPIRE Trial Evaluating Carfilzomib, Lenalidomide, and Dexamethasone Versus Lenalidomide and Dexamethasone in Patients With Relapsed Multiple Myeloma. *J Clin Oncol*. 2016;32:3921-3930.

61. Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *J Clin Oncol*. 2015;33(26):2863-2869.

62. Palumbo A, Bringhen S, Mateos MV, et al. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. *Blood*. 2015;125(13):2068-2074.

63. Fairclough DL, Thijs H, Huang IC, Finnern HW, Wu AW. Handling missing quality of life data in HIV clinical trials: what is practical? *Qual Life Res*. 2008;17(1):61-73.

64. Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *BMJ Open*. 2016;6(6):e010938.

65. McCarthy PL, Owzar K, Hofmeister CC, et al. Lenalidomide after stem-cell transplantation for multiple myeloma. *N Engl J Med*. 2012;366(19):1770-1781.

66. Wang Y, Yang F, Shen Y, et al. Maintenance therapy with immunomodulatory drugs in multiple myeloma: a meta-analysis and systematic review. *J Natl Cancer Inst*. 2016;108(3):1-10.

67. Acaster S, Gaugris S, Velikova G, Yong K, Lloyd AJ. Impact of the treatment-free interval on health-related quality of life in patients with multiple myeloma: a UK cross-sectional survey. *Support Care Cancer*. 2013;21(2):599-607.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

SUPPLEMENTERY APPENDIX

Table of context

FIRST-LINE TREATMENT STUDIES INCLUDING INTRODUCTION TREATMENT AND AUTOLOGOUS HAEMATOPOIETIC STEM CELL TRANSPLANTATION

Study F1:
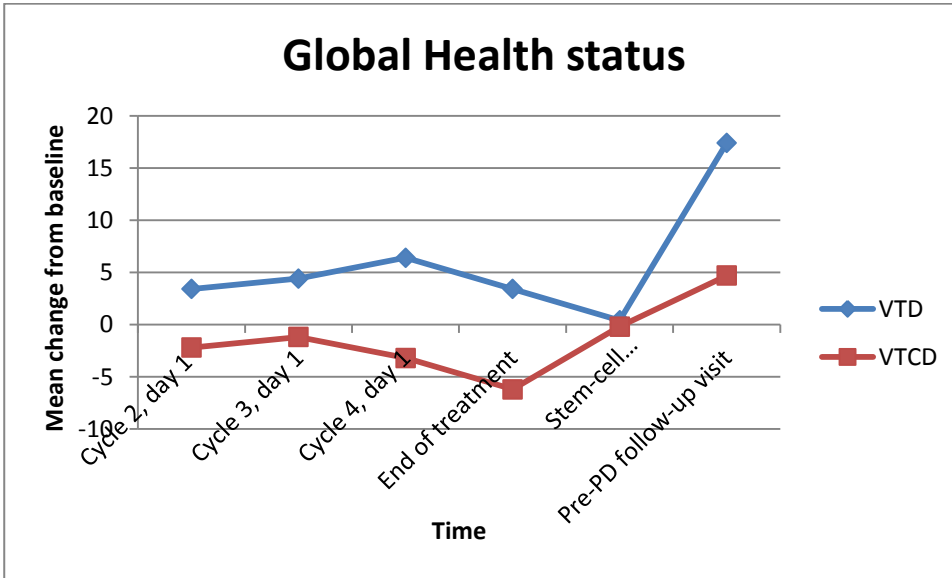
VTD; Bortezomib-Thalidomide-Dexamethason, VTCD; Bortezomib-Thalidomide-Cyclophosphamide-Dexamethasone, Pre-PD; Pre progressive disease

Figure F1b. Mean change of score from baseline – Global Health status

Study F2:
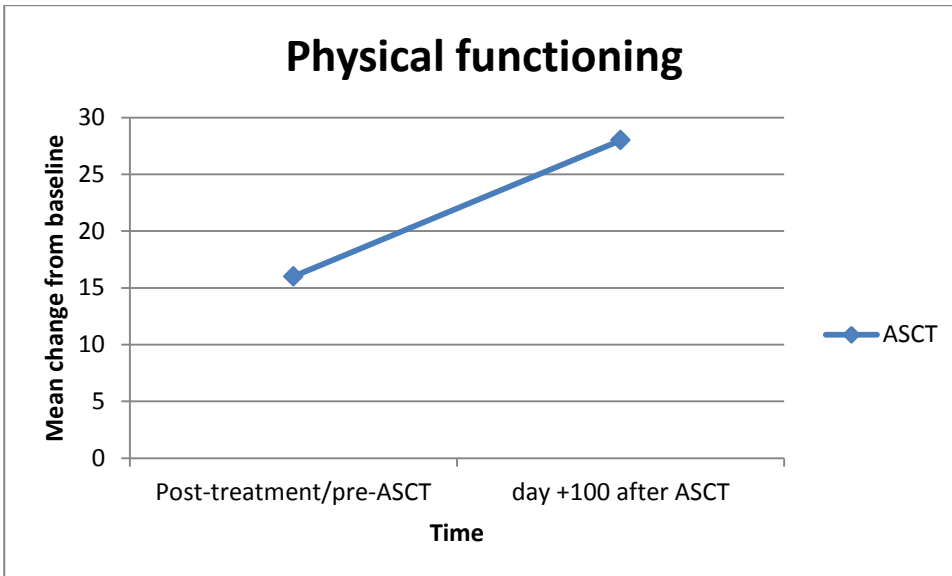
ASCT; Autologous haematopoietic stem cell transplantation

Figure F2a. Mean change of score from baseline – Physical functioning
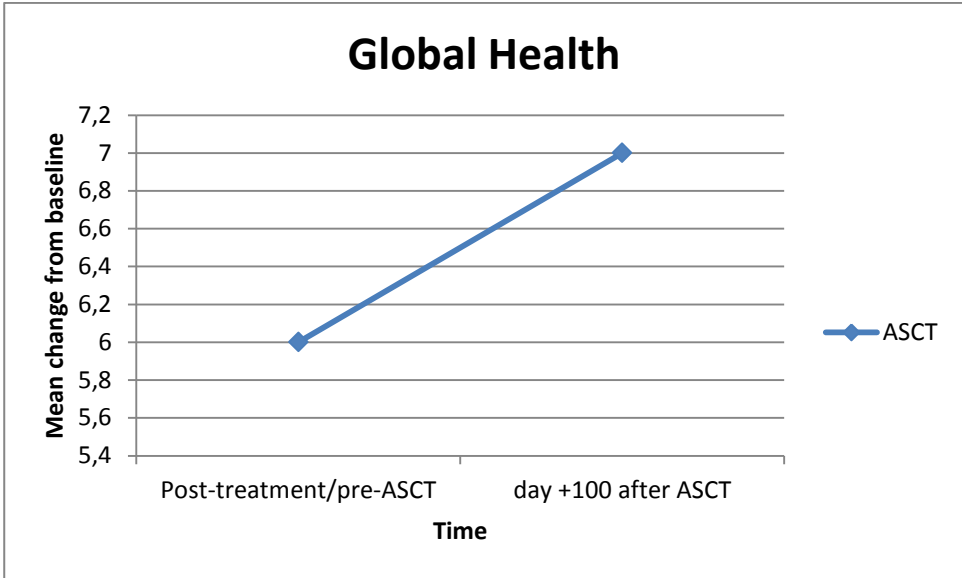
ASCT; Autologous haematopoietic stem cell transplantation

Figure F2b. Mean change of score from baseline – Global Health
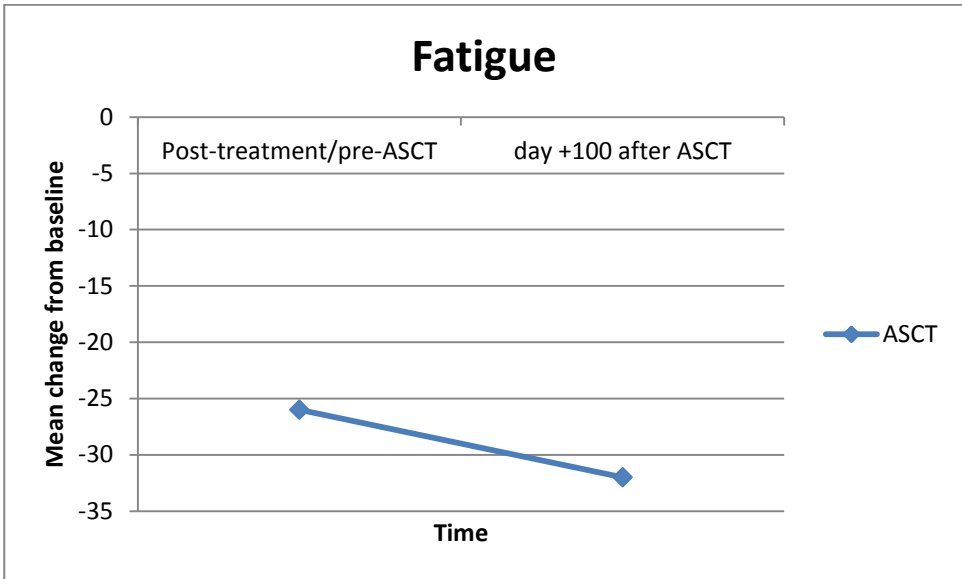
ASCT; Autologous haematopoietic stem cell transplantation

Figure F2c. Mean change of score from baseline – Fatigue
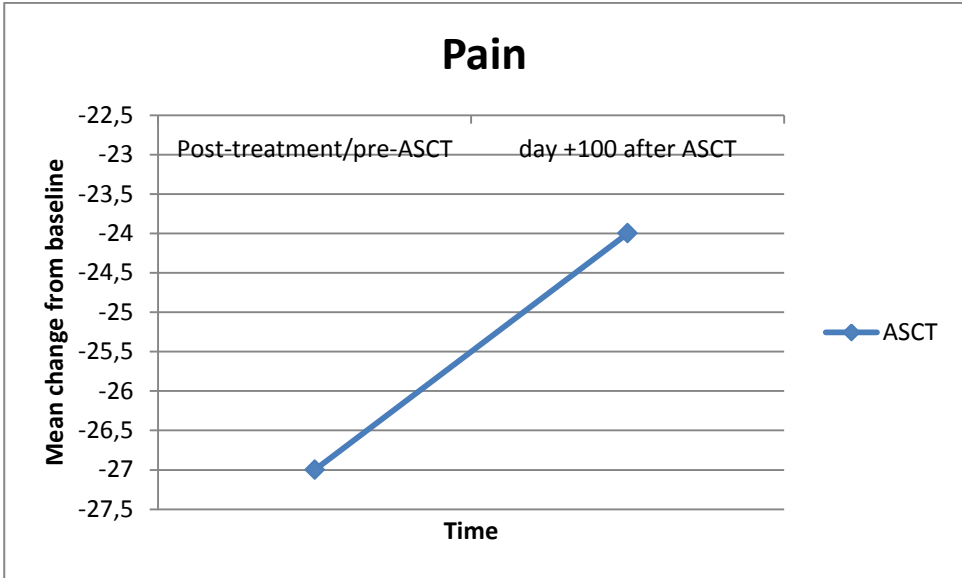
ASCT; Autologous haematopoietic stem cell transplantation

Figure F2d. Mean change of score from baseline – Pain

Study F3:
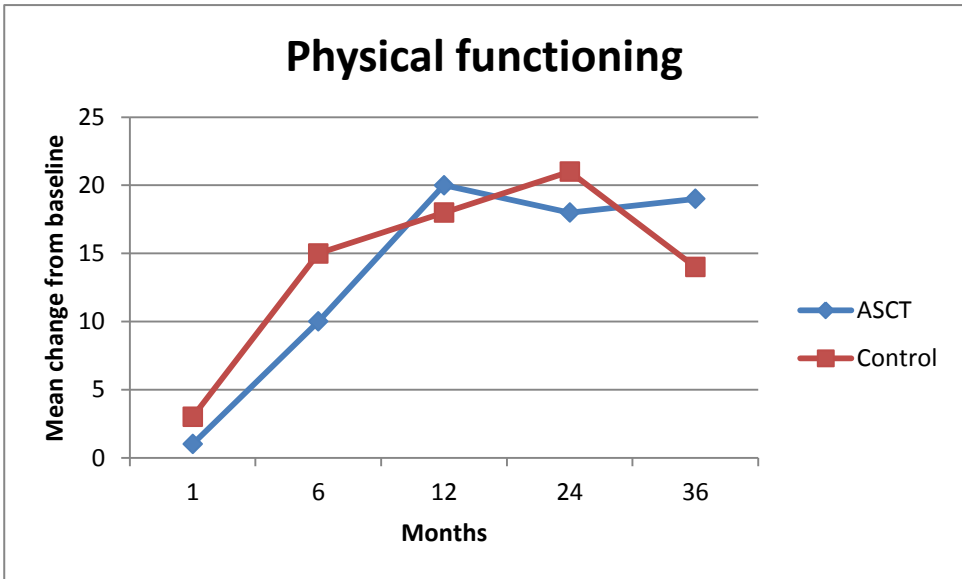
ASCT; Autologous haematopoietic stem cell transplantation

F3a. Mean change of score from baseline – Physical functioning
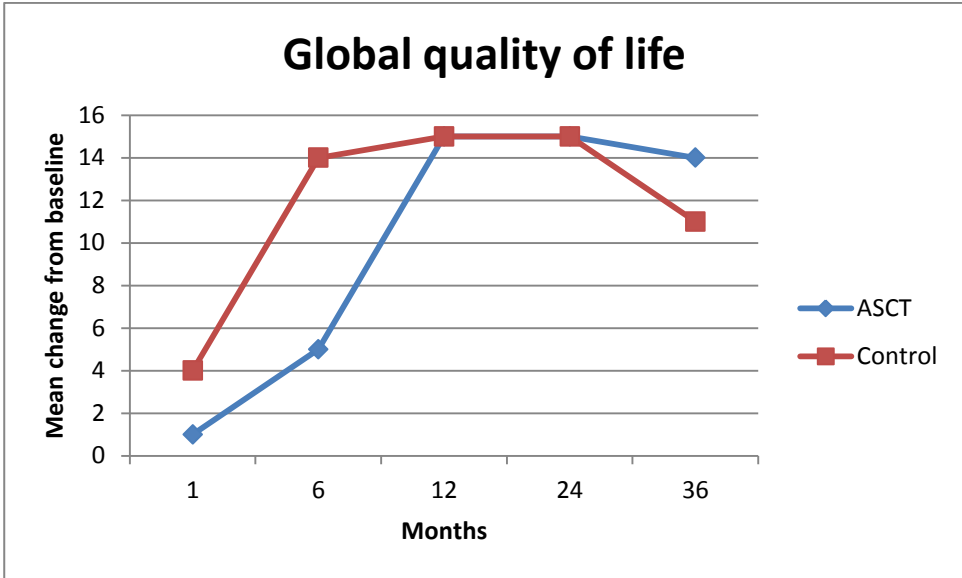
ASCT; Autologous haematopoietic stem cell transplantation

Figure F3b. Mean change of score from baseline – Global quality of life
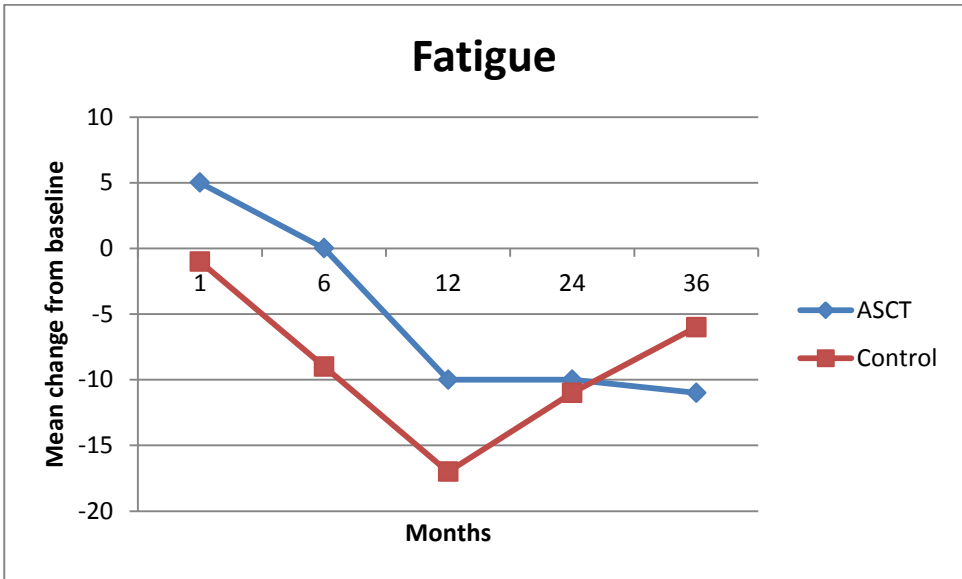
ASCT; Autologous haematopoietic stem cell transplantation

Figure F3c. Mean change of score from baseline – Fatigue
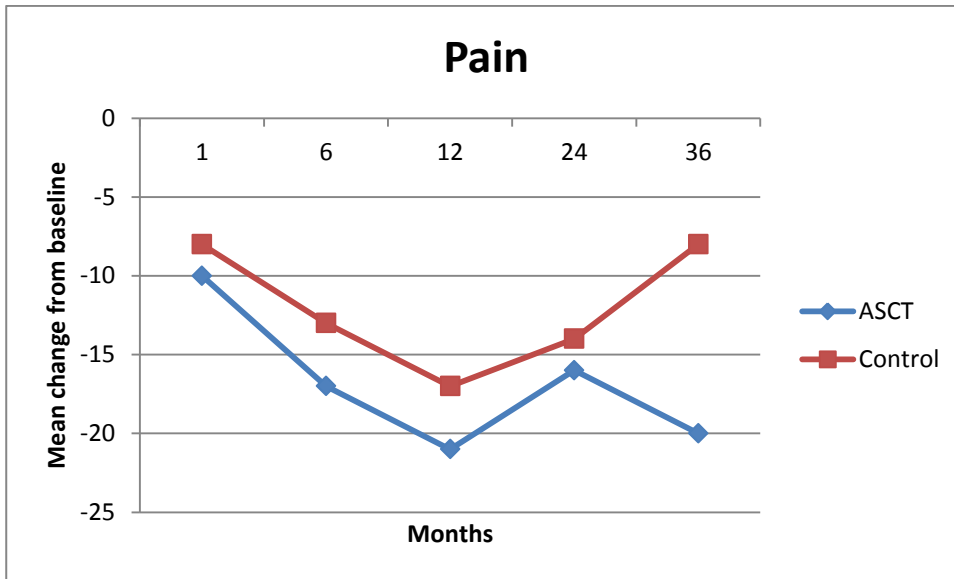
ASCT; Autologous haematopoietic stem cell transplantation

Figure F3d. Mean change of score from baseline – Pain

FIRST-LINE TREATMENT STUDIES WITHOUT AUTOLOGOUS HAEMOTOPOIETIC STEM CELL TRANSPLANTATION

Study P1:

The mean change of score from baseline for evaluation of mean change from baseline is calculated by the baseline score and the score for each month listed in the supplementary appendix of the publication.
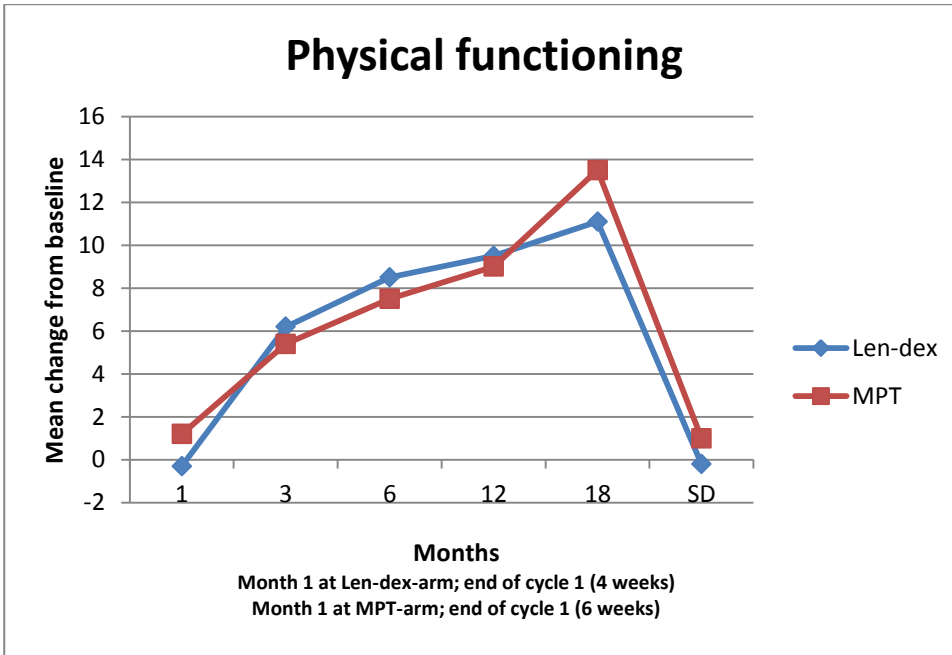
Len-dex; Lenalidomid-Dexamethasone, MPT; Melphalan-Prednisolone-Thalidomide, SD: Study discontinuation, which occur at any time point.

Figure P1a. Mean change of score from baseline – Physical function
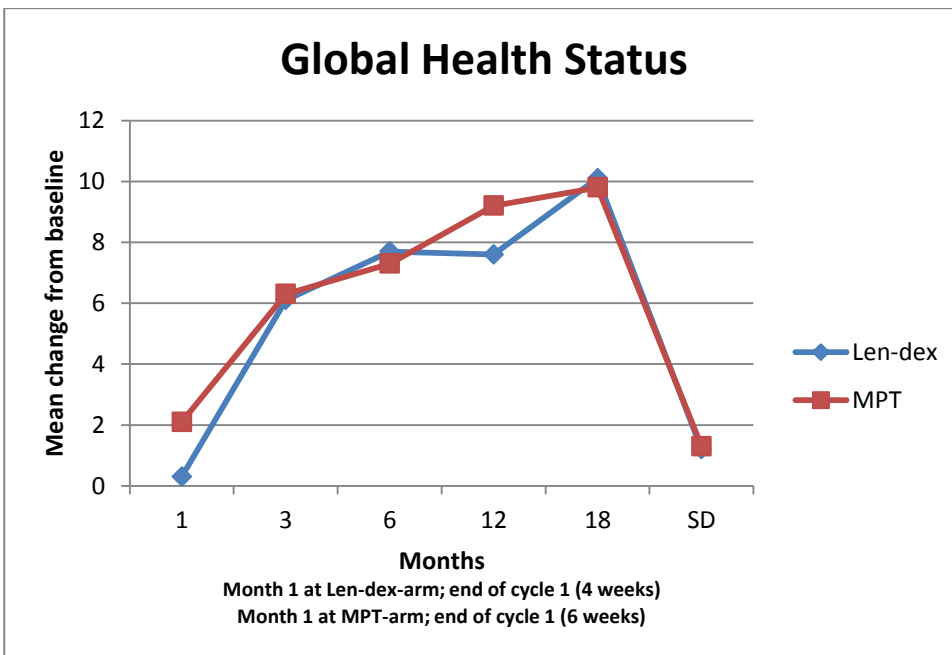
Len-dex; Lenalidomid-dexamethasone, MPT; Melphalan-Prednisolone-Thalidomide, SD: Study discontinuation, which occur at any time point.

Figure P1b. Mean change of score from baseline – Global health status/QoL
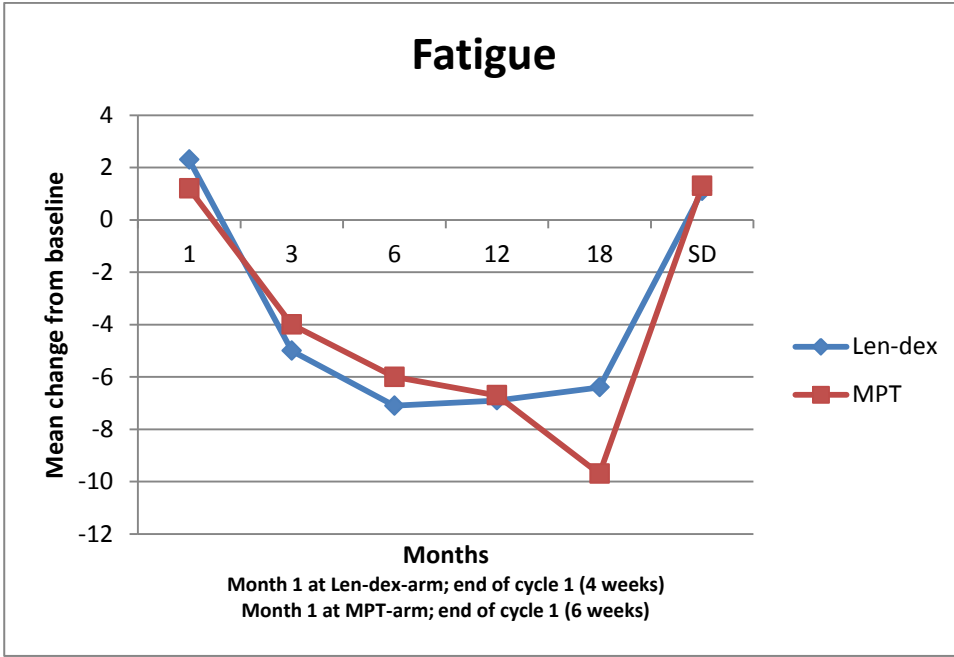
7

Fatigue

Len-dex; Lenalidomid-Dexamethasone, MPT; Melphalan-Prednisolone-Thalidomide, SD: Study discontinuation, which occur at any time point.

Figure P1c. Mean change of score from baseline – Fatigue
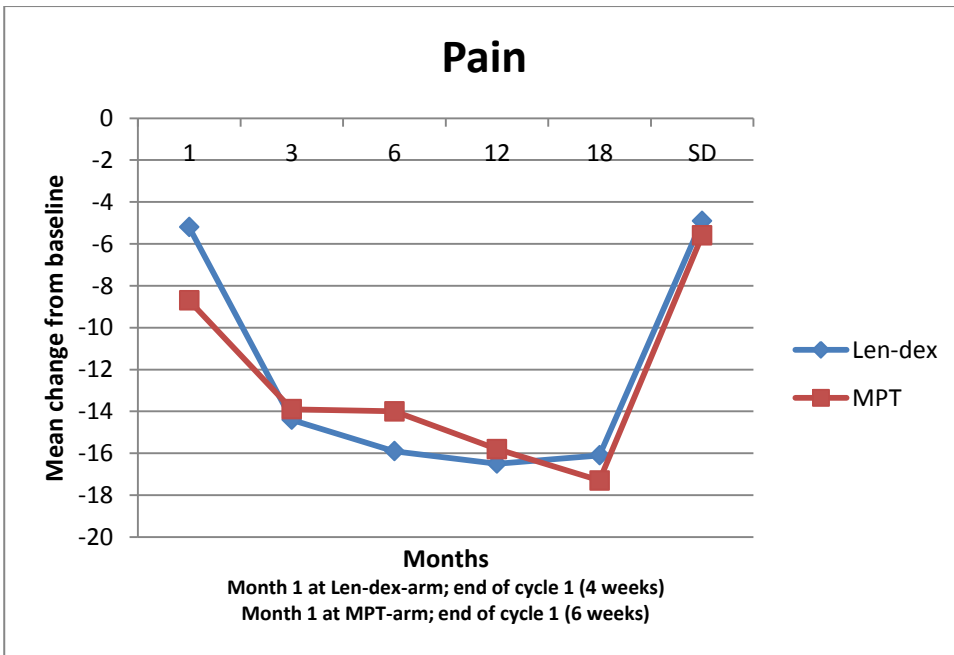
Pain

Len-dex; Lenalidomid-Dexamethasone, MPT; Melphalan-Prednisolone-Thalidomide, SD: Study discontinuation, which occur at any time point.

Figure P1c. Mean change of score from baseline – Pain

Study P2:

The evaluations of mean scores from baseline are based on the mean scores in Table 1 in the publication. End of induction measurement is day 1 in cycle 10, which is the day of start of maintenance. In this evaluation only the induction therapies are maintained.
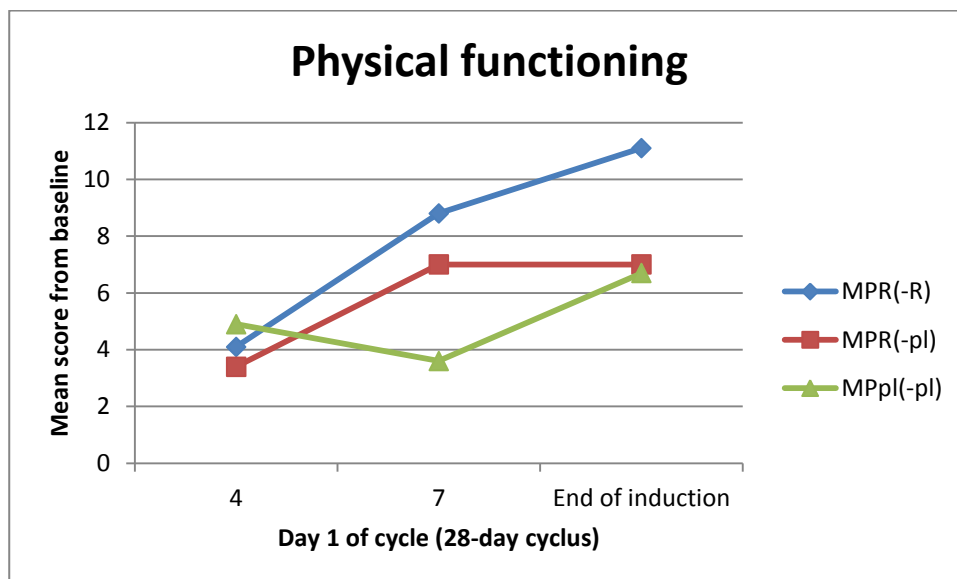
MPR (-R); Melphalan-Prednisolone-Lenalidomid introduction phase with planned Lenalidomid maintenance, MPR(-pl); Melphalan-Prednisolone-Lenalidomid induction phase with planned placebo maintenance, MP(-pl): Melphalan-Prednisolone induction phase with planned placebo maintenance.

Figure P2a. Mean change of score from baseline – Physical functioning
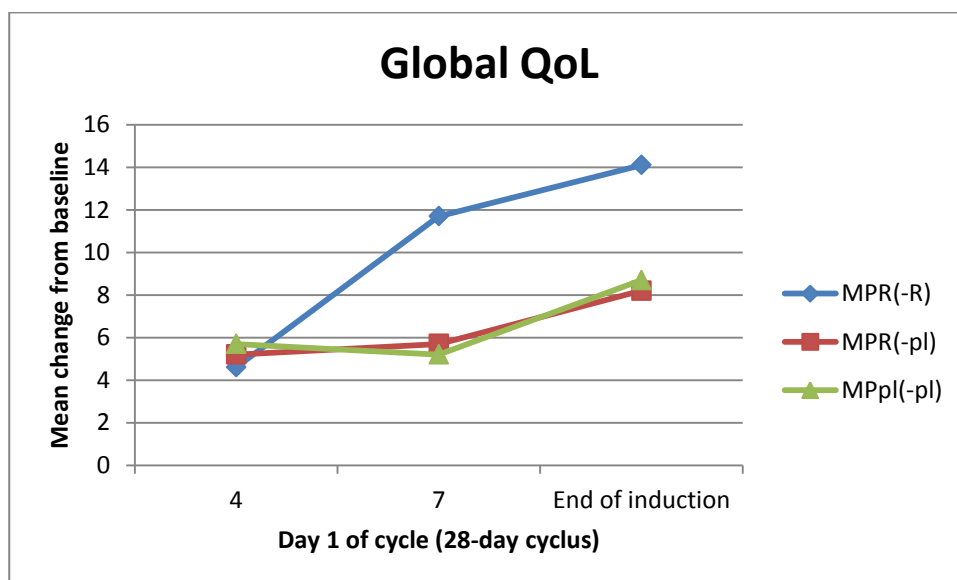
MPR (-R); Melphalan-Prednisolone-Lenalidomid introduction phase with planned Lenalidomid maintenance, MPR(-pl); Melphalan-Prednisolone-Lenalidomid induction phase with planned placebo maintenance, MP(-pl): Melphalan-Prednisolone induction phase with planned placebo maintenance.

Figure P2b. Mean change of score from baseline – Global health status/QoL
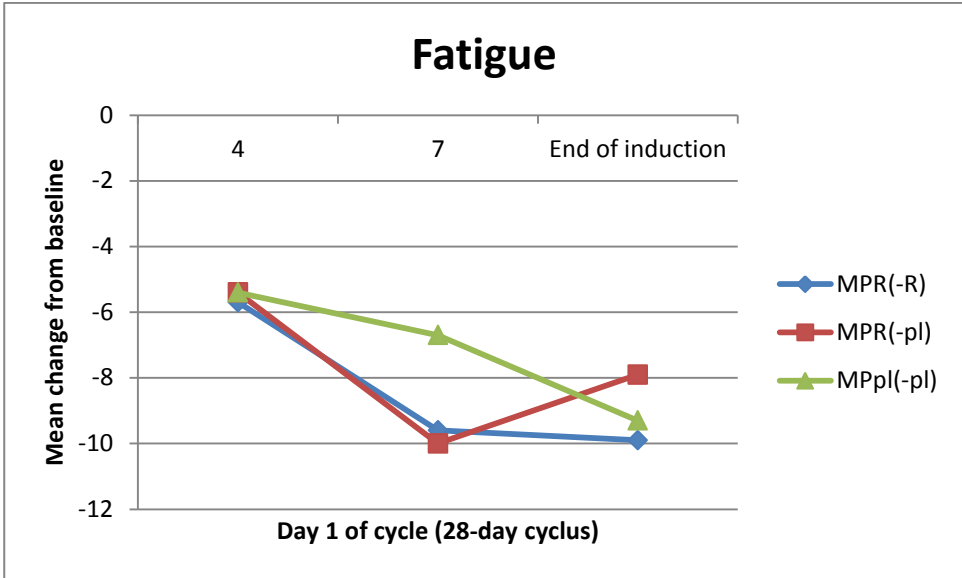
Fatigue

MPR (-R); Melphalan-Prednisolone-Lenalidomid introduction phase with planned Lenalidomid maintenance, MPR(-pl); Melphalan-Prednisolone-Lenalidomid induction phase with planned placebo maintenance, MP(-pl): Melphalan-Prednisolone induction phase with planned placebo maintenance.

Figure P2c. Mean change of score from baseline – Fatigue
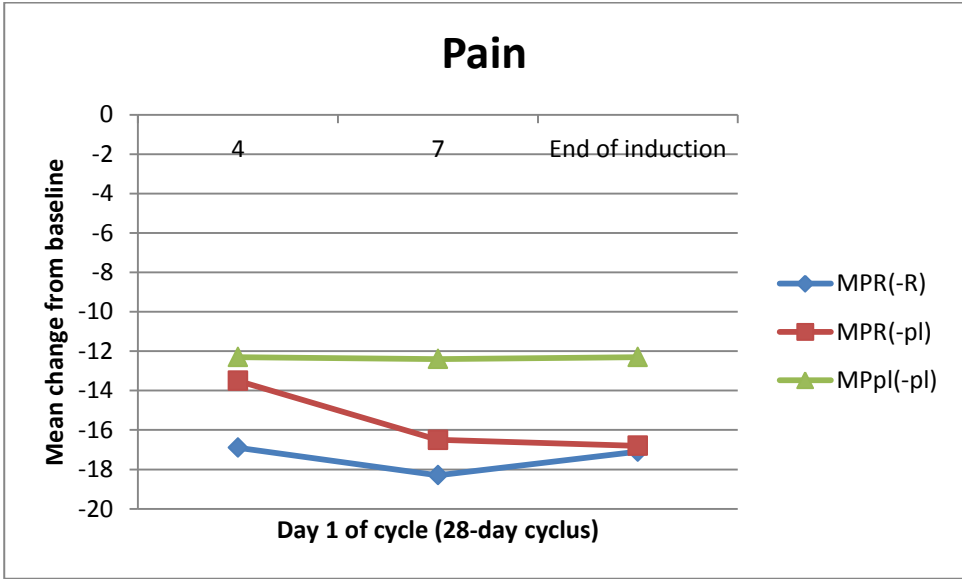
Pain

MPR (-R); Melphalan-Prednisolone-Lenalidomid introduction phase with planned Lenalidomid maintenance, MPR(-pl); Melphalan-Prednisolone-Lenalidomid induction phase with planned placebo maintenance, MP(-pl): Melphalan-Prednisolone induction phase with planned placebo maintenance.

Figure P2d. Mean change of score from baseline – Pain
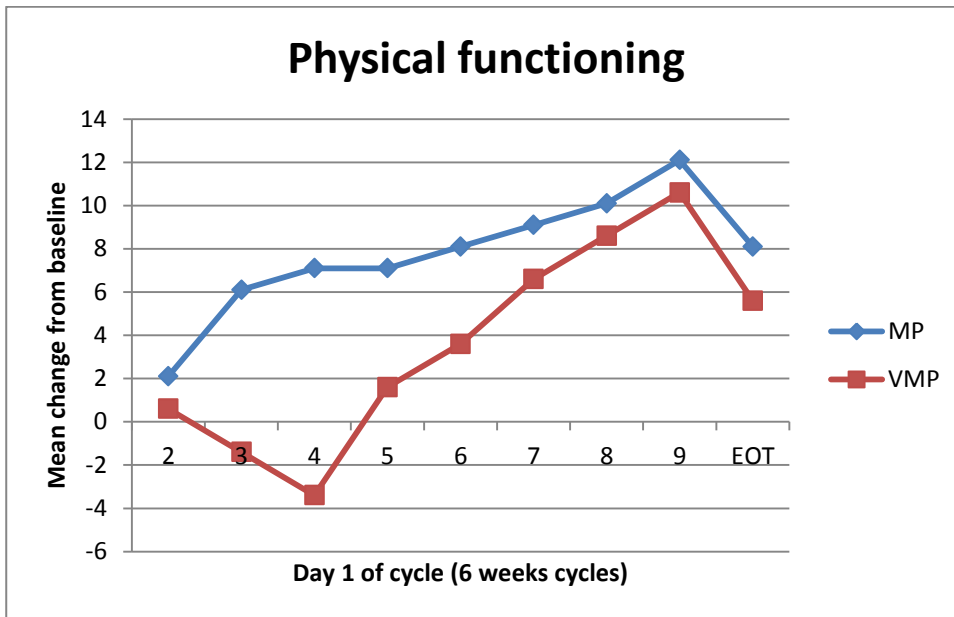
VMP; Melphalan-Prednisone-Bortezomib, MP; Melphalan-Predisolone, EOT; End of treatment

Figure P3a. Mean change of score from baseline – Physical functioning
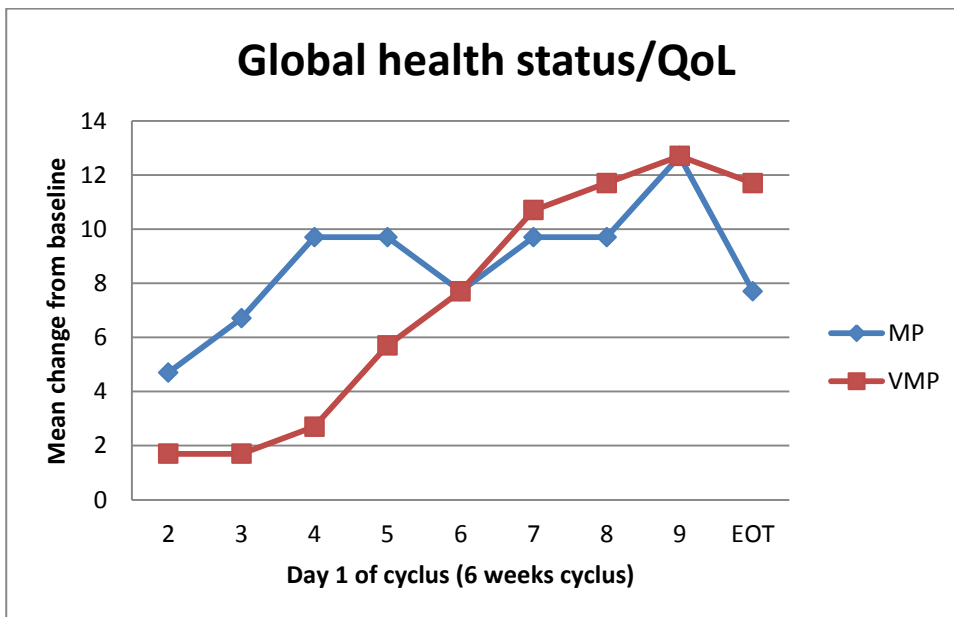
VMP; Melphalan-Prednisone-Bortezomib, MP; Melphalan-Predisolone, EOT; End of treatment

Figure P3b. Mean change of score from baseline – Global health status/QoL
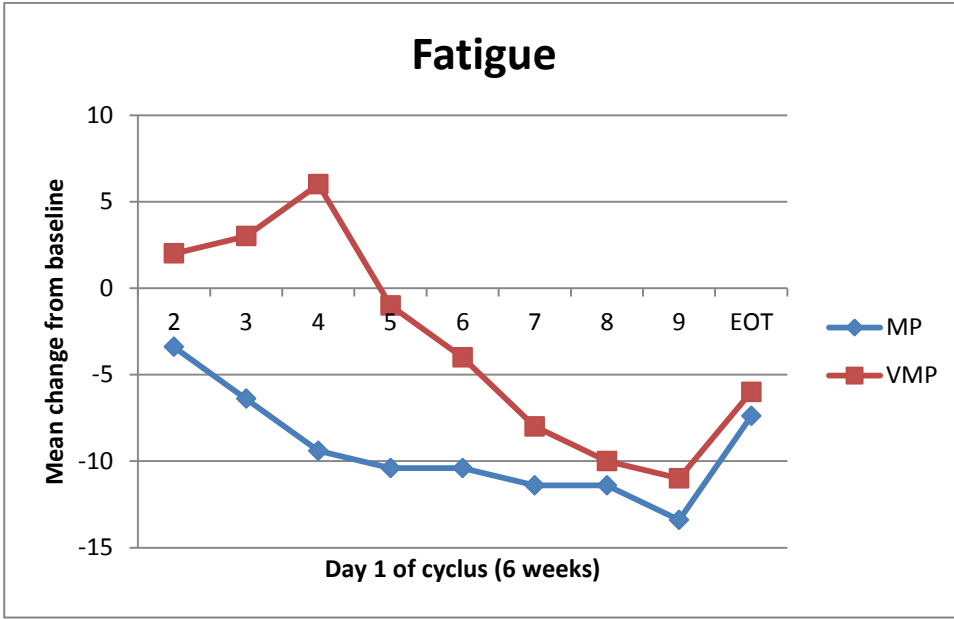
VMP; Melphalan-Prednisone-Bortezomib, MP; Melphalan-Predisolone, EOT; End of treatment

Figure P3c. Mean change of score from baseline – Fatigue
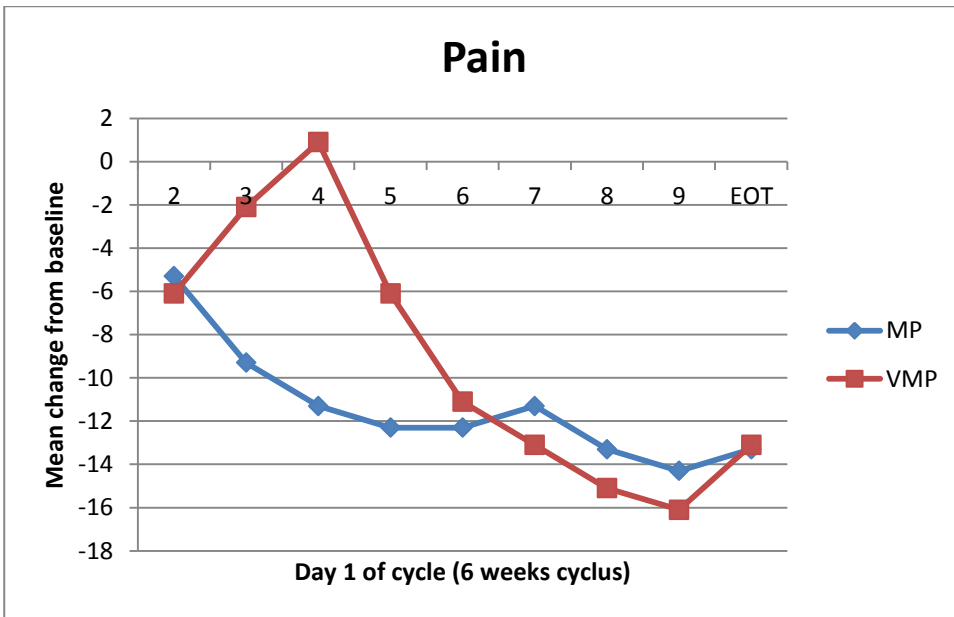
VMP; Melphalan-Prednisone-Bortezomib, MP; Melphalan-Predisolone, EOT; End of treatment

Figure P3d. Mean change of score from baseline – Pain

Study P4:

Follow-up time points
8; at the end of induction treatment > 8 months, 12; at the start of the episode post-induction <12 months
18; at the end of the episode post-induction, Off; after going off protocol
The evaluation of mean change from baseline is based on calculations of mean score listed in Table 2 in the publication.
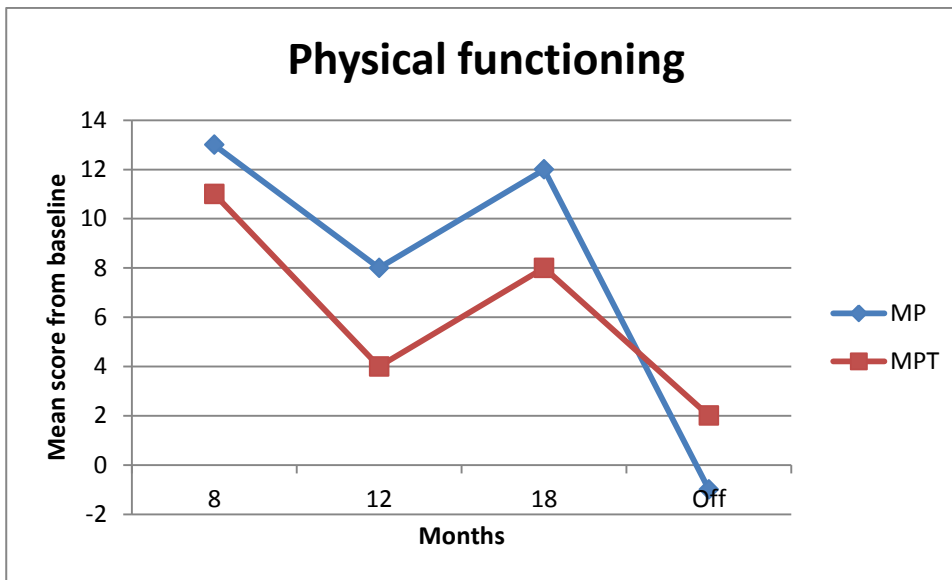
MP; Melphalan-Predisolone, MPT; Melphalan-Prednisolone-Thalidomide, Off: Off protocol
Figure P4a. Mean change of score from baseline – Physical functioning
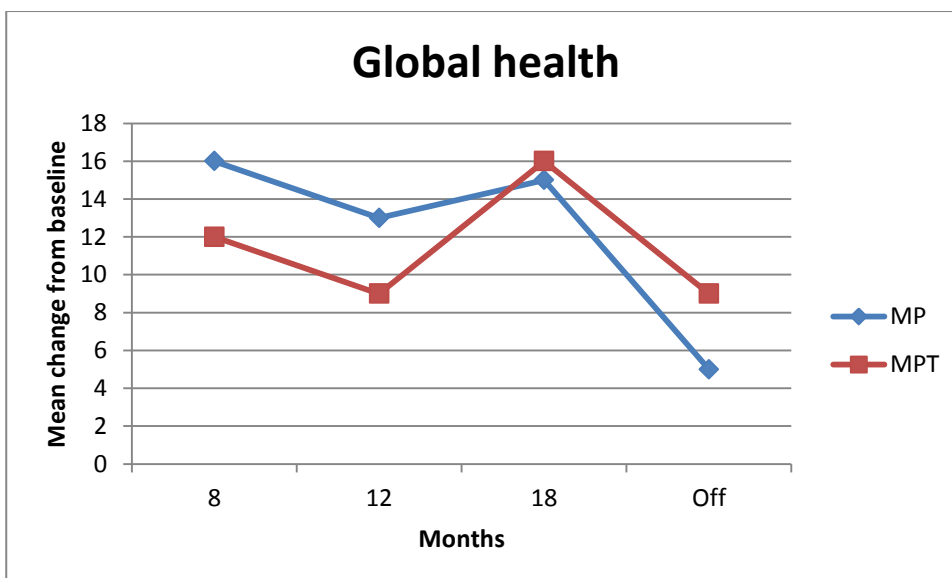
MP; Melphalan-Predisolone, MPT; Melphalan-Prednisolone-Thalidomide, Off: Off protocol
Figure P4b. Mean change of score from baseline – Global health
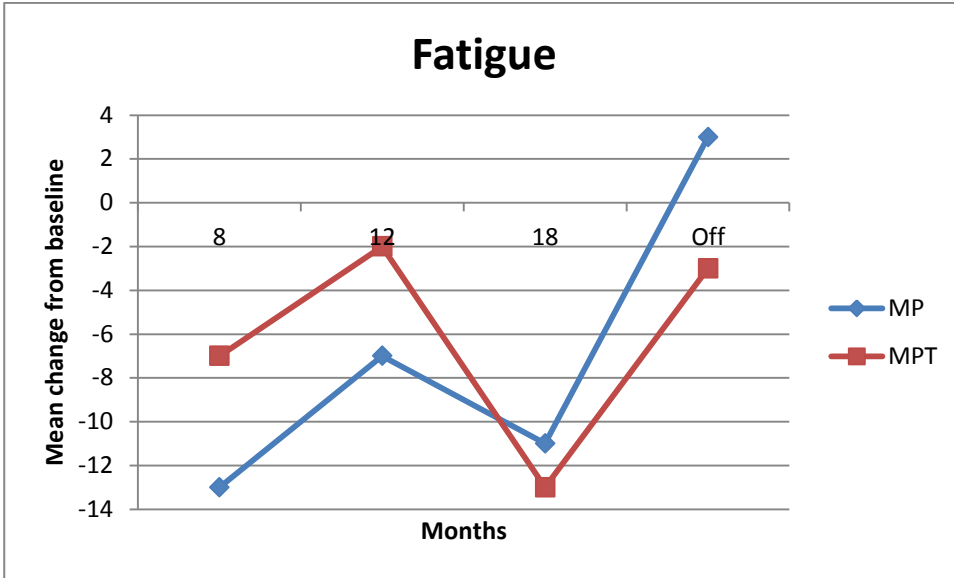
MP; Melphalan-Predisolone, MPT; Melphalan-Prednisolone-Thalidomide, Off: Off protocol

Figure P4c. Mean change of score from baseline – fatigue
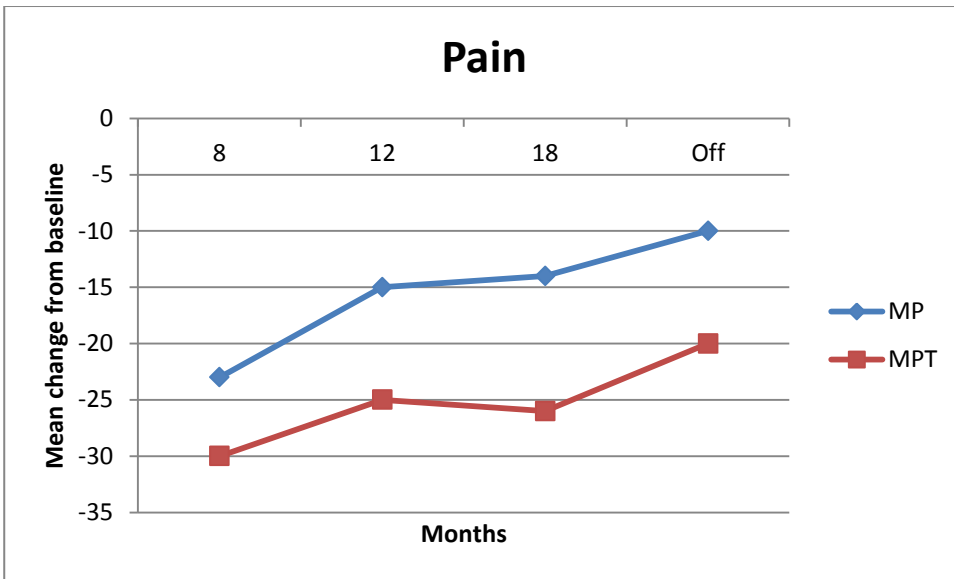
MP; Melphalan-predisolone, MPT; Melphalan-prednisolone-thalidomide, Off: Off protocol

Figure P4d. Mean change of score from baseline – pain

Study P5:
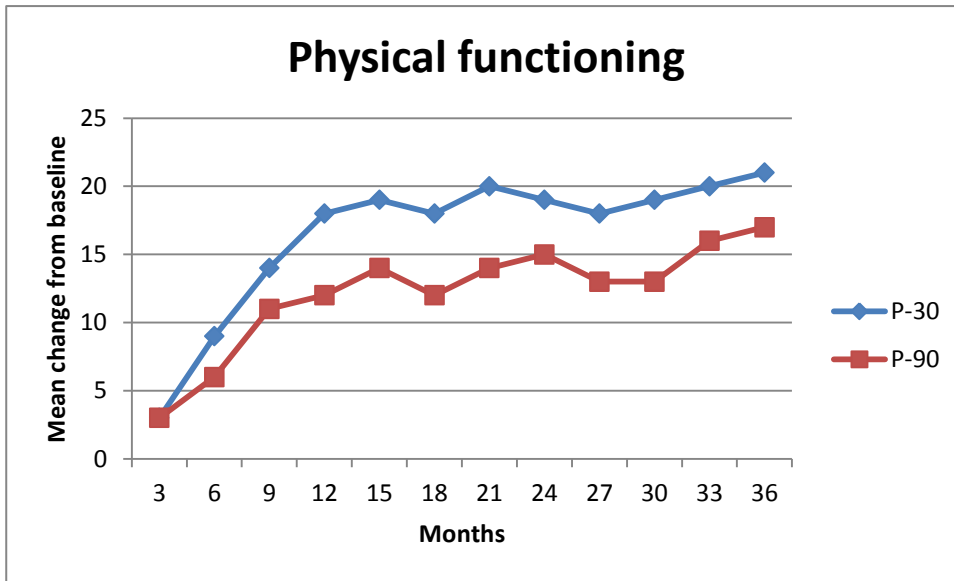
P-30; Pamidronate 30 mg, P-90; Pamidronate 90 mg

Figure P5a. Mean change of score from baseline – Physical functioning
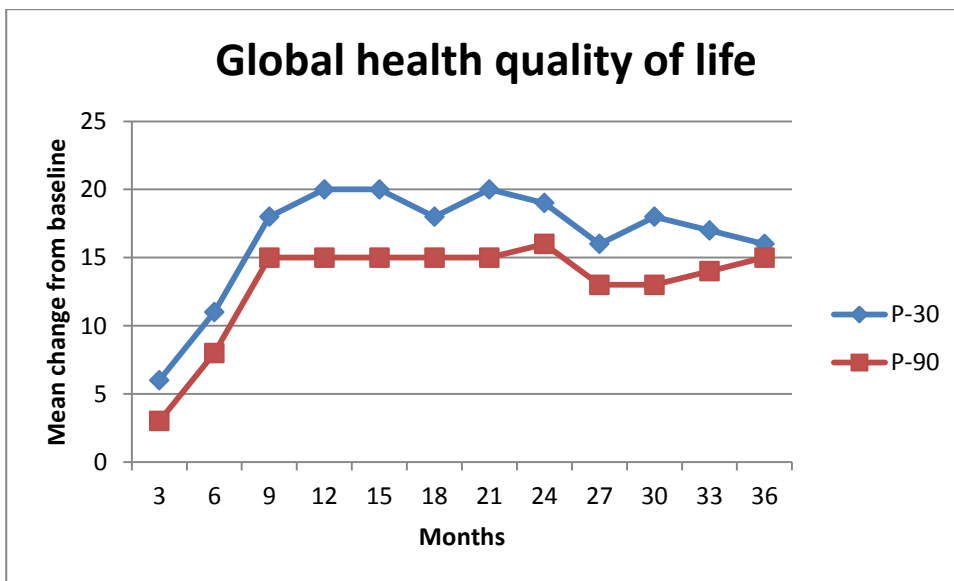
P-30; Pamidronate 30 mg, P-90; Pamidronate 90 mg

Figure P5b. Mean change of score from baseline – Global health quality of life
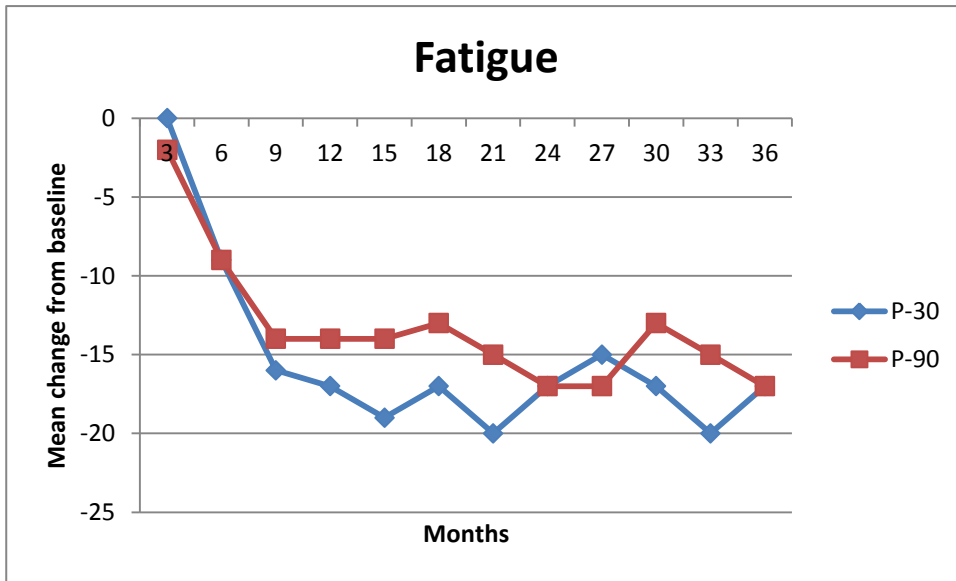
P-30; Pamidronate 30 mg, P-90; Pamidronate 90 mg

Figure P5c. Mean change of score from baseline – Fatigue
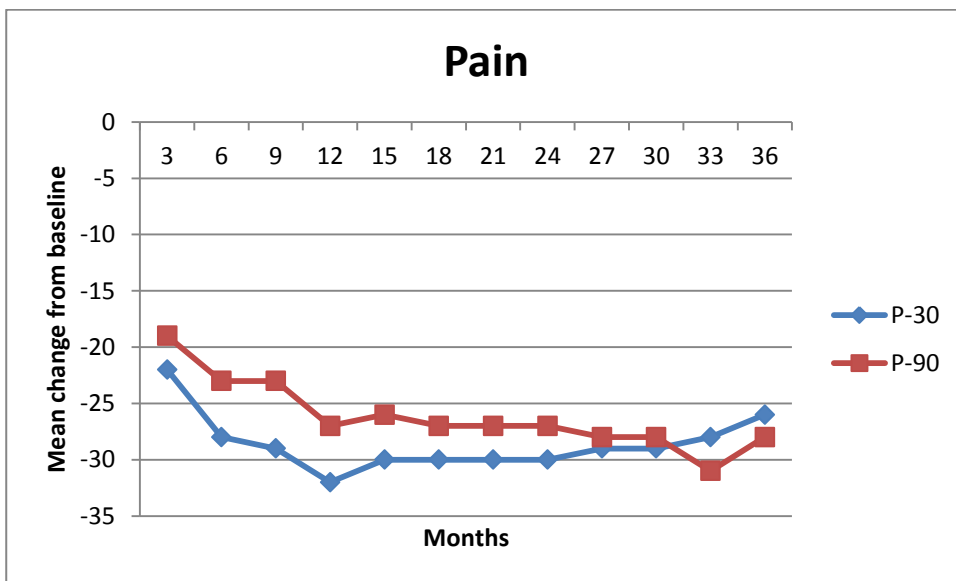
P-30; Pamidronate 30 mg, P-90; Pamidronate 90 mg

Figure P5d. Mean change of score from baseline – pain

Study P6:
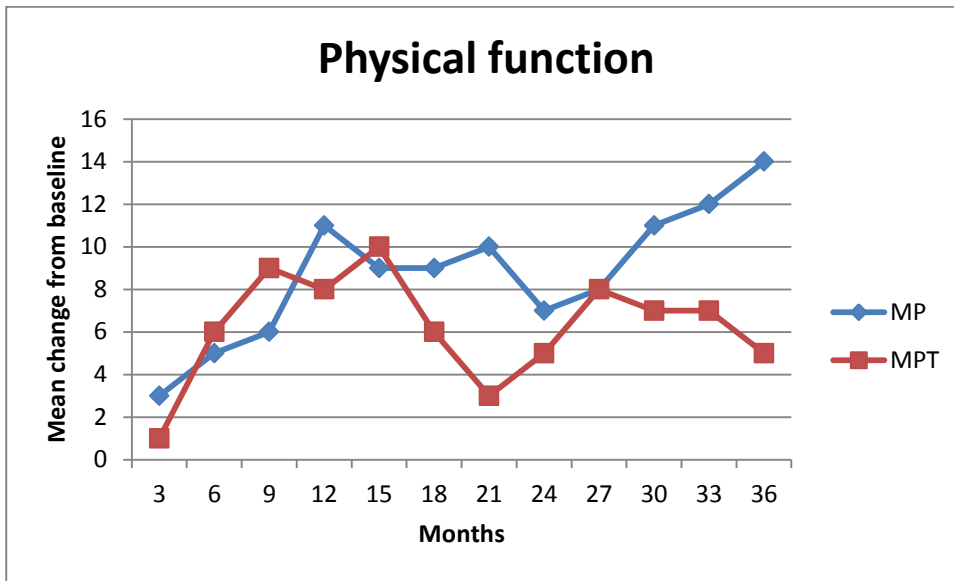
MP; Melphalan-Predisolone, MPT; Melphalan-Prednisolone-Thalidomide

Figure P6a. Mean change of score from baseline – Physical function
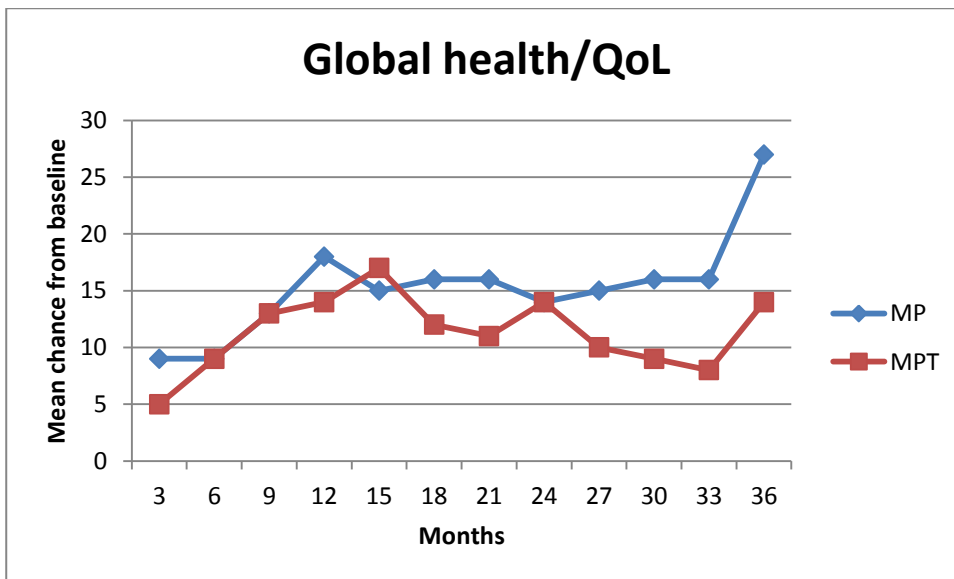
MP; Melphalan-Predisolone, MPT; Melphalan-Prednisolone-Thalidomide

Figure P6b. Mean change of score from baseline – Global health/QoL
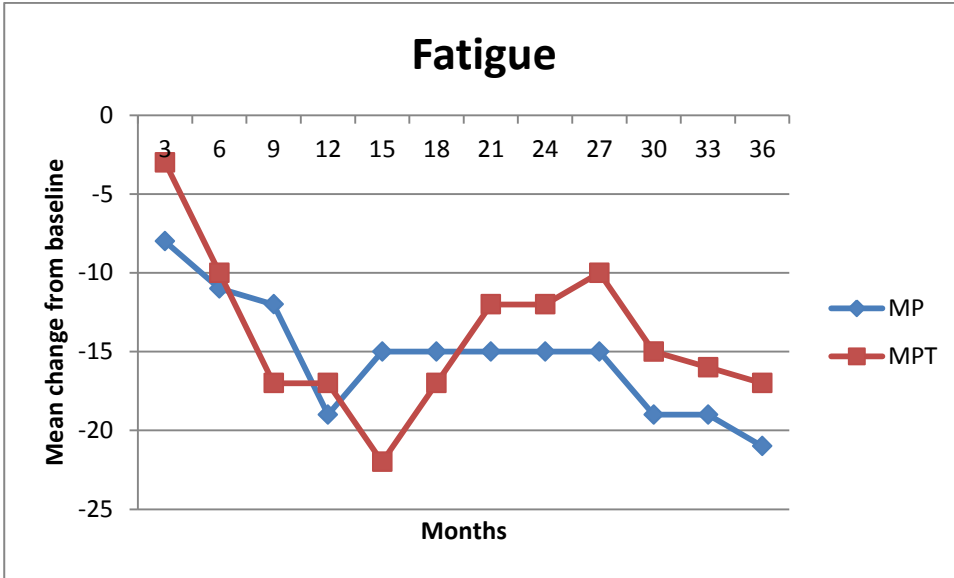
MP; Melphalan-Predisolone, MPT; Melphalan-Prednisolone-Thalidomide

Figure P6c. Mean change of score from baseline – Fatigue
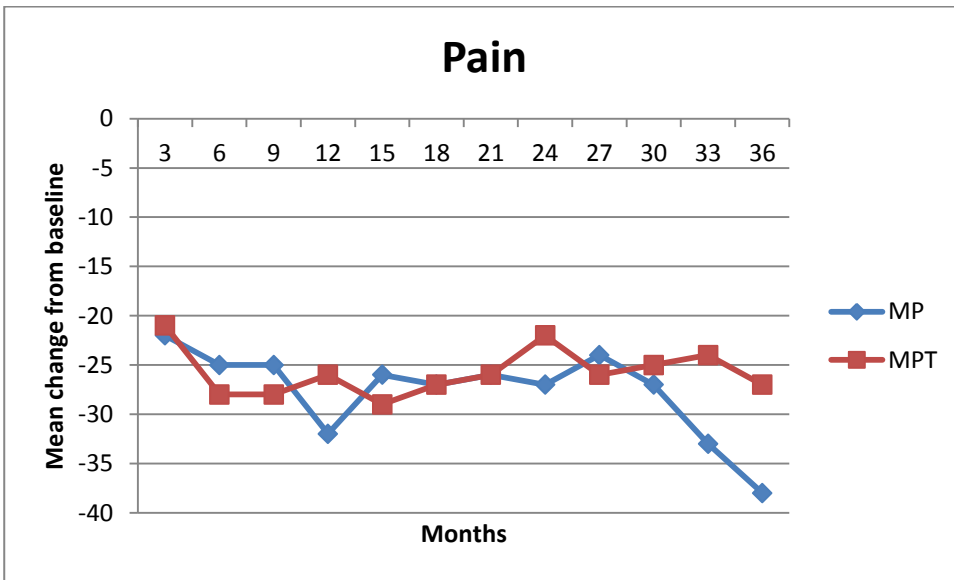
MP; Melphalan-Predisolone, MPT; Melphalan-Prednisolone-Thalidomide

Figure P6d. Mean change of score from baseline – pain

Study P7

	Physical function	Global quality of life	Fatigue	Pain
Mean change from baseline to 6 month	13,8	13,8	-11,4	-18,5

Figure P7

Study P8
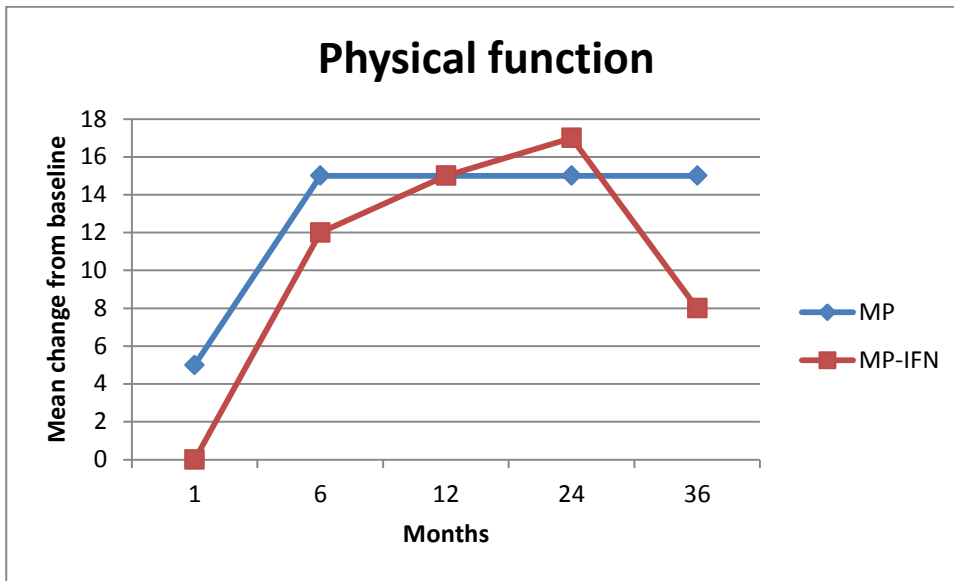
MP; Melphalan-Predisolone, MP-IFN; Melphalan-Prednisone-Interferon alfa-2b

Figure P8a. Mean change of score from baseline – Physical function
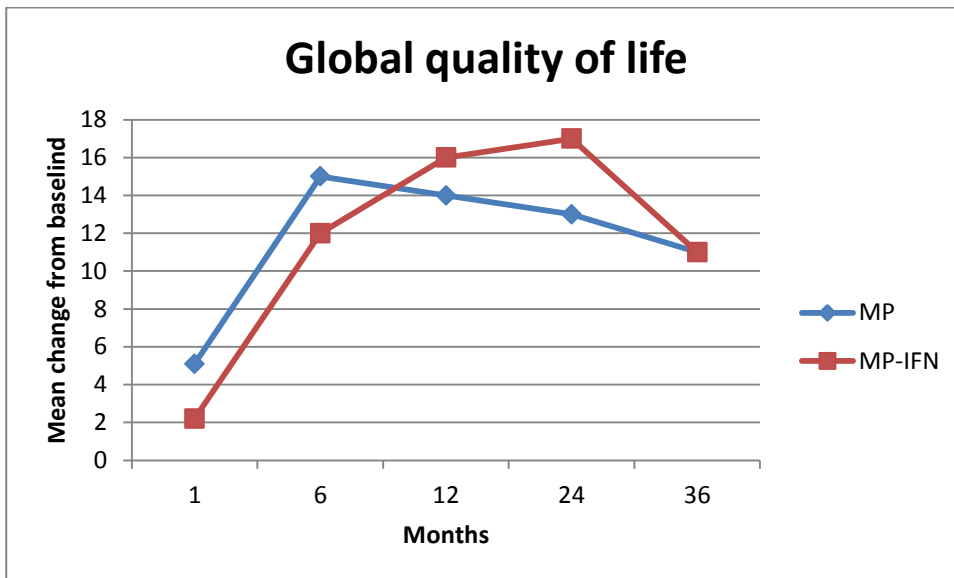
MP; Melphalan-Predisolone, MP-IFN; Melphalan-Prednisone-Interferon alfa-2b

Figure P8b. Mean change of score from baseline – Global quality of life
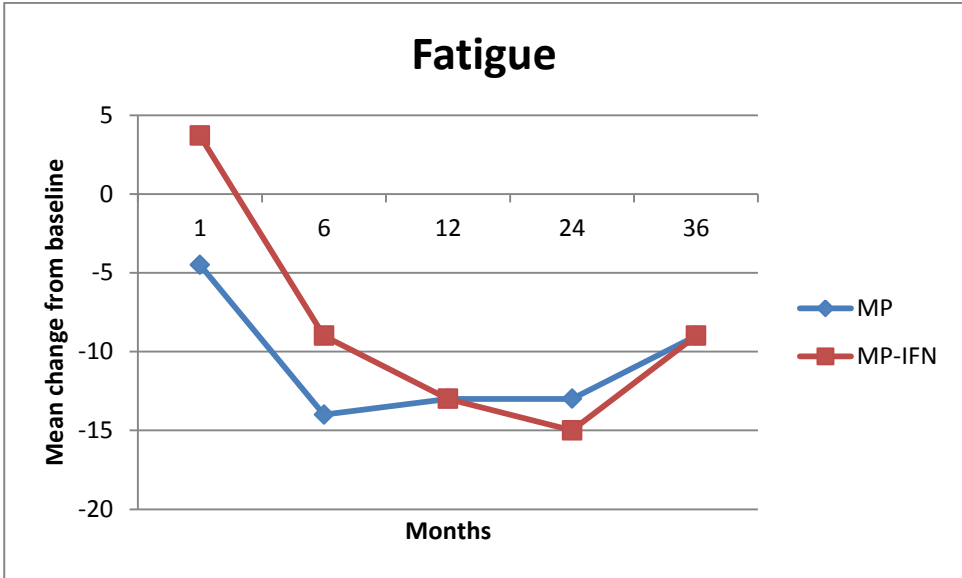
MP; Melphalan-Predisolone, MP-IFN; Melphalan-Prednisone-Interferon alfa-2b

Figure P8c. Mean change of score from baseline – Fatigue
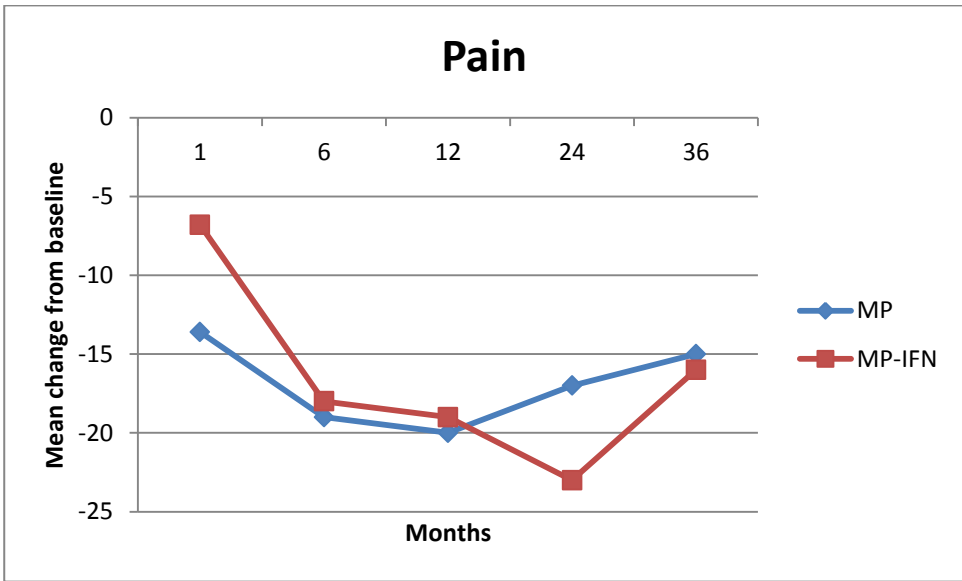
MP; Melphalan-Predisolone, MP-IFN; Melphalan-Prednisone-Interferon alfa-2b

Figure P8d. Mean change of score from baseline – pain

CONSOLIDATION TREATMENT STUDIES

Study C1:
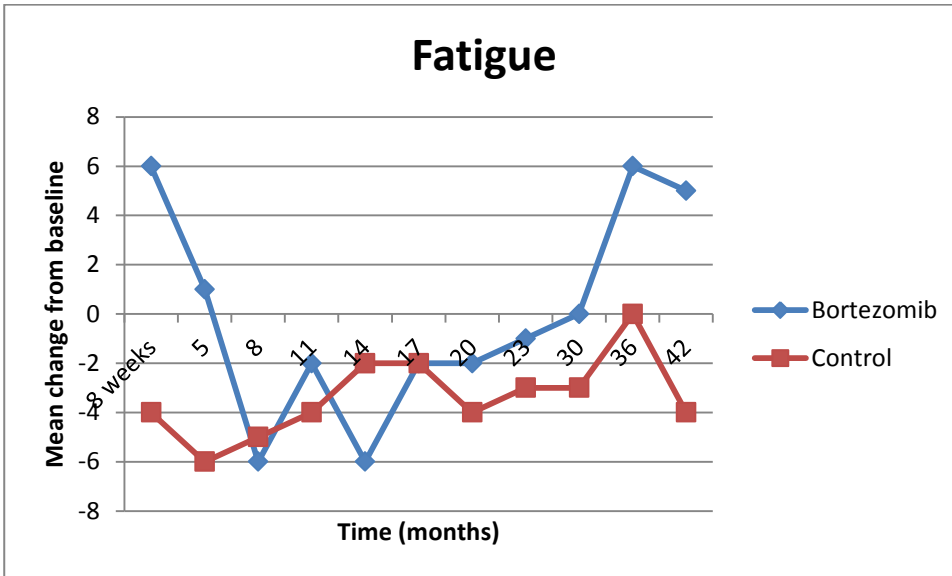
Figure C1c. Mean change of score from baseline – Fatigue

Study C2:
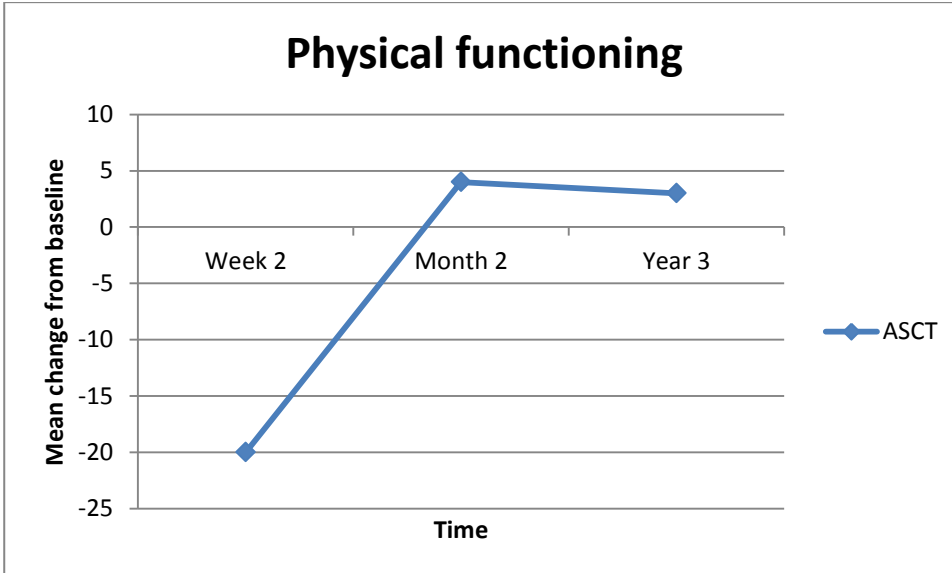
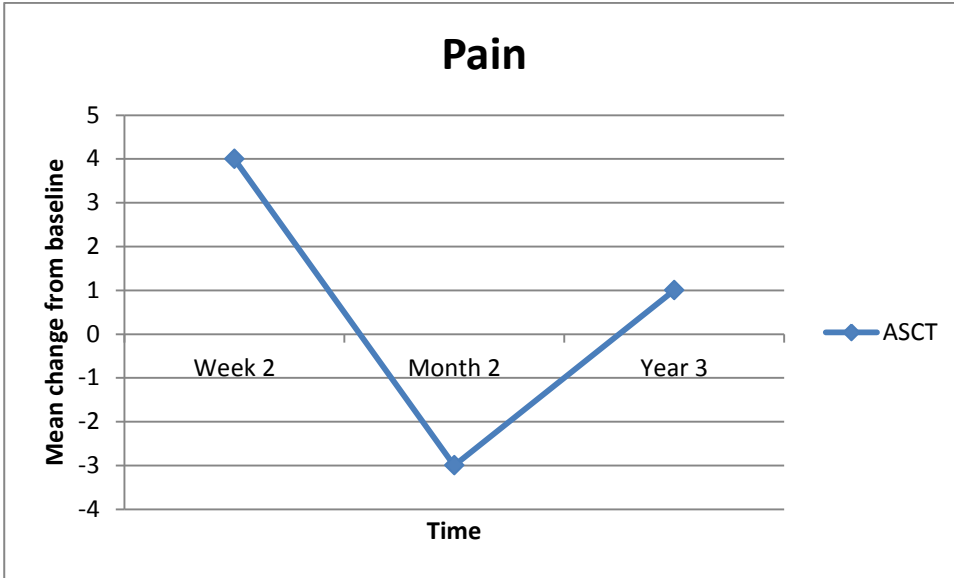

ASCT; Autologous haematopoietic stem cell transplantation

Figure C2a. Mean change of score from baseline – Physical functioning
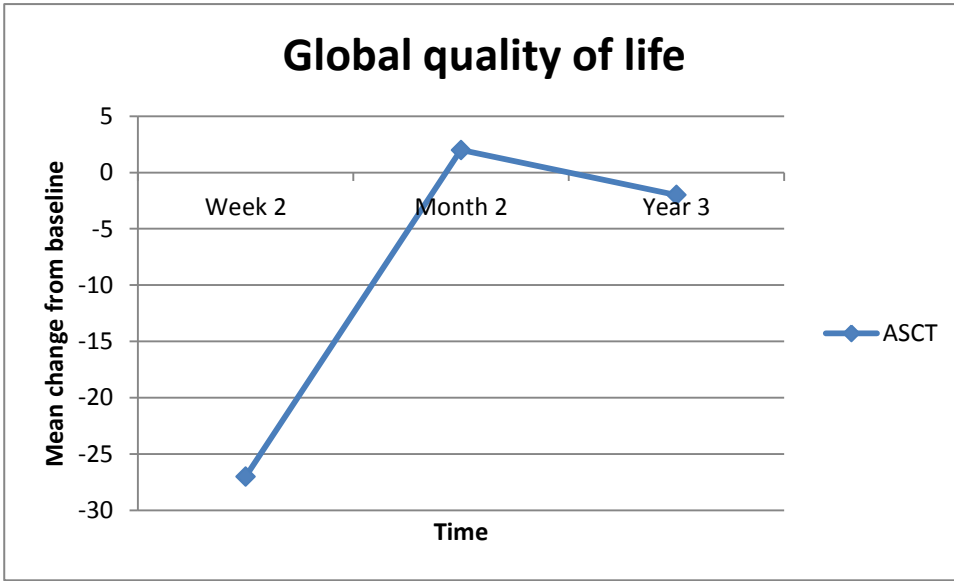
ASCT; Autologous haematopoietic stem cell transplantation

Figure C2b. Mean change of score from baseline – Global quality of life
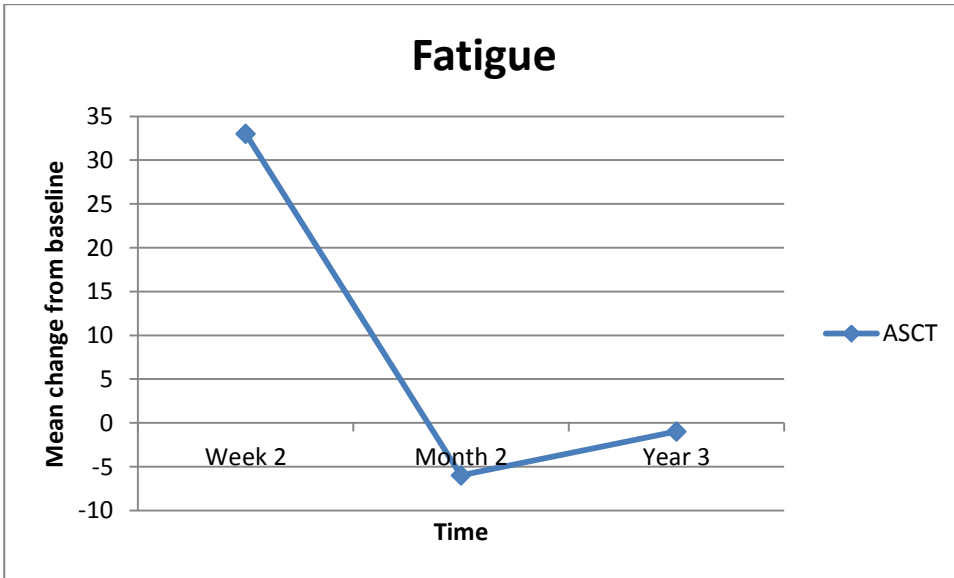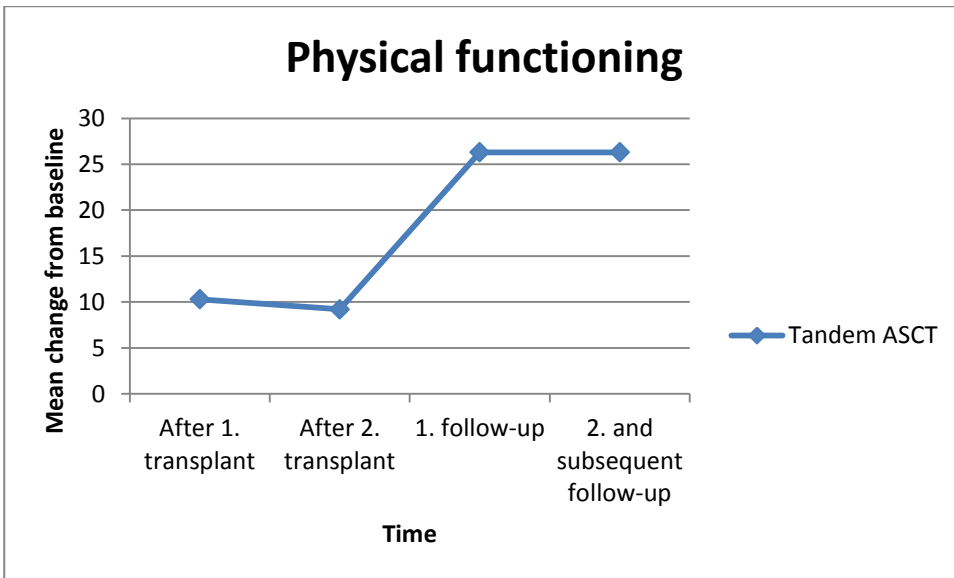

ASCT; Autologous haematopoietic stem cell transplantation

Figure C2c. Mean change of score from baseline – Fatigue

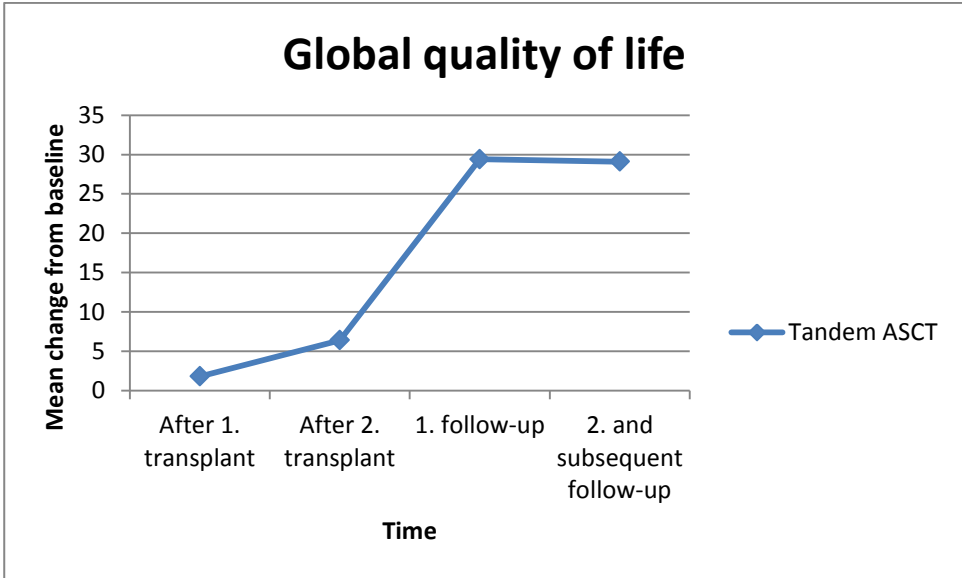ASCT; Autologous haematopoietic stem cell transplantation

Figure C2d. Mean change of score from baseline – Pain

Study C3:

ASCT; Autologous haematopoietic stem cell transplantation

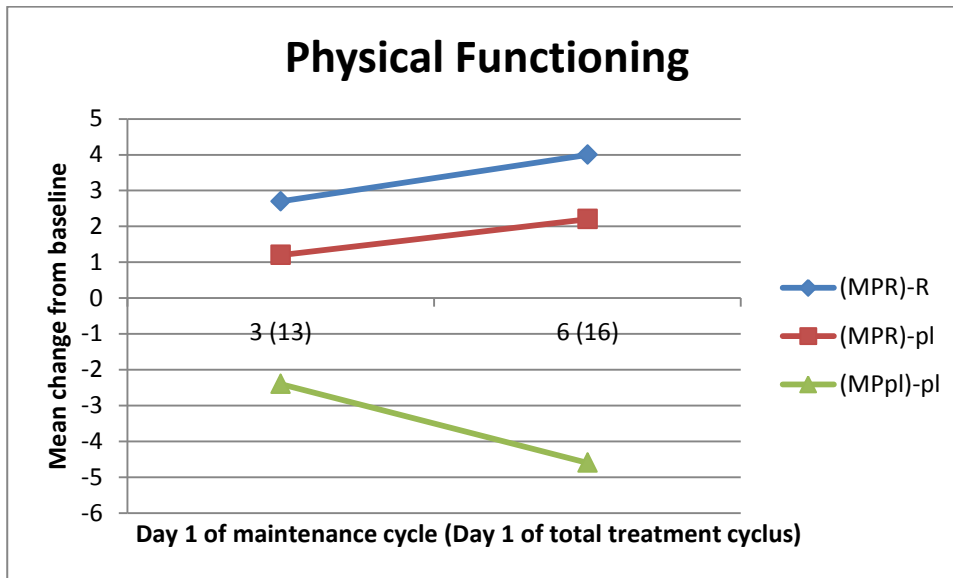Figure C3a. Mean change of score from baseline – Physical functioning

ASCT; Autologous haematopoietic stem cell transplantation

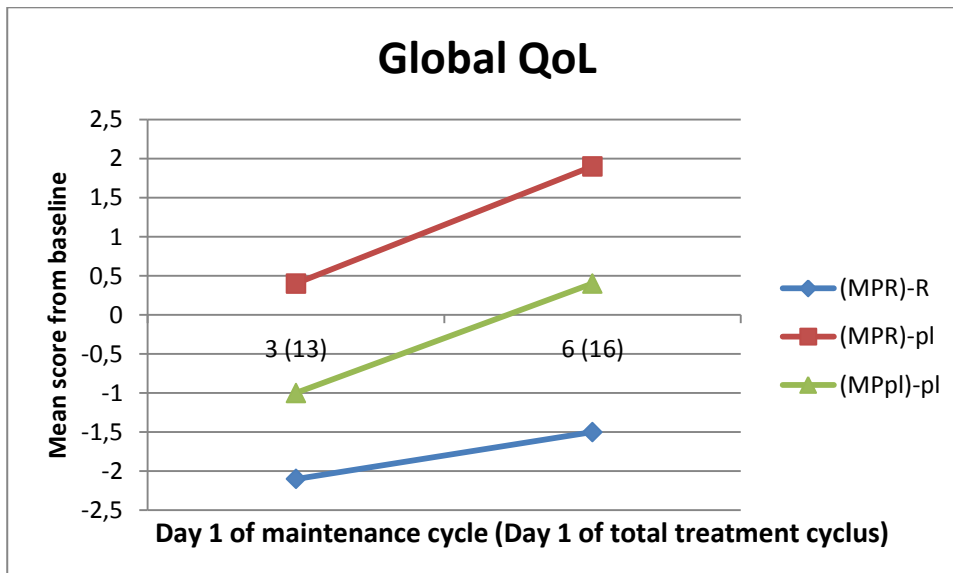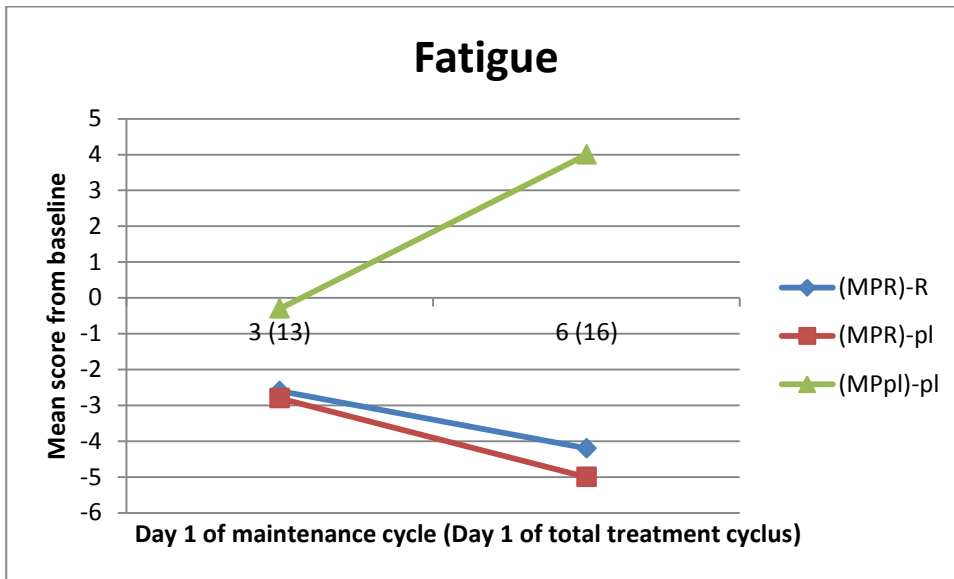Figure C3b. Mean change of score from baseline – Global quality of life

Study M1

(MPR)-R; Lenalidomid maintenance after Melphalan-prednisolone-lenalidomid induction, (MPR)-pl; Placebo maintenance after Melphalan-prednisolone-lenalidomid induction, (MPpl)-pl: Placebo maintenance after Melphalan-prednisolone-placebo induction

Figure M1a. Mean change of score from baseline – Physical functioning

(MPR)-R; Lenalidomid maintenance after Melphalan-prednisolone-lenalidomid induction, (MPR)-pl; Placebo maintenance after Melphalan-prednisolone-lenalidomid induction, (MPpl)-pl: Placebo maintenance after Melphalan-prednisolone-placebo induction

Figure M1b. Mean change of score from baseline – Global QoL

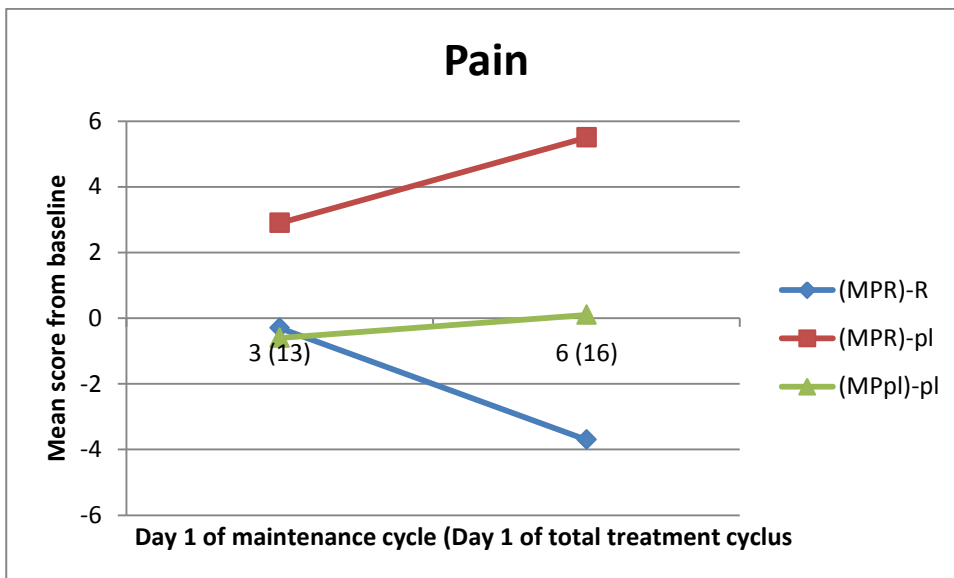**Figure M1c. Mean change of score from baseline – Fatigue**

(MPR)-R; Lenalidomid maintenance after Melphalan-prednisolone-lenalidomid induction, (MPR)-pl; Placebo maintenance after Melphalan-prednisolone-lenalidomid induction, (MPpl)-pl: Placebo maintenance after Melphalan-prednisolone-placebo induction

Figure M1d. Mean change of score from baseline – Pain

(MPR)-R; Lenalidomid maintenance after Melphalan-prednisolone-lenalidomid induction, (MPR)-pl; Placebo maintenance after Melphalan-prednisolone-lenalidomid induction, (MPpl)-pl: Placebo maintenance after Melphalan-prednisolone-placebo induction

Study M2

	Physical functioning	Global quality of life	Fatigue	Pain
Mean change from baseline to 3 months P-IFN	3,33	1,67	-4,96	-9,09
Mean change from baseline to 3 months IFN	-0,52	-7,18	5,84	-1,44

P-IFN; Pegylated interferon-alfa2b, INF; Interferon-a2
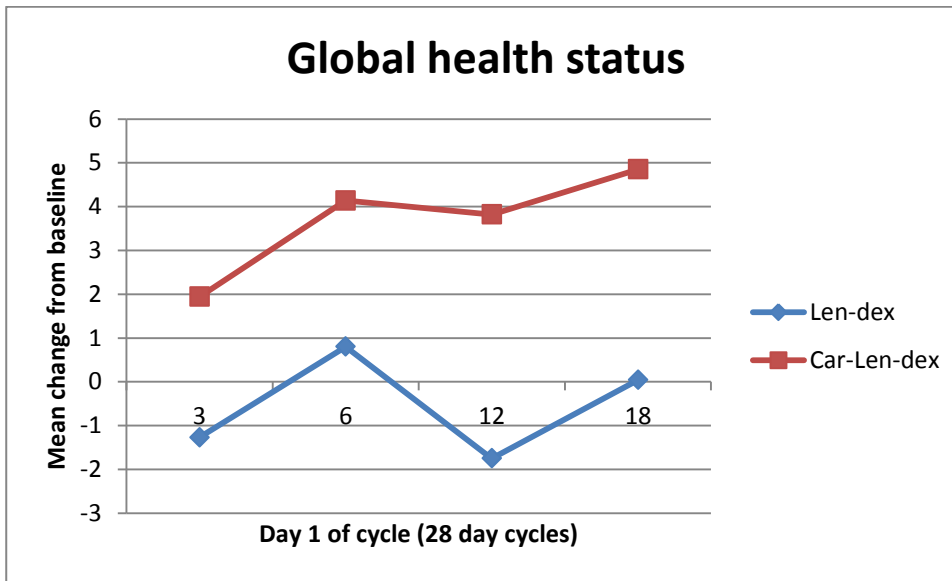
Figure M2

Study R1:
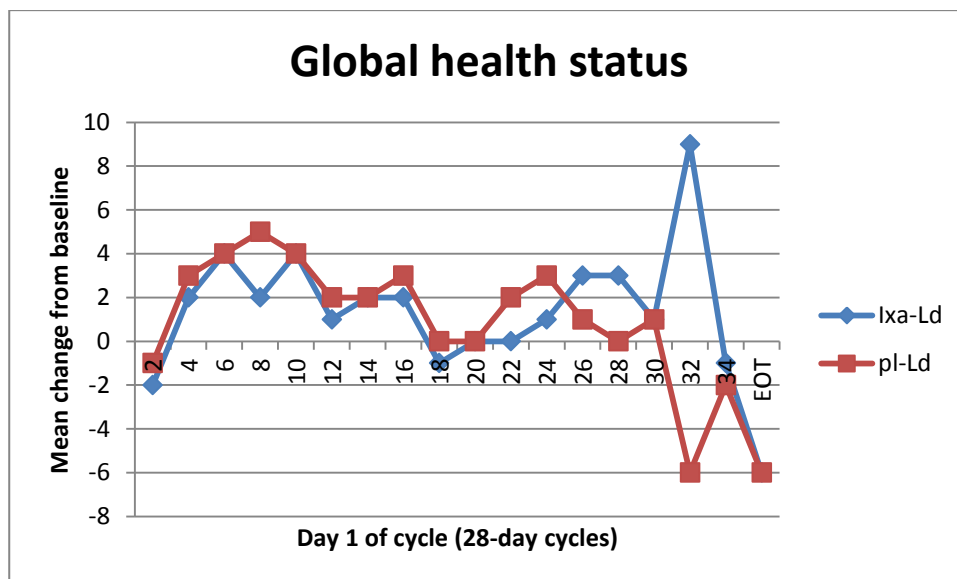
Ixa-Ld; Ixazomib-Lenalidomid-Dexamethasone, pl-Ld; placebo plus Lenalidomid-Dexamethasone, EOT; End of treatment

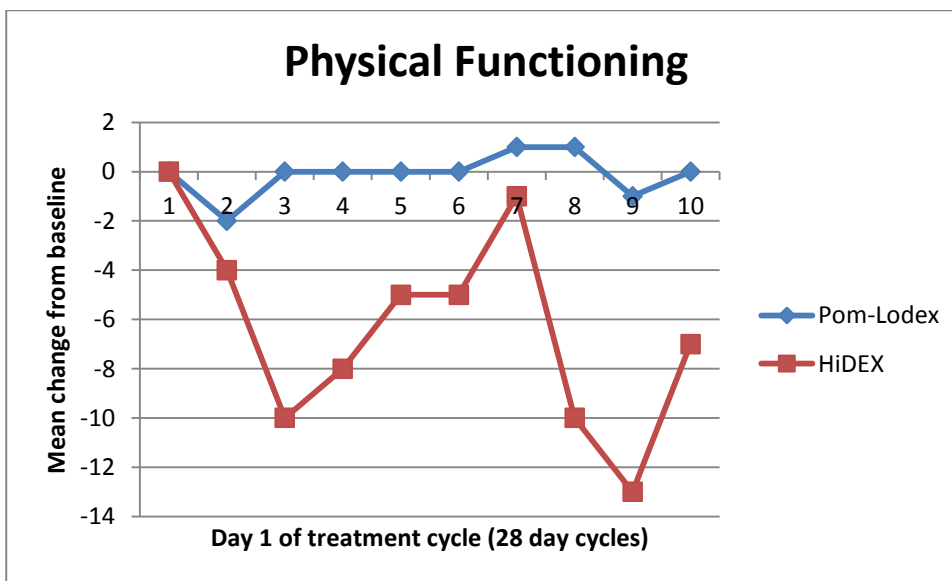Figure R1b. Mean change of score from baseline – Global Health status

The last evaluation of the mean change from baseline was at 22 months for both groups, because the compliance hereafter was under 25% from baseline.

Study R2:

Car-Len-dex; Carfilzomib-Lenalidomid-dexamethasone, Len-dex; Lenalidomid-dexamethasone

Figure R2b. Mean change of score from baseline – Global Health status

Study R3:

Pom-Lodex; Pomalidomide low-dose Dexamethasone, HiDEX; High-dose Dexamethasone

Figure R3a. Mean change of score from baseline – Physical Functioning

The last evaluation of the mean change from baseline for physical functioning was at treatment cycle 8 for Pom-Lodex and cycle 3 for HIDEX, because the compliance hereafter was under 25% from baseline.
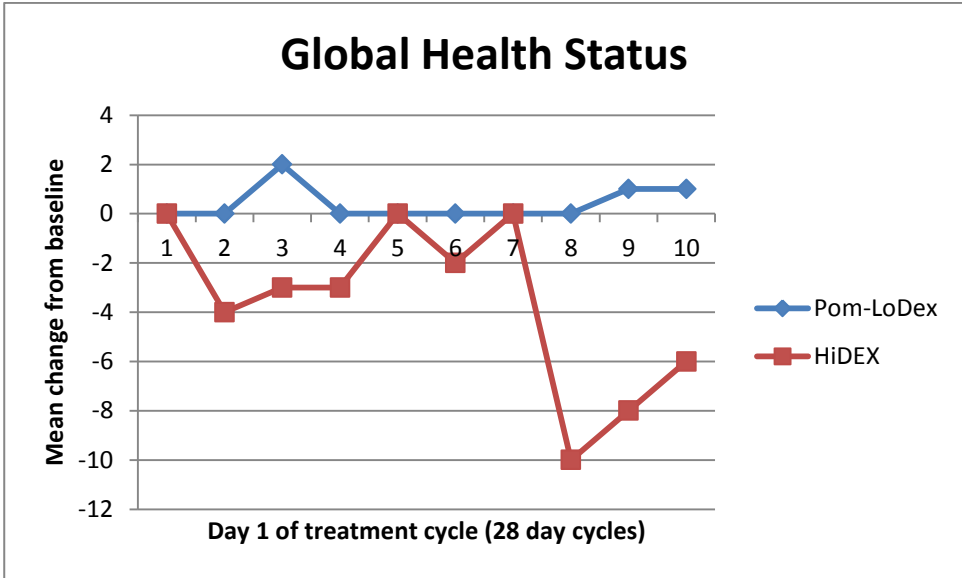
Pom-Lodex; Pomalidomide low-dose Dexamethasone, HiDEX; High-dose Dexamethasone

Figure R3b. Mean change of score from baseline – Global Health status

The last evaluation of the mean change from baseline for global health status was at treatment cycle 8 for Pom-Lodex and cycle 4 for HIDEX, because the compliance hereafter was under 25% from baseline.

Pom-Lodex; Pomalidomide low-dose Dexamethasone, HiDEX; High-dose Dexamethasone

Figure R3c. Mean change of score from baseline – Fatigue

The last evaluation of the mean change from baseline for fatigue was at treatment cycle 8 for Pom-Lodex and cycle 4 for HIDEX, because the compliance hereafter was under 25% from baseline.

Pom-Lodex; Pomalidomide low-dose Dexamethasone, HiDEX; High-dose Dexamethasone

Figure R3d. Mean change of score from baseline – Pain
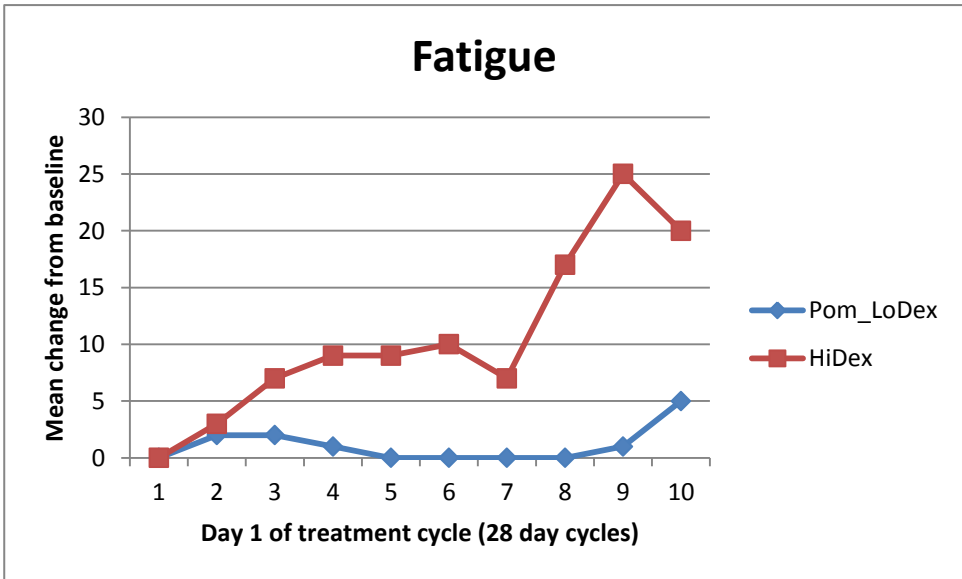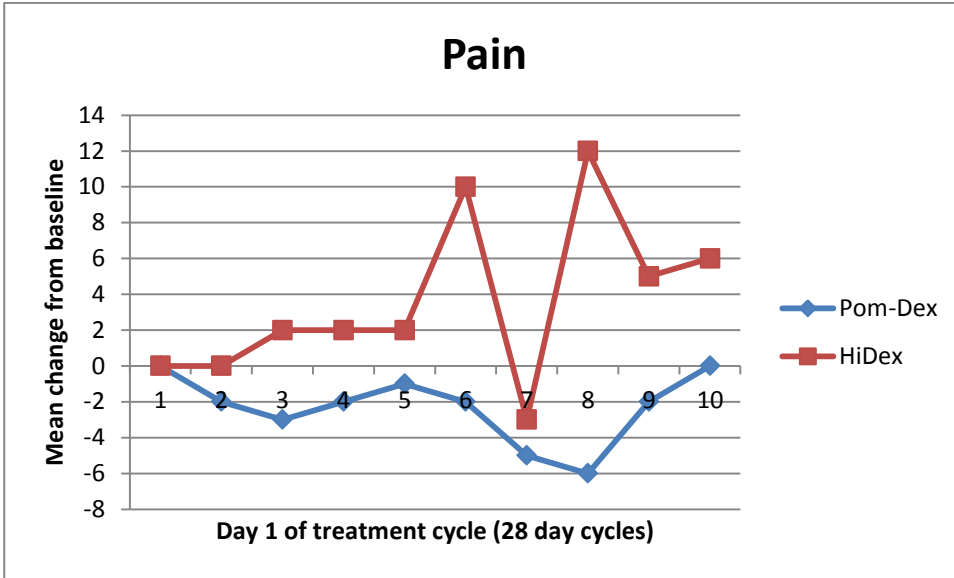
The last evaluation of the mean change from baseline for pain was at treatment cycle 8 for Pom-Lodex and cycle 4 for HIDEX, because the compliance hereafter was under 25% from baseline.

Study R4:
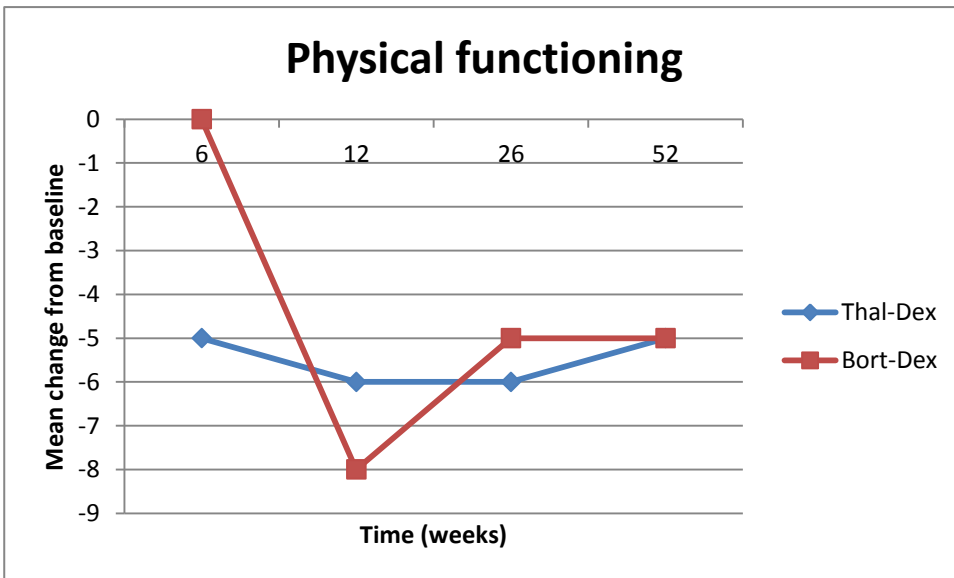
Thal-Dex; Thalidomide-Dexamethasone, Bort-Dex; Bortezomib-Dexamethasone

Figure R4a. Mean change of score from baseline – Physical Functioning
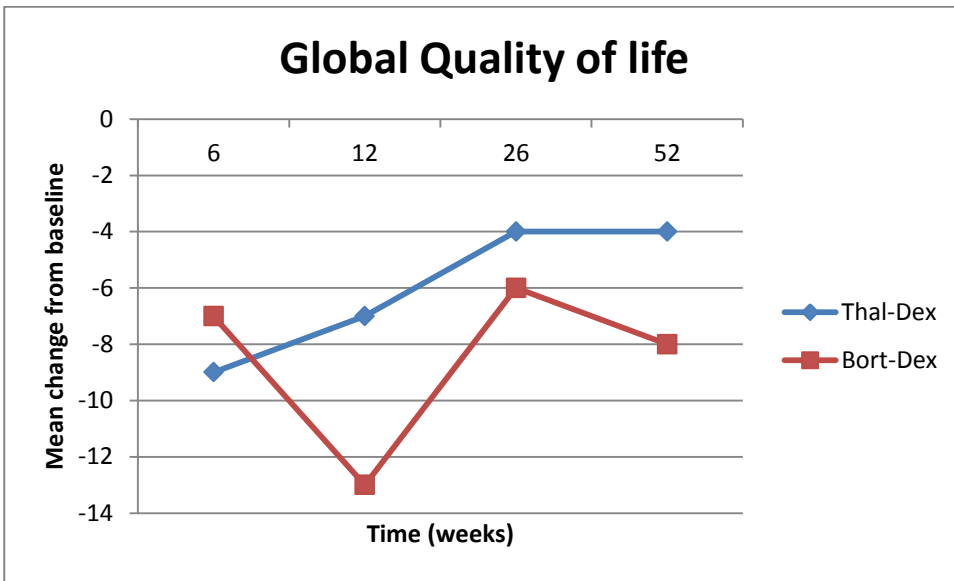
Thal-Dex; Thalidomide-Dexamethasone, Bort-Dex; Bortezomib-Dexamethasone

Figure R4b. Mean change of score from baseline – Global Quality of life
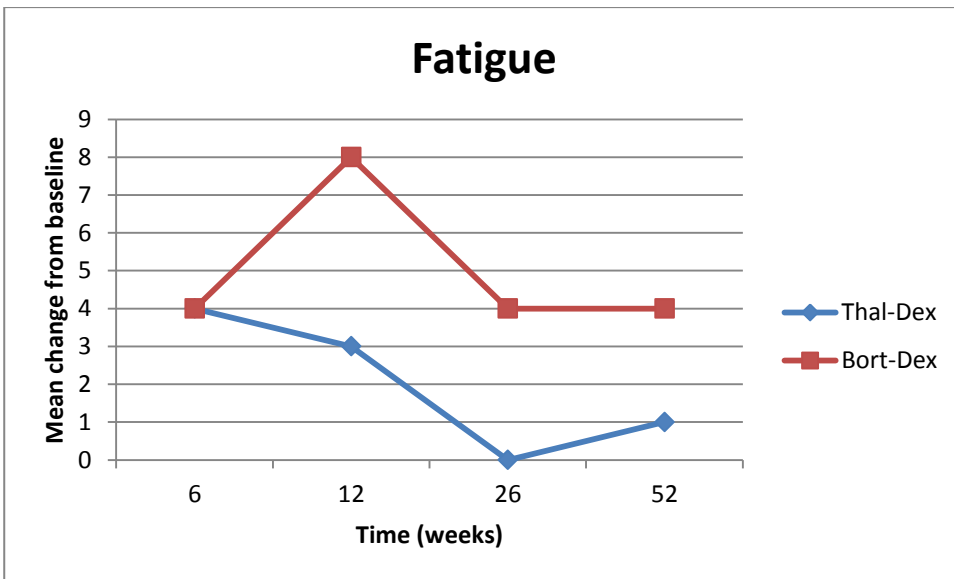
Thal-Dex; Thalidomide-Dexamethasone, Bort-Dex; Bortezomib-Dexamethasone

Figure R4c. Mean change of score from baseline – Fatigue
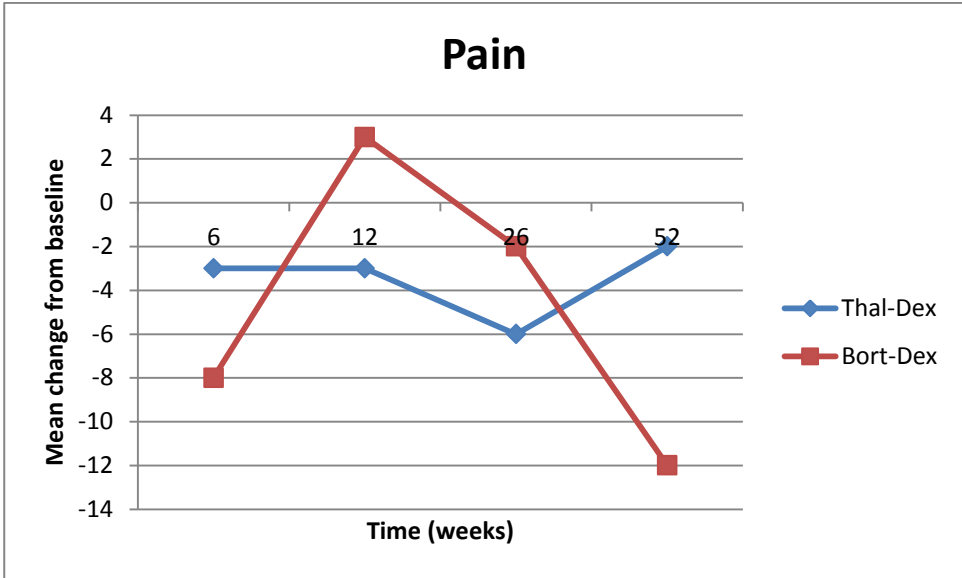
Thal-Dex; Thalidomide-Dexamethasone, Bort-Dex; Bortezomib-Dexamethasone

Figure R4d. Mean change of score from baseline – Pain

Study R5:
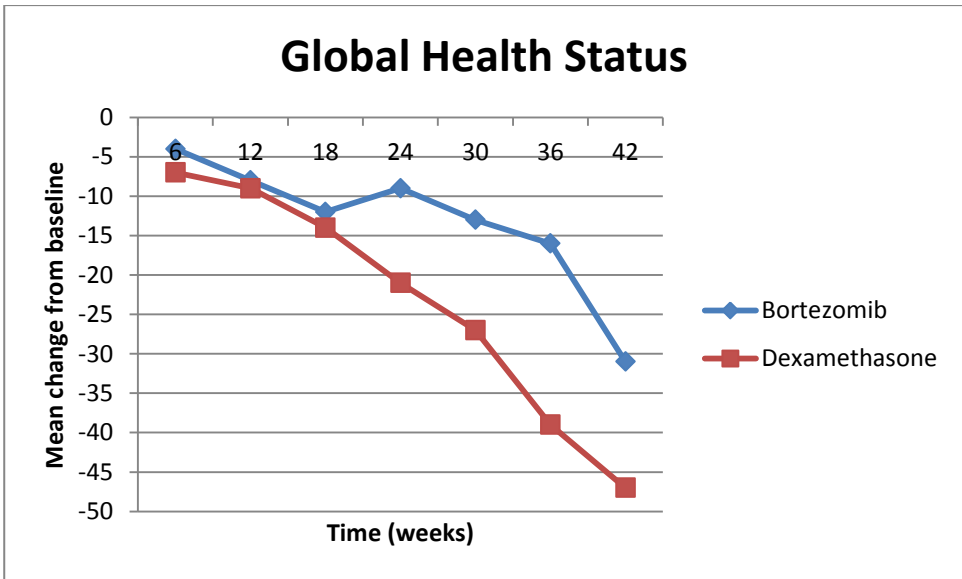
Figure R5b. Mean change of score from baseline – Global Health Status

The last evaluation of the mean change from baseline for global health status was at week 36 for the Bortezomib group, because the compliance hereafter was under 25% from baseline.

Study R6:

	Physical functioning	Global Quality of life	Fatigue	Pain
Mean change from baseline to end of study Bortezomib	-1	2	-1	-2

Figure R6

Study R7:
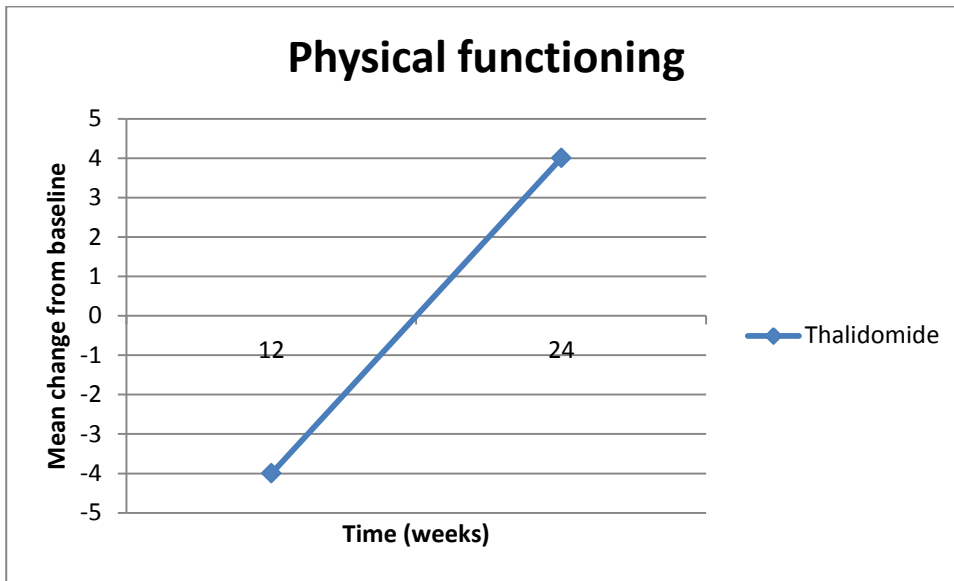
R7a. Mean change of score from baseline – physical Functioning
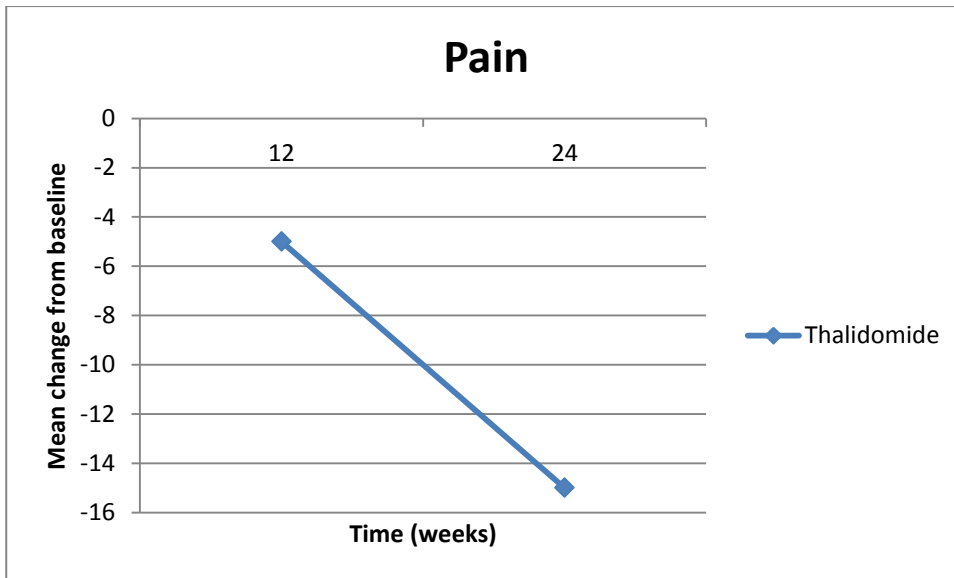
R7b. Mean change of score from baseline – Pain

POPULATION-BASED STUDY

Study B1

	Physical functioning	Quality of life	Fatigue	Pain
Mean change from baseline to one year	-2	-13	6	10

Figure B1

Appendix II

Paper II and supplementary material

Health-related quality of life in transplant ineligible newly diagnosed multiple myeloma patients treated with either thalidomide or lenalidomide-based regimen until progression (HOVON87/NMSG18 study): a prospective, open-label, multicenter, randomized, phase 3 study

AUTHORS

Lene Kongsgaard Nielsen*, MD, Quality of Life Research Center, Department of Haematology, Odense University Hospital, Kløvervænget 10, 12. Floor, 5000 Odense, Denmark

Claudia Stege*, MD, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Hematology, Cancer Center Amsterdam, Amsterdam, The Netherlands,

Birgit Lissenberg-Witte, PhD, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam, The Netherlands,

Bronno van der Holt, PhD, HOVON Data Center, Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Ulf-Henrik Mellqvist, PhD, Section of Hematology and Coagulation, Department of Medicine, Sahlgrenska University Hospital, Gotheborg, Sweden

Morten Salomo, MD, Department of Haematology, Rigshospitalet, Copenhagen, Denmark

Gerard Bos, PhD, Department of Haematology, Maastricht University Medical Center, Maastricht, The Netherlands

Mark-David Levin, PhD, Department of Internal Medicine, Albert Schweitzer Hospital, Dordrecht, The Netherlands

Heleen Visser-Wisselaar, PhD, HOVON Data Center, Erasmus MC Cancer Institute, Rotterdam , The Netherlands

Markus Hansson, MD, Department of Haematology and Wallenberg center for molecular medicine, Skåne University Hospital, Lund University, Lund, Sweden

Annette van der Velden, MD, Department of Internal Medicine, Martini Ziekenhuis, Groningen, The Netherlands

Wendy Deenik, MD, Department of Internal Medicine, Tergooi Ziekenhuis, Hilversum, The Netherlands

Juleon Coenen, Department of Internal Medicine, Isala, Zwolle, The Netherlands

Maja Hinge, PhD, Department of Internal Medicine, Division of Hematology, Vejle Hospital, Vejle, Denmark

Saskia Klein, PhD, Department of Internal Medicine, Meander Medisch Centrum, Amersfoort, The Netherlands

Bea Tanis, Department of Internal Medicine, Groene Hart Ziekenhuis, Gouda, The Netherlands

Damian Szatkowski, PhD, Department of Oncology, Haematology and Palliative Care, Førde Central Hospital, Førde, Norway

Rolf Brouwer, MD, Department of Internal Medicine, Reinier de Graaf Ziekenhuis, Delft, The Netherlands

Matthijs Westerman, MD, Department of Internal Medicine, Northwest Clinics, Alkmaar, The Netherlands

Rineke Leys, MD, Department of Internal Medicine, Maasstad Ziekenhuis, Rotterdam, The Netherlands

Harm Sinnige, PhD, Department of Internal Medicine, Jeroen Bosch Ziekenhuis, Den Bosch, The Netherlands

Einar Haukås, MD, Department of Haematology, Stavanger University Hospital, Stavanger, Norway

Klaas van der Hem, PhD, Department of Internal Medicine, Zaans Medisch Centrum, Zaandam, The Netherlands

Marc Durian, MD, Department of Internal Medicine, Tweesteden Ziekenhuis, Tilburg, The Netherlands

Peter Gimsing, Associate professor, Department of Haematology, Rigshospitalet, Copenhagen, Denmark

Niels W.C.J. van de Donk, PhD, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Haematology, Cancer Center Amsterdam, Amsterdam, The Netherlands

Pieter Sonneveld, Professor, Department of Haematology, Erasmus Medical Center Cancer Center, Rotterdam, The Netherlands

Anders Waage, Department of Haematology, St Olavs Hospital and Norwegian University of Science and Technology, Trondheim, Norway

Niels Abildgaard, Professor, Quality of Life Research Center, Department of Haematology, Odense University Hospital and the Academy of Geriatric Cancer Research (AgeCare), Odense University Hospital, Denmark

Sonja Zweegman, Professor, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Haematology, Cancer Center Amsterdam, Amsterdam, The Netherlands

[*] Contributed equally

Corresponding author: Lene Kongsgaard Nielsen, Quality of Life Research Center, Department of Haematology, Odense University Hospital, Kløvervænget 10, 12. Floor, 5000 Odense, Denmark, Email lene.kongsgaard.nielsen@rsyd.dk, phone: +45 2324 1605.

Summary

Background The overall survival of patients with multiple myeloma has improved due to effective treatment options, often including maintenance therapy. However, the impact of long term treatment on health-related quality of life (HRQoL) is largely unknown.

Methods The HOVON87/NMSG18 study was a prospective, randomized, phase 3 study of newly diagnosed transplant ineligible multiple myeloma patients, comparing melphalan-prednisolone in combination with thalidomide or lenalidomide, followed by thalidomide or lenalidomide maintenance therapy (MPT-T or MPR-R). The EORTC QLQ-C30 and QLQ-MY20 questionnaires were completed by the patients at baseline, after 3 and 9 induction cycles and after 6 and 12 months of maintenance therapy. Both linear mixed models and minimal important differences were used for within and between arms HRQoL evaluation. The study was registered at www.trialregister.nl NTR1630.

Findings 596 of 637 included patients participated in HRQoL reporting. Patients in both arms reported improved global QoL, future perspective, role and emotional functioning and summary score and less fatigue, pain, insomnia and appetite loss. MPR-R treated patients reported worsening in diarrhoea, less side effect of treatment and peripheral neuropathy than patients treated with MPT-T. MPT-T treated patients reported worsening in peripheral neuropathy, but less pain, insomnia and diarrhoea compared to the patients treated with MPR-R.

Interpretation Treatment with both MPT-T and MPR-R improved HRQoL, mainly with differences in symptomatic toxicity profile without impact on global QoL between the two treatment arms. This highlights the need for capturing symptomatic toxicities with validated patient-reported instruments when assessing benefits and harms in HRQoL during treatments.

Funding Dutch Cancer Society KWF VU-2008 4246, Celgene

Article type: Article (Clinical Trials)

KEYWORDS: Health-related quality of life, Multiple myeloma, clinical trials

RESEARCH IN CONTEXT

Evidence before study

Before analysing the data from this study, we performed a systematic review of longitudinal health-related quality of life (HRQoL) studies of patients with multiple myeloma (MM). The literature search was performed in May 2016 in PubMed, EMBASE, PsycINFO and CINAHL using the MeSH terms "Multiple Myeloma", "Myelomatosis", "Quality of life" and "Life quality". Only English language studies using the EORTC QLQ-C30 instrument for longitudinal HRQoL measurement in patients diagnosed with multiple myeloma were eligible. When interpreting the published mean scores according to thresholds for minimal important difference for within arm change, we found a general trend; Patients with newly diagnosed MM report clinically meaningful improvement in global QoL, physical functioning, fatigue and pain, irrespective of treatment regimen. Less data is available on the HRQoL impact of maintenance therapy. This is of importance as especially after reaching disease control with induction therapy, long-term continuation of maintenance could affect HRQoL.

Added value of this study

We here confirm the improvement in global QoL, fatigue and pain during first line treatment with melphalan-prednisolone in combination with thalidomide and lenalidomide, followed by thalidomide or lenalidomide maintenance therapy. The differences in HRQoL we found between the two treatment arms were mainly in symptomatic toxicities, which did not turn into a difference in global QoL score between arms. Together with findings from the systematic review, this suggests the multidimensional global QoL scale is not an informative measure for considering benefits and harms of adverse events to therapies from the patients´ perspective.

Implications of all the available evidence

Our findings support the use of validated flexible patient-reported instruments to assess symptomatic toxicities during treatment and not being accounted by global QoL or summary scores only. This will provide more insight into the benefits and harms of different kinds of therapies from clinical trials. In addition, PRO measurements will hopefully close the gap of underreporting subjective toxicities by physicians, providing specified profiles from a patient's view.

INTRODUCTION

Multiple myeloma (MM) is a malignancy of plasma cells in the bone marrow. Patients with MM are at high risk of developing bone destructions and fractures, hypercalcaemia, renal failure and anemia.[1,2] Compared to patients with other haematological malignancies, myeloma patients report a higher incidence and severity of symptoms with a reduced health-related quality of life (HRQoL) as a consequence.[3-5] There are only six studies describing the effect of first line treatment on HRQoL in transplant ineligible newly diagnosed MM (NDMM) patients.[6] In several of these trials the immunomodulatory drugs (IMiDs) thalidomide and lenalidomide were investigated.[7-11] HRQoL during treatment with thalidomide and lenalidomide were compared head to head in the FIRST and the ECOG E1A06 trials.[10,12]

The FIRST trial compared continuous therapy with lenalidomide and dexamethasone (Rd), with Rd for 18 months, and Melphalan-Prednisone-Thalidomide (MPT) for 18 months.[12] Here, HRQoL were measured by the European Organisation for Research and Treatment of Cancer (EORTC) Quality of life Questionnaire Core 30 (QLQ-C30)[13] and the Myeloma specific module QLQ-MY20.[14] Six HRQoL scales were published, which were preselected because they were perceived to be clinically relevant.[9] Both Rd and MPT resulted in a statistically significant improvement in all subscales, except side effects of treatment that worsened over time in both arms. There were no differences between arms in global QoL, physical functioning, pain and fatigue, although Rd treated patients reported significantly less side effects of treatment and less disease symptoms at 3 months as compared to MPT. A post hoc prediction model was developed suggesting that HRQoL was at least maintained or further improved beyond 18 months, unfortunately, the effect of Rd continuous versus 18 months only on HRQoL cannot be deduced from this study with certainty, as HRQoL data beyond 18 months was lacking.[15]

In the ECOG E1A06 trial, HRQoL was evaluated using the FACT-Ntx TOI score and it was shown that melphalan-prednisone-lenalidomide (MPR) followed by lenalidomide maintenance (MPR-R) resulted in a superior HRQoL after 12 months, as compared to MPT followed by thalidomide maintenance (MPT-T).[10] However, the HRQoL effects of lenalidomide and thalidomide maintenance therapies were not investigated separately. The effect of lenalidomide maintenance on HRQoL has been investigated in the MM-015 trial comparing MPR-R, MPR and melphalan-prednisolone (MP).[11,16] Within six months no clinically relevant differences in HRQoL were found between MPR-R and MPR and HRQoL data beyond six months of maintenance was lacking.[6,11]

In conclusion, data on the impact of treatment with IMiDs on HRQoL in patients with transplant ineligible NDMM are both limited and heterogeneously measured. In addition, the data on the effect of maintenance therapy on HRQoL are scarce. However, it is important to know the impact of maintenance treatment on HRQoL because this is given after achieving disease control with induction therapy and thereby in a situation where the patients do not have a definite need for further treatment.

We here report data on all the collected HRQoL scales from the HOVON87/NMSG18 study. In this study we compared MPT-T and MPR-R in 637 transplant ineligible NDMM patients, both during induction and maintenance therapy.[17] The aim of the analysis was to assess HRQoL and to investigate the impact of side effects on HRQoL, especially peripheral neuropathy, in more detail, and particularly, to evaluate the effects of long-term maintenance on HRQoL.

METHODS

Study design

Study details have been published previously[17]. Briefly, newly diagnosed symptomatic MM patients > 65 years of age or transplant ineligible patients ≤65 years, with a World Health Organization (WHO) performance status 0 to 3 (WHO 0-2 for patients above 75 years) were included. The patients were recruited from 98 community hospitals in The Netherlands, Belgium, Sweden, Norway and Denmark. Patients were randomized between nine 28-day induction cycles of MPT followed by thalidomide maintenance (MPT-T) or nine 28-day induction cycles of MPR followed by lenalidomide maintenance (MPR-R). MP induction was administered as oral melphalan 0.18 mg/kg on days 1-4 and prednisolone 2 mg/kg on days 1-4. The patients randomized to MPT were treated with oral thalidomide 200 mg/day until 4 weeks after the last induction cycle and continued with thalidomide 100 mg/day. The patients randomized to MPR received MP in combination with oral lenalidomide 10 mg on days 1-21 independent of age and after the end of induction, continued lenalidomide (10 mg day 1-21 in 28-day cycles). Maintenance treatment was given until progression, intolerable side effects or other conditions that required treatment discontinuation. The study protocol was approved by the Ethics Committee and conducted in accordance with the ethical principles of the Declaration of Helsinki and the Good Clinical Practice Guidelines. We obtained written informed consent from all participants. The study was registered at www.trialregister.nl as NTR1630.

Health-related quality of life assessments

Participation in the HRQoL reporting was optional. The questionnaires were given to the patients at baseline (T0), after induction cycle three (T1) and nine (T2), and after six (T3) and twelve (T4) months of maintenance therapy.

For HRQoL assessment, the QLQ-C30[13] and QLQ-MY20[14] were used, which are both validated instruments for HRQoL measurements in myeloma patients.[14,18] The QLQ-C30 contains five functional scales, nine symptom scales, one global quality of life (QoL) scale and a summary score based on all but two QLQ-C30 scales; global QoL and financial difficulties.[19,20] The QLQ-MY20 contains two functional and two symptom scales. For the evaluation of peripheral neuropathy question 13 of the QLQ-MY20 *"Did you have tingling hands or feet?"* was used. The EORTC manual[20] was used to calculate all HRQoL scales. Each scale was scored from 0-100; for global QoL, functional scales and the summary score a higher score means a better functioning, whereas for symptom scales, a higher score means higher degree of symptoms. A detailed

description of the QLQ-C30, the QLQ-MY20 and the peripheral neuropathy scale, data collection and assignment of the questionnaires to T0-T4 is described in Appendix A.

Statistical analyses

Change in HRQoL over time was assessed by linear mixed models and a p-value $<0\cdot005$ was considered statistically significant in order to account for multiple testing. Changes in HRQoL over time were investigated both "*within arms*" and "*between arms*". Model estimates were used for post-hoc comparisons of changes from baseline. We analysed changes in HRQoL from T0 to T4, as well as from T2 to T4 for patients who had at least 3 months of maintenance therapy. The EORTC manual was followed to handle item non-response.[20] A difference or change in mean HRQoL score was defined clinically relevant or -meaningful if it was above the threshold for minimal important difference (MID).[21] The percentage of patients improving or deteriorating by more than the MID from baseline was calculated. Details are described in Appendix B1. We explored the impact of missing questionnaires by comparing the HRQoL courses over time between patients that discontinued early versus late/never and performed sensitivity analyses by multiple imputation. See for detailed description Appendix B2. For statistical analysis SPSS version $22\cdot0$ (IBM Corp., Armonk, NY, USA) was used.

ROLE OF FUNDING SOURCE

The funder had no role in study design, data collection, data analyses, data interpretation and writing of the report. LKN, SZ, CS, BLW and BvdH had full access to all trial data.

RESULTS

Between March 12, 2009 and October 19, 2012, 637 eligible patients were included in the HOVON87/NMSG18 study. Ninety-four percent of the included patients (596 patients) gave their informed consent for participation in the HRQoL study. Only patients who completed a baseline questionnaire were included in the HRQoL analysis; 272 patients in MPT-T versus 281 patients in MPR-R. The patient and disease characteristics of the HRQoL cohort (Table 1) were comparable to the original study population. The patient flow and drop-out during study are presented in the CONSORT diagram in Figure 1. The number of patients who discontinued induction therapy was comparable: 67/272 patients in the MPT arm versus 66/281 patients in the MPR arm. However, less patients in the MPT-T arm started maintenance as compared to patients in the MPR-R arm, 146 (54%) versus 174 (62%). In addition, more patients had to discontinue thalidomide maintenance, when compared to discontinuation of lenalidomide maintenance (68% vs 30%). The main reason for discontinuation of thalidomide treatment was peripheral neuropathy.[17]

Compliance of answering questionnaires at the scheduled time points ranged from 69 to 87 percent (Figure 2). Missing items in individual questionnaires were low, with a mean of 1·2% missing QLQ C30 items and a mean of 2·2% for QLQ-MY20 items over all time points.

HRQoL analyses within treatment arms

We analysed whether there was a statistically significant change in HRQoL *within* the treatment arms during induction and maintenance phase. Patients reported improvement in HRQoL over time for the majority of scales, irrespective of received treatment (Appendix C). Body image and financial difficulties remained stable in both arms. Cognitive functioning and dyspnoea only improved in MPR-R treated patients. MPR-R treated patients reported an increase in diarrhoea (p<0·001) and MPT-T treated patients in constipation and side effects of treatment (p=0·003 and p<0·001 respectively). Peripheral neuropathy worsened in both arms (both p<0·001). The changes *within* treatment arms were clinically meaningful for global QoL, role and emotional functioning, pain, fatigue, summary score and future perspective in both treatment arms (Figure 3 and Appendix C). In addition, the MPT-T treated patients reported decreased insomnia and appetite loss, whereas the MPR-R treated patients reported improved physical functioning. The statistical significant worsening in peripheral neuropathy was only clinically meaningful to the MPT-T treated patients and not in the MPR-R treated patients (Figure 3). The MPR-R treated patients reported clinically meaningful worsening in diarrhoea (Figure 3). In general, clinically meaningful improvement occurred from T3 and T4 onwards only (i.e. after 6 to 12 months of maintenance therapy). In contrast, global QoL, future perspective and pain improved already during induction therapy and sustained during maintenance treatment (Figure 3 and Appendix Figure C1).

HRQoL analyses between treatment arms

We analysed whether there was a difference in HRQoL *between* treatment arms for the induction and maintenance phase together. The baseline mean scores of the QLQ-C30 and the QLQ-MY20 scales were comparable between treatment arms (Table 2). There was no statistically significant difference *between* treatment arms for the induction and maintenance phase together in the majority of HRQoL scales over time between the two treatment arms, including global QoL (p=0·79, Appendix C1). However, in 5 out of 21 HRQoL scales there were significant differences between the two treatment arms. Side effects of treatment (p=0·003) and peripheral neuropathy (p<0·001) were reported less in the MPR-R arm as compared to the MPT-T arm. In contrast, patients treated with MPT-T reported less pain (p=0·004), insomnia (p=0·004) and diarrhoea (p<0·001) as compared to the patients in the MPR-R arm (Figure 3).

In addition, clinical meaningful differences *between* the two arms occurred in 12 out of 21 HRQoL scales (Figure 3 and supplementary Figure C1). MPT-T treated patients reported less diarrhoea at all follow-up time points, less pain at T1, less fatigue at T2, and less insomnia and appetite loss at T1 and T4, whereas they reported more side effects of treatment at T3 and T4 and more peripheral neuropathy at all follow-up time points. In contrast, MPR-R treated patients reported better future perspective, physical and role functioning

at T4, better cognitive functioning at T1 and T4, and body image at T3 as compared to the MPT-T treated patients.

HRQoL during maintenance treatment

A total of 346 patients started the maintenance phase and received at least 3 months of maintenance therapy, of which 242 also had filled in a T2 questionnaire. We analysed whether there was a difference in HRQoL *within* or *between* the lenalidomide and thalidomide treated patients specifically in the maintenance phase (95 and 147 patients for thalidomide and lenalidomide, respectively). At the start of maintenance therapy, the MPR treated patients reported less side effects of treatment, constipation and peripheral neuropathy than the MPT treated patients. However, the MPT treated patients reported less diarrhoea than the MPR treated patients (Appendix Table D1).

HRQoL evaluation *within* maintenance treatment arms revealed statistically significant improvement mainly in the lenalidomide treated patients. In this arm, improvement was observed in global QoL (p=0·003, clinical relevant at T3), physical (p<0.001) and role functioning (p<0·001, clinical relevant at both T3 and T4), fatigue (p<0·001), dyspnoea (p=0·004) and the summary score (p<0·001, clinically relevant at T3). In both arms a statistically significant reduction in appetite loss was reported (thalidomide p=0·003, lenalidomide p<0·001). Thalidomide treated patients reported statistically significant worsening of peripheral neuropathy symptoms (p<0·001, clinically relevant at both T3 and T4) (Appendix Table D2 and Figure D1).

During maintenance, HRQoL change over time was statistically significantly different *between* arms in only one of the 21 scales: peripheral neuropathy (p<0·001) increased more in thalidomide treated patients (Appendix Table D2 and Figure D1). In addition, we investigated whether clinically meaningful differences occurred. Lenalidomide treated patients reported better physical functioning at T4 and role functioning at both T3 and T4 as compared to thalidomide treated patients. Thalidomide treated patients reported less appetite loss at T4 and more peripheral neuropathy at both T3 and T4 as compared to the lenalidomide treated patients.

Figure 4 describes the percentage of patients who achieved clinically relevant (>MID) improvement or deterioration in HRQoL from baseline, both after induction (T2) and after 1 year (T4) of maintenance. The percentage of patients, who reported change in global QoL after 1 year maintenance as compared to baseline did not statistically significantly differ between the arms: clinical relevant improvement in 54% versus 61% and deterioration in 32% versus 19% in MPT-T and MPR-R respectively (p=0·26). Significantly more MPT-T treated patients reported clinical relevant worsening in peripheral neuropathy at T2 (55 vs 27%, p<0·001) and T4 (63 vs 31%, p=0·003) as compared patients receiving MPR-R. Significantly less MPT-T treated patients reported clinical relevant worsening of diarrhoea at T2 (9 vs 31%, p<0·001).

Impact of missing questionnaires

In the investigation of the impact of missing questionnaires, we only found a statistically significant difference between patients who discontinued therapy early versus late/never for insomnia (being more

pronounced in patients who discontinued MPT-T early versus late) and emotional functioning (being inferior in patients discontinuing MPR-R early versus late) (Appendix Table E1 and Figure E1). Results from the sensitivity analysis using multiple imputation showed no statistically significant differences when comparing with results from the linear mixed models method.

DISCUSSION

In this large multicenter randomized phase III study, no difference in progression free survival between transplant ineligible newly diagnosed MM patients treated with MPT-T versus MPR-R was observed.[17] Therefore, evaluation of HRQoL is important in the choice between the two investigated regimens.

All patients reported statistically significant and clinically meaningful improved global QoL, future perspective, role and emotional functioning and the summary score, and less fatigue, pain, insomnia and appetite loss. Two differences in change over time *within* group were seen; MPR-R treated patients reported increased diarrhoea during induction therapy, whereas the MPT-R treated patients reported increased peripheral neuropathy during induction therapy and thalidomide maintenance. The only differences in HRQoL between arms was that the MPT-T treated patients reported statistically and clinically relevant less pain, insomnia and diarrhoea compared to the MPR-R treated patients, whereas the MPR-R treated patients reported less side effects of treatment and peripheral neuropathy compared to the MPT-R treated patients. Our findings are in line with previously published data on HRQoL for patients during primary treatments, who in general report clinical meaningful improvement in global QoL, physical functioning, fatigue and pain.[6] Only exception we found from this general finding was that the patients in both treatment arms reported unchanged physical functioning. A possible reason for the general finding is, that HRQoL is a multidimensional construct of the patients´ subjective perception of positive and negative aspects of health as well as non-medical factors. [22] Although HRQoL can be affected by toxicities, non-medical factors might play a substantial role in the way patients report their HRQoL over time, which has to be considered in the interpretation of longitudinal HRQoL results. There is a probability that a patient's standards and values are changing over time, which is a well-known phenomenon called response shift. [23] Patients with MM might adapt to their worsening function and increased symptoms and thereby not allow these aspects to affect their global QoL. [24]

We did find well-known differences in patient-reported toxicities between the two treatment arms of diarrhoea during lenalidomide treatment and insomnia, peripheral neuropathy and reduced diarrhoea during thalidomide treatment.[7-9,25] We found a limitation in the results for insomnia, since the investigation of the impact of missing questionnaires revealed that the score for insomnia was impacted by the patients with early study discontinuation. Most pronounced difference in patient-reported toxicity was peripheral neuropathy, which was reported by 55% of the MPT treated patients after end of induction and 63% of the thalidomide treated patients after 12 months of maintenance. Moreover, 27% of the patients treated with

MPR also reported clinical meaningful peripheral neuropathy at the end of induction therapy and 31% of the lenalidomide treated patients after 12 months of maintenance. This difference in patient-reported peripheral neuropathy might have influenced the result for between group differences for the multi-item domain of side effects of treatment. The patients treated with MPR-R reported less side effect of treatment during induction treatment compared to the MPT-T treated patients, which is similar to the findings in the FIRST study, where the patients treated with Rd reported less side effects of treatment compared to the patients treated with MPT. An important limitation in our results of peripheral neuropathy is that it is not a validated scale, but calculated on basis on the patients´ answer to one question: *"Did you have tingling hands or feet?"*. The psychometric validity of the scale is therefore unknown and might not cover all parts of patient experienced peripheral neuropathy. Peripheral neuropathy might be underestimated when using this scale, since only 77 % of cancer patients with chemotherapy-induced peripheral neuropathy report tingling hands or feet.[26] The use of a validated questionnaire covering peripheral neuropathy would have strengthened our results.

For the maintenance phase separately, the lenalidomide treated patients reported clinical meaningful improvement in global QoL, role functioning and summary score. Also we found that that improvement in several HRQoL subscales reached clinical relevance after 6 to 12 months of maintenance therapy only, with the exception of global QoL, future perspective and pain, which improved already during induction phase and sustained hereafter. Interpretation of the results during maintenance phase might be compromised by low sample size due to study discontinuation, and a selection of patients with a better outcomes might have occurred. Handling of missing data is one of the identified and unsolved challenges cancer trials.[27,28] We performed two analyses to investigate the impact of missing questionnaires. For insomnia and emotional functioning only, we found a difference between patients who discontinued therapy early versus late/never.

In conclusion, we found that differences in HRQoL between the two treatment regimens were mainly found in symptomatic toxicities scales with reporting of worsening in diarrhoea during MPR-R treatment and worsening in peripheral neuropathy and side effects during MPT-T treatment. We also found less insomnia during MPT-T treatment. It is noteworthy that peripheral neuropathy was investigated by a single-item domain from EORTC QLQ-MY20 and the findings could thereby be inaccurate and underestimated. None of the observed differences in symptomatic toxicities had consequences for the reported global QoL score, which was similar between the two treatment groups. This highlights the need for capturing symptomatic toxicities with PRO instruments when assessing benefits and harms in HRQoL during treatments. This has also recently been advocated in a The Lancet Haematology Commission by Thanarajasingam and colleagues.[22] The patient-reported version of the Common Terminologies Criteria for Adverse Events (PRO-CTCAE) for self-reported toxicities and the EORTC item bank together with the computer adaptive tests version of the QLQ-C30 questionnaire might make this possible in future studies.[29,30]

REFERENCE

1.   Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet Oncology* 2014; **15**(12): e538-48.

2.   Kyle RA, Rajkumar SV. Multiple myeloma. *Blood* 2008; **111**(6): 2962-72.

3.   Jordan K, Proskorovsky I, Lewis P, et al. Effect of general symptom level, specific adverse events, treatment patterns, and patient characteristics on health-related quality of life in patients with multiple myeloma: results of a European, multicenter cohort study. *Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer* 2014; **22**(2): 417-26.

4.   Johnsen AT, Tholstrup D, Petersen MA, Pedersen L, Groenvold M. Health related quality of life in a nationally representative sample of haematological patients. *European journal of haematology* 2009; **83**(2): 139-48.

5.   Priscilla D, Hamidin A, Azhar MZ, Noorjan KON, Salmiah MS, Bahariah K. Quality of life among patients with hematological cancer in a Malaysian hospital. *Medical Journal of Malaysia* 2011; **66**(2): 117.

6.   Nielsen LK, Jarden M, Andersen CL, Frederiksen H, Abildgaard N. A systematic review of health-related quality of life in longitudinal studies of myeloma patients. *European journal of haematology* 2017; **99**(1): 3-17.

7.   Waage A, Gimsing P, Fayers P, et al. Melphalan and prednisone plus thalidomide or placebo in elderly patients with multiple myeloma. *Blood* 2010; **116**(9): 1405-12.

8.   Verelst SG, Termorshuizen F, Uyl-de Groot CA, et al. Effect of thalidomide with melphalan and prednisone on health-related quality of life (HRQoL) in elderly patients with newly diagnosed multiple myeloma: a prospective analysis in a randomized trial. *Annals of hematology* 2011; **90**(12): 1427-39.

9.   Delforge M, Minuk L, Eisenmann JC, et al. Health-related quality-of-life in patients with newly diagnosed multiple myeloma in the FIRST trial: lenalidomide plus low-dose dexamethasone versus melphalan, prednisone, thalidomide. *Haematologica* 2015; **100**(6): 826-33.

10.   Stewart AK. Melphalan , prednisone , and thalidomide vs melphalan , prednisone , and lenalidomide (ECOG E1A06) in untreated multiple myeloma. *Blood* 2015; **126**(11): 1294.

11.   Dimopoulos MA, Delforge M, Hajek R, et al. Lenalidomide, melphalan, and prednisone, followed by lenalidomide maintenance, improves health-related quality of life in newly diagnosed multiple myeloma patients aged 65 years or older: results of a randomized phase III trial. *Haematologica* 2013; **98**(5): 784-8.

12.   Benboubker L, Dimopoulos MA, Dispenzieri A, et al. Lenalidomide and dexamethasone in transplant-ineligible patients with myeloma. *The New England journal of medicine* 2014; **371**(10): 906-17.

13.   Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute* 1993; **85**(5): 365-76.

14.   Cocks K, Cohen D, Wisløff F, et al. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. *European Journal of Cancer* 2007; **43**(11): 1670-8 9p.

15.   Vogl DT, Delforge M, Song K, et al. Long-term health-related quality of life in transplant-ineligible patients with newly diagnosed multiple myeloma receiving lenalidomide and dexamethasone. *Leuk Lymphoma* 2018; **59**(2): 398-405.

16.   Dimopoulos MA, Palumbo A, Hajek R, et al. Factors that influence health-related quality of life in newly diagnosed patients with multiple myeloma aged > 65 years treated with melphalan, prednisone and lenalidomide followed by lenalidomide maintenance: Results of a randomized trial. *Leukemia and Lymphoma* 2014; **55**(7): 1489-97.

17.   Zweegman S, van der Holt B, Mellqvist UH, et al. Melphalan, prednisone, and lenalidomide versus melphalan, prednisone, and thalidomide in untreated multiple myeloma. *Blood* 2016; **127**(9): 1109-16.

18.   Wisloff F, Eika S, Hippe E, et al. Measurement of health-related quality of life in multiple myeloma. Nordic Myeloma Study Group. *British journal of haematology* 1996; **92**(3): 604-13.

19.   Giesinger JM, Kieffer JM, Fayers PM, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *Journal of clinical epidemiology* 2016; **69**: 79-88.

20.   Fayers P, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). European Organisation for Research and Treatment of Canceer, Brussels 2001.

21.   King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1996; **5**(6): 555-67.

22.   Thanarajasingam G, Minasian LM, Baron F, et al. Beyond maximum grade: modernising the assessment and reporting of adverse events in haematological malignancies. *The Lancet Haematology* 2018.

23.   Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Social science & medicine (1982)* 1999; **48**(11): 1507-15.

24.   Kvam AK, Wisloff F, Fayers PM. Minimal important differences and response shift in health-related quality of life; a longitudinal study in patients with multiple myeloma. *Health and quality of life outcomes* 2010; **8**: 79-7525-8-79.

25.   Tacchetti P, Terragna C, Galli M, et al. Bortezomib- and thalidomide-induced peripheral neuropathy in multiple myeloma: clinical and molecular analyses of a phase 3 study. *Am J Hematol* 2014; **89**(12): 1085-91.

26.   Wolf SL, Barton DL, Qin R, et al. The relationship between numbness, tingling, and shooting/burning pain in patients with chemotherapy-induced peripheral neuropathy (CIPN) as measured by the EORTC QLQ-CIPN20 instrument, N06CA. *Support Care Cancer* 2012; **20**(3): 625-32.

27.   Pe M, Dorme L, Coens C, et al. Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *The Lancet Oncology* 2018; **19**(9): e459-e69.

28.   Bottomley A, Pe M, Sloan J, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *The Lancet Oncology* 2016; **17**(11): e510-e4.

29.   Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst* 2014; **106**(9).

30.   Petersen MA, Aaronson NK, Arraras JI, et al. The EORTC CAT Core-The computer adaptive version of the EORTC QLQ-C30 questionnaire. *Eur J Cancer* 2018; **100**: 8-16.

Table 1: Demographic characteristics of the patients included in the health-related quality of life analysis

Demographic characteristics	MPT-T (N=272)	MPR-R (N=281)
Median age, years (IQR)	72 (69-77)	73 (69-77)
Age ≥ 76 years, N (%)	90 (33%)	98 (35%)
Sex, N (%)		
Male	133 (49%)	164 (58%)
Female	139 (51%)	117 (42%)
WHO performance, N (%)		
0	89 (33%)	107 (38%)
1	132 (49%)	124 (44%)
2	39 (14%)	40 (14%)
3	5 (2%)	5 (2%)
Unknown	6 (2%)	5 (2%)
M-protein subtype, N (%)		
IgG	177 (65%)	176 (63%)
IgA	73 (27%)	69 (25%)
IgD	4 (2%)	1 (<0.5%)
Light chain only	17 (6%)	34 (12%)
Unknown	0 (0%)	1 (<0.5%)
ISS, N (%)		
I	61 (23%)	78 (28%)
II	134 (49%)	136 (48%)
III	74 (27%)	65 (23%)
Unknown/missing	3 (1%)	2 (1%)
Lytic bone lesions, N (%)		
None	86 (32%)	89 (32%)
1	25 (9%)	19 (7%)
2	15 (6%)	19 (7%)
3 or more	141 (52%)	150 (53%)
Unknown/missing	5 (2%)	4 (1%)
FISH performed, N (%)		
Yes	206 (76%)	220 (78%)
FISH abnormality if performed, N (%)		
17p13 loss	23/188 (12%)	16/196 (8%)
t(4;14)(p16;q32)	18/199 (9%)	17/216 (8%)
t(14;16)(q32;q23)	2/170 (1%)	10/192 (5%)
1q21 gain	56/146 (38%)	58/165 (35%)

MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy; N, number of patients; IQR; interquartile range, WHO, World Health Organization; ISS, International Staging System

Table 2. Mean baseline scores for all EORTC QLQ-C30 and EORTC QLQ-MY20 scales for each treatment arm.

Health-related quality of life scales	MPT-T Baseline mean scores (SD) (N=272)	MPR-R Baseline mean scores (SD) (N=281)
EORTC QLQ-C30		
Global health status/QoL	52.9 (24.1)	55.5 (25.3)
Physical functioning	57.0 (27.7)	58.6 (28.2)
Role functioning	45.8 (35.4)	49.2 (37.1)
Emotional functioning	70.8 (21.7)	71.2 (21.7)
Cognitive functioning	80.2 (23.1)	80.4 (23.2)
Social functioning	68.3 (30.4)	68.2 (31.4)
Fatigue	48.2 (29.1)	43.8 (28.5)
Nausea and vomiting	10.5 (19.2)	8.7 (17.7)
Pain	51.1 (36.3)	47.2 (34.7)
Dyspnoea	32.5 (32.1)	25.7 (28.6)
Insomnia	30.4 (31.6)	28.1 (31.4)
Appetite loss	26.0 (33.6)	23.0 (32.1)
Constipation	22.4 (30.4)	22.1 (30.8)
Diarrhoea	10.6 (22.6)	8.8 (21.2)
Financial difficulties	5.2 (15.4)	4.5 (14.5)
Summary score	64.7 (16.0)	66.8 (16.2)
EORTC QLQ-MY20		
Disease Symptoms	31.2 (22.6)	31.1 (22.4)
Side Effects of Treatment	19.4 (13.9)	18.8 (13.6)
Future Perspective	52.3 (25.2)	52.3 (25.5)
Body Image	80.8 (29.3)	81.4 (28.1)
Peripheral neuropathy	10.9 (20.7)	11.7 (21.4)

EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30; EORTC QLQ-MY20, European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20, SD, standard deviation, MPT-T; melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy, N; number

Figure 1. Consort diagram of the numbers of patients participating in the HRQoL study and the numbers of answered questionnaire. The presented numbers of patients off protocol and specified reasons are related to the numbers of patients going out of the protocol.

N; number of patients, MPT; melphalan-prednisone-thalidomide, MPR; melphalan-prednisone-lenalidomide, PD, progressive disease, PNP; peripheral neuropathy

Figure 2. Number of patients on protocol and completed questionnaires at each time point.

MPT-T   MPR-R

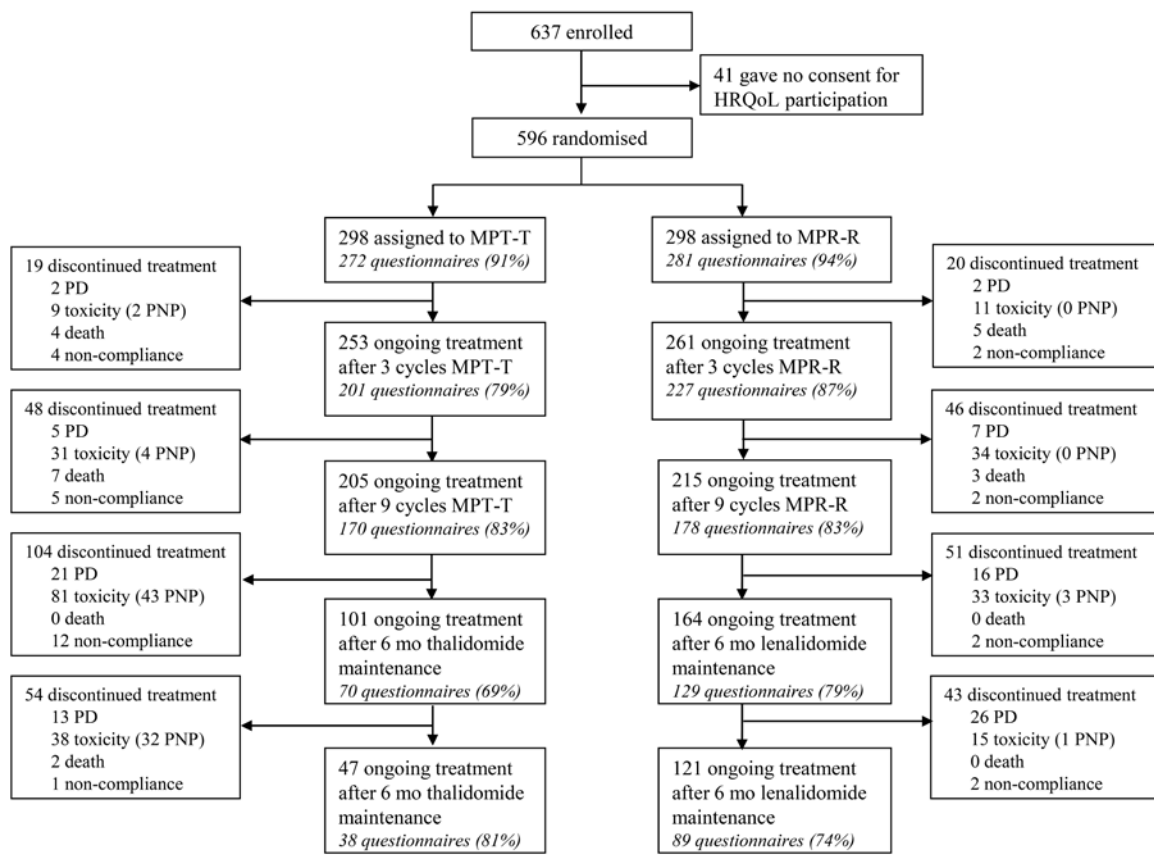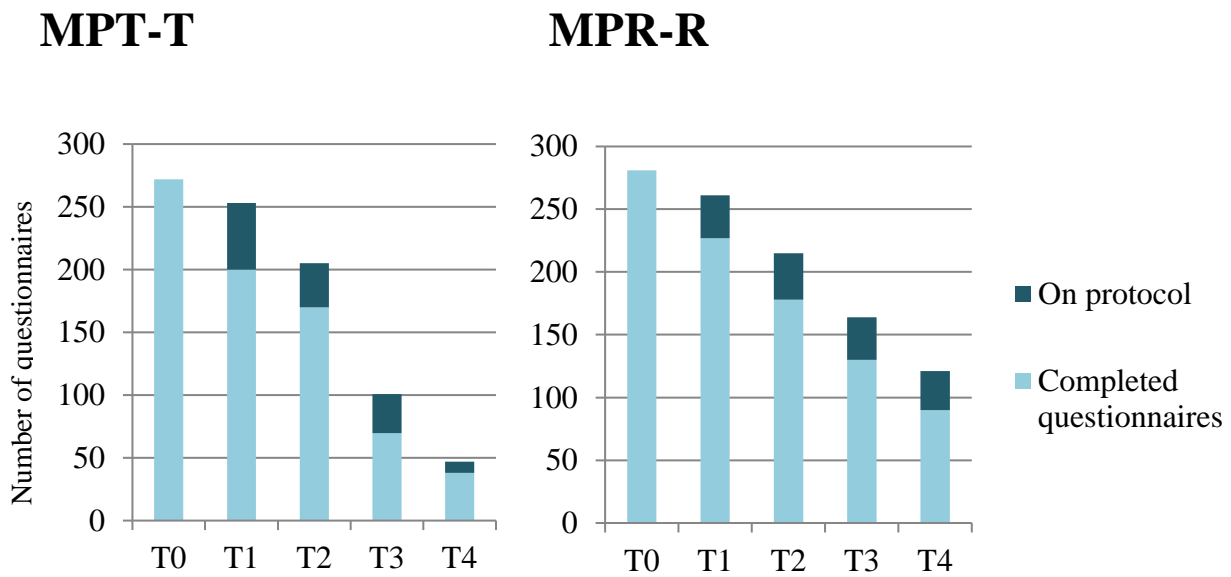MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy; T0: baseline; T1: after 3 induction cycles; T2: after 9 induction cycles; T3: after 6 months maintenance treatment; T4 after 12 months maintenance treatment
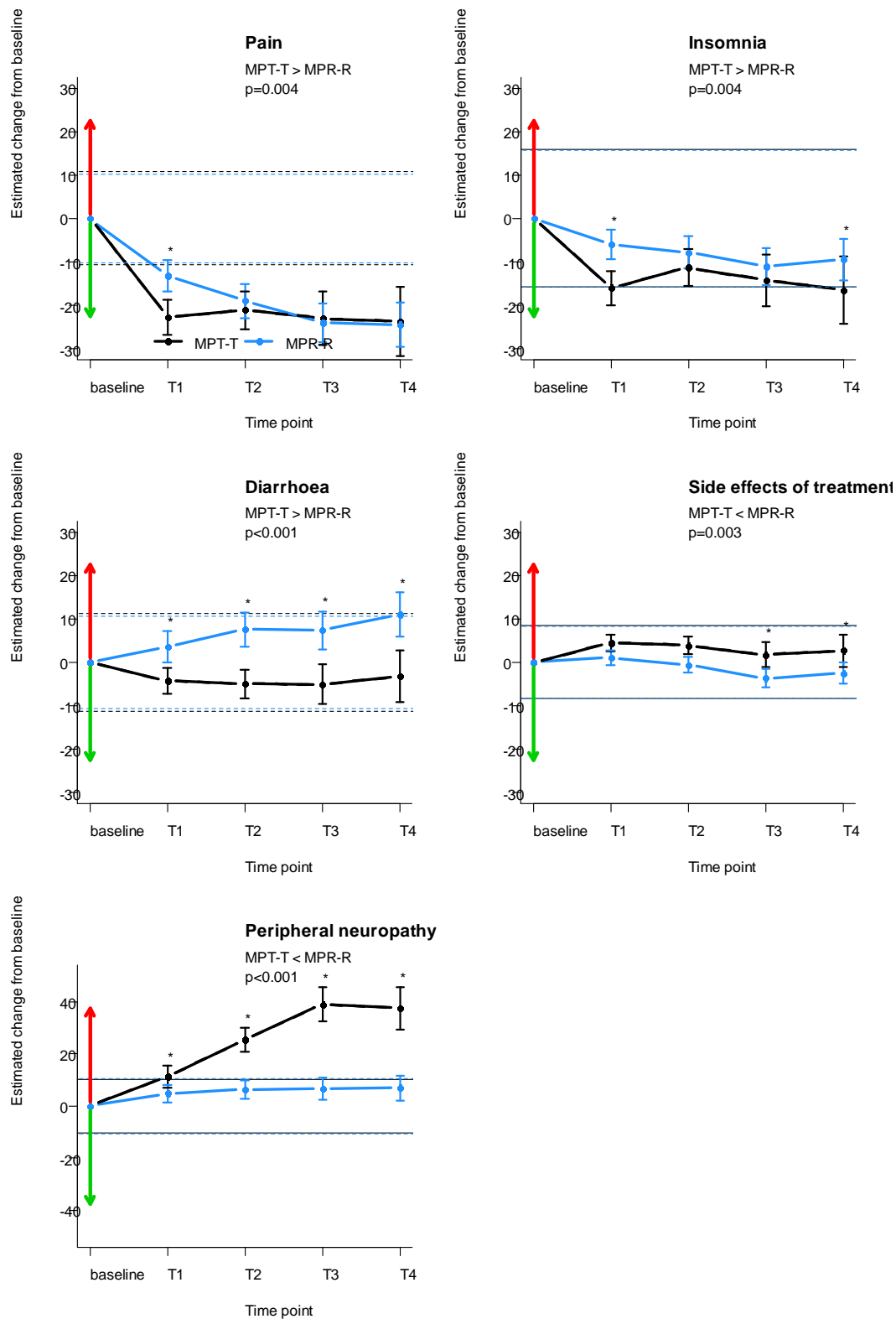
Figure 3. Estimated change in HRQoL score from baseline with corresponding 95% confidence intervals and p-value for the five scales with statistically significant difference in change over time between treatment arms. Time points with clinical relevant difference between arms are marked with *. The dashed horizontal line represents the calculated threshold for minimal important difference, the black for MPT-T and the blue for the MPR-R treatment. The green arrows indicate the direction of improvement in functional scales or reduction of symptom scales. The red arrows indicate the direction of deterioration in functional scales or increase of symptom scales.

MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy
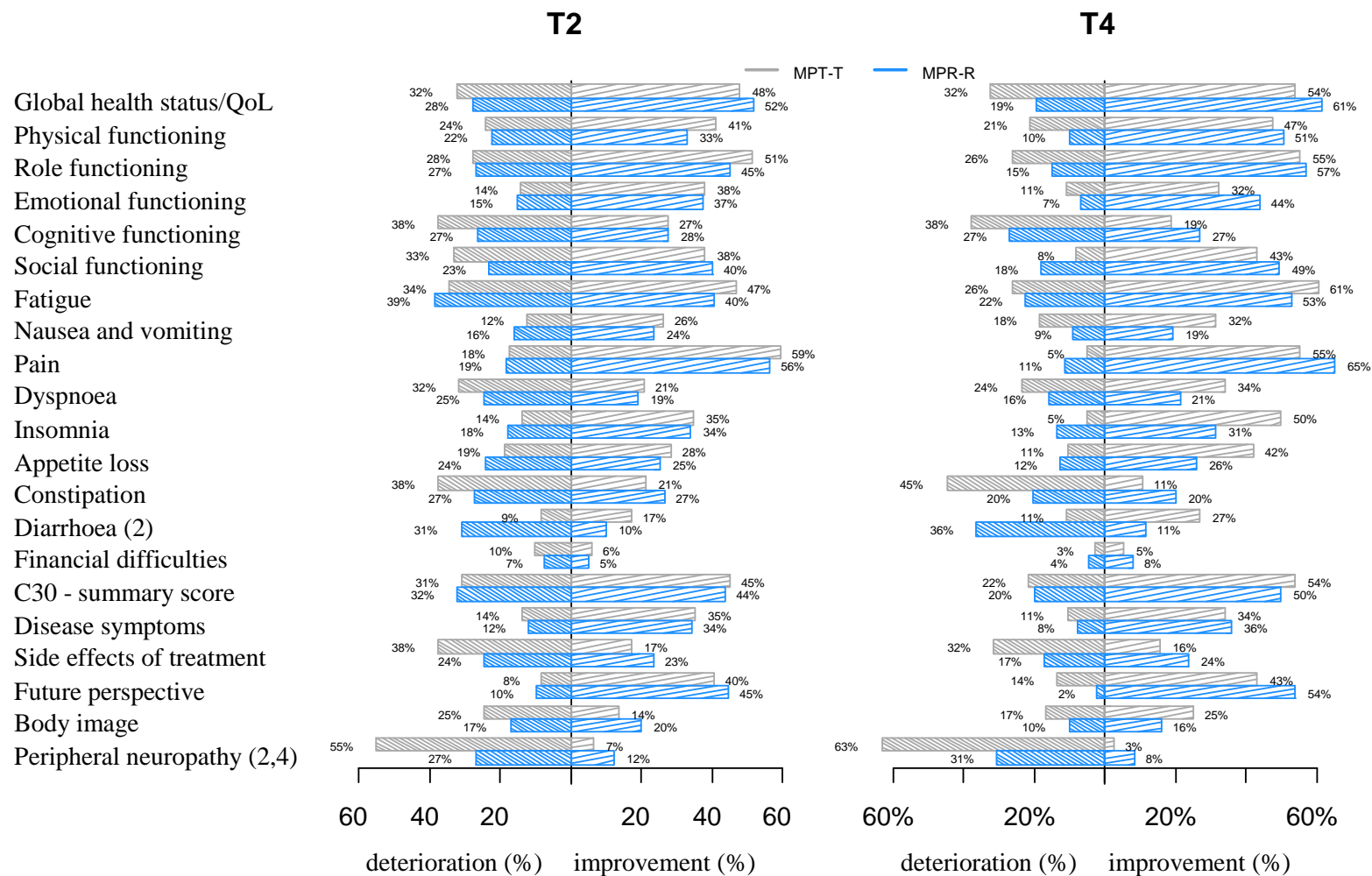
Figure 4. Responders. The percentage of the patients reaching a clinical relevant change in HRQoL, eg. the minimal important difference (MID) threshold for within group change during the induction phase (T2) and induction phase and maintenance phase together (T4). We found a significant difference between the arms with respect to the number of patients improving or deteriorating by more than MID for diarrhoea and peripheral neuropathy at T2 and for peripheral neuropathy at T4.

Index of the Supplementary Appendix

A. HRQoL data collection and categorization of questionnaires

Health related quality of life (HRQoL) was assessed by using two questionnaires: the European Organisation for Research and Treatment of Cancer (EORTC) Quality of life EORTC QLQ-C30 and the Myeloma specific EORTC QLQ-MY20 module. The EORTC QLQ-C30 consists of 15 scales: one global quality of life (QoL) scale, five functional scales (physical, role, emotional, cognitive and social functioning), nine symptoms scales (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties) and the summary score (containing all scales minus global QoL and financial difficulties). The EORTC QLQ-MY20 consists of four scales: two symptom scales (disease symptoms and side effects of treatment) and two functional scales (future perspective and body image). Neither of these questionnaires has a separate scale for peripheral neuropathy. Since results from the adverse event registration form showed a high number of patients developing peripheral neuropathy, a symptom scale "peripheral neuropathy" was assessed by item 13 of the QLQ-MY20 *"Did you have tingling hands or feet?"*. The peripheral neuropathy scale was an additional single-item scale, transforming it to a 0 to 100 score, according to the EORTC manual. In this HRQoL study all scales were evaluated.

Patients received a paper version of the questionnaires. For patients recruited by an investigator from the HOVON, the local QoL coordinators sent the answered questionnaires to the HOVON Data Center, Amsterdam, and for patients recruited by an investigator from NMSG the local QoL coordinators e-mailed it to the QoL Center, Ullevål Hospital Oslo. If a scheduled questionnaire was not received by HOVON Data Center or QoL Center, an e-mail was sent to the local QoL coordinator as a reminder in order to obtain the questionnaires from the patient.

Ideally the patient completed the questionnaire at the exact evaluation times according to the HRQoL study protocol: at baseline (T0), after induction cycles 3 (T1) and 9 (T2), and after 6 (T3) and 12 (T4) months of maintenance therapy. However, the questionnaires were not always answered exactly at these times and were therefore assigned to one of the five time points according to the following criteria:
- T0: questionnaire was completed between randomization and 28 days after start of the first induction cycle;
- T1: questionnaire was completed between start of the second induction cycle and 2x28 days after start of the third induction cycle;
- T2: questionnaire was completed between start of the sixth cycle and 30 days after start maintenance therapy;
- T3: questionnaire was completed between 3 and 9 months of maintenance therapy; and
- T4: questionnaire was completed between 9 and 15 months of maintenance therapy.

Questionnaires not answered in the time frames were excluded from the analysis. The timing of questionnaires is illustrated in Figure A1. All questionnaires had to be completed in at most 1 month after going off protocol. HRQoL assessment was terminated, from the time a patient went off protocol.
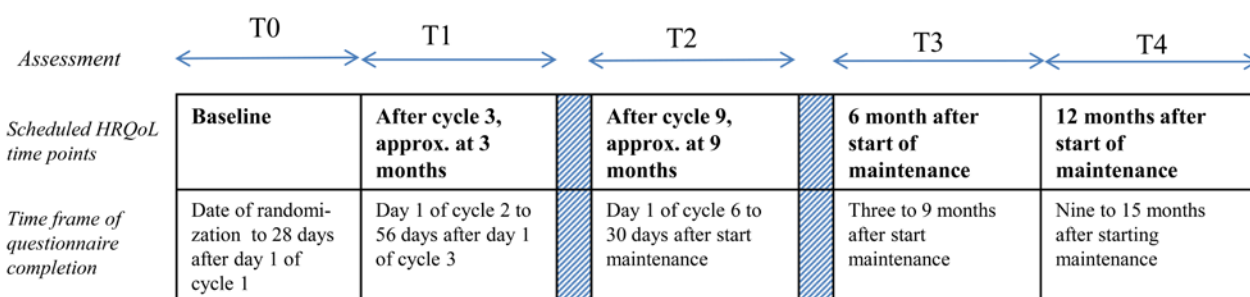

Figure A1. Time of the collected HRQoL questionnaires
The collected HRQoL questionnaires were assigned to T0, T1, T2, T3 and T4.
HRQoL;health-related quality of life,

B. Statistical analysis

Change in HRQoL over time was assessed by linear mixed models. For differences *within* the whole study population or each arm separately, the linear mixed model only included a fixed effect for time and random intercept for subject. For differences *between* arms, the model included fixed effects for time, treatment arm and their two-way interaction and a random intercept for subject. Model estimates were used for post-hoc comparisons of change from baseline, within and between arms.

B1. Minimal important difference

Clinical relevance of the differences and clinically meaningful changes in HRQoL over time were assessed by minimal important difference (MID). The MID threshold was defined as > 5 points for clinically relevant differences *between* treatment arms, which is a general accepted threshold.[1] Clinically meaningful change from baseline *within* arms was based on the standard error of measurements for multi-item scales[2] and Cohen´s criteria for medium effect size for single-item scales.[3,4] More specifically, for multi-item scales the MID equals the Cronbach's α times the standard deviation (SD) at baseline/start of maintenance and for single-item scales the MID equals 0·5×SD at baseline/start of maintenance.

In addition, for each arm separately, we calculated the percentage of patients who improved and deteriorated by more than MID in HRQoL from their baseline and T2.[5] Percentages of improvement or deterioration were compared between arms with the chi-square test.

B2. Impact of missing questionnaires

No systematic HRQoL data collection was done from the time a patient discontinued treatment. Since patients discontinuing treatment might do so because of excessive toxicity, they consequently might have a worse HRQoL than patients continuing treatment.[6,7] Therefore, HRQoL data could be missing not at random (MNAR) and informative since the mechanism might be due to the missing HRQoL result.[8,9]

As recommended by Bell and colleagues[6], we explored the impact of missing data by comparing the HRQoL course over time between patients that discontinued treatment "early" (e.g. during induction therapy) and patients discontinuing treatment "late" (e.g. after starting maintenance therapy) or never. A linear mixed model was used, including fixed effect for time, timing off protocol (early vs late/never) and their two-way interaction and a random intercept for subject. A significant interaction (p-value <0·005) was considered an indication against missing (completely) at random (M(C)AR).

For the multiple imputation analysis, each missing data point multiple (in our study m=5) possible values are imputed: several versions of the dataset are being created. Imputation was based on the following 8 variables: gender, treatment arm, WHO performance, disease status, timing and reason for going off protocol, starting maintenance yes/no, and evaluation time point.

C. Induction and maintenance phase together

Table C1. P-values for the comparison of HRQoL course over time between arms, within the whole study population and each arm separately.
The **bold** p-values represent a statistically significant change over time where the change from baseline was above the threshold for minimal important difference at least at one time point.

Health-related quality of life scales	Change over time *between* arms	Change over time *within* whole study population	Change over time *within* arm MPT-T	Change over time *within* arm MPR-R
EORTC QLQ-C30				
Global health status/QoL	0·79	**<0·001**	**<0·001**	**<0·001**
Physical functioning	0·29	**<0·001**	<0·001	<0·001
Role functioning	0·41	**<0·001**	**<0·001**	**<0·001**
Emotional functioning	0·12	**<0·001**	**<0·001**	**<0·001**
Cognitive functioning	0·013	0·093	0·34	0·003
Social functioning	0·84	<0·001	<0·001	<0·001
Fatigue	0·17	**<0·001**	**<0·001**	**<0·001**
Nausea and vomiting	0·86	<0·001	<0·001	0·001
Pain	**0·004**	**<0·001**	**<0·001**	**<0·001**
Dyspnoea	0·56	<0·001	0·18	<0·001
Insomnia	**0·004**	<0·001	**<0·001**	**<0·001**
Appetite loss	0·29	<0·001	**<0·001**	**<0·001**
Constipation	0·008	<0·001	0·003[#]	0·002
Diarrhoea	**<0·001**	0·016	0·012	<0·001[#]
Financial difficulties	0·84	0·50	0·57	0·77
Summary score	0·76	**<0·001**	**<0·001**	**<0·001**
EORTC QLQ-MY20				
Disease symptoms	0·38	<0·001	<0·001	<0·001
Side effects of treatment	**0·003**	<0·001	<0·001[#]	<0·001
Future perspective	0·32	**<0·001**	**<0·001**	**<0·001**
Body image	0·11	0·54	0·50	0·13
Peripheral neuropathy	**<0·001**	**<0·001**	**<0·001**[#]	0·001[#]

EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20, MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy.
[#] *represents a statistically significant worsening in HRQoL over time.*

Table C2. MID thresholds for clinical meaningful HRQoL change from baseline/start maintenance.

Health-related quality of life scales	MID since start induction		MID since start maintenance	
	MPT-T	**MPR-R**	**Thalidomide**	**Lenalidomide**
EORTC QLQ-C30				
Global health status/QoL	6·9	7·2	5·1	5·5
Physical functioning	9·7	9·9	9·5	9·6
Role functioning	10·7	11·2	10·0	10·3
Emotional functioning	9·6	9·6	7·4	8·6
Cognitive functioning	14·8	14·9	13·3	13·0
Social functioning	14·4	14·8	11·8	12·0
Fatigue	10·4	10·2	9·1	9·4
Nausea and vomiting	10·6	9·7	8·0	9·7
Pain	10·7	10·3	11·3	10·7
Dyspnoea	16·1	14·3	14·9	14·3
Insomnia	15·8	15·7	10·8	12·3
Appetite loss	16·8	16·1	14·1	13·4
Constipation	15·2	15·4	15·0	11·9
Diarrhoea	11·3	10·6	6·7	12·1
Financial difficulties	7·7	7·3	8·6	7·4
Summary score	4·0	4·1	3·5	3·6
EORTC QLQ-MY20				
Disease Symptoms	11·8	11·6	9·4	9·5
Side effects of treatment	8·4	8·2	9·5	7·8
Future Perspective	11·3	11·5	8·7	8·8
Body Image	14·6	14·1	14·9	13·0
Peripheral neuropathy	10·4	10·7	13·9	13·3

MID; minimal important difference, EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20, MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy.
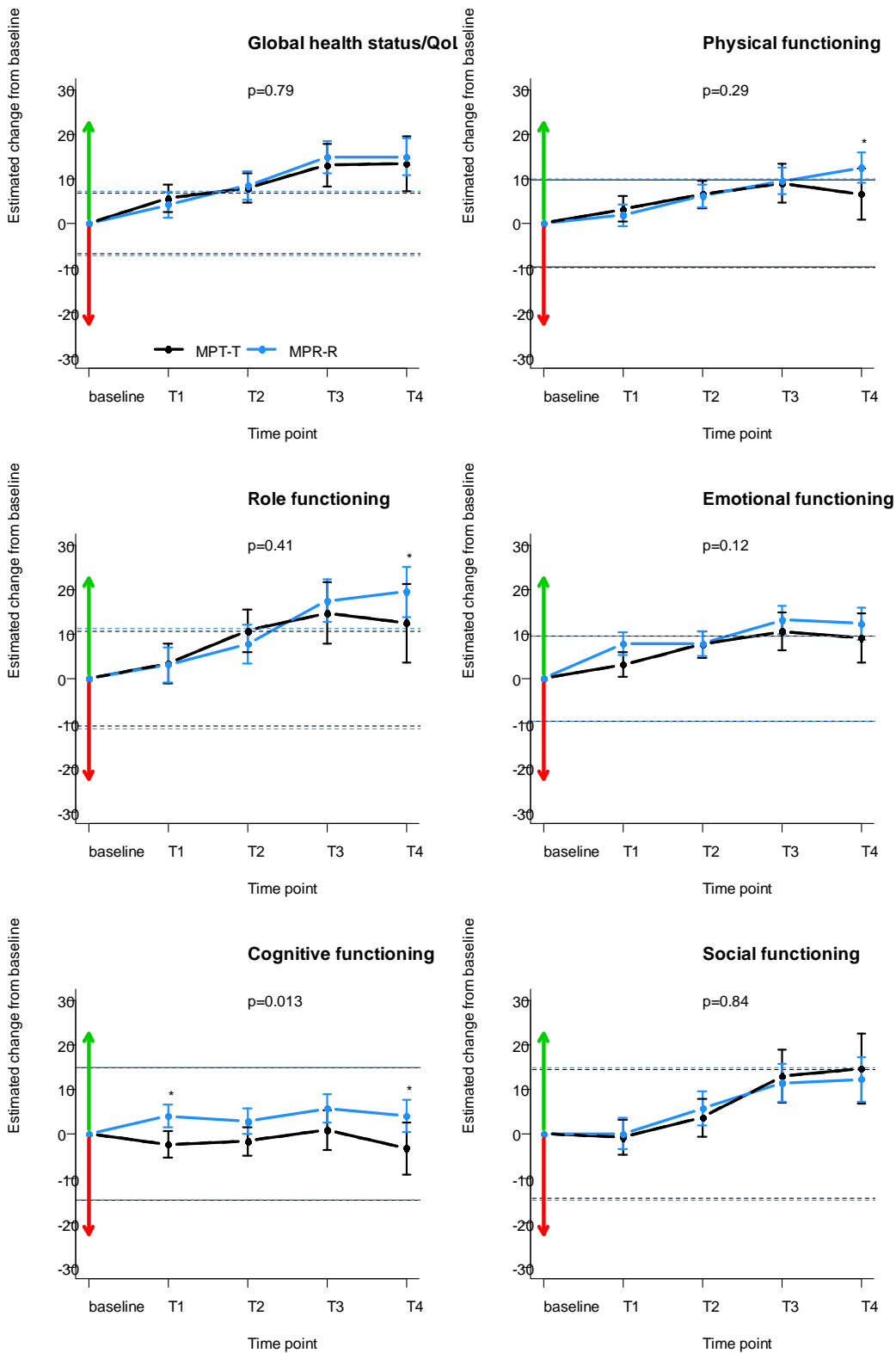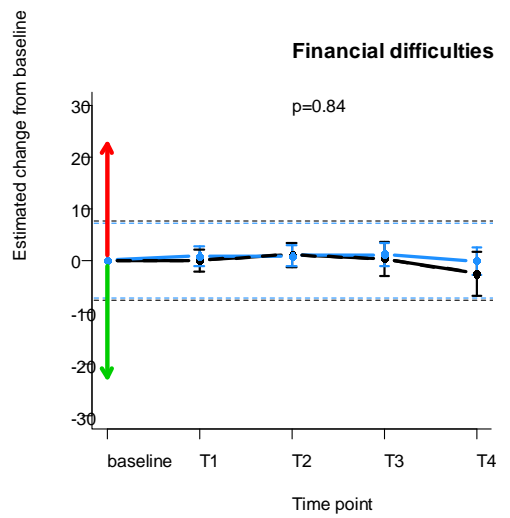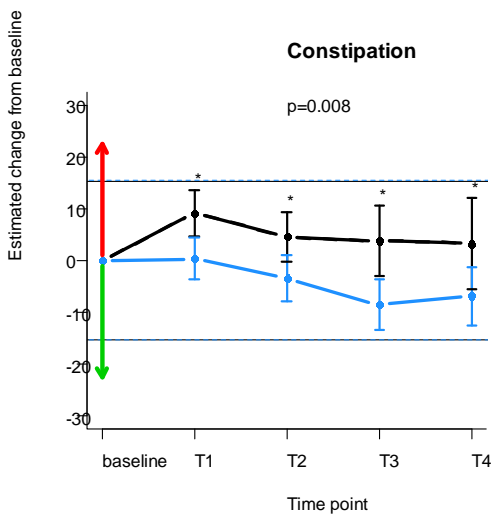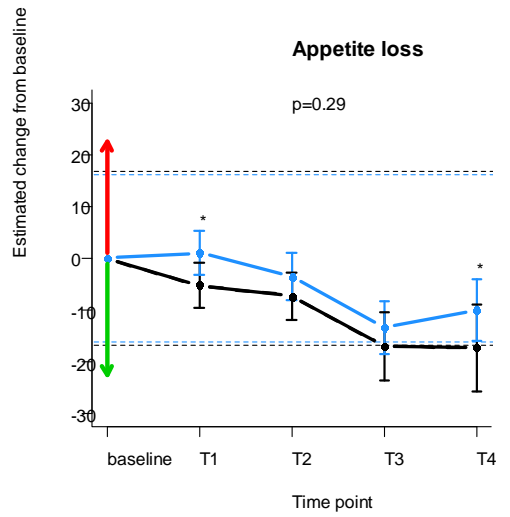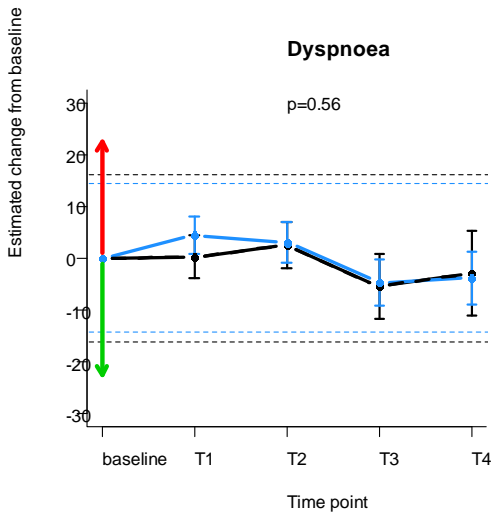
HRQoL over time within and between treatment arms
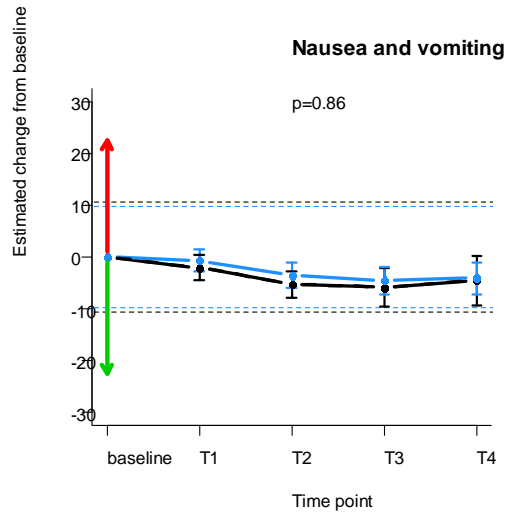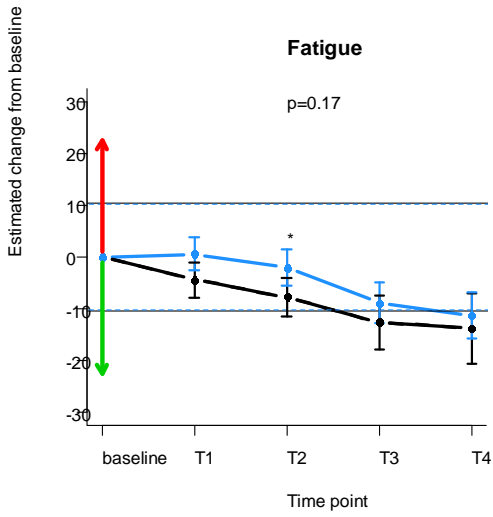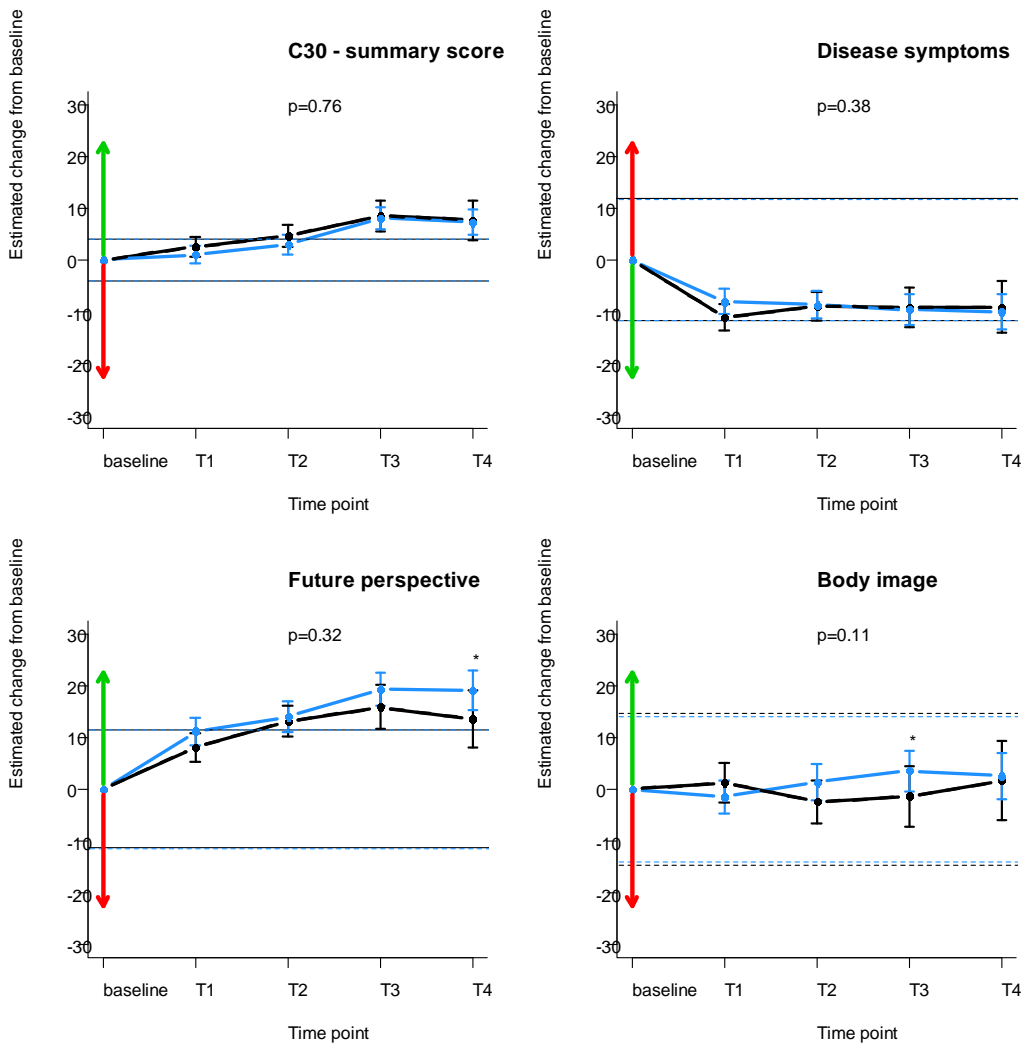
Figure C1. Estimated change in HRQoL score from baseline with corresponding 95% confidence intervals for the 16 scales with no statistically significant difference in change over time between treatment arms.
The p-values correspond to the difference in change over time *between* arms. Time points with clinical relevant difference between arms score are marked with *. The dashed horizontal line represents the calculated threshold for minimal important difference (MID), the black for MPT-T and the blue for the MPR-R treatment. The green arrows indicate the direction of improvement in functioning or reduction of symptoms. The red arrows indicate the direction of deterioration in functioning or increasing in symptoms.
MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy.

D. Maintenance phase separately

Table D1. Mean HRQoL scores at T2, e.g. before start of maintenance treatment with thalidomide or lenalidomide.
P-values correspond to the comparison of the mean values between arms at T2.

Health-related quality of life scales	Thalidomide maintenance baseline mean scores (SD) (N=95)	Lenalidomide maintenance baseline mean scores (SD) (N=147)	P-values
EORTC QLQ-C30			
Global health status/QoL	61·6 (18·1)	67·3 (19·5)	0·024
Physical functioning	65·1 (20·9)	68·9 (21·3)	0·18
Role functioning	58·7 (29·4)	61·5 (30·1)	0·48
Emotional functioning	79·4 (18·8)	81·6 (21·8)	0·41
Cognitive functioning	78·4 (20·8)	84·5 (20·2)	0·025
Social functioning	74·2 (24·8)	77·1 (25·1)	0·39
Fatigue	36·1 (24·0)	36·4 (24·9)	0·95
Nausea and vomiting	3·7 (9·2)	4·9 (11·1)	0·38
Pain	26·1 (27·5)	26·6 (25·9)	0·89
Dyspnoea	31·2 (29·9)	24·9 (28·6)	0·10
Insomnia	11·2 (21·5)	19·5 (24·6)	0·006
Appetite loss	17·5 (28·3)	15·6 (26·8)	0·60
Constipation	32·6 (30·0)	17·8 (23·9)	**<0·001**
Diarrhoea	4·6 (13·5)	15·1 (24·2)	**<0·001**
Financial difficulties	5·6 (17·3)	4·3 (14·8)	0·54
Summary score	71·0 (12·4)	72·3 (12·7)	0·42
EORTC QLQ-MY20			
Disease symptoms	20·5 (16·7)	20·9 (16·7)	0·87
Side effects of treatment	22·1 (16·0)	16·2 (13·1)	**0·002**
Future perspective	66·5 (22·5)	68·7 (22·7)	0·47
Body image	77·0 (29·7)	85·3 (26·0)	0·028
Peripheral neuropathy	30·1 (27·8)	16·2 (26·7)	**<0·001**

EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20.

Table D2. P-values for the comparison of HRQOL course over time during maintenance between arms and within each arm separately.

The **bold** p-values represent a significant change over time where the change from start maintenance was above the threshold for minimal important difference at least at one time point.

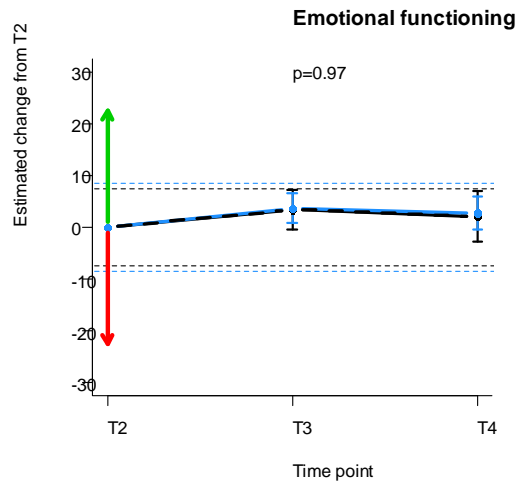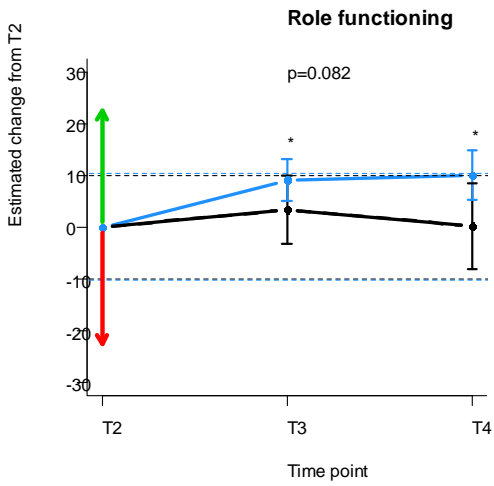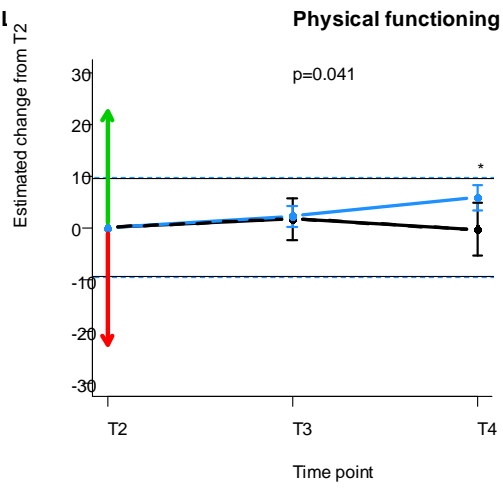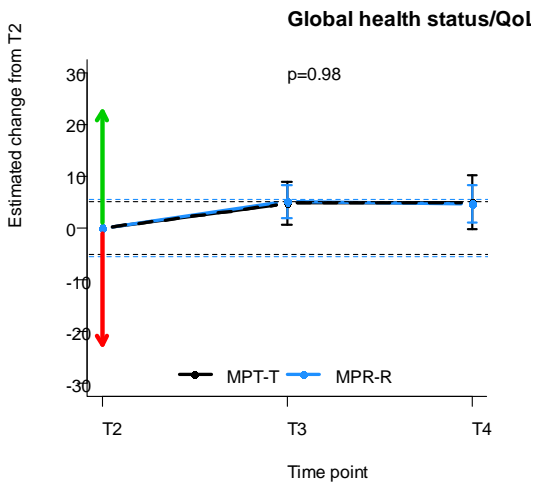| Health-related quality of life scales	Change over time *between* maintenance arms	Change over time *within* whole population	Change over time *within* arm Thalidomide	Change over time *within* arm Lenalidomide
EORTC QLQ-C30				
Global health status/QoL	0·98	**<0·001**	0·042	**0·003**
Physical functioning	0·041	0·002	0·64	<0·001
Role functioning	0·082	<0·001	0·58	**<0·001**
Emotional functioning	0·97	0·006	0·20	0·032
Cognitive functioning	0·80	0·003	0·053	0·066
Social functioning	0·43	<0·001	0·003	0·024
Fatigue	0·58	<0·001	0·095	<0·001
Nausea and vomiting	0·39	0·89	0·53	0·59
Pain	0·26	0·069	0·95	0·028
Dyspnoea	0·92	<0·001	0·051	0·004
Insomnia	0·31	0·31	0·82	0·12
Appetite loss	0·28	<0·001	0·003	<0·001
Constipation	0·49	0·005	0·036	0·15
Diarrhoea	0·85	0·045	0·52	0·12
Financial difficulties	0·48	0·26	0·22	0·76
Summary score	0·85	**<0·001**	0·005	**<0·001**
EORTC QLQ-MY20				
Disease symptoms	0·77	0·97	0·71	0·98
Side effects of treatment	0·86	0·006	0·34	0·019
Future perspective	0·84	0·003	0·28	0·011
Body image	0·70	0·82	0·74	0·80
Peripheral neuropathy	**<0·001**	<0·001[#]	**<0·001**[#]	0·89

EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20, [#] represents a statistically significant worsening in HRQoL over time.

Global health status/QoI — p=0.98

Physical functioning — p=0.041

Role functioning — p=0.082

Emotional functioning — p=0.97

Cognitive functioning — p=0.80

Social functioning — p=0.43

MPT-T   MPR-R

Fatigue

p=0.58

Nausea and vomiting

p=0.39

Pain

p=0.26

Dyspnoea

p=0.92

Insomnia

p=0.31

Appetite loss

p=0.28

Constipation

p=0.49

Diarrhoea

p=0.85

Financial difficulties

p=0.48

C30 - summary score

p=0.85

Disease symptoms

p=0.77

Side effects of treatment

p=0.86

13

Figure D1. Graphs of the estimated mean score HRQoL change from start of maintenance with corresponding 95% confidence intervals and p-value for change over time between treatment arms.
Time points with ≥5 difference in mean change from start of maintenance are marked with *, which represents a clinical relevant difference between the two treatment arms. The dotted horizontal line represents the calculated threshold for minimal important difference for the maintenance phase, the blue for the lenalidomide and the black for thalidomide maintenance. The green arrows indicate the direction of improvement in functional scales or reduction of symptom scales. The red arrows indicate the direction of deterioration in functional scales or increasing of symptom scales.
MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy

E. Impact of missing questionnaires

The results of the comparison of the course of HRQoL between patients who discontinued treatment early (before or at T2) versus late/never (after T2 or never) are presented in table F2 and figure F2.

Table E1. P-values of two-way interaction for the comparison of HRQoL course over time between patients discontinuing treatment early versus late.
The **bold** p-values represent a significant change (<0.005).

Health related quality of life scale	Within whole population	Within arm MPT-T	Within arm MPR-R
EORTC QLQ-C30			
Global health status/QoL	0·36	0·49	0·31
Physical functioning	0·18	0·11	0·65
Role functioning	0·092	0·14	0·36
Emotional functioning	**<0·001**	0·027	**0·002**
Cognitive functioning	0·19	0·12	0·97
Social functioning	0·033	0·072	0·18
Fatigue	0·24	0·27	0·077
Nausea and vomiting	0·44	0·48	0·53
Pain	0·12	0·14	0·85
Dyspnoea	0·66	0·91	0·93
Insomnia	0·059	**<0·001**	0·67
Appetite loss	0·23	0·13	0·015
Constipation	0·34	0·011	0·31
Diarrhoea	0·77	0·20	0·65
Financial difficulties	0·66	1·00	0·25
Summary score	0·17	0·061	0·13
EORTC QLQ-MY20			
Disease symptoms	0·096	0·19	0·54
Side effects of treatment	0·13	0·51	0·23
Future perspective	0·19	0·60	0·23
Body image	0·60	0·61	0·24
Peripheral neuropathy	**<0·001**	0·016	0·36

EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20, MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy.
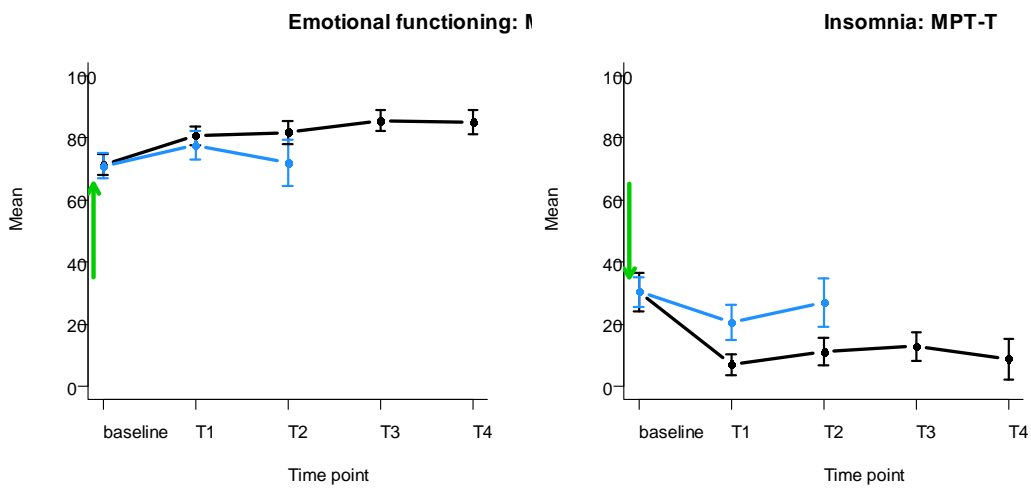
Figure E1. Course of HRQoL for scales with a statistically significant difference between patients that discontinued treatment early (before or at T2) versus late (after T2) or never.

The green arrows are indicating the direction of improvement in functioning or reduction of symptoms.

MPT-T, melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R, melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy

F. References

1.	Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *J Clin Oncol* 2011; **29**(1): 89-96.
2.	Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; **52**(9): 861-73.
3.	Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale N.J.: Lawrence Erlbaum Associates; 1988.
4.	Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical care* 2003; **41**(5): 582-92.
5.	Osoba D, Bezjak A, Brundage M, Zee B, Tu D, Pater J. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. *Eur J Cancer* 2005; **41**(2): 280-7.
6.	Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Statistical methods in medical research* 2014; **23**(5): 440-59.
7.	Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *The BMJ* 2013; **346**: e8668.
8.	Rubin DB. Inference and Missing Data. *Biometrika* 1976; **63**(3): 581-92.
9.	Fairclough DL. Design and analysis of quality of life studies in clinical trials. 2. ed. ed. Boca Raton: Chapman & Hall/CRC; 2010.

Appendix III

Paper III and supplementary material

ORIGINAL ARTICLE

Clarithromycin added to bortezomib-cyclophosphamide-dexamethasone impairs health-related quality of life in multiple myeloma patients

Lene Kongsgaard Nielsen[1] (iD) | Tobias Wirenfeldt Klausen[2] | Mary Jarden[3] |
Henrik Frederiksen[1,4] | Annette Juul Vangsted[3] (iD) | Trung Do[2] | Ida Bruun Kristensen[5] |
Ulf Christian Frølund[6] | Christen Lykkegaard Andersen[6] (iD) | Niels Abildgaard[1,4] |
Henrik Gregersen[7]

[1]Quality of Life Research Center, Department of Haematology, Odense University Hospital, Odense, Denmark

[2]Department of Haematology, University Hospital of Copenhagen at Herlev, Copenhagen, Denmark

[3]Department of Haematology, University Hospital of Copenhagen at Rigshospitalet, Copenhagen, Denmark

[4]Department of Haematology, Odense University Hospital, Odense, Denmark

[5]Department of Haematology, Vejle Hospital, Vejle, Denmark

[6]Department of Haematology, Zealand University Hospital, Roskilde, Denmark

[7]Department of Haematology, Aalborg University Hospital, Aalborg, Denmark

Correspondence
Lene Kongsgaard Nielsen, Quality of Life Research Center, Department of Haematology, Odense University Hospital, Odense C, Denmark.
Email: lene.kongsgaard.nielsen@rsyd.dk

Funding information
Obel Family Foundation, Karen Elise Jensen Foundation and Danish Myeloma Study Group.

Abstract

Objectives: The Danish Myeloma Study Group initiated a randomized, placebo-controlled, double-blinded phase II study to investigate the efficacy of adding clarithromycin to cyclophosphamide-bortezomib-dexamethasone (VCD) induction therapy in transplant eligible, newly diagnosed multiple myeloma patients. The study was prematurely terminated due to severe complications, and no effect of adding clarithromycin was found. The aim of this study was to compare health-related quality of life (HRQoL) between the two groups and to explore the coherence hereof with adverse event (AE) registration by clinicians.

Methods: Patients completed three validated HRQoL questionnaires at inclusion, before cyclophosphamide priming, and two months after high-dose therapy (HDT). The mean score difference was interpreted by clinically relevant differences between groups. Spearman's correlation analysis was used to compare patient-reported toxicities with AEs.

Results: Of 58 included patients, 55 participated in the HRQoL reporting. Before cyclophosphamide priming, patients in the clarithromycin group reported clinically relevant reduced HRQoL for eleven domains with persistent reduction in four domains two months after HDT. Poor correlation between patient-reported toxicities and clinician-reported AEs was observed.

Conclusions: Despite the premature study termination, our data demonstrate impaired HRQoL when clarithromycin was added to the VCD regimen. We found clear underreporting of toxicities by clinicians. ClinicalTrials.gov number NCT02573935.

KEYWORDS
clinical trials, multiple myeloma, quality of life, transplantation

1 | INTRODUCTION

Analyses of health-related quality of life (HRQoL) captured by patient-reported outcomes (PRO) are incorporated in most randomized phase II and III clinical cancer studies.[1]

Patient-experienced benefits and toxicities are valuable parameters for shared treatment decision-making in daily practice.[2-4] Also, PRO data results are important from a regulatory perspective in the evaluation of medicinal products, which has been stated by the US Food and Drug Administration (FDA)

and European Medicine Agency (EMA) when new drugs or drug combinations are approved.[5,6]

Health-related quality of life during induction therapy and high-dose chemotherapy with stem cell support (HDT) in newly diagnosed multiple myeloma (MM) patients has been reported in more studies.[7-11] In general, the patients report unchanged global quality of life (QoL) during induction therapy with clinically meaningful deterioration in global QoL, physical functioning and increased degree of pain and fatigue two weeks after HDT. Two months after HDT the patients report full recovery and further improvement until 12 months after HDT.[12]

The Danish Myeloma Study Group (DMSG) initiated a randomized, placebo-controlled double-blinded phase II study to investigate the efficacy and safety of adding clarithromycin to bortezomib-cyclophosfamide-dexamethason (VCD) induction therapy prior to HDT in newly diagnosed MM patients.[13] Clarithromycin in combination with lenalidomide and low-dose dexamethasone is been found to be an effective treatment regimen with manageable side effects in treatment naïve symptomatic MM patients.[14] The rationale for this study, entitled the CLAIM study, was to test these previous findings using a randomized placebo-controlled study design with addition of patient-reported HRQoL captured by validated PRO questionnaires. In fact, a valid investigation of HRQoL during an anti-myeloma regimen with addition of clarithromycin has to our knowledge never been published.

The CLAIM study was prematurely terminated on 16 September 2016, after inclusion of 58 patients, due to a high incidence of serious adverse events (AE) in the intervention group. Response data did not suggest any effect of adding clarithromycin to the VCD regimen.[13] The primary objective of this analysis was to evaluate the patient-reported HRQoL in patients receiving clarithromycin added to the VCD induction therapy. The secondary objective was to compare patient-reported toxicities to AEs reported by clinicians.

2 | PATIENTS AND METHODS

2.1 | Study design

Study details have been published previously.[13] Newly diagnosed transplant-eligible MM patients with treatment-demanding disease according to the International Myeloma Working Group criteria were eligible for inclusion.[15] The patients were randomized (1:1 ratio) to treatment with clarithromycin 500 mg orally twice daily or a matching placebo tablet for 63 days in combination with VCD induction therapy. The VCD regimen consisted of three 21-day cycles of subcutaneous bortezomib 1.3 mg/m^2 day 1, 4, 8, 11, intravenous cyclophosphamide 500 mg/m^2 on days 1 and 8, and 40 mg dexamethasone orally on days 1, 2, 4, 5, 8, 9, 11 and 12 of each cycle. After initiating the protocol, an amendment was approved to include a fourth VCD cycle prior to stem cell harvest according to an update of the Danish National Multiple Myeloma guidelines. No changes were made in relation to dosage or duration of study medication or placebo with the amendment. The study was approved by the

Danish Ethical Committee, the patients provided written informed consent before participation and the trial was conducted in accordance with the principles of the Helsinki Declaration.

2.2 | Health-related quality of life assessment

For HRQoL assessment, two "European Organisation for Research and Treatment of Cancer Quality of life" (EORTC) questionnaires were used; the cancer specific QoL instrument QLQ-C30[16] and the Multiple Myeloma module QLQ-MY20.[17] The EORTC QLQ-C30 is a validated, reliable and the most commonly used instrument for HRQoL measurement in clinical trials with MM patients.[12,18] The EORTC QLQ-C30 contains one global QoL domain, five functional domains (physical, role, emotional, cognitive and social) and nine symptom domains (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties).[19] The EORTC QLQ-MY20 contains two functional domains (future perspective and body image) and two symptom domains (disease symptoms and side effects of treatment). Each domain was scored from 0 to 100 and for the functional and global QoL domains, a higher score means better functioning/global QoL, and for the symptom domains, a higher score means a higher degree of symptoms.

For evaluation of peripheral neuropathy, the "Functional Assessment of Cancer Therapy/Gynecologic Oncology Group Neurotoxicity" (FACT/GOG-ntx) subscale was used, which is a single domain 11-item questionnaire.[20] The questionnaire has been validated and used previously in myeloma patients for evaluation of treatment-related peripheral neuropathy.[21,22] The domain was scored from 0 to 44 and a higher score means a lower degree of peripheral neuropathy.

2.3 | Health-related quality of life data collection procedure

The three questionnaires were scheduled to be completed by the patients at baseline (inclusion), before cyclophosphamide priming and two months after HDT. The patients were encouraged to complete the questionnaires electronically at home via a link sent by e-mail. The Internet-based tool of Electronic Data Capture platform has been well accepted by haematological patients.[23] The email with a link was sent to patients at baseline, at day 60 and 180 after inclusion. If patients did not complete the questionnaire within 24 hr, a reminder was sent, and in case of non-response seven days after the target date, the link to the questionnaire was blocked. Patients, who were not willing or able to answer the questionnaires electronically, completed the questionnaires by paper at study visits before cyclophosphamide priming and two months after HDT.

2.4 | Adverse events reported by clinicians

AEs were evaluated according to "Common Terminology Criteria of Adverse Events" (CTCAE) version 4.0[24] at day 1 of each VCD induction cycle, at study visits before cyclophosphamide priming and two

TABLE 1 Patient demographics

Characteristics	Clarithromycin group N = 25	Placebo group N = 30
Median age, years (IQR) [range]	64 (55; 67) [40; 70]	62 (55; 66) [37; 70]
Sex, Male (N)	19 (63.3%)	19 (76.0%)
Type of myeloma (N)		
IgA	3 (12.0%)	9 (30.0%)
IgG	18 (72.0%)	16 (53.3%)
Light chain	4 (16.0%)	5 (16.7%)
Disease stage according to ISS (N)		
I	7 (29.2%)	7 (24.1%)
II	9 (37.5%)	18 (62.1%)
III	8 (33.3%)	4 (12.9%)
Missing	1	1
B-2 microglobulin, mg/L (SD) [range]	3.4 (2.4; 7.2) [1.6; 27.1]	3.6 (2.6; 4.6) [1.9; 23.4]
Missing values	1	1
LDH, units/L (SD) [range]	164 (146; 212) [101; 267]	178 (158; 215) [110; 487]
Missing values	1	3
Serum creatinine, μmol/L (SD) [range]	81 (69; 92) [50; 271]	84 (67; 97) [45; 167]
WHO performance status scale (N)		
0	13 (52.0%)	17 (56.7%)
≥1	12 (48.0%)	13 (43.3%)

IQR, Interquartile range; ISS, International Staging System; LDH, Lactate dehydrogenase; WHO, World Health Organization.

months after HDT by clinicians. All unresolved AEs at the visit before cyclophosphamide priming were followed by the responsible clinician until the AEs were resolved.

2.5 | Statistical methods and handling of missing data

Calculation of domain scores and handling of missing items were performed as described in EORTC and FACT scoring manuals.[19,25] For the analysis of the HRQoL mean scores results, mixed model for repeated measures with an unstructured covariance matrix was used. A baseline constrained model where baseline values are constrained to be equal across treatment groups was chosen.[26] Due to early study, termination sample size was lower than planned. Therefore, the HRQoL results were primarily interpreted by thresholds of clinical relevance between treatment groups.[27] A treatment group difference of ≥5 point was defined as clinically relevant for the EORTC domains and ≥11.8 points for the FACT/GOG-ntx subscale.[28,29]

To explore the impact of non-responses to scheduled questionnaires, sensitivity analyses of the results of the global QoL domain were performed using two methods (A and B). First, variables predicting non-responses were explored using odds ratio analyses.[30-32] Variables tested for baseline non-responses were creatinine, haemoglobin, C-reactive protein and World Health Organization Performance Status

(WHO PS) at baseline. For non-responses to the follow-up questionnaires, grade 1-2 AE, grade 3-4 AE, postponed induction cycle (more than 42 days from day 1 cycle 1 to day 1 cycle 3), dose reduction of bortezomib, dexamethasone or cyclophosphamide, were tested as predictors for non-responses. In sensitivity analysis method A, multiple imputations were used. Missing scores were imputed using each patients´ creatinine, haemoglobin, C-reactive protein, WHO PS, grade 3-4 AE, information on dose reduction of bortezomib, dexamethasone or cyclophosphamide, postponed induction cycle or other values of global QoL.[33-35] In sensitivity analysis method B, we identified the non-responses in the dataset, which were timely coincident 7 days before to 30 days after with the previously found predictive variables for non-responses. The timely coincident missing scores for the non-responses were replaced by the worst possible score for global QoL in the dataset, and the analysis was repeated.

Spearman's correlation analysis was used to compare AEs assessed by clinicians with patient-reported toxicities. Cohen´s criteria for medium effect size was used to calculate the minimal important difference (MID) for the clinically meaningful change (0.5 × SD at baseline), and a score change above the MID was determined as clinical meaningful to the patient.[36-38] We used Fleiss thresholds for agreement to interpret the rho score; rho values of <0.40 were poor agreement, values between 0.40 and 0.75 were moderate to good agreement, and values >0.75 were excellent agreement.[39]

Health-related quality of life domains	Clarithromycin group Mean score (SD) N = 25	Placebo group Mean score (SD) N = 30
EORTC QLQ-C30		
Global QoL	51.7 (25.3)	60.1 (28.0)
Physical functioning	64.6 (26.5)	63.9 (30.9)
Role functioning	48.6 (35.8)	48.3 (41.9)
Emotional functioning	75.7 (17.9)	72.3 (21.2)
Cognitive functioning	86.8 (17.0)	81.0 (19.1)
Social functioning	78.5 (25.8)	75.0 (35.3)
Fatigue	39.4 (30.0)	39.5 (29.0)
Nausea and vomiting	9.0 (12.0)	9.2 (17.0)
Pain	45.8 (39.7)	55.7 (40.2)
Dyspnoea	16.7 (19.7)	13.8 (24.4)
Insomnia	27.8 (27.2)	39.1 (30.9)
Appetite loss	13.9 (25.9)	16.1 (30.4)
Constipation	34.7 (37.4)	28.6 (32.3)
Diarrhoea	8.3 (17.7)	4.8 (11.9)
Financial difficulties	5.8 (12.9)	7.1 (18.9)
EORTC QLQ-MY20		
Disease symptoms	32.1 (21.2)	39.8 (30.4)
Side effects of treatment	13.6 (9.9)	13.4 (13.4)
Future perspective	76.8 (34.0)	84.0 (26.7)
Body image	41.5 (29.1)	41.2 (38.4)
FACT/GOG-Ntx subscale	8.7 (9.3)	8.0 (8.6)

EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire core questionnaire; EORTC QLQ-MY20, European Organisation for Research and Treatment of Cancer Multiple Myeloma module; FACT/GOG-Ntx subscale, Functional Assessment of Cancer Therapy/Gynecologic Oncology Group Neurotoxicity subscale; SD, Standard deviation.

P-values below 0.05 were considered significant. R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) and SAS version 9.4 (SAS institute, Cary, NC, USA). SAS was used for mixed model for repeated measures, whereas R package "mice" was used for multiple imputations.

3 | RESULTS

3.1 | Patient population and compliance

From the time of inclusion of the first patient on 16 November 2015 until termination of the study on 16 September 2016, 58 patients were included. Three patients did not answer any of the questionnaires and were excluded from the HRQoL analysis. Of the patients included in the analysis, 25 patients were allocated to clarithromycin and 30 patients to placebo. Patient baseline characteristics are presented in Table 1.

The completeness rates of questionnaires were 84% in the clarithromycin group and 89% in the placebo group. Mean scores at baseline and SD for each domain and treatment group are presented
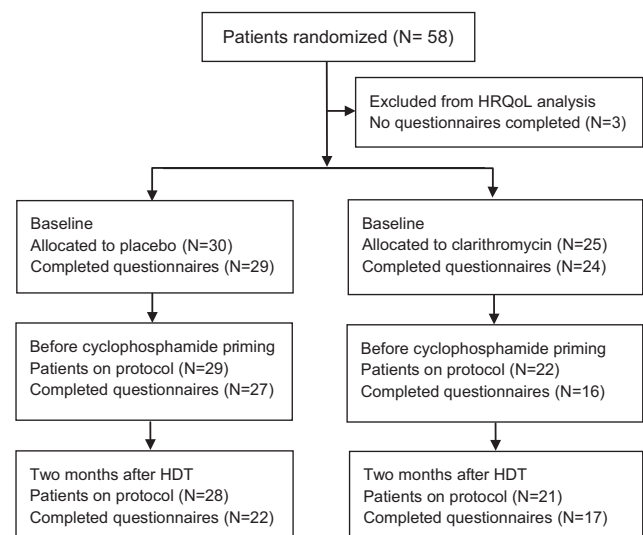
FIGURE 1 CONSORT flow diagram of the number of patients in follow-up and number of completed questionnaires. HRQoL; health-related quality of life, HDT; high-dose chemotherapy with stem cell support

(A) Functional domains – in favour of placebo
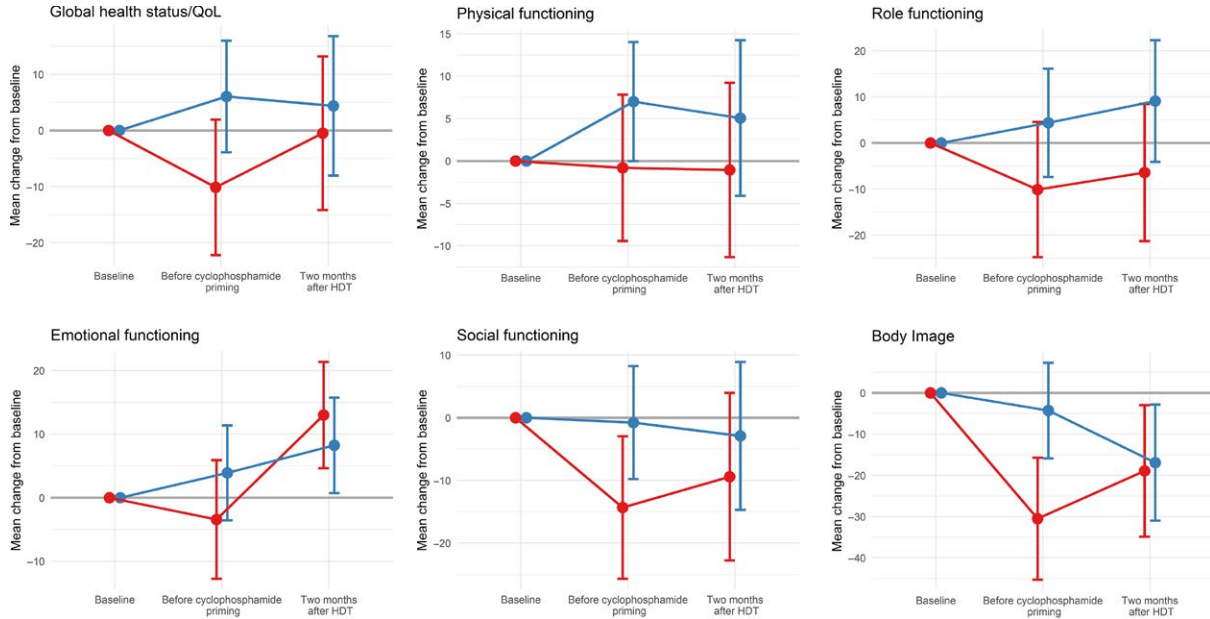
(B) Symptom domains – in favour of placebo
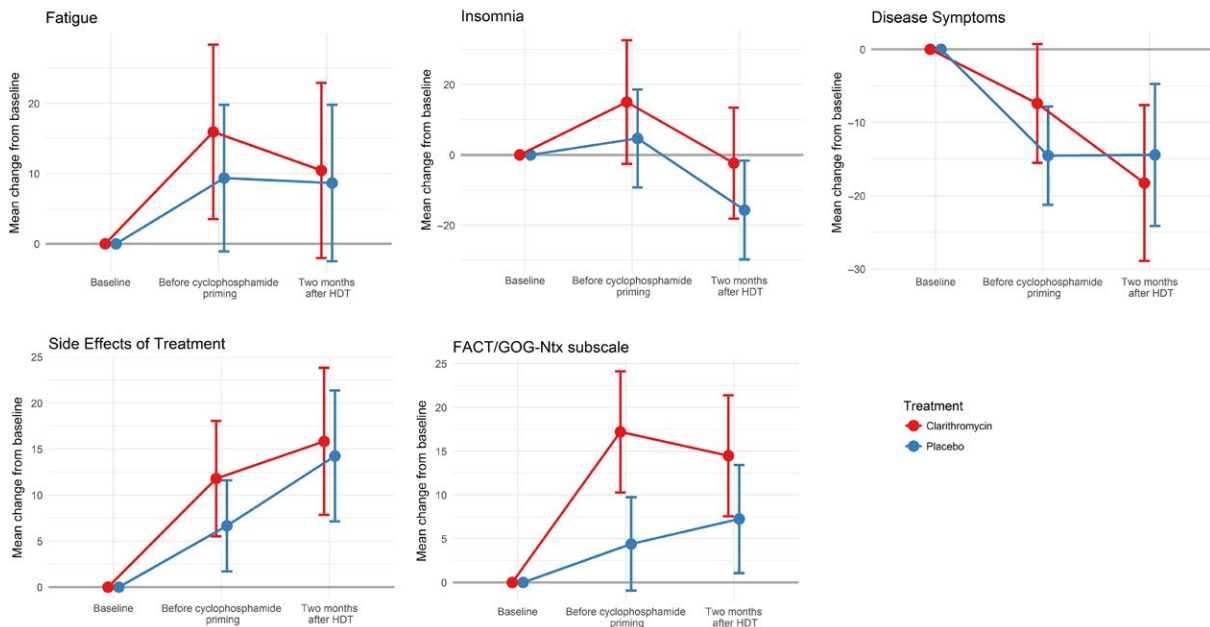
(C) Symptom domains – in favour of clarithromycin
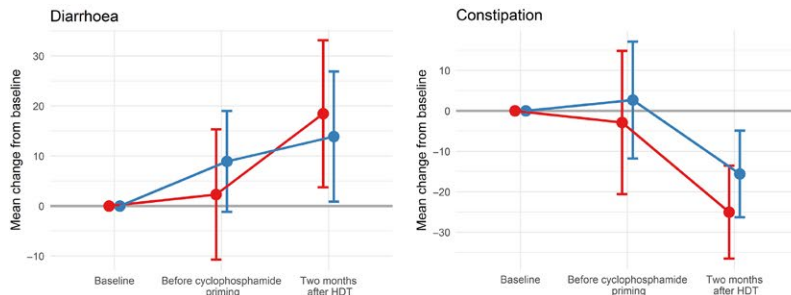
FIGURE 2 (A-C) Graphs of the domains with a clinical relevant difference between the two treatment groups before cyclophosphamide priming. For physical, role and social functioning and insomnia and constipation the clinical relevant differences were persistent two months after HDT. For the functional domains including global health status/QoL, a higher score means better functioning/QoL, and for the symptom domains, a higher score means a higher degree of symptoms.

in Table 2. The mean baseline scores in global QoL were imbalanced with a difference of 8.4 points between the two groups and a graph of change in global QoL score over time is presented in the supplementary file Figure S1. The number of patients in the study at baseline, before cyclophosphamide priming and two months after HDT and the number of completed questionnaires are presented in the CONSORT diagram in Figure 1. The main reason for early patient drop out was serious AEs, which was the case for four patients in the clarithromycin group and one patient in the placebo group.

Thirty-four patients (62%) completed the questionnaires electronically, and 21 patients (38%) chose paper questionnaires. Since some VCD induction cycles were postponed due to complications and some patients were treated with four cycles of VCD, not all patients completed the follow-up questionnaires at the scheduled time points before cyclophosphamide priming and two months after HDT. The follow-up questionnaires before cyclophosphamide priming were completed with a median of nine days too early (range −51 to 11) for the clarithromycin group and 12 days too early (range −41 to 1) for the placebo group. Also, the two months after HDT assessments were completed with a median of four days too early (range −36 to 45) for the clarithromycin group and one day too early (range −38 to 19) for the placebo group.

3.2 | Comparison of HRQoL between treatment groups

HRQoL domains with a clinical relevant difference in mean change of score before cyclophosphamide priming are presented in Figure 2, and the domains with no clinical relevant difference are presented in the supplementary file Figure S2.

Before cyclophosphamide priming, the patients in the clarithromycin group reported clinically relevant reduced global QoL, physical, role, emotional and social functioning, body image and increasing fatigue, insomnia, disease symptoms, side effects of treatment and peripheral neuropathy compared to the patients in the placebo group. Two months after HDT, the clinical relevant reduced HRQoL was persistent for physical, role and social functioning, and insomnia. Only for diarrhoea and constipation before cyclophosphamide priming and for constipation two months after HDT, the patients receiving clarithromycin reported clinically relevant reduced symptoms compared to the patients receiving placebo. The mean score difference for global QoL between the two groups was −16.2 points (95% CI −2.6; −29.8, $P = 0.021$) before cyclophosphamide priming and −4.9 (95% CI −11.1; 20.8, $P = 0.54$) two months after HDT. The p-values for comparison of mean change in score from baseline between the two treatment groups are presented in the supplementary file Table S1.

The only statistical significant predictor for non-responses to scheduled questionnaires was registration of grade 3 or 4 AEs with an odds ratio of 4.2 ($P = 0.03$) before cyclophosphamide priming and 3.5 ($P = 0.04$) two months after HDT. A table of grade 3 or 4 AEs is presented in the Table S2. Using multiple imputation for non-responses coincident with registration of grade 3 and 4 AEs (method A), the mean score differences for global QoL were −15.8 (95% CI
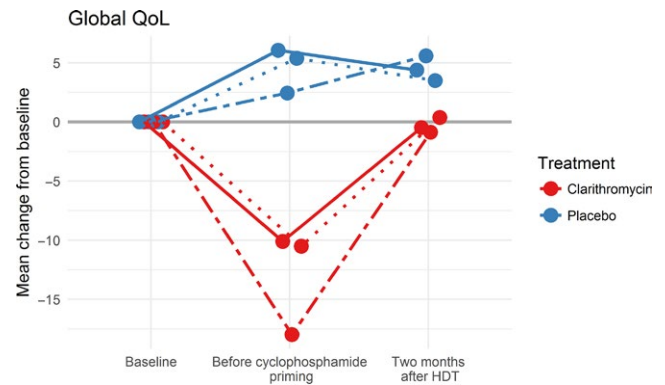
FIGURE 3 Impact of non-responses to scheduled questionnaires. The analysis using mixed model repeated measure (solid lines), sensitivity analysis method A using multiple imputations (dotted lines) and sensitivity analysis methods B using missing score replacement with zero (spotted lines).

−29.1; −2.6 $P = 0.019$) before cyclophosphamide priming and −3.1 (95% CI −17.9; 11.7, $P = 0.68$) two months after HDT. For method B, we replaced the score of non-responses with the worst possible reported score for global QoL in the dataset, which was zero. We found mean score differences of −20.4 (95% −35.5; −5.3, $P = 0.009$) before cyclophosphamide priming and −6.4 (95% −22.6; 9.7, $P = 0.009$) two months after HDT. The results of and the sensitivity analyses method A and B are illustrated in Figure 3.

3.3 | Adverse events registered by clinicians and patient-reported toxicities

In the correlation analysis, we compared clinician registered AEs to the patient-reported toxicities for the eight toxicity domains of fatigue, nausea and vomiting, dyspnoea, insomnia, appetite loss, constipation, diarrhoea and peripheral neuropathy. Since some discrepancies were observed between the time points of AE evaluation by clinicians at the study visits and the time points of answered questionnaires, time effect correlation analyses were carried out. For constipation, we observed a statistically significant time effect before cyclophosphamide priming (rho = −0.39; $P = 0.012$) and two months after HDT (rho = 0.47; $P = 0.005$). Also, for diarrhoea, there was a statistically significant time effect two months after HDT (rho = −0.34; $P = 0.045$). Therefore, correlation analyses were not performed for constipation at the two follow-up time points and for diarrhoea two months after HDT. Overall, poor correlations between the patient-reported toxicity and clinician registered AEs for all six toxicities were found with rho values <0.4 (Table 3).

4 | DISCUSSION

Our data demonstrate that MM patients report a clinically relevant reduced HRQoL, when clarithromycin is added to the VCD regimen

TABLE 3 Correlation between registered adverse events by clinicians and patient-reported toxicities. Any grade of toxicity according to Common Terminology Criteria for Adverse Events (CTCAE) version 4.0, patient-reported change from baseline above the threshold for minimal important difference calculated by Cohen's medium effect size (0.5 × SD of mean baseline score) for the domain

Patient-reported toxicities	Before cyclophosphamide priming Correlation		Two months after HDT Correlation	
	Rho	*P*-value	Rho	*P*-value
Diarrhoea	−0.10	0.53	NA	NA
Constipation	NA	NA	NA	NA
Nausea and vomiting	−0.11	0.51	−0.31	0.06
Fatigue	0.35	0.023	−0.01	0.97
Insomnia	0.20	0.20	−0.09	0.61
Peripheral neuropathy	0.29	0.075	−0.02	0.90
Dyspnoea[a]	—	—	—	—
Appetite loss	0.02	0.92	0.26	0.13

[a]Correlation calculation was not possible, since none of the patients had dyspnoea registered as an adverse event, NA; correlation analysis was not performed since a statistically significant time effect was found.

with persisting HRQoL sequelae two months after HDT. Using registered toxicities by CTCAE this knowledge could not be concluded from the clinicians' AE evaluation, since they underreported symptomatic toxicities.

A limitation of our results is that it is based on an underpowered study due to premature study termination with a poor questionnaire completion rate, which made us unable to obtain a valid statistical result. Still, when comparing our results to existing literature of HRQoL during induction therapy and HDT in MM patients, it is noteworthy that the patients in the clarithromycin group reported decreased HRQoL after induction phase.[7,9] Our findings could be explained by the pharmacokinetics of bortezomib and clarithromycin. Bortezomib is primarily metabolized by the cytochrome P450 enzyme CYP3A4, which is known to be inhibited by clarithromycin. Thus, the reduced HRQoL could be a result of increased biological effect of bortezomib in the clarithromycin group.[13] Clarithromycin has been used in other treatment regimens for MM often in combination with lenalidomide and low-dose dexamethasone, which is found to have favourable toxicity profile.[14] This discrepancy in AE findings compared to our study supports the explanation of being caused by the pharmacokinetic interaction between bortezomib and clarithromycin, when those two drugs are administrated in parallel. In the CLAIM study, special precaution was made for the potential risk of QT prolongation, ventricular tachycardia and sudden death caused by clarithromycin. Severe cardiac disease or QT prolongation was exclusion criteria, and ECG was performed at screening, on day 4 and before start of VCD cycle 2. If the patient developed QT prolongation

(QTc interval > 500 ms), the clarithromycin/placebo treatment was permanently discontinued. However, no serious cardiovascular events were reported during the study.[13]

In clinical studies, AEs are traditionally collected as described in CTCAE guideline by clinicians.[24] Drug efficacy and toxicity profile analyses are included in the process where a given drug is considered for approval by the FDA and EMA. In earlier studies comparison of CTCAE and patient-reported toxicities revealed underreporting of toxicities by the clinicians as compared to patient-reported toxicities.[40,41] Our study confirmed this discrepancy, thereby emphasizing the importance of including HRQoL as an endpoint in clinical trials. Also, it highlights the potentially important role of integrating PRO data in real-time safety monitoring in clinical trials as well as in the daily clinical practice.[42] A limitation in the interpretation of our results when comparing patient-reported toxicities and clinician reported AEs is the lack of synchronous registration of toxicities by clinicians and patients and the retrospective nature of the analysis. Still, we believe the results are convincing since clinicians may tend to underreport AEs.[40]

In this current study, we observed that there were non-responses to scheduled questionnaires, which is a common challenge in PRO data collection, analysis and interpretation.[32,43,44] The potential consequences of non-responses are decreased precision and power, and more seriously, the introduction of bias to the PRO data results, when a patient fails to complete a questionnaire because of severe illness or other reasons. It is recommended to design clinical studies with PRO data collection with focus on minimization of non-responses and to perform sensitivity analysis to explore the impact of non-responses on the PRO data results.[44,45] In our study, more patients in the clarithromycin group dropped out early due to serious AEs resulting in a lower questionnaire completion compared to patients in the placebo group. Therefore, the analyses performed are hypothetically fragile for biased results. We performed analyses to explore the impact of non-responses of being "missing not at random".[46] We examined the mechanisms of non-responses and found that registration of a grade 3-4 AEs was a predictor of non-responses, which confirms that some of the non-responses were "missing not at random." When integrating this information into the sensitivity analysis method B, it was confirmed that non-responses to questionnaires do impact the results of the global QoL domain and that our results might be conservative. However, in the sensitivity analysis method A using multiple imputations, we found no impact of non-responses on the global QoL results. Limitations in using the multiple imputation method in our study are the low sample size and a limited number of patients with grade 3 or 4 AEs reporting a global QoL score. Also, the global QoL domain is described as a "distal" measure with limitations in interpretability due to greater mediation by personal and environmental characteristics rather than disease and treatment-related chances.[47]

In conclusion, the CLAIM study demonstrated that adding clarithromycin to the VCD regimen in MM patients resulted in impaired HRQoL during the VCD induction phase continuing up to two months after HDT. The study emphasizes that well-designed randomized, double-blinded and placebo-controlled studies with PRO data

collection is necessary to determine drug risk benefit assessment, and also to test well-known drugs in new combinations. Treatment with clarithromycin and VCD in parallel cannot be recommended because of a higher risk of complications and reduced HRQoL. The PRO data in the CLAIM study played a key role in explaining the causality link between the observed complications and the possible interaction between clarithromycin and bortezomib. In addition, the study demonstrates that "real-time" monitoring of patient-reported toxicities as a supplement to CTCAE registration should be included in clinical trials. The National Cancer Institute´s PRO version of the common terminology criteria for adverse events (PRO-CTCAE) has been validated and found feasible in clinical trials for documentation of symptomatic toxicities.[48] With this tool, PRO data can be incorporated into future clinical cancer studies. Moreover, PRO data will most likely be a useful tool in shared treatment decision making in clinical practice. Studies designed to validate the use of PRO data in daily practice are warranted.

ACKNOWLEDGEMENTS

ORCID

Lene Kongsgaard Nielsen iD http://orcid.org/0000-0003-2083-3870

Annette Juul Vangsted iD http://orcid.org/0000-0002-2131-731X

Christen Lykkegaard Andersen iD http://orcid.org/0000-0001-7753-6768

REFERENCES

1. Basch E, Rogak LJ, Dueck AC. Methods for implementing and reporting patient-reported outcome (PRO) measures of symptomatic adverse events in cancer clinical trials. *Clin Ther*. 2016;38(4):821-830.
2. Efficace F, Gaidano G, Lo-Coco F. Patient-reported outcomes in hematology: is it time to focus more on them in clinical trials and hematology practice? *Blood*. 2017;130(7):859-866.
3. Deber RB, Kraetschmer N, Urowitz S, Sharpe N. Do people want to be autonomous patients? Preferred roles in treatment decision-making in several patient populations. *Health Expect*. 2007;10(3):248-258.
4. Tariman JD, Doorenbos A, Schepp KG, Singhal S, Berry DL. Older adults newly diagnosed with symptomatic myeloma and treatment decision making. *Oncol Nurs Forum*. 2014;41(4):411-419.
5. Speight J, Barendse SM. FDA guidance on patient reported outcomes. *BMJ*. 2010;340:c2921.
6. European Medicine Agency. The use of patient reported outcome (PRO) measures in oncology studies. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man; 2016. https://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf
7. Gulbrandsen N, Wisloff F, Brinch L, et al. Health-related quality of life in multiple myeloma patients receiving high-dose chemotherapy with autologous blood stem-cell support. *Med Oncol*. 2001;18(1):65-77.
8. Etto LY, Morelli VM, Silva VC, et al. Autologous stem cell transplantation improves quality of life in economically challenged, Brazilian multiple myeloma patients. *Clinics (Sao Paulo, Brazil)*. 2011;66(11):1855-1859.
9. Ludwig H, Viterbo L, Greil R, et al. Randomized phase II study of bortezomib, thalidomide, and dexamethasone with or without cyclophosphamide as induction therapy in previously untreated multiple myeloma. *J Clin Oncol*. 2013;31(2):247-255.
10. Segeren CM, Sonneveld P, van der Holt B, et al. Overall and event-free survival are not improved by the use of myeloablative therapy following intensified chemotherapy in previously untreated patients with multiple myeloma: a prospective randomized phase 3 study. *Blood*. 2003;101(6):2144-2151.
11. Frodin U, Borjeson S, Lyth J, Lotfi K. A prospective evaluation of patients' health-related quality of life during auto-SCT: a 3-year follow-up. *Bone Marrow Transplant*. 2011;46(10):1345-1352.
12. Nielsen LK, Jarden M, Andersen CL, Frederiksen H, Abildgaard N. A systematic review of health-related quality of life in longitudinal studies of myeloma patients. *Eur J Haematol*. 2017;99(1):3-17.
13. Gregersen H, Do T, Kristensen IB, et al. A randomized placebo-controlled phase II study of clarithromycin or placebo combined with VCD induction therapy prior to high-dose melphalan with stem cell support in patients with newly diagnosed multiple myeloma. *Exp Hematol Oncol*. 2018;7:18.
14. Niesvizky R, Jayabalan DS, Christos PJ, et al. BiRD (Biaxin [clarithromycin]/Revlimid [lenalidomide]/dexamethasone) combination therapy results in high complete- and overall-response rates in treatment-naive symptomatic multiple myeloma. *Blood*. 2008;111(3):1101-1109.
15. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 2014;15(12):e538-e548.
16. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365-376.
17. Cocks K, Cohen D, Wisløff F, et al. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. *Eur J Cancer*. 2007;43(11):1670-1678. 9p.
18. Wisloff F, Hjorth M, Kaasa S, Westin J. Effect of interferon on the health-related quality of life of multiple myeloma patients: results of a Nordic randomized trial comparing melphalan-prednisone to melphalan-prednisone+alpha-interferon. *Br J Haematol*. 1996;94(2):324-332.
19. Fayers P, Aaronson NK, Bjordal K, et al. *The EORTC QLQ-C30 Scoring Manual*, 3rd edn. Brussels, Belgium: European Organisation for Research and Treatment of Cancer; 2001.
20. Huang HQ, Brady MF, Cella D, Fleming G. Validation and reduction of FACT/GOG-Ntx subscale for platinum/paclitaxel-induced neurologic symptoms: a gynecologic oncology group study. *Int J Gynecol Cancer*. 2007;17(2):387-393.
21. Cho J, Kang D, Lee JY, Kim K, Kim SJ. Impact of dose modification on intravenous bortezomib-induced peripheral neuropathy in multiple myeloma patients. *Support Care Cancer*. 2014;22(10):2669-2675.
22. Lee SJ, Richardson PG, Sonneveld P, 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):511-519.
23. Brochmann N, Zwisler AD, Kjerholt M, Flachs EM, Hasselbalch HC, Andersen CL. A new internet-based tool for reporting and analysing patient-reported outcomes and the feasibility of repeated data

European Journal of
Haematology

WILEY

collection from patients with myeloproliferative neoplasms. *Qual Life Res*. 2015;25(4):835-846.

24. National Cancer Institute NIoH. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. US Department of Health and Human Services; 2009. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm

25. Administration and Scoring Guidelines www.facit.org: functional assessment of chronic illness therapy; [FACT/GOG-Ntx Scoring Guidelines (Version 4).

26. Coffman CJ, Edelman D, Woolson RF. To condition or not condition? Analysing 'change' in longitudinal randomised controlled trials. *BMJ Open*. 2016;6(12):e013096.

27. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171-184.

28. Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *J Clin Oncol*. 2011;29(1):89-96.

29. Calhoun EA, Welshman EE, Chang CH, et al. Psychometric evaluation of the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (Fact/GOG-Ntx) questionnaire for patients receiving systemic chemotherapy. *Int J Gynecol Cancer*. 2003;13(6):741-748.

30. Biering K, Hjollund NH, Frydenberg M. Using multiple imputation to deal with missing data and attrition in longitudinal studies with repeated measures of patient-reported outcomes. *Clin Epidemiol*. 2015;7:91-106.

31. Wang C, Hall CB. Correction of bias from non-random missing longitudinal data using auxiliary information. *Stat Med*. 2010;29(6):671-679.

32. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res*. 2014;23(5):440-459.

33. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: J. Wiley; 1987:258.

34. Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*, 2nd edn. Boca Raton, FL: Chapman & Hall/CRC; 2010:404.

35. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1-67.

36. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

37. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582-592.

38. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res*. 1996;5(6):555-567.

39. Grove WA. Statistical methods for rates and proportions, 2nd ed. *Am J Psychiatry*. 1981;138(12):1644-a-5.

40. Di Maio M, Gallo C, Leighl NB, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. *J Clin Oncol*. 2015;33(8):910-915.

41. Atkinson TM, Ryan SJ, Bennett AV, et al. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. *Support Care Cancer*. 2016;24(8):3669-3676.

42. Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med*. 2010;362(10):865-869.

43. Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res*. 2016;25(12):2977-2983.

44. Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *BMJ Open*. 2016;6(6):e010938.

45. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA*. 2018;319(5):483-494.

46. Dufouil C, Brayne C, Clayton D. Analysis of longitudinal studies with death and drop-out: a case study. *Stat Med*. 2004;23(14):2215-2226.

47. Luckett T, King MT. Choosing patient-reported outcome measures for cancer clinical research–practical principles and an algorithm to assist non-specialist researchers. *Eur J Cancer*. 2010;46(18):3149-3157.

48. Dueck AC, Mendoza TR, Mitchell SA, et al. Validity and reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *JAMA Oncol*. 2015;1(8):1051-1059.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary appendix
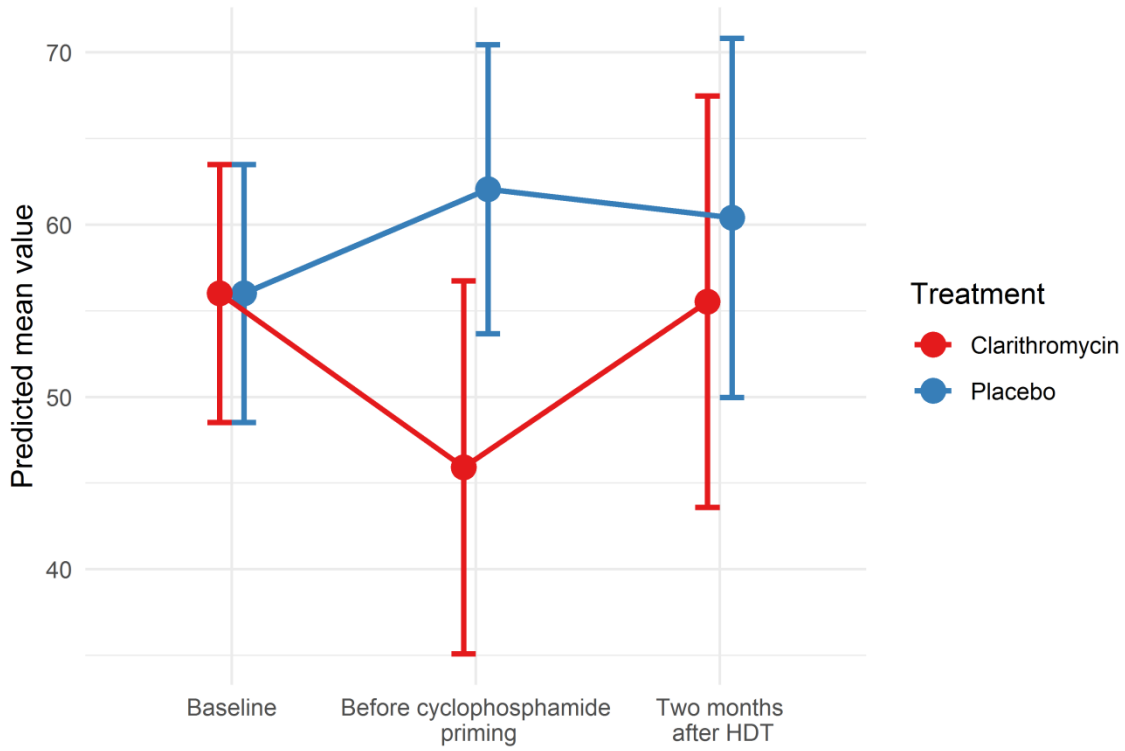
Figure 1S. Graph of the predicted mean value for the global quality of life domain and 95% confidence intervals.

Functional domains – with no clinical relevant differences
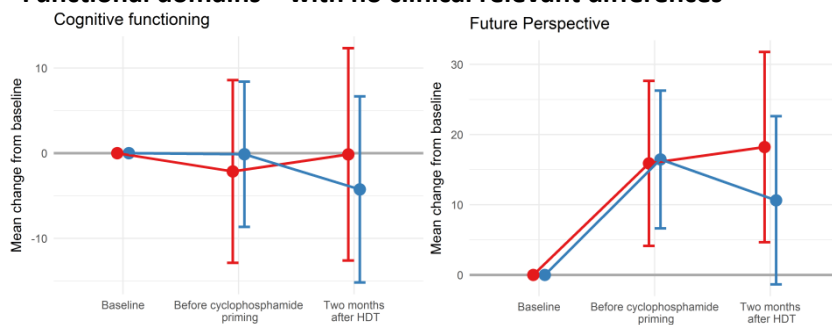
Symptom domains – with no clinical relevant differences
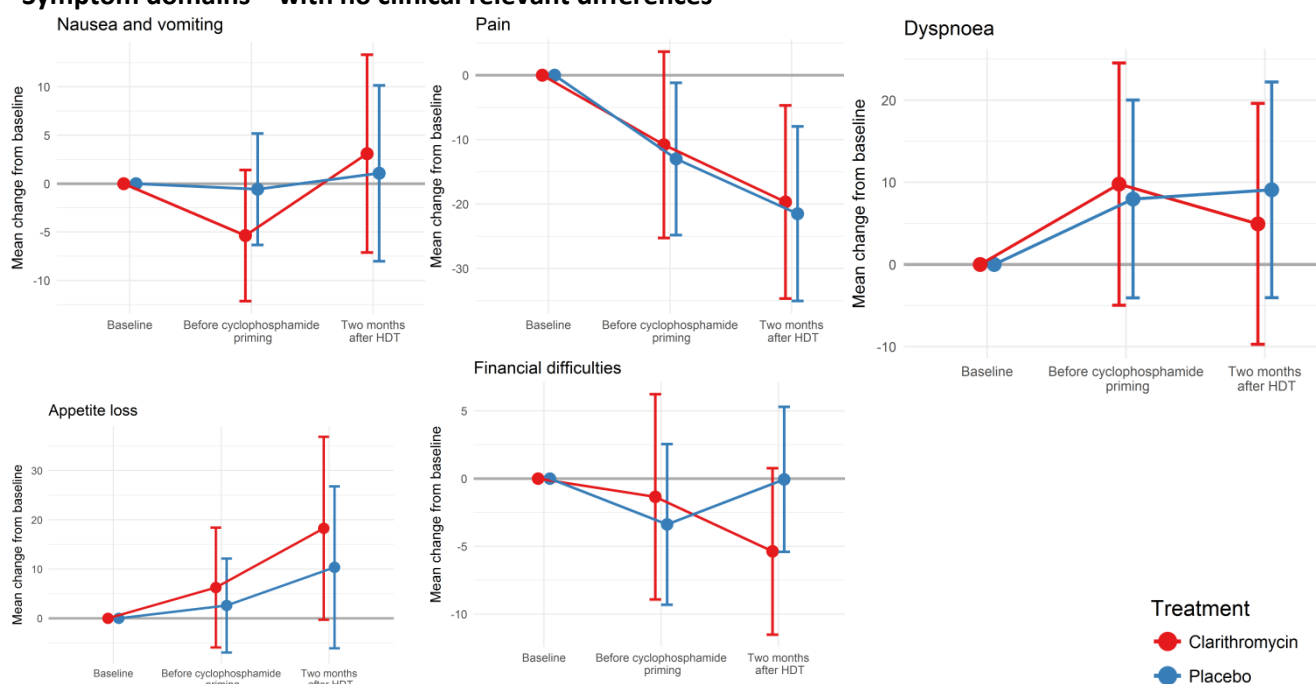
Figure 2S. Graphs of the domains with a no clinical relevant difference between the two treatment groups before cyclophosphamide priming or two months after HDT.

For the functional domains including global health status/QoL, a higher score means better functioning/QoL, and for the symptom domains, a higher score means a higher degree of symptoms.

Table 1S. P-values for comparison of mean change in score from baseline between the clarithromycin group compared to the placebo group before cyclophosphamide priming and two months after high-dose chemotherapy with stem cell support. Statistical significant p-values are marked in bold (p<0.05).

Health-related quality of life domains	Before cyclophosphamide priming	Two months after HDT
	P-value	**P-value**
EORTC QLQ-C30		
Global QoL	**0.02**	0.54
Physical Functioning	0.12	0.30
Role Functioning	0.10	0.10
Emotional Functioning	0.19	0.32
Cognitive Functioning	0.76	0.60
Social Functioning	0.05	0.43
Fatigue	0.34	0.80
Nausea and vomiting	0.20	0.74
Pain	0.79	0.82
Dyspnoe	0.83	0.65
Insomnia	0.34	0.17
Appetite loss	0.62	0.50
Constipation	0.60	0.09
Diarrhoea	0.41	0.64
Financial difficulties	0.66	0.15
EORTC QLQ-MY20		
Disease symptoms	0.13	0.53
Side effects of treatment	0.19	0.76
Future Perspective	0.94	0.34
Body image	**0.01**	0.83
FACT/GOG-Ntx subscale	**<0.01**	0.11

HDT; high-dose chemotherapy with stem cell support, EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire core questionnaire, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Multiple Myeloma module, FACT/GOG-Ntx subscale; Functional Assessment of Cancer Therapy/Gynecologic Oncology Group Neurotoxicity subscale

Table 2S. Most common adverse events grade 3-4 in the clarithromycin and placebo group reported by clinicians.

	Clarithromycin group N=25	Placebo group N=30
Haematological events		
Thrombocytopenia	2 (7.4%)	0
Anemia	2 (7.4%)	0
Neutropenia	1 (3.7%)	0
Gastrointestinal events		
Typhlitis and colon perforation	2 (7.4%)	1 (3.2%)
Paralytic ileus	2 (7.4%)	0
Constipation	1 (3.7%)	0
Diarrhea	3 (11.1)	1 (3.2%)
Dyspepsia	0	0
Nausea	0	0
Infections		
Respiratory tract infection	1 (3.7%)	5 (16.1%)
Urinary tract infection	0	1 (3.2%)
Septicaemia	5 (18.5%)	1 (3.2%)
Other infections	4 (14.8%)	5 16.1%)
Other candidiasis	1 (3.7%)	0
Nervous system disorders		
Peripheral sensory neuropathy	0	0
Dizziness	1 (3.7%)	0
Other conditions		
Peripheral oedema	1 (3.7%)	1 (3.2%)
Hypotension	2 (7.4%)	0
Fatigue	1 (3.7%)	0
Rash	0	0
Insomnia	0	0
Weight loss	0	0
Mucositis	0	0
Psychiatric symptoms	1 (3.7%)	1 (3.2%)

Appendix IV

Paper IV

Methodological aspects of health-related quality of life measurement and analysis in patients with multiple myeloma

Running title: HRQoL measurement and analysis in patients with MM

Lene Kongsgaard Nielsen[1,2], Niels Abildgaard[1,2], Mary Jarden[3] and Tobias Wirenfeld Klausen[4]

[1]Quality of Life Research Center, Department of Haematology, Odense University Hospital, Denmark; [2]The Academy of Geriatric Cancer Research (AgeCare), Odense University Hospital, Denmark; [3]Department of Haematology, Copenhagen University Hospital, Copenhagen, Denmark; [4]Department of Haematology, Herlev Hospital, Denmark

Corresponding author: Lene Kongsgaard Nielsen, Quality of Life Research Center, Department of Haematology, Odense University Hospital, Kløvervænget 10, 12. Floor, 5000 Odense, Denmark, email lene.kongsgaard.nielsen@rsyd.dk phone: +45 2324 1605.

Summary

Multiple myeloma (MM) is an incurable but treatment sensitive cancer. For most patients, this mean treatment with multiple lines of anti-myeloma therapy and a life with disease and treatment-related symptoms and complications. Health-related quality of life (HRQoL) issues play an important role in treatment decision-making. Methodological challenges in longitudinal HRQoL measurements and analyses have been identified including non-responses (NR) to scheduled questionnaires. Building upon publications we identified for a systematic review of longitudinal HRQoL studies in MM, we here focused on methodological aspects of HRQoL measurement and analysis. Diversity in timing of HRQoL data collection and applied statistical methods were noticed. We observed high rate of NR and only in 8/23 studies investigation of the impact of NR was performed. Thus, evidence-based knowledge of HRQoL in patients with MM is compromised. To improve quality of HRQoL results and their implementation in daily practice, future studies should follow established guidelines.

4-6 keywords Health-related Quality of Life, Missing data, statistical analysis, review, multiple myeloma.

Introduction

Health-related quality of life (HRQoL) and other Patient Reported Outcomes (PRO) have become increasingly used as endpoints in clinical cancer studies to measure patient experienced benefits and toxicities of treatments (Basch, *et al* 2016, Vodicka, *et al* 2015). PRO results are important in the approval of new drugs, as well as in shared decision-making in the daily care of patients (EMA 2016, Speight and Barendse 2010).

Multiple myeloma (MM) is an incurable malignancy derived from plasma cells in the bone marrow. MM is the second most common haematological cancer, and worldwide, it is estimated that 86,000 patients annually are diagnosed with MM (Becker 2011). The prognosis of MM has improved markedly over the past 20 years and is expected to improve further in the coming years due to new treatment options (Kumar, *et al* 2014, Kumar, *et al* 2008). The median survival of MM patients under the age of 70 at the time of diagnosis now exceeds 6-7 years (Kumar, *et al* 2014). MM is associated with severe morbidity caused by bone destruction/bone fractures, renal dysfunction, bone marrow failure, high infection rates and potential physical disability (Kyle and Rajkumar 2008, Rajkumar, *et al* 2014).

Current treatment of newly diagnosed MM patients involves induction treatment with repeated cycles of two or three drug combinations followed by stem cell harvest and high dose chemotherapy with stem cell support (HDT) in younger, eligible patients (Laubach, *et al* 2016, Lenhoff, *et al* 2000, Mateos, *et al* 2014). After HDT patients are treated with consolidation and/or long-term maintenance treatment or enter a drug free period (McCarthy, *et al* 2017, Richardson, *et al* 2018). Elderly and HDT ineligible patients are treated with longer induction treatment for 8-12 months and in some patients, treatment is continued until relapse (Benboubker, *et al* 2014, Stewart, *et al* 2015a). However, eventually, the MM disease will progress or relapse, and initiation of a rescue treatment will be necessary (Laubach, *et al* 2016). Chemotherapy as well as treatment with proteasome inhibitors (PI) and immune-regulatory agents (IMID) cause a high risk of acute adverse events, e.g. suppression of the bone marrow with risk of infections and hospitalization as well as other side effects, e.g. peripheral neuropathy and fatigue (Boland, *et al* 2013, Mateos 2010, Molassiotis, *et al* 2011, Richardson, *et al* 2012). Treatment decision-making in MM is complex and involves factors such as disease stage, prognostic risk stratification, severity of myeloma symptoms and complications, expected progression free survival and toxicity profiles of available treatment regimens, as well as the patients´ comorbidities and preferences concerning goals in life, convenience in drug administration and HRQoL during and after treatment (Deber, *et al* 2007, Laubach, *et al* 2016, Leleu, *et al* 2015, Mikhael, *et al* 2013, Tariman, *et al* 2014a, Tariman, *et al* 2014b).

In order to use PRO results in clinical decision-making and the process of drug approval, valid PRO results from clinical trials are essential. Methodological challenges in integrating PRO measurements in clinical cancer trials have been identified (Calvert, *et al* 2018). PRO specific content has often been omitted from clinical trial protocols and trial coordinators have been found to lack training and support in PRO administration practices (Kyte, *et al* 2014, Mercieca-Bebber, *et al* 2018). The statistical analysis approaches for longitudinal PRO data analysis differ and reporting and handling of missing data have been found inadequate in general (Brundage, *et al* 2011a, Fielding, *et al* 2016, Hamel, *et al* 2017). The clinical implications of PRO results from clinical trials have been disappointing, declaring that there is a need for international standards for PRO measurements (Bottomley, *et al* 2002, Bottomley, *et al* 2016, Bottomley, *et al* 2018, Kvam, *et al* 2009, Lee and Chi 2000, Sonneveld, *et al* 2013, Sprangers 2010). Specific recommendations for PRO implementation and reporting in clinical cancer and haematological trials have been compiled (Blade, *et al* 2018, Calvert, *et al* 2018, Efficace, *et al* 2017, EHA 2013, Mercieca-Bebber, *et al* 2016). Among important steps are identifying optimal PRO measuring time points, strategies to minimize missing PRO data, and to use appropriate statistical analysis methods to handle missing data (Bell and Fairclough 2014, Calvert, *et al* 2013, Efficace, *et al* 2017, EHA 2013, Hamel, *et al* 2017, Mercieca-Bebber, *et al* 2018, Mercieca-Bebber, *et al* 2016).

A fully missing scheduled PRO questionnaire, defines a non-response (NR), and can be subdivided into patterns of NR as monotone, intermittent or mixed (Fielding, *et al* 2009, Little, *et al* 2012). A monotone pattern of NR is a pattern of complete responses until NR occurs by e.g. drop-out, intermittent NR is a pattern of one or more NRs between completed questionnaires, and a mixed pattern of NR occurs when a patient first has an intermittent and later a monotone pattern. If a patient participating in a clinical trial or cohort study does not complete any scheduled questionnaires or is excluded from the PRO data analysis, the patient is defined as a complete non-responder.

Three different mechanisms for NR have been described (Rubin 1976), and each is exemplified here. "*Missing completely at random*" (MCAR) occurs, for example, if the questionnaires are not given to the patient. "*Missing at random*" (MAR) occurs, for example, if a specific subgroup of patients with similar outcomes e.g. poorer PRO scores has a higher proportion of NR. "*Missing not at random*" (MNAR) occurs, for example, if the patient does not complete the questionnaire due to experiencing adverse events or complications (Fielding, *et al* 2009, Palmer, *et al* 2018).

The objectives of this review were to investigate applied PRO measuring time points, statistical analysis methods and the magnitude and ways of handling NR in longitudinal PRO studies of patients with MM. Based on our findings we

will discuss the quality of existing evidence of HRQoL in patients with MM and provide recommendations for future clinical trial investigators.

Material and methods

Publication selection

We used the earlier identified corpus of publications reported in Nielsen*, et al* (2017) which is based on a systematic literature search with the primary objective to identify longitudinal HRQoL studies in MM patients. The literature search and publication selection are described in detail in Nielsen*, et al* (2017). In brief, publications were eligible if the following criteria were met; patients were diagnosed with MM, and the study applied a longitudinal study design using the European Organisation for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30) (Aaronson*, et al* 1993) instrument for HRQoL measurement of physical function, global QoL, fatigue and/or pain. Articles in languages other than English were excluded. There was no time limit set for the literature search. After the systematic literature search, separate publications with additional reporting of the HRQoL data from the ASPIRE trial and TOURMALINE-MM1 study were published (Leleu*, et al* 2018, Stewart*, et al* 2016). These two additional publications were included in the data extraction process for this review. When the PRO results from a clinical trial were presented in a separate publication, the first publication from the trial, including reporting of primary study endpoint, was identified and included in the data extraction process.

Data extraction

Information extracted from the publications was 1) whether the HRQoL data collection was a primary or secondary endpoint in the study or clinical trial, 2) scheduled timing of follow-up HRQoL assessment, 3) the statistical analysis method applied for between group differences and/or within group change estimation and the predefined statistical significance level, 4) a description of reasons for exclusion of patients from HRQoL analysis and 5) the statistical handling of NRs. Numbers extracted from the publications were 6) the number of patients included in the study at baseline, 7) the number of PRO assessments at baseline, 8) the number of participating patients from whom a completed PRO assessments were expected at each scheduled PRO assessment time point, 9) the number of completed PRO assessments at each scheduled PRO assessment time point, and 10) the number of PRO assessments at the last presented follow-up time point. In the case of the last published PRO follow-up time point was an end of study/treatment discontinuation assessment, the number of available questionnaires at the former time point was used.

The total intermittent NR rate for each study was estimated by calculating the proportion who did not complete scheduled PRO assessments of those from whom a completed PRO assessment were expected for each presented follow-up time point together. We calculated the magnitude of monotone NR by the proportion of NR at last presented follow-up time point compared to the number of patients enrolled in the study.

The statistical methods applied were divided into group A-D. A) Descriptive analyses, B) Non-parametric tests (Mann Whitney U-test, Wilcoxon signed rank test), C) Parametric tests, subdivided into C1) T-test, one-way ANOVA and C2) linear mixed model of repeated measures, generalized estimating equations, and D) Ordinal logistic regression, generalized mixed model with ordinal outcome.

The data extraction was done independently by three of the authors, LKN, MJ and TWK. Disagreements were discussed to achieve consensus.

Results

Twenty-three longitudinal HRQoL datasets were identified for data extraction. The PRISMA flow diagram for the study selection and subdivision of publications into five treatment categories are presented in Nielsen*, et al* (2017). The longitudinal PRO data from MM-015 studies were divided into two treatment categories; first line treatment without autologous stem cell transplantation and maintenance therapy, respectively, since the HRQoL data from those two treatments could be extracted separately from the studies (Dimopoulos*, et al* 2013, Dimopoulos*, et al* 2014). The included publications with references are presented in Table 1 with an additional reference to the publication reporting the primary study endpoint.

Endpoint and timing of PRO data collection

HRQoL was the secondary endpoint in 16 studies and the primary endpoint in seven studies. The most frequent time point for PRO assessment was at predefined *calendar time points* in ten studies, the second most common PRO assessment time point was at *day 1 of a new treatment cycle* in eight studies and at a predefined *clinical time point* in four studies.

Statistical analyses method and evaluation strategy applied

For all 23 studies, the statistical methods used for analyzing the longitudinal PRO data were described. The most frequently applied statistical method was C2) parametric statistical method of mixed model of repeated measures or generalized estimation equations of 11 studies, and the second most used was C1) parametric statistical methods of t-

test or one-way ANOVA in eight studies. Non-parametric statistics were used in six studies, ordinal logistic regression in three studies and descriptive statistics in two studies. Adjustment of statistical significance level to avoid multiplicity testing and type I error was performed in eight of the studies. The PRO data was evaluated by *between* group differences in nine studies and *within* group change in eight studies. Both strategies of between group differences and within group changes were used for evaluation of longitudinal PRO data in four studies, and in two studies, no strategy for evaluation was applied.

Magnitude and handling of complete non-responders

For 17 of the 23 longitudinal PRO study results, the number of PRO assessments at baseline was lower than the number of patients included in the clinical trial, leaving some patients as complete non-responders. In Table 1, the number of patients included in the clinical trial and number of PRO assessments at baseline are presented. The lowest proportion of all studied patients was included in the PRO analysis of the population-based PROFILES registry (Mols, *et al* 2012) of 51%, since the analyzed cohort was limited to the patients with a completed one year follow-up questionnaire. The second lowest studied cohort was the two randomized groups in the study of Gimsing, *et al* (2010) with 65% of the included patients in pamidronate 90 mg and 68% in the pamidronate 30 mg group. The analyzed cohort in that study was the patients who returned questionnaires at 12 month follow-up and who were still on study treatment. The third lowest proportion was the cohort of the SUMMIT study (Dubois, *et al* 2006) of 71%. Here the analysis cohort was limited to the patients with PRO information available and with a clinical response to bortezomib. For the remaining studies, the proportions of patients included in the PRO data analyses compared to the number of patients included in the clinical trial were between 81and 96%.

In 10 of the 17 publications, the selection strategy for the reduced number of baseline PRO assessments patients included in the PRO data analysis compared to the clinical trial was described. This description is presented in Table 1. The most frequently used strategy was to include only participants with a non-missing baseline questionnaire and minimum one follow-up questionnaire. In three publications, characteristics of participants and non-participants were compared. In the Table 1 of the paper by Gulbrandsen, *et al* (2001) the baseline characteristics of the participants and non-participants are presented. Wisloff, *et al* (1996a) found participants likely to be younger, female and to have longer survival than non-participants. In the PROFILES registry (Mols, *et al* 2012), the participants were diagnosed more recently and often treated with other regimens than chemotherapy only, compared to non-participants.

Magnitude of intermittent non-responses

In six of the studies, the number of completed PRO assessments together with the number of participating patients from whom a completed PRO assessment were expected at each scheduled PRO assessment time point was presented. The total number of participating patients and the total number of completed PRO assessments for the six studies are presented in Table 2. The lowest presented total intermittent NR rate was 2% in the control group of the study by Gulbrandsen, *et al* (2001) and the highest was 22% in the study by Waage, *et al* (2004).

Magnitude of monotone non-responses

In 16 studies, the number of completed questionnaires at last follow-up was presented and for all of them, the number was lower than at baseline. The number of questionnaires is presented in Table 1. The highest proportion of monotone NR was seen in the TOURMALINE study (Leleu, *et al* 2018) of 99% in the lenalidomide-dexamethasone group and 98% in the ixazomib-lenalidomide-dexamethasone group. The PRO data results in that study were collected at a specified cut-off date and some patients are still in follow-up. The second highest proportion of monotone NR was in the MM-003 trial (Song, *et al* 2015, Weisel, *et al* 2015) of 96% in the high dose dexamethasone group and 82% in the pomalidomide-dexamethasone group. The third highest proportion of monotone NR was in the APEX study (Lee, *et al* 2008), which was 88% in the bortezomib group and 83% in the dexamethasone group. For the remaining studies, the proportion of monotone NR were between 28% and 75%.

Statistical handling of intermittent and monotone non-responses

Statistical methods for handling intermittent and monotone NR or methods to investigate the impact of NRs were used in eight of the 23 studies. The methods applied are presented in Table 1. In the ASPIRE and TOURMALINE studies, a graphical approach was used to explore patterns of monotone NR (Leleu, *et al* 2018, Stewart, *et al* 2016). Multiple imputation was used to test the robustness of the PRO results in the study of Gimsing, *et al* (2010), Lee, *et al* (2008), Waage, *et al* (2010). Other methods of exploring the missing data mechanisms or robustness of PRO data results were by confirming the results by standardized area under the curve, mixed method of repeated measures or comparing mean scores for patients with available questionnaires at follow-up to patients with early study discontinuation (Delforge, *et al* 2015, Stewart, *et al* 2016, Waage, *et al* 2004).

Discussion

In this review, we have investigated methodological aspects of PRO data measurements and analyses in 23 published longitudinal PRO studies of patients with MM, identified in a previously published systematic review (Nielsen, *et al* 2017). We observed diversity in the timing of PRO data collection and statistical methods for analysing the longitudinal

PRO data among the studies. In 17/23 studies the number of PRO assessments at baseline was lower than the number of patients included in the clinical trial or study with proportions of complete non-responders being up to 51%. Reporting of intermittent NR rate was in general lacking and when reported being up to 22%. For studies where every patient had reached the date of cut-off for the analyses, we found proportions of monotone NR between 28% and 96%. Despite the high proportions of complete, intermittent and monotone NR, only in 8/23 studies investigation of the impact of NRs was performed.

Diversity in applied PRO methodologies and poor quality of PRO reporting from clinical trials has previously been reported (Brundage, *et al* 2011a, Bylicki, *et al* 2015, Dirven, *et al* 2014, Efficace, *et al* 2015, Efficace, *et al* 2014, Hamel, *et al* 2017, Lemieux, *et al* 2011). However, to our knowledge, this is the first review with the objective to investigate methodological aspects in PRO measurements and analyses in studies of patients with MM.

In 2013, the CONSORT PRO guideline for reporting PRO from randomized clinical trials was published to improve the accuracy and validity of PRO data reporting (Calvert, *et al* 2013). Recently, in 2018, a guideline for inclusion of PROs in clinical trial protocols became available (Calvert, *et al* 2018). This guideline, together with international standards for analysing PRO data, have been pointed out as supporting the increased application of PRO data results from clinical trials to daily clinical practice (Bottomley, *et al* 2016, Bottomley, *et al* 2018, Brundage, *et al* 2011b).

Clinical cancer trials are often designed with termination of PRO data collection if the patient drops out. Due to the nature of MM with risk of treatment failure, unacceptable adverse events to treatment and shortness of the patients´ life expectancy, PRO data collection in MM studies is at high risk of monotone NR. Also, patients with MM often experience disease complications and significant adverse events during treatment, particularly during HDT, which increases the risk of intermittent NR.

Missing PRO data can lead to a variety of problems, including loss of study power and precision (Bell and Fairclough 2014, Fairclough 2010). Participants who drop out early may have a poor HRQoL (Bell and Fairclough 2014, Mercieca-Bebber, *et al* 2017). Specific strategies to minimize missing PRO data should be implemented in the study design and data collection procedure (Calvert, *et al* 2018, Little, *et al* 2012, Mercieca-Bebber, *et al* 2016) and transparent reporting of the number of completed questionnaires at baseline and at subsequent time points are recommended (Calvert, *et al* 2013). NR to questionnaires might cause biased results if appropriate statistical handling guided by the missing data mechanisms is not performed (Bell and Fairclough 2014, Bell, *et al* 2013, Fairclough 2010). Handling of NR by simple imputation of "last observation carried forward" is not recommended for longitudinal data (Lavori, *et al* 2008). In our

review, the most frequently used statistical analysis method was linear mixed model of repeated measures or generalized estimation equations including multiple imputation, where NR are handled as MAR. In three studies, multiple imputation was used as the statistical analysis method, and the authors found that the MAR assumption was not the correct missing data mechanism in all cases.

When using descriptive statistics, non-parametric or parametric methods of t-test and one-way ANOVA, the missing data are handled as MCAR, which however, is rarely the case for the majority of missing PRO data (Bell and Fairclough 2014, Fairclough 2010). The assumption for MCAR missing data mechanism could not be confirmed in the included studies, where this aspect was investigated. In the two studies where a graphical approach was used, a slightly different pattern of change in global quality of life score over time was found for patients dropping out earlier compared to patients staying in the study for the whole period, but the differences were not statistically significant or clinically meaningful.

We observed variations in the timing of PRO data measurements among the included studies. The scheduled PRO data assessments time point should ensure capturing the patient experienced effect of the intervention aimed at PRO data collection. Using *day 1 of a new treatment cycle* for PRO data measurement has a clear advantage of reduced risk of missing response, since the patient can complete the questionnaire in hospital with assistance from a study coordinator. Most PRO data instruments for clinical research have a seven day recall period, and most anti-myeloma regimens are administrated in 21 or 28 days cycles with the last week being drug free. Also, a general principle is rescheduling the next treatment cycle, if the patient experiences severe toxicity or complication, such as admission with neutropenic fever, too much fatigue etc. Therefore, when PRO data collection is scheduled on *day 1 of a new treatment cycle*, the PRO data measurement during periods with complication is missed. In addition, the patients might have completed the questionnaires with reflection of a drug free week. This might lead to overestimation of HRQoL and underestimation of toxicities (Giesinger*, et al* 2014).

A limitation in our review is the restriction in selection criteria for the systematic review that comprised the publications using the EORTC QLQ-C30 instrument only. The EORTC instruments are traditionally used in European clinical trials, and applied PRO data research methods might be different in other parts of the world. An important aspect of the generalisability of PRO data results is that most PRO data deviates from patients included in clinical trials. This is also the case for patients with MM since 22 of the 23 studies in the systematic review are clinical trials (Nielsen*, et al* 2017). Newly diagnosed patients with MM included in clinical trials are not representative for the general MM population and PRO data from clinical trials might not be generally applicable (Klausen*, et al* 2018). HRQoL data from population-

based studies of patients with MM with high focus on minimization of NR are needed. In this review, we focus on how the EORTC QLQ-C30 has been used for HRQoL measurement in studies of patients with MM, since it is the most used HRQoL instrument in this population. Another relevant consideration is whether the EORTC QLQ-C30 questionnaire was the most suitable tool to measure what matters to patients with MM in each identified study. We did not review the identified studies and protocols to investigate, whether there was a specific research question and rationale for choosing the EORTC QLQ-C30 instrument to elucidate HRQoL in each study (Calvert, *et al* 2018).

In conclusion, we found diversity in the PRO data measurements and analyses applied in clinical studies of patients with MM and we observed a large fraction of NR. We found no transparent reporting of NR, and the missing data mechanisms were rarely investigated, which resulted in use of statistical methods of PRO data analysis based on untested assumptions. Based on the publications investigated, these findings suggest that the evidence-based knowledge of HRQoL in patients with MM is compromised by significant rates of complete, intermittent and monotone NR. This threatens the generalizability of PRO data results in MM and their application to daily clinical practice. In order to improve quality of PRO data and translation of PRO data results in patients with MM, we recommend PRO data investigators to follow the SPIRIT-PRO Extension Checklist during clinical trial protocol writing (Calvert, *et al* 2018) and the CONSORT PRO Extension Checklist Item when reporting PRO results from randomized trials (Calvert, *et al* 2013). Strategies to reduce NR that are suitable for the investigated cohort should be integrated in the study design, PRO data collection and procedures. Linear mixed models of repeated measures have been found to be the most suitable for analysing longitudinal PRO data and multiple imputation is considered the best method for sensitivity analyses, but these are not general recommendations (Hamel, *et al* 2017, Rombach, *et al* 2018). International standards for analysing PRO data from clinical trials are currently being developed (Bottomley, *et al* 2016). Therefore, being aware of potential pitfalls in PRO methodology when international standards are not available is important.

Author contributions

LKN and NA planned the study, performed literature search, data extraction, data analyses, interpretation and wrote the first manuscript draft. TWK and MJ contributed with data extraction, data analysis, and interpretation. All authors approved the submitted and final version.

References

Basch, E., Rogak, L.J. & Dueck, A.C. (2016) Methods for Implementing and Reporting Patient-reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. *Clinical therapeutics,* **38,** 821-830.

Becker, N. (2011) Epidemiology of multiple myeloma. *Recent Results Cancer Research,* **183,** 25-35.

Bell, M.L. & Fairclough, D.L. (2014) Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Statistical Methods in Medical Research,* **23,** 440-459.

Bell, M.L., Kenward, M.G., Fairclough, D.L. & Horton, N.J. (2013) Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *The BMJ,* **346,** e8668.

Benboubker, L., Dimopoulos, M.A., Dispenzieri, A., Catalano, J., Belch, A.R., Cavo, M., Pinto, A., Weisel, K., Ludwig, H., Bahlis, N., Banos, A., Tiab, M., Delforge, M., Cavenagh, J., Geraldes, C., Lee, J.J., Chen, C., Oriol, A., de la Rubia, J., Qiu, L., White, D.J., Binder, D., Anderson, K., Fermand, J.P., Moreau, P., Attal, M., Knight, R., Chen, G., Van Oostendorp, J., Jacques, C., Ervin-Haynes, A., Avet-Loiseau, H., Hulin, C. & Facon, T. (2014) Lenalidomide and dexamethasone in transplant-ineligible patients with myeloma. *The New England journal of medicine,* **371,** 906-917.

Blade, J., Calleja, M.A., Lahuerta, J.J., Poveda, J.L., de Paz, H.D. & Lizan, L. (2018) Defining a set of standardised outcome measures for newly diagnosed patients with multiple myeloma using the Delphi consensus method: the IMPORTA project. *The BMJ Open,* **8,** e018850.

Boland, E., Eiser, C., Ezaydi, Y., Greenfield, D.M., Ahmedzai, S.H. & Snowden, J.A. (2013) Living with advanced but stable multiple myeloma: a study of the symptom burden and cumulative effects of disease and intensive (hematopoietic stem cell transplant-based) treatment on health-related quality of life. *Journal of Pain Symptom Manage,* **46,** 671-680.

Bottomley, A., Efficace, F. & Fayers, P.M. (2002) Standards are needed for quality of life clinical trials. *The BMJ,* **324,** 1156.

Bottomley, A., Pe, M., Sloan, J., Basch, E., Bonnetain, F., Calvert, M., Campbell, A., Cleeland, C., Cocks, K., Collette, L., Dueck, A.C., Devlin, N., Flechtner, H.H., Gotay, C., Greimel, E., Griebsch, I., Groenvold, M., Hamel, J.F., King, M., Kluetz, P.G., Koller, M., Malone, D.C., Martinelli, F., Mitchell, S.A., Moinpour, C.M., Musoro, J., O'Connor, D., Oliver, K., Piault-Louis, E., Piccart, M., Pimentel, F.L., Quinten, C., Reijneveld, J.C., Schurmann, C., Smith, A.W., Soltys, K.M., Taphoorn, M.J., Velikova, G. & Coens, C. (2016) Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *The Lancet Oncology,* **17,** e510-e514.

Bottomley, A., Pe, M., Sloan, J., Basch, E., Bonnetain, F., Calvert, M., Campbell, A., Cleeland, C., Cocks, K., Collette, L., Dueck, A.C., Devlin, N., Flechtner, H.H., Gotay, C., Greimel, E., Griebsch, I., Groenvold, M., Hamel, J.F., King, M., Kluetz, P.G., Koller, M., Malone, D.C., Martinelli, F., Mitchell, S.A., Moinpour, C.M., Musoro, J.Z., O'Connor, D., Oliver, K., Piault-Louis, E., Piccart, M., Pimentel, F.L., Quinten, C., Reijneveld, J.C., Schurmann, C., Smith, A.W., Soltys, K.M., Sridhara, R., Taphoorn, M.J.B., Velikova, G. & Coens, C. (2018) Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. *Clinical trials*, 1740774518795637.

Brundage, M., Bass, B., Davidson, J., Queenan, J., Bezjak, A., Ringash, J., Wilkinson, A. & Feldman-Stewart, D. (2011a) Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. *Quality of life research Journal,* **20,** 653-664.

Brundage, M., Bass, B., Jolie, R. & Foley, K. (2011b) A knowledge translation challenge: clinical use of quality of life data from cancer clinical trials. *Quality of life research Journal,* **20,** 979-985.

Bylicki, O., Gan, H.K., Joly, F., Maillet, D., You, B. & Peron, J. (2015) Poor patient-reported outcomes reporting according to CONSORT guidelines in randomized clinical trials evaluating systemic cancer therapy. *Annals of Oncology,* **26,** 231-237.

Calvert, M., Blazeby, J., Altman, D.G., Revicki, D.A., Moher, D. & Brundage, M.D. (2013) Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *The Journal of the American Medical Association,* **309,** 814-822.

Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A.W., King, M.T., Hunn, A., Bottomley, A., Regnault, A., Chan, A.W., Ells, C., O'Connor, D., Revicki, D., Patrick, D., Altman, D., Basch, E., Velikova, G., Price, G., Draper, H., Blazeby, J., Scott, J., Coast, J., Norquist, J., Brown, J., Haywood, K., Johnson, L.L., Campbell, L., Frank, L., von Hildebrand, M., Brundage, M., Palmer, M., Kluetz, P., Stephens, R., Golub, R.M., Mitchell, S. & Groves, T. (2018) Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *The Journal of the American Medical Association,* **319,** 483-494.

Deber, R.B., Kraetschmer, N., Urowitz, S. & Sharpe, N. (2007) Do people want to be autonomous patients? Preferred roles in treatment decision-making in several patient populations. *Health Expect,* **10,** 248-258.

Delforge, M., Dhawan, R., Robinson, D., Jr., Meunier, J., Regnault, A., Esseltine, D.L., Cakana, A., van de Velde, H., Richardson, P.G. & San Miguel, J.F. (2012) Health-related quality of life in elderly, newly diagnosed multiple myeloma patients treated with VMP vs. MP: results from the VISTA trial. *European journal of haematology,* **89,** 16-27.

Delforge, M., Minuk, L., Eisenmann, J.C., Arnulf, B., Canepa, L., Fragasso, A., Leyvraz, S., Langer, C., Ezaydi, Y., Vogl, D.T., Giraldo-Castellano, P., Yoon, S.S., Zarnitsky, C., Escoffre-Barbe, M., Lemieux, B., Song, K., Bahlis, N.J., Guo, S., Monzini, M.S., Ervin-Haynes, A., Houck, V. & Facon, T. (2015) Health-related quality-of-life in patients with newly diagnosed multiple myeloma in the FIRST trial: lenalidomide plus low-dose dexamethasone versus melphalan, prednisone, thalidomide. *Haematologica,* **100,** 826-833.

Dimopoulos, M.A., Delforge, M., Hajek, R., Kropff, M., Petrucci, M.T., Lewis, P., Nixon, A., Zhang, J., Mei, J. & Palumbo, A. (2013) Lenalidomide, melphalan, and prednisone, followed by lenalidomide maintenance, improves health-related quality of life in newly diagnosed multiple myeloma patients aged 65 years or older: results of a randomized phase III trial. *Haematologica,* **98,** 784-788.

Dimopoulos, M.A., Palumbo, A., Hajek, R., Kropff, M., Petrucci, M.T., Lewis, P., Millar, S., Zhang, J., Mei, J. & Delforge, M. (2014) Factors that influence health-related quality of life in newly diagnosed patients with multiple myeloma aged > 65 years treated with melphalan, prednisone and lenalidomide followed by lenalidomide maintenance: Results of a randomized trial. *Leukemia and Lymphoma,* **55,** 1489-1497.

Dirven, L., Taphoorn, M.J., Reijneveld, J.C., Blazeby, J., Jacobs, M., Pusic, A., La Sala, E., Stupp, R., Fayers, P. & Efficace, F. (2014) The level of patient-reported outcome reporting in randomised controlled trials of brain tumour patients: a systematic review. *European Journal of Cancer,* **50,** 2432-2448.

Dubois, D., Dhawan, R., van de Velde, H., Esseltine, D., Gupta, S., Viala, M. & de la Loge, C. (2006) Descriptive and prognostic value of patient-reported outcomes: the bortezomib experience in relapsed and refractory multiple myeloma. *Journal of Clinical Oncology,* **24,** 976-982.

Efficace, F., Fayers, P., Pusic, A., Cemal, Y., Yanagawa, J., Jacobs, M., la Sala, A., Cafaro, V., Whale, K., Rees, J. & Blazeby, J. (2015) Quality of patient-reported outcome reporting across cancer randomized controlled trials according to the CONSORT patient-reported outcome extension: A pooled analysis of 557 trials. *Cancer,* **121,** 3335-3342.

Efficace, F., Feuerstein, M., Fayers, P., Cafaro, V., Eastham, J., Pusic, A. & Blazeby, J. (2014) Patient-reported outcomes in randomised controlled trials of prostate cancer: methodological quality and impact on clinical decision making. *European Urology,* **66,** 416-427.

Efficace, F., Gaidano, G. & Lo-Coco, F. (2017) Patient-reported outcomes in hematology: is it time to focus more on them in clinical trials and hematology practice? *Blood,* **130,** 859-866.

EHA (2013) Guidelines Patient-Reported Outcomes in Hematology. Scientific Working Group Quality of Life and Symptoms European Hematology Association, https://ehaweb.org/organization/theme-of-the-year/2012-2013-quality-of-life-in-hematology/.

EMA (2016) The use of patientreported outcome (PRO) measures in oncology studies. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. European Medicine Agency., Vol. 2017, http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf.

Etto, L.Y., Morelli, V.M., Silva, V.C., Hungria, V.T., Ciconelli, R.M., Almeida, M.S., Oliveira, J.S., Barros, J.C., Durie, B.G. & Colleoni, G.W. (2011) Autologous stem cell transplantation improves quality of life in economically challenged, Brazilian multiple myeloma patients. *Clinics (Sao Paulo),* **66,** 1855-1859.

Fairclough, D.L. (2010) *Design and analysis of quality of life studies in clinical trials.* Chapman & Hall/CRC, Boca Raton.

Fielding, S., Fayers, P.M. & Ramsay, C.R. (2009) Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. *Health and quality of life outcomes,* **7,** 57.

Fielding, S., Ogbuagu, A., Sivasubramaniam, S., MacLennan, G. & Ramsay, C.R. (2016) Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Quality of life research Journal,* **25,** 2977-2983.

Frodin, U., Borjeson, S., Lyth, J. & Lotfi, K. (2011) A prospective evaluation of patients' health-related quality of life during auto-SCT: a 3-year follow-up. *Bone marrow transplantation,* **46,** 1345-1352.

Giesinger, J.M., Wintner, L.M., Zabernigg, A., Gamper, E.M., Oberguggenberger, A.S., Sztankay, M.J., Kemmler, G. & Holzner, B. (2014) Assessing quality of life on the day of chemotherapy administration underestimates patients' true symptom burden. *BMC cancer,* **14,** 758.

Gimsing, P., Carlson, K., Turesson, I., Fayers, P., Waage, A., Vangsted, A., Mylin, A., Gluud, C., Juliusson, G., Gregersen, H., Hjorth-Hansen, H., Nesthus, I., Dahl, I.M.S., Westin, J., Nielsen, J.L., Knudsen, L.M., Ahlberg, L., Hjorth, M., Abildgaard, N., Andersen, N.F., Linder, O. & Wisloff, F. (2010) Effect of pamidronate 30 mg versus 90 mg on physical function in patients with newly diagnosed multiple myeloma (Nordic Myeloma Study Group): A double-blind, randomised controlled trial. *The Lancet Oncology,* **11,** 973-982.

Gulbrandsen, N., Wisloff, F., Brinch, L., Carlson, K., Dahl, I.M., Gimsing, P., Hippe, E., Hjorth, M., Knudsen, L.M., Lamvik, J., Lenhoff, S., Lofvenberg, E., Nesthus, I., Nielsen, J.L., Turesson, I., Westin, J. & Nordic Myeloma Study, G. (2001) Health-related quality of life in multiple myeloma patients receiving high-dose chemotherapy with autologous blood stem-cell support. *Medical oncology (Northwood, London, England),* **18,** 65-77.

Hamel, J.F., Saulnier, P., Pe, M., Zikos, E., Musoro, J., Coens, C. & Bottomley, A. (2017) A systematic review of the quality of statistical methods employed for analysing quality of life data in cancer randomised controlled trials. *European Journal of Cancer,* **83,** 166-176.

Hjorth, M., Hjertner, O., Knudsen, L.M., Gulbrandsen, N., Holmberg, E., Pedersen, P.T., Andersen, N.F., Andreasson, B., Billstrom, R., Carlson, K., Carlsson, M.S., Flogegard, M., Forsberg, K., Gimsing, P., Karlsson, T., Linder, O., Nahi, H., Othzen, A., Swedin, A. & Nordic Myeloma Study, G. (2012) Thalidomide and dexamethasone vs. bortezomib and dexamethasone for melphalan refractory myeloma: a randomized study. *European journal of haematology,* **88,** 485-496.

Khalafallah, A., McDonnell, K., Dawar, H.U., Robertson, I. & Woods, D. (2011) Quality of life assessment in multiple myeloma patients undergoing dose-reduced tandem autologous stem cell transplantation. *Mediterranean Journal Hematology and Infectious Diseases,* **3,** e2011057.

Klausen, T.W., Gregersen, H., Abildgaard, N., Andersen, N.F., Frolund, U.C., Gimsing, P., Helleberg, C. & Vangsted, A.J. (2018) The majority of newly diagnosed myeloma patients do not fulfill the inclusion criteria in clinical phase III trials. *Leukemia.*

Kumar, S.K., Dispenzieri, A., Lacy, M.Q., Gertz, M.A., Buadi, F.K., Pandey, S., Kapoor, P., Dingli, D., Hayman, S.R., Leung, N., Lust, J., McCurdy, A., Russell, S.J., Zeldenrust, S.R., Kyle, R.A. & Rajkumar, S.V. (2014) Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia,* **28,** 1122-1128.

Kumar, S.K., Rajkumar, S.V., Dispenzieri, A., Lacy, M.Q., Hayman, S.R., Buadi, F.K., Zeldenrust, S.R., Dingli, D., Russell, S.J., Lust, J.A., Greipp, P.R., Kyle, R.A. & Gertz, M.A. (2008) Improved survival in multiple myeloma and the impact of novel therapies. *Blood,* **111,** 2516-2520.

Kvam, A.K., Fayers, P., Hjermstad, M., Gulbrandsen, N. & Wisloff, F. (2009) Health-related quality of life assessment in randomised controlled trials in multiple myeloma: a critical review of methodology and impact on treatment recommendations. *European journal of haematology,* **83,** 279-289.

Kyle, R.A. & Rajkumar, S.V. (2008) Multiple myeloma. *Blood,* **111,** 2962-2972.

Kyte, D., Duffy, H., Fletcher, B., Gheorghe, A., Mercieca-Bebber, R., King, M., Draper, H., Ives, J., Brundage, M., Blazeby, J. & Calvert, M. (2014) Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. *PLoS One,* **9,** e110229.

Laubach, J., Garderet, L., Mahindra, A., Gahrton, G., Caers, J., Sezer, O., Voorhees, P., Leleu, X., Johnsen, H.E., Streetly, M., Jurczyszyn, A., Ludwig, H., Mellqvist, U.H., Chng, W.J., Pilarski, L., Einsele, H., Hou, J., Turesson, I., Zamagni, E., Chim, C.S., Mazumder, A., Westin, J., Lu, J., Reiman, T., Kristinsson, S., Joshua, D., Roussel, M., O'Gorman, P., Terpos, E., McCarthy, P., Dimopoulos, M., Moreau, P., Orlowski, R.Z., Miguel, J.S., Anderson, K.C., Palumbo, A., Kumar, S., Rajkumar, V., Durie, B. & Richardson, P.G. (2016) Management of relapsed multiple myeloma: recommendations of the International Myeloma Working Group. *Leukemia,* **30,** 1005-1017.

Lavori, P.W., Brown, C.H., Duan, N., Gibbons, R.D. & Greenhouse, J. (2008) Missing Data in Longitudinal Clinical Trials Part A: Design and Conceptual Issues. *Psychiatric Annnals,* **38,** 784-792.

Lee, C.W. & Chi, K.N. (2000) The standard of reporting of health-related quality of life in clinical cancer trials. *Journal of clinical epidemiology,* **53,** 451-458.

Lee, S.J., Richardson, P.G., Sonneveld, P., Schuster, M.W., Irwin, D., San Miguel, J.F., Crawford, B., Massaro, J., Dhawan, R., Gupta, S. & Anderson, K.C. (2008) 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. *British journal of haematology,* **143,** 511-519.

Leleu, X., Masszi, T., Bahlis, N.J., Viterbo, L., Baker, B., Gimsing, P., Maisnar, V., Samoilova, O., Rosinol, L., Langer, C., Song, K., Izumi, T., Cleeland, C., Berg, D., Lin, H.M., Zhu, Y., Skacel, T., Moreau, P. & Richardson, P.G. (2018) 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*.

Leleu, X., Mateos, M., Delforge, M., Lewis, P., Schindler, T., Bibson, C., Yang, M. & Weisel, K. (2015) Assessment of Multiple Myeloma Patient Preferences on Treatment Choices: An International Discrete Choice Study. *Blood,* **126,** 2086.

Lemieux, J., Goodwin, P.J., Bordeleau, L.J., Lauzier, S. & Theberge, V. (2011) Quality-of-life measurement in randomized clinical trials in breast cancer: an updated systematic review (2001-2009). *Journal of the National Cancer Institute,* **103,** 178-231.

Lenhoff, S., Hjorth, M., Holmberg, E., Turesson, I., Westin, J., Nielsen, J.L., Wisloff, F., Brinch, L., Carlson, K., Carlsson, M., Dahl, I.M., Gimsing, P., Hippe, E., Johnsen, H.E., Lamvik, J., Lofvenberg, E., Nesthus, I. & Rodjer, S. (2000) Impact on survival of high-dose therapy with autologous stem cell support in patients younger than 60 years with newly diagnosed multiple myeloma: a population-based study. Nordic Myeloma Study Group. *Blood,* **95,** 7-11.

Little, R.J., D'Agostino, R., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., Neaton, J.D., Rotnitzky, A., Scharfstein, D., Shih, W.J., Siegel, J.P. & Stern, H. (2012) The prevention and treatment of missing data in clinical trials. *The New England journal of medicine,* **367,** 1355-1360.

Ludwig, H., Viterbo, L., Greil, R., Masszi, T., Spicka, I., Shpilberg, O., Hajek, R., Dmoszynska, A., Paiva, B., Vidriales, M.B., Esteves, G., Stoppa, A.M., Robinson, D., Jr., Ricci, D., Cakana, A., Enny, C., Feng, H., van de Velde, H. & Harousseau, J.L. (2013) Randomized phase II study of bortezomib, thalidomide, and dexamethasone with or without cyclophosphamide as induction therapy in previously untreated multiple myeloma. *Journal of Clinical Oncology,* **31,** 247-255.

Mateos, M.V. (2010) Management of treatment-related adverse events in patients with multiple myeloma. *Cancer Treatment Reviews,* **36,** S24-S32.

Mateos, M.V., Leleu, X., Palumbo, A. & Miguel, J.F.S. (2014) Initial treatment of transplant-ineligible patients in multiple myeloma. *Expert review of hematology,* **7,** 67-77.

Mateos, M.V., Richardson, P.G., Schlag, R., Khuageva, N.K., Dimopoulos, M.A., Shpilberg, O., Kropff, M., Spicka, I., Petrucci, M.T., Palumbo, A., Samoilova, O.S., Dmoszynska, A., Abdulkadyrov, K.M., Schots, R., Jiang, B., Esseltine, D.L., Liu, K., Cakana, A., van de Velde, H. & San Miguel, J.F. (2010) Bortezomib plus melphalan and prednisone compared with melphalan and prednisone in previously untreated multiple myeloma: updated follow-up and impact of subsequent therapy in the phase III VISTA trial. *Journal of Clinical Oncology,* **28,** 2259-2266.

McCarthy, P.L., Holstein, S.A., Petrucci, M.T., Richardson, P.G., Hulin, C., Tosi, P., Bringhen, S., Musto, P., Anderson, K.C., Caillot, D., Gay, F., Moreau, P., Marit, G., Jung, S.H., Yu, Z., Winograd, B., Knight, R.D., Palumbo, A. & Attal, M. (2017) Lenalidomide Maintenance After Autologous Stem-Cell Transplantation in Newly Diagnosed Multiple Myeloma: A Meta-Analysis. *Journal of Clinical Oncology,* **35,** 3279-3289.

Mellqvist, U.H., Gimsing, P., Hjertner, O., Lenhoff, S., Laane, E., Remes, K., Steingrimsdottir, H., Abildgaard, N., Ahlberg, L., Blimark, C., Dahl, I.M., Forsberg, K., Gedde-Dahl, T., Gregersen, H., Gruber, A., Guldbrandsen, N., Haukas, E., Carlson, K., Kvam, A.K., Nahi, H., Lindas, R., Andersen, N.F., Turesson, I., Waage, A., Westin, J. & Nordic Myeloma Study, G. (2013) Bortezomib consolidation after autologous stem cell transplantation in multiple myeloma: a Nordic Myeloma Study Group randomized phase 3 trial. *Blood,* **121,** 4647-4654.

Mercieca-Bebber, R., Calvert, M., Kyte, D., Stockler, M. & King, M.T. (2018) The administration of patient-reported outcome questionnaires in cancer trials: Interviews with trial coordinators regarding their roles, experiences, challenges and training. *Contemporary Clinical Trials Communications,* **9,** 23-32.

Mercieca-Bebber, R., Palmer, M.J., Brundage, M., Calvert, M., Stockler, M.R. & King, M.T. (2016) Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *The BMJ Open,* **6,** e010938.

Mercieca-Bebber, R.L., Price, M.A., Bell, M.L., King, M.T., Webb, P.M. & Butow, P.N. (2017) Ovarian cancer study dropouts had worse health-related quality of life and psychosocial symptoms at baseline and over time. *Asia-Pacific Journal of Clinical Oncology,* **13,** e381-e388.

Miguel, J.S., Weisel, K., Moreau, P., Lacy, M., Song, K., Delforge, M., Karlin, L., Goldschmidt, H., Banos, A., Oriol, A., Alegre, A., Chen, C., Cavo, M., Garderet, L., Ivanova, V., Martinez-Lopez, J., Belch, A., Palumbo, A., Schey, S., Sonneveld, P., Yu, X., Sternas, L., Jacques, C., Zaki, M. & Dimopoulos, M. (2013) 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. *The Lancet Oncology,* **14,** 1055-1066.

Mikhael, J.R., Dingli, D., Roy, V., Reeder, C.B., Buadi, F.K., Hayman, S.R., Dispenzieri, A., Fonseca, R., Sher, T., Kyle, R.A., Lin, Y., Russell, S.J., Kumar, S., Bergsagel, P.L., Zeldenrust, S.R., Leung, N., Drake, M.T., Kapoor, P., Ansell, S.M., Witzig, T.E., Lust, J.A., Dalton, R.J., Gertz, M.A., Stewart, A.K., Rajkumar, S.V., Chanan-Khan, A. & Lacy, M.Q. (2013) Management of newly diagnosed symptomatic multiple myeloma: updated Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART) consensus guidelines 2013. *Mayo Clinic proceedings,* **88,** 360-376.

Molassiotis, A., Wilson, B., Blair, S., Howe, T. & Cavet, J. (2011) Unmet supportive care needs, psychological well-being and quality of life in patients living with multiple myeloma and their partners. *Psycho-oncology,* **20,** 88-97.

Mols, F., Oerlemans, S., Vos, A.H., Koster, A., Verelst, S., Sonneveld, P. & van de Poll-Franse, L.V. (2012) Health-related quality of life and disease-specific complaints among multiple myeloma patients up to 10 yr after diagnosis: results from a population-based study using the PROFILES registry. *European journal of haematology,* **89,** 311-319.

Moreau, P., Masszi, T., Grzasko, N., Bahlis, N.J., Hansson, M., Pour, L., Sandhu, I., Ganly, P., Baker, B.W., Jackson, S.R., Stoppa, A.M., Simpson, D.R., Gimsing, P., Palumbo, A., Garderet, L., Cavo, M., Kumar, S., Touzeau, C., Buadi, F.K., Laubach, J.P., Berg, D.T., Lin, J., Di Bacco, A., Hui, A.M., van de Velde, H. & Richardson, P.G. (2016) Oral Ixazomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. *The New England journal of medicine,* **374,** 1621-1634.

Nielsen, L.K., Jarden, M., Andersen, C.L., Frederiksen, H. & Abildgaard, N. (2017) A systematic review of health-related quality of life in longitudinal studies of myeloma patients. *European journal of haematology,* **99,** 3-17.

NMSG (1996) Interferon-alpha 2b added to melphalan-prednisone for initial and maintenance therapy in multiple myeloma. A randomized, controlled trial. The Nordic Myeloma Study Group. *Annals of Internal Medicine,* **124,** 212-222.

Palmer, M.J., Mercieca-Bebber, R., King, M., Calvert, M., Richardson, H. & Brundage, M. (2018) A systematic review and development of a classification framework for factors associated with missing patient-reported outcome data. *Clinical trials,* **15,** 95-106.

Palumbo, A., Hajek, R., Delforge, M., Kropff, M., Petrucci, M.T., Catalano, J., Gisslinger, H., Wiktor-Jedrzejczak, W., Zodelava, M., Weisel, K., Cascavilla, N., Iosava, G., Cavo, M., Kloczko, J., Bladé, J., Beksac, M., Spicka, I., Plesner, T., Radke, J., Langer, C., Ben Yehuda, D., Corso, A., Herbein, L., Yu, Z., Mei, J., Jacques, C. & Dimopoulos, M.A. (2012) Continuous lenalidomide treatment for newly diagnosed multiple myeloma. *The New England journal of medicine,* **366,** 1759.

Rajkumar, S.V., Dimopoulos, M.A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.V., Kumar, S., Hillengass, J., Kastritis, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., Sezer, O., Kristinsson, S.Y., Caers, J., Usmani, S.Z., Lahuerta, J.J., Johnsen, H.E., Beksac, M., Cavo, M., Goldschmidt, H., Terpos, E., Kyle, R.A., Anderson, K.C., Durie, B.G. & Miguel, J.F. (2014) International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet Oncology,* **15,** e538-548.

Richardson, P.G., Barlogie, B., Berenson, J., Singhal, S., Jagannath, S., Irwin, D., Rajkumar, S.V., Srkalovic, G., Alsina, M., Alexanian, R., Siegel, D., Orlowski, R.Z., Kuter, D., Limentani, S.A., Lee, S., Hideshima, T., Esseltine, D.L., Kauffman, M., Adams, J., Schenkein, D.P. & Anderson, K.C. (2003) A phase 2 study of Bortezomib in relapsed, refractory myeloma. *New England Journal of Medicine,* **348,** 2609-2617.

Richardson, P.G., Delforge, M., Beksac, M., Wen, P., Jongen, J.L., Sezer, O., Terpos, E., Munshi, N., Palumbo, A., Rajkumar, S.V., Harousseau, J.L., Moreau, P., Avet-Loiseau, H., Lee, J.H., Cavo, M., Merlini, G., Voorhees, P., Chng, W.J., Mazumder, A., Usmani, S., Einsele, H., Comenzo, R., Orlowski, R., Vesole, D., Lahuerta, J.J., Niesvizky, R., Siegel, D., Mateos, M.V., Dimopoulos, M., Lonial, S., Jagannath, S., Blade, J., Miguel, J.S., Morgan, G., Anderson, K.C., Durie, B.G. & Sonneveld, P. (2012) Management of treatment-emergent peripheral neuropathy in multiple myeloma. *Leukemia,* **26,** 595-608.

Richardson, P.G., Laubach, J., Gandolfi, S., Facon, T., Weisel, K. & O'Gorman, P. (2018) Maintenance and continuous therapy for multiple myeloma. *Expert Review Anticancer Therapy,* **18,** 751-764.

Richardson, P.G., Sonneveld, P., Schuster, M.W., Irwin, D., Stadtmauer, E.A., Facon, T., Harousseau, J.L., Ben-Yehuda, D., Lonial, S., Goldschmidt, H., Reece, D., San-Miguel, J.F., Blade, J., Boccadoro, M., Cavenagh, J., Dalton, W.S., Boral, A.L., Esseltine, D.L., Porter, J.B., Schenkein, D. & Anderson, K.C. (2005) Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. *The New England journal of medicine,* **352,** 2487-2498.

Rombach, I., Gray, A.M., Jenkinson, C., Murray, D.W. & Rivero-Arias, O. (2018) Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. *BMC Medical Research Methodology,* **18,** 87.

Rubin, D.B. (1976) Inference and Missing Data. *Biometrika,* **63,** 581-592.

Sirohi, B., Powles, R., Lawrence, D., Treleaven, J., Kulkarni, S., Leary, A., Rudin, C., Horton, C. & Morgan, G. (2007) An open, randomized, controlled, phase II, single centre, two-period cross-over study to compare the quality of life and toxicity experienced on PEG interferon with interferon-alpha2b in

patients with multiple myeloma maintained on a steady dose of interferon-alpha2b. *Annals of Oncology,* **18,** 1388-1394.

Song, K.W., Dimopoulos, M.A., Weisel, K.C., Moreau, P., Palumbo, A., Belch, A., Schey, S., Sonneveld, P., Sternas, L., Yu, X., Amatya, R., Monzini, M.S., Zaki, M., Jacques, C. & San Miguel, J. (2015) 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,* **100,** e63-67.

Sonneveld, P., Verelst, S.G., Lewis, P., Gray-Schopfer, V., Hutchings, A., Nixon, A. & Petrucci, M.T. (2013) Review of health-related quality of life data in multiple myeloma patients treated with novel agents. *Leukemia,* **27,** 1959-1969.

Speight, J. & Barendse, S.M. (2010) FDA guidance on patient reported outcomes. *The BMJ,* **340,** c2921.

Sprangers, M.A. (2010) Disregarding clinical trial-based patient-reported outcomes is unwarranted: Five advances to substantiate the scientific stringency of quality-of-life measurement. *Acta Oncologica,* **49,** 155-163.

Stewart, A.K., Dimopoulos, M.A., Masszi, T., Spicka, I., Oriol, A., Hajek, R., Rosinol, L., Siegel, D.S., Niesvizky, R., Jakubowiak, A.J., San-Miguel, J.F., Ludwig, H., Buchanan, J., Cocks, K., Yang, X., Xing, B., Zojwalla, N., Tonda, M., Moreau, P. & Palumbo, A. (2016) Health-Related Quality of Life Results From the Open-Label, Randomized, Phase III ASPIRE Trial Evaluating Carfilzomib, Lenalidomide, and Dexamethasone Versus Lenalidomide and Dexamethasone in Patients With Relapsed Multiple Myeloma. *Journal of Clinical Oncology*.

Stewart, A.K., Jacobus, S., Fonseca, R., Weiss, M., Callander, N.S., Chanan-Khan, A.A. & Rajkumar, S.V. (2015a) Melphalan, prednisone, and thalidomide vs melphalan, prednisone, and lenalidomide (ECOG E1A06) in untreated multiple myeloma. *Blood,* **126,** 1294-1301.

Stewart, A.K., Rajkumar, S.V., Dimopoulos, M.A., Masszi, T., Spicka, I., Oriol, A., Hajek, R., Rosinol, L., Siegel, D.S., Mihaylov, G.G., Goranova-Marinova, V., Rajnics, P., Suvorov, A., Niesvizky, R., Jakubowiak, A.J., San-Miguel, J.F., Ludwig, H., Wang, M., Maisnar, V., Minarik, J., Bensinger, W.I., Mateos, M.V., Ben-Yehuda, D., Kukreti, V., Zojwalla, N., Tonda, M.E., Yang, X., Xing, B., Moreau, P., Palumbo, A. & Investigators, A. (2015b) Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. *The New England journal of medicine,* **372,** 142-152.

Tariman, J.D., Doorenbos, A., Schepp, K.G., Becker, P.S. & Berry, D.L. (2014a) Patient, Physician and Contextual Factors Are Influential in the Treatment Decision Making of Older Adults Newly Diagnosed with Symptomatic Myeloma. *Cancer Treatment Communication,* **2,** 34-47.

Tariman, J.D., Doorenbos, A., Schepp, K.G., Singhal, S. & Berry, D.L. (2014b) Older adults newly diagnosed with symptomatic myeloma and treatment decision making. *Oncology Nursing Forum,* **41,** 411-419.

Verelst, S.G., Termorshuizen, F., Uyl-de Groot, C.A., Schaafsma, M.R., Ammerlaan, A.H., Wittebol, S., Sinnige, H.A., Zweegman, S., van Marwijk Kooy, M., van der Griend, R., Lokhorst, H.M., Sonneveld, P. & Wijermans, P.W. (2011) Effect of thalidomide with melphalan and prednisone on health-related quality of life (HRQoL) in elderly patients with newly diagnosed multiple myeloma: a prospective analysis in a randomized trial. *Annals of Hematology,* **90,** 1427-1439.

Vodicka, E., Kim, K., Devine, E.B., Gnanasakthy, A., Scoggins, J.F. & Patrick, D.L. (2015) Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). *Contemporary clinical trials,* **43,** 1-9.

Weisel, K., Dimopoulos , M., Song , K.W., Moreau , P., Palumbo , A., Belch , A., Schey , S., Sonneveld , P., Sternas , L. & Yu , X. (2015) 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. *Clinical Lymphoma, Myeloma and Leukemia,* **15,** 519.

Wijermans, P., Schaafsma, M., Termorshuizen, F., Ammerlaan, R., Wittebol, S., Sinnige, H., Zweegman, S., van Marwijk Kooy, M., van der Griend, R., Lokhorst, H., Sonneveld, P. & Dutch-Belgium Cooperative Group, H. (2010) Phase III study of the value of thalidomide added to melphalan plus prednisone in

elderly patients with newly diagnosed multiple myeloma: the HOVON 49 Study. *Journal of Clinical Oncology,* **28,** 3160-3166.

Wisloff, F., Eika, S., Hippe, E., Hjorth, M., Holmberg, E., Kaasa, S., Palva, I. & Westin, J. (1996a) Measurement of health-related quality of life in multiple myeloma. Nordic Myeloma Study Group. *British journal of haematology,* **92,** 604-613.

Wisloff, F., Hjorth, M., Kaasa, S. & Westin, J. (1996b) Effect of interferon on the health-related quality of life of multiple myeloma patients: results of a Nordic randomized trial comparing melphalan-prednisone to melphalan-prednisone + alpha-interferon. The Nordic Myeloma Study Group. *British journal of haematology,* **94,** 324-332.

Waage, A., Gimsing, P., Fayers, P., Abildgaard, N., Ahlberg, L., Bjorkstrand, B., Carlson, K., Dahl, I.M., Forsberg, K., Gulbrandsen, N., Haukas, E., Hjertner, O., Hjorth, M., Karlsson, T., Knudsen, L.M., Nielsen, J.L., Linder, O., Mellqvist, U.H., Nesthus, I., Rolke, J., Strandberg, M., Sorbo, J.H., Wisloff, F., Juliusson, G., Turesson, I. & Nordic Myeloma Study, G. (2010) Melphalan and prednisone plus thalidomide or placebo in elderly patients with multiple myeloma. *Blood,* **116,** 1405-1412.

Waage, A., Gimsing, P., Juliusson, G., Turesson, I., Gulbrandsen, N., Eriksson, T., Hjorth, M., Nielsen, J.L., Lenhoff, S., Westin, J., Wisloff, F. & Nordic Myeloma Study, G. (2004) Early response predicts thalidomide efficiency in patients with advanced multiple myeloma. *British journal of haematology,* **125,** 149-155.

Aaronson, N.K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N.J., Filiberti, A., Flechtner, H., Fleishman, S.B. & de Haes, J.C. (1993) The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute,* **85,** 365-376.

Table 1. Methodological and statistical aspects of measurement, analyses and interpretation of PRO data extracted from publications

Trial name, study design and references	PRO data measurement		PRO data statistical analyses and handling of non-responses				
	Endpoint	Follow-up measurement time points	Statistical evaluation strategy and statistical significance level	Number of patients at baseline	Number of PRO assessments at baseline	Number of PRO assessments at last presented follow-up[1]	Handling of complete, intermittent and monotone non-responders
First-line treatment studies including induction therapy and ASCT							
Randomized phase II study (Ludwig, *et al* 2013)	Secondary	Day 1 of treatment cycles	Mixed model repeated measures for *between* group differences.	49 and 49	46 and 46	29 and 26	Last observation carried forward
Phase II study (Etto, *et al* 2011)	Primary	Clinical time points	Oneway ANOVA, t-test (p<0.05) for *within* group change.	29	29	14[2]	Not described
Evaluation, phase II study (Gulbrandsen, *et al* 2001, Lenhoff, *et al* 2000)	Secondary	Calendar time points	Mann Whitney U-test and Wilcoxon signed rank test (p<0.01) for *between* group differences.	274 and 120	221 and 113	72 and 38	Participants and non-participants comparison.
First-line treatment studies without ASCT							
FIRST trial, randomized phase III study (Benboubker, *et al* 2014, Delforge, *et al* 2015)	Secondary	Clinical time points	1-sample t-test for *within* group change (p<0.05). 2-sample t-test for *between* group difference (p<0.05).	1,076 (=535+541) and 547	1025 and 509[3]	506 and 189	Results of t-tests were confirmed by mixed model repeated measures
MM-015 study, randomized phase III study (Dimopoulos, *et al* 2013, Dimopoulos, *et al* 2014, Palumbo, *et al* 2012)	Secondary	Day 1 of treatment cycles	Paired t-test (p<0.05, p<0.01 and p<0.001) and Mixed model repeated measures for *within* group change.	152 and 153 and 154	140 and 146 and 148[3]	91 and 99 and 99[3,4]	Only patients with a non-missing baseline and minimum one follow-up assessment were included. HRQoL observations at PD/DC, if occurred earlier than cycle 16 were carried forward to the next measurement time point.
VISTA trial, randomized phase III study	Secondary	Day 1 of treatment cycles	T-test (p<0.05) for *between* group differences	344 and 338	331 and 318	164 and 136[3]	Only patients with a valid baseline and at least one follow-up HRQoL assessment.

(Delforge, *et al* 2012, Mateos, *et al* 2010)							
HOVON 49, randomized phase III study (Verelst, *et al* 2011, Wijermans, *et al* 2010)	Secondary	Clinical time points	Mixed model repeated measures for multi-item domains and ordinal logistic regression repeated measures for one-item domains *between* group differences	168 and 165	149 and 135	Not reported	Participants and non-participants group comparison. Patients with at least one HRQoL assessment, either at baseline, during treatment or follow-up, were included in the analysis.
Randomized phase III study (Gimsing, *et al* 2010)	Primary	Calendar time points	T-test for *within* group change. Generalised estimating equations for *within* group change.	252 and 252	164 and 171	164 and 171[5] at 12 months	Only patients who returned questionnaires at 12 months and who were still on study treatment were included in the HRQoL analyses. Multiple imputation.
Randomized phase III study (Waage, *et al* 2010)	Secondary	Calendar time points	Generalised estimating equations confirmed by standardized area under the curve for *within* group change	182 and 175	Not reported	Not reported[6]	Analyses were based on patients who returned completed questionnaires. Multiple imputation.
NMSG 4/90, cohort study (Wisloff, *et al* 1996a)	Secondary	Calendar time points	Descriptive	583	524	424	Only eligible patients completing the first questionnaire were included. Comparison of characteristics of participants and non-participants
NMSG 4/90, randomized phase III study (NMSG 1996, Wisloff, *et al* 1996b)	Secondary	Calendar time points	Mann Whitney U-test (p<0.01) for *between* group differences	297 and 286	271 and 253	67 and 74	Only eligible patients completing the first questionnaire were included.
Consolidation treatment studies							
Randomized phase II study (Mellqvist, *et al* 2013)	Primary	Calendar time points	Mann Whitney U-test (p<0.01) for *between* group differences	187 and 183	311 in total	Not reported	Only patients completing the baseline questionnaire were included in the analysis
Phase II study (Frodin, *et al* 2011)	Primary	Clinical time points	Descriptive	56	56	25	Not described
Phase II study (Khalafallah, *et al* 2011)	Primary	Calendar time points	Ordinal logistic regression repeated measures for *within* group change	18	18	Not reported	Not described
Maintenance treatment studies							
Randomized phase III study, MM-015 study (Dimopoulos, *et al* 2013, Dimopoulos, *et al* 2014, Palumbo, *et al* 2012)	Secondary	Day 1 of treatment cycles	Paired t-test (p<0.05, p<0.01 and p<0.001) and Mixed model repeated measures for *within* group change.	88 and 94 and 102	91 and 99 and 99[3,4]	65 and 56 and 65[3]	Only patients with a non-missing baseline and minimum one follow-up assessment were included. HRQoL observations at PD/DC, if occurred earlier than cycle 16 were carried forward to the next measurement time point.

Randomized phase II study (Sirohi, *et al* 2007)	Primary	Calendar time points	Two-sample t-test Mixed model repeated measures for *between* group differences.	30 and 30	30 and 30	Not reported	90% completed all three questionnaires and all patients completed at least two.
Relapse treatment studies							
TOURMALINE-MM1, randomized phase III study (Leleu, *et al* 2018, Moreau, *et al* 2016)	Secondary	Day 1 of treatment cycles	Mixed model repeated measures for *between* group difference and *within* group change (p<0.05)	360 and 362	337 and 349	7 and 2[7]	Graphical examination stratified by time of last assessment and pattern mixture model as sensitivity analysis.
ASPIRE trial, randomized phase III study (Stewart, *et al* 2016, Stewart, *et al* 2015b)	Secondary	Day 1 of treatment cycles	Mixed model repeated measures for *between* group difference and *within* group change (p<0.01 and p<0.001)	396 and 396	348 and 348	227 and 148	Analysis includes patients with at least one HRQoL assessment. Graphical examination stratified by time of last assessment. Results were confirmed with standardized area under the curve and pattern mixture model as sensitivity analysis.
MM-003, randomized phase III study (Miguel, *et al* 2013, Song, *et al* 2015, Weisel, *et al* 2015)	Secondary	Day 1 of treatment cycles	Mixed model repeated measure. Paired t-test (p<0.05) for *within* group change. Unpaired t-test (p<0.05) for *between* group differences confirmed. Logistic regression analysis for analysis of responders	302 and 153	289 and 144	51 and 6[3]	Analysis includes patients who received at least one study drug and had one HRQoL assessment. Missing data were categorically evaluated for all HRQoL assessments.
NMSG 17/07, randomized phase III study (Hjorth, *et al* 2012)	Secondary	Day 1 of treatment cycles	Mann Whitney U-test (p<0.01) for *between* group differences	67 and 64	67 and 61	Not reported[8]	Not described
APEX study, randomized phase III study (Lee, *et al* 2008, Richardson, *et al* 2005)	Secondary	Calendar time points	Generalised estimating equations for *between* group difference (p<0.05)	333 and 336	288 and 287	65 and 81	Analyses are based on patients with a valid HRQoL assessment and at least one post-baseline questionnaire. Multiple imputation taking deaths into account and other parameters as sensibility. analyses
SUMMIT study, randomized phase III study (Dubois, *et al* 2006, Richardson, *et al* 2003)	Secondary	Day 1 of treatment cycles	Wilcoxon signed rank test (p<0.05) for *within* group change	202	144	144[9]	Analysis includes 144 patients with both clinical response and PRO information available

Phase II study (Waage, *et al* 2004)	Secondary	Calendar time points	Wilcoxon signed rank test[10] (p<0.01) for related samples for *within* group change	65	62	20	Comparing mean score for patients with available questionnaire at 24 weeks to patients with early study discontinuation
Non-interventional study							
PROFILES registry, cohort study (Mols, *et al* 2012)	Primary	Calendar time point	Paired t-test for *within* group change (p<0.01)	156	156	80	Non-responders and responders were compared

PRO; patient-reported outcomes, PD; Progressive disease, DC; discontinuation for other reasons

[1] In case of HRQoL measurement at study discontinuation, the number of available questionnaires at the former time point is presented unless another time point is specified.

[2] The patients at follow-up are not all the same as at diagnosis

[3] Based on mean score of physical functioning

[4] Patients in each group at cycle 10, which was at start of maintenance

[5] The number of questionnaires used for later follow-up time point evaluation is not reported

[6] 50% of the patients in the melphalan-prednisone-thalidomide arm and 62% of the patients in the melphalan-prednisone arm. The exact numbers could not be extracted.

[7] The results of the PRO data in the study was made at a specified cut-off date and some patients were still in follow-up after

[8] 29 vs. 29 patients were alive at the time of last follow-up. Study design with crossover at treatment failure.

[9] The change in PRO over time was assessed by comparing the change in scores according to clinical response between baseline and best end point

[10] The non-parametric Mann Whitney U-test is used for related samples were used to compare the score at different time points, which is interpreted as a Wilcoxon signed rank test

Table 2. The total number of participating patients and the total number of completed PRO assessments at each scheduled PRO assessment time point for the six studies with the numbers presented.

| Author and year of publication | Study arm/cohort | Total sum for all presented PRO assessment time points | | Intermittent non-responses rate |
		Sum of *expected* PRO assessments	Sum of *completed* PRO assessments	
Gulbrandsen, *et al* 2001[a]	Induction therapy and HDT	1076	966	10%
	Control	541	528	2%
Delforge, *et al* 2015	Lenalidomide-dexamethasone	5166	4743	8%
	Melphalan-prednisone-thalidomide	2492	2179	13%
Wisloff, *et al* 1996	Cohort	2541	2055	19%
Stewart, *et al* 2016	Carfilzomib-lenalidomide-dexamethasone	1706	1543	10%
	Lenalidomide-dexamethasone	1556	1351	13%
Leleu, *et al* 2018[b]	Ixazomib-lenalidomide-dexamethasone	3242	3007	7%
	Lenalidomide-dexamethasone	3209	2991	7%
Waage *et al* 2004	Thalidomide	153	120	22%

PRO; Patient-reported outcomes, HDT; high dose chemotherapy with stem cell support; [a]Based on Figure 1 in the paper, [b]Calculated on basis of table S2 in the supplementary file

Appendix V

Paper V

Strategies to improve patient-reported outcome completion rates in longitudinal studies

Lene Kongsgaard Nielsen[1,2], Madeleine King[1,3], Sören Möller[2,4], Mary Jarden[5], Christen Lykkegaard Andersen[5], Henrik Frederiksen[1], Henrik Gregersen[6], Anja Klostergaard[7], Morten Saaby Steffensen[8], Per Trøllund Pedersen[9], Maja Hinge[10], Mikael Frederiksen[11], Bo Amdi Jensen[12], Carsten Helleberg[13], Anne Kærsgaard Mylin[5], and Niels Abildgaard[1,14]

[1]Quality of Life Research Center, Department of Haematology, Odense University Hospital, Denmark;

[2]OPEN, Odense Patient data Explorative Network, Odense University Hospital, Denmark;

[3]University of Sydney, Faculty of Science, School of Psychology, Australia;

[4]Department of Clinical Research, University of Southern Denmark, Denmark;

[5]Department of Haematology, Copenhagen University Hospital, Copenhagen, Denmark;

[6]Department of Haematology, Aalborg University Hospital, Aalborg, Denmark;

[7]Department of Haematology, Aarhus University Hospital, Denmark;

[8]Department of Haematology, Regional Hospital West Jutland, Denmark;

[9]Department of Haematology, South West Jutland Hospital;

[10]Department of Internal Medicine, Section of Haematology, Vejle Hospital, Denmark;

[11]Department of Haematology, Hospital of Southern Jutland, Denmark;

[12]Department of Haematology, Zealand University Hospital, Denmark;

[13]Department of Haematology, Herlev Hospital, Denmark;

[14]The Academy of Geriatric Cancer Research (AgeCare), Odense University Hospital, Denmark;

Corresponding author; Lene Kongsgaard Nielsen, lene.kongsgaard.nielsen@rsyd.dk, Phone: +45 2324 1605, ORCID ID; 0000-0003-2083-3870

Abstract

Purpose The quality of patient-reported outcome (PRO) data can be compromised by non-response (NR) to scheduled questionnaires, particularly if reasons for NR relate to health problems as this may bias results. We aimed to investigate whether education of study nurses, electronic reminders and real-time monitoring improve patient-reported outcome (PRO) completion rates.

Methods The ongoing population-based study "Quality of life in Danish multiple myeloma patients" is a longitudinal, multicenter study with consecutive inclusion of treatment-demanding newly diagnosed or relapsed patients with multiple myeloma (MM). Education of study nurses in the avoidance of NR, electronic reminders at predefined time points, pre-planned seven-day windows for answering and real-time monitoring of NR were integrated in the study design, conduct and procedures. Patients were expected to answer a set of PRO questionnaires at study entry and 12 times during follow-up, by either electronic or paper method. We investigated the effect of the initiatives on the PRO completion rate.

Results 271 included patients constituted the study cohort in the analyses; of those, 249 (85%) patients chose electronic completion. A total of 1441 scheduled follow-up PRO assessments were reached at data cut-off for analyses. Eighty-four percent of the scheduled PRO assessments were completed within the pre-planned time window, and another 11% were completed after real-time monitoring, equivalent to a PRO completion rate of 95%.

Conclusions The applied strategies achieved a very high completion rate in our study. We propose this kind of strategy in PRO studies, noting that staff resources are required for implementation.

Introduction

Multiple myeloma (MM) is an incurable malignancy of plasma cells in the bone marrow. MM is associated with severe morbidity, specifically caused by bone destruction and pathological bone fractures, renal dysfunction, high infection rate and potential physical disability [1,2]. The prognosis of MM has improved markedly over the past 20 years, and the median survival of patients with MM under the age of 70 has increased from 3 years to 6-7 years [3-5]. The improved prognosis is mediated by the introduction of high dose chemotherapy with autologous stem cell support (HDT) in the 1990s, new treatment options with immunomodulatory drugs (IMiDs), such as thalidomide, lenalidomide and pomalidomide [6-8], and the proteasome inhibitors, bortezomib, carfilzomib and ixazomib [9-11]. Most recently, the monoclonal antibodies elotuzumab and daratumumab [12,13] have been introduced, and the prognosis is expected to improve even further in coming years [14].

Treatment choice in MM depends on several factors including patient age, disease complications, existing comorbidity, and whether the patient is judged fit for specific regimens, such as HDT. Treatment usually involves repeated cycles of a 2-3 drug combination therapy with a proteasome inhibitor, IMiD, cytostatic agent or monoclonal antibodies and steroid. Treatment implies a risk of both acute adverse events, such as infections, as well as late effects, such as peripheral neuropathy and fatigue [15-18].

Patients with MM report a high symptom burden. Common symptoms include fatigue, pain, constipation, insomnia and tingling hand/feet, with consequent decrease in physical and cognitive functioning [19-21]. Compared to patients with other hematological malignancies, patients with MM report a lower health-related quality of life (HRQoL) [21,22]. Longitudinal HRQoL studies of patients with MM suggest that clinically beneficial improvements in HRQoL are more likely during primary treatments than during treatment for relapse [23].

Patients' experience of symptoms and impact on HRQoL can be validly and reliably captured with patient-reported outcome (PRO) questionnaires [24]. Typically, a schedule of assessment times is specified for capture of PROs at key time-points. If a patient fails to complete a questionnaire at a scheduled time, this is termed non-response (NR). The consequent missing data can lead to a variety of problems, more so as NR rates increase, including loss of study power and precision [25,26]. If the reason for NR is related to the patient's poor health status, this may lead to bias, if not handled appropriately. For example, if only complete case analysis methods are used, there is a risk of overestimated HRQoL and underestimated toxicity [25,27,26]. This is because the analysis would be based on patients with complete PRO data available who presumably have better HRQoL outcomes, since patients who drop out might have more toxicity and a worse HRQoL [28,25,29-32]. Thus, NRs represent a threat to internal and external validity and

is one of the inherent barriers in establishing high quality PRO data for use in patient-centred care [33-35]. Several strategies designed to minimize NR have been proposed, and these can be integrated into the study design, protocol and implementation procedures for the PRO study [27]. These include ensuring that staff are aware of the importance of reducing NR and have access to written guidance and support [29,27]. Also, given the time-sensitive nature of PRO-data, real-time monitoring of PRO completion rates during study conduct is recommended [36]. However, we are not aware of any studies that have assessed and documented the effectiveness of such strategies in reducing NR.

The study of "Quality of life in Danish multiple myeloma patients" (QoL-MM) is a Danish multicenter, prospective, observational and primarily web-based survey with real-time monitoring of NRs. In QoL-MM, we are implementing several strategies to reduce NR: education of study nurses, electronic reminders and real-time monitoring of PRO completion rates. The primary aim of this analysis was to investigate whether the applied strategies reduced NRs.

Methods

The QoL-MM study includes newly diagnosed or relapsed, treatment demanding patients with MM who, according to International Myeloma Working Group (IMWG) criteria, are eligible for inclusion [2,37]. The inclusion criteria are very broad to ensure inclusion of a population based, representative cohort of MM patients. Only patients who are not able to understand the Danish language or who are diagnosed with a psychiatric condition are ineligible. All 10 Danish departments of hematology participate. The goal is to recruit 800 patients, and each patient is followed for 24 months or until early drop-out due to withdrawal of consent, death, or permanent lack of ability to fill out the questionnaires. The patients are introduced to the study by their treating physicians or nurses, and written informed consent is obtained before inclusion. Demographic data are collected as part of an inclusion interview performed by a local study nurse. Moreover, the patients provide information related to activity of daily living, instrumental activity of daily living and self-reported diseases, summarized by us into the Charlson Comorbidity Index [38-40]. This information is used for calculating the IMWG myeloma frailty score, which divides the patients into categories of "Frail", "Intermediate Fitness" or "Fit" and the Freiburger Comorbidity Index of 0-3 [41,42] . Each patient's Karnofsky Performance status was assessed by the local study nurse[43]. The patients´ clinical data, e.g. date of diagnosis, MM subtype and the prognostic score, International Staging System are collected from The Danish Multiple Myeloma Registry [44]. Data on admissions, discharges and other hospital procedures are captured from The National Registry of Patients [45].

PRO study design

Patients complete questionnaires at study entry and at 12 follow-up time points during the 24 months study follow-up. The target dates of completion of the follow-up questionnaires are every four weeks for the first 6 months and thereafter every 3 months until 24 months. Depending on the PRO assessment time point, the patient completes between two and four PRO instruments, equivalent to 50-85 items. Each set of questionnaires starts with the cancer specific instrument of European Organisation for Research and Treatment of Cancer Quality of life QLQ-C30 (EORTC QLQ-C30) [24]. Other PRO instruments used are the Multiple Myeloma module QLQ-MY20 (EORTC QLQ-MY20) [46], the Chemotherapy-Induced Peripheral Neuropathy module (EORTC QLQ-CIPN20) [47] and the Short-form health survey version 2 (SF12v2) [48].

PRO data collection procedures

Participants are asked to complete the entire set of questionnaires at one time, preferably on the target date and no later than seven days hereafter, which is the pre-planned completion time window for all follow-up questionnaires. The patients are furthermore encouraged to use a web-based answering method, where a link is sent to the patients´ e-mail box on the target date. However, patients can choose a paper-and-pencil method, if preferred. A REDCap database is set to automatically send the e-mails on the target date as well as to send the e-mail reminders [49].

At baseline, the patient completes the questionnaires alone or with a study nurse present, using a tablet or paper. Alternatively, if the patient chooses to complete the follow-up questionnaires electronically, the baseline questionnaire can be completed at home by computer or tablet. However, the study nurse must ensure that the baseline questionnaire has been completed no later than on the day the patient starts anti-myeloma treatment. If a patient has an uncompleted baseline questionnaire or missed completion of one or more of the four baseline PRO instruments, the patient is excluded as a screening failure, if the answers are not provided no later than day 3 after start of anti-myeloma treatment.

Patients who decide to use paper questionnaires receive three sets of questionnaires at the inclusion interview to complete at 4, 8 and 12 weeks follow-up. The target date of each set of questionnaires is written on the front page, and a letter with the local study nurses' contact information is added. The patients are asked to bring the completed questionnaires into the outpatient clinic at scheduled appointments related to treatment. To ensure that the patient remembers to complete the first follow-up paper questionnaire, the local study nurse contacts the patient at 4 weeks, in order to remind the patient about completion. After week 12, it is the responsibility of the local study nurses to provide the next three

questionnaires to the patient for completion at 16, 20 weeks and 6 months etc. Completed paper questionnaires are uploaded to the database by the local study nurse and centrally entered continuously during the study period.

Strategies to minimize non-responses

All local study nurses are trained by the project leader in the importance of minimizing missing items and NR to a questionnaire or set of questionnaires. Guidance by the local study nurse to frail patients or patients with temporary lack of ability to independently complete the questionnaires is allowed. In this case, the local study nurse reads the items and response categories aloud and mark the answer on behalf of the patient. The aim is for this to be done within the seven-day window after target date. If the patient does not have an appointment in the outpatient clinic within this time frame, but has a scheduled appointment a few days before the target date, the study nurse provides the patient with the questionnaire at that appointment. Otherwise, the questionnaire is completed after the seven-day window.

For patients using the web-based method but who have not completed the electronic questionnaire at day four, a reminder is automatically sent to the patient on day four. If a patient has still not answered the questionnaire on day seven after the target date, the local study nurse is notified by the central study office during weekdays, as part of real-time monitoring of NR. In this situation, the local study nurse has two weekdays to contact the patient, ascertain and document the reason for NR and to invite the patient to complete the questionnaire. The study nurses have access to a written guideline of all project related tasks, and in case of a need for further clarification of a project procedure, the study office can be contacted by telephone or e-mail during weekdays.

Real-time monitoring of PRO completion or NR for both web-based and paper-based PRO questionnaire completion is carried out by the study office. NR is defined as non-completion of an entire set of questionnaires within the seven-day window of a scheduled PRO assessment time-point. If a patient has completed the EORTC QLQ-C30, which is the first questionnaire at every scheduled PRO assessment time point, the follow-up PRO assessment is defined as completed. Missing items and partly completed questionnaires or sets of questionnaires are not part of the real-time monitoring.

The participating departments are financially compensated for handling a NR, providing guidance to a patient to complete a questionnaire and when they collect three completed paper questionnaires and provide three new paper questionnaires to a patient.

Information to the participants

As part of the inclusion interview, all patients are informed about the importance of completing the follow-up questionnaires within the seven-day window, and that the study nurse will contact them if they have not completed a questionnaire by the seventh day. Patients choosing the electronic platform are informed that they will receive a reminder if they have not completed a scheduled questionnaire within four days. The patients are also informed about the importance of completing the questionnaires for the study, and therefore told they will be contacted if the questionnaires is not completed within the seven-day window. It is made clear to the patient that participation and completing questionnaires in the study are voluntary and that the patient is allowed to skip a scheduled questionnaire without this having any consequences for the patient. All patients receive the study nurses´ contact information and are encouraged to seek support in case of questions, technical challenges or a wish of changing the method of completion.

Patient cohort and data analysis

The patient cohort for this paper is all patients who consented to QoL-MM from study initiation at the 20th September 2016 to 16th August 2018 and who had reached at least the first follow-up PRO assessment time point at week four. Questionnaires, which are completed before or within the pre-planned seven-day time window are defined as "on-time responses". In case the patient completes the questionnaire at day seven after the target date or later, the response is defined as "salvage response", the remainder were categorized as a "never response". The PRO completion rate was calculated as the number of completed on-time and salvage responses as a proportion of the number of scheduled PRO assessments expected to be completed. Also, we calculated the rate of on-time and salvage responses separately. The effect of the reminder was investigated by the proportion completing the questionnaire at day three compared to day four of the remaining incomplete questionnaire. Data is presented by descriptive analyses using STATA version 15.

Results

As of August 16th 2018, 481 patients were found eligible for the QoL-MM study, and hereof 292 provided written consent for participation and inclusion in the study. Of the 292 patients included, 271 had reached at least the first follow-up PRO assessment time point at 4 weeks and were included in the analyses. Patient and disease characteristics are presented in Table 1. 23% of the patients were 76 years or older, and 13% of the patients were characterized as "Frail" according to the IMWG myeloma frailty score. Of all patients, 55% had newly diagnosed symptomatic MM, 31% started an induction regimen with planned HDT at study entry, and 11% started fifth or later line of therapy.

Electronic completion of follow-up questionnaires was chosen by 85% of the 271 patients, and 15% chose the paper-and-pencil method. Three patients changed mode of answering method during follow-up, two of them from electronic method to paper method, since the electronic method was found to be too complicated.

PRO completion rate

Per protocol, for the study cohort (n=271), 1441 scheduled follow-up questionnaires were expected to be completed at the time of analysis. The number of patients and completed questionnaires (on-time and salvage) and never-responses at each follow-up time point are presented in Figure 1. The reasons for reduction in number of patients during follow-up are early drop-out or end of follow-up. The largest proportion of never responses with the first year of follow-up was at four weeks at 7% (19 questionnaires out of 271 expected).

1214 of the 1441 scheduled questionnaires (84%) were completed on-time. Of the 227 questionnaires that were not completed on-time, 153 (67%) were salvaged responses and 74 (33%) were never completed. When adding the salvage responses to the on-time responses, a total 1367 of the scheduled questionnaires were completed, equivalent to a PRO completion rate of 95%.

Pattern of response

Questionnaire completion patterns are presented in Figure 2 and Table 2. Of the 1367 scheduled questionnaires, 553 (40%) were answered on the target date, 471 (31%) were completed on days one to three. After the reminder was sent on day 4 to the patients completing questionnaires electronically, a further 207/1367 (15%) were answered between days four and six. A higher number of questionnaires were completed at day four (by 101 patients) compared to day three (by 77 patients). Also, a higher number of questionnaires were completed at day seven (by 60 patients) compared to day 6 (by 36 patients).

Discussion

The primary aim of this analysis was to investigate whether the applied strategies reduced NRs. The initiatives included education of study nurses by the project leader, electronic reminders and real-time monitoring of NR. Using these strategies, we achieved a very high PRO completion rate of 95%, with just 5% non-response.

Comparing this PRO completion rate to reported PRO completion rates in other longitudinal HRQoL studies of patients with MM [50], only one study has reported a higher PRO completion rate. This was the NMSG 4/90 study by Gulbrandsen *et al.*, where the PRO completion rate of the historical control group of newly

diagnosed patients with MM was 98% [51]. This historical control group originates from the EORTC QLQ-C30 validation study of patients with MM [52]. One of the aims in the validation study was to evaluate the applicability of the questionnaire in a cohort of patients with MM and included sampling of data concerning the patient´s need for assistance in completing the questionnaires. The authors found that up to 30% of the MM patients reported need of assistance in completing the questionnaire. Other strategies of how this high PRO completion rate was achieved are not described in the paper [52].

When we designed the QoL-MM study, we had particular focus on how we could minimize NR. We introduced real-time monitoring of NR and provided the patients with reminders. Staff resources were dedicated for this purpose, software as well as a high proportion of the patients choosing the web-based answering method made it possible, and we succeeded in reaching a high PRO completion rate. Still, some NR could not be avoided, and as part of the study, we collected reasons for NR. Information about the clinical status of patients when they fail to complete a scheduled questionnaire and the reason for NR might assist the PRO researcher in making the correct assumption for the underlying mechanisms of missing data [25,53,26]. The link between the documented reasons and the missing data mechanisms as well as estimation of the impact of patient drop-out will be investigated in future analyses of the QoL-MM study.

The overall aim of the QoL-MM study is to describe the quality of life of the general population of patients with MM from diagnosis to late, advanced disease throughout different anti-myeloma therapies [54]. Methodological considerations concerning PRO assessment time points were included as part of the study planning. Clinical visits are a frequently chosen time point for PRO assessment in clinical trials of MM patients. This decision has the advantage of reducing the risk of NR, since patients have the opportunity to complete the questionnaires at the hospital with assistance from the study nurse. Disadvantages in using day 1 of treatment cycles include a potential risk of underestimation of toxicities that occur after day 1 and not capturing periods with temporary decline in HRQoL, resulting in rescheduling of chemotherapy [55]. We chose to collect the PRO data in QoL-MM at predefined calendar time points to meet the overall study aim and thereby capture HRQoL at regular non-clinic time-points throughout the MM patients´ diverse disease trajectories. This decision could have made the study vulnerable for low PRO completion rates, since the general population of patient with MM can be frail and are at risk of adverse events, hospital admissions, as well as risk of physical and mental disabilities caused by the disease and therapy. Therefore, we implemented the educational and procedural strategies to reduce NRs.

One of the procedural strategies we chose was to use a pre-planned time window of seven days for each scheduled questionnaire. This allowed the study nurses to clearly communicate the expectations to the

patients participating in the study and systematically capture reasons for NR from every patient who failed to complete the questionnaire within the pre-planned time window [56,57]. Whether or not patients who were not able to complete the questionnaires within the time window have a poorer HRQoL will be investigated further as part of the QoL-MM study.

Conclusions

We evaluated strategies to maximize PRO completion in a longitudinal cohort study of patients with MM receiving cancer treatment. Real-time monitoring of NR, mailing electronic reminders to patients, and education of study nurses are effective strategies that resulted in a questionnaire completion rate of 95%. To our knowledge, the QoL-MM study is the first study to provide insight into how to ensure high PRO completion rates in a cohort of cancer patients receiving chemotherapy. We propose our applied strategies as a model for improving PRO completion rates in clinical trials and registries to increase the quality and value of PRO data in patient-centred care.

Compliance with ethical standards

Conflict of interest. The authors declare that they have no conflict of interest. The study was approved by the Danish Data Protection Agency, registered at ClinicalTrials.gov by number NCT02892383 and carried out in accordance with the Helsinki Declaration and Good Clinical Practice guidelines. Informed consent was obtained from all participants included in the study.

References

1. Kyle, R. A., & Rajkumar, S. V. (2008). Multiple myeloma. *Blood, 111*(6), 2962-2972, doi:10.1182/blood-2007-10-078022 [doi].
2. Rajkumar, S. V., Dimopoulos, M. A., Palumbo, A., Blade, J., Merlini, G., Mateos, M. V., et al. (2014). International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet Oncology, 15*(12), e538-548, doi:10.1016/S1470-2045(14)70442-5 [doi].
3. Kumar, S. K., Dispenzieri, A., Lacy, M. Q., Gertz, M. A., Buadi, F. K., Pandey, S., et al. (2014). Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia, 28*(5), 1122-1128, doi:10.1038/leu.2013.313.
4. Blade, J., & Rosinol, L. (2009). Changing paradigms in the treatment of multiple myeloma. *Haematologica, 94*(2), 163-166, doi:10.3324/haematol.2008.002766 [doi].
5. Kumar, S. K., Rajkumar, S. V., Dispenzieri, A., Lacy, M. Q., Hayman, S. R., Buadi, F. K., et al. (2008). Improved survival in multiple myeloma and the impact of novel therapies. *Blood, 111*(5), 2516-2520, doi:blood-2007-10-116129 [pii].
6. Waage, A., Gimsing, P., Juliusson, G., Turesson, I., Gulbrandsen, N., Eriksson, T., et al. (2004). Early response predicts thalidomide efficiency in patients with advanced multiple myeloma. *British journal of haematology, 125*(2), 149-155, doi:10.1111/j.1365-2141.2004.04879.x [doi].
7. Miguel, J. S., Weisel, K., Moreau, P., Lacy, M., Song, K., Delforge, M., et al. (2013). 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. *The Lancet Oncology, 14*(11), 1055-1066, doi:10.1016/s1470-2045(13)70380-2.
8. Weber, D. M., Chen, C., Niesvizky, R., Wang, M., Belch, A., Stadtmauer, E. A., et al. (2007). Lenalidomide plus dexamethasone for relapsed multiple myeloma in North America. *N Engl J Med, 357*(21), 2133-2142, doi:10.1056/NEJMoa070596.
9. Moreau, P., Masszi, T., Grzasko, N., Bahlis, N. J., Hansson, M., Pour, L., et al. (2016). Oral Ixazomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. *The New England journal of medicine, 374*(17), 1621-1634, doi:10.1056/NEJMoa1516282.
10. Stewart, A. K., Rajkumar, S. V., Dimopoulos, M. A., Masszi, T., Spicka, I., Oriol, A., et al. (2015). Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. *The New England journal of medicine, 372*(2), 142-152, doi:10.1056/NEJMoa1411321 [doi].
11. Richardson, P. G., Barlogie, B., Berenson, J., Singhal, S., Jagannath, S., Irwin, D., et al. (2003). A phase 2 study of Bortezomib in relapsed, refractory myeloma. *New England Journal of Medicine, 348*(26), 2609-2617, doi:http://dx.doi.org/10.1056/NEJMoa030288.
12. Palumbo, A., Chanan-Khan, A., Weisel, K., Nooka, A. K., Masszi, T., Beksac, M., et al. (2016). Daratumumab, Bortezomib, and Dexamethasone for Multiple Myeloma. *N Engl J Med, 375*(8), 754-766, doi:10.1056/NEJMoa1606038.
13. Lonial, S., Dimopoulos, M., Palumbo, A., White, D., Grosicki, S., Spicka, I., et al. (2015). Elotuzumab therapy for relapsed or refractory multiple myeloma. *New England Journal of Medicine, 373*(7), 621-631, doi:http://dx.doi.org/10.1056/NEJMoa1505654.
14. Ocio, E. M., Richardson, P. G., Rajkumar, S. V., Palumbo, A., Mateos, M. V., Orlowski, R., et al. (2014). New drugs and novel mechanisms of action in multiple myeloma in 2013: a report from the International Myeloma Working Group (IMWG). *Leukemia, 28*(3), 525-542, doi:10.1038/leu.2013.350 [doi].
15. Boland, E., Eiser, C., Ezaydi, Y., Greenfield, D. M., Ahmedzai, S. H., & Snowden, J. A. (2013). Living with advanced but stable multiple myeloma: a study of the symptom burden and cumulative effects of disease and intensive (hematopoietic stem cell transplant-based) treatment on health-related quality of life. *J Pain Symptom Manage, 46*(5), 671-680, doi:10.1016/j.jpainsymman.2012.11.003.
16. Holmstrom, M. O., Gimsing, P., Abildgaard, N., Andersen, N. F., Helleberg, C., Clausen, N. A., et al. (2015). Causes of early death in multiple myeloma patients who are ineligible for high-dose therapy

with hematopoietic stem cell support: A study based on the nationwide Danish Myeloma Database. *American Journal of Hematology, 90*(4), E73-74, doi:10.1002/ajh.23932 [doi].

17. Mateos, M. V. (2010). Management of treatment-related adverse events in patients with multiple myeloma. *Cancer Treatment Reviews, 36*(SUPPL. 2), S24-S32, doi:http://dx.doi.org/10.1016/S0305-7372%2810%2970009-8.

18. Richardson, P. G., Delforge, M., Beksac, M., Wen, P., Jongen, J. L., Sezer, O., et al. (2012). Management of treatment-emergent peripheral neuropathy in multiple myeloma. *Leukemia, 26*(4), 595-608, doi:10.1038/leu.2011.346 [doi].

19. Ramsenthaler, C., Kane, P., Gao, W., Siegert, R. J., Edmonds, P. M., Schey, S. A., et al. (2016). Prevalence of symptoms in patients with multiple myeloma: a systematic review and meta-analysis. *Eur J Haematol, 97*(5), 416-429, doi:10.1111/ejh.12790.

20. Jordan, K., Proskorovsky, I., Lewis, P., Ishak, J., Payne, K., Lordan, N., et al. (2014). Effect of general symptom level, specific adverse events, treatment patterns, and patient characteristics on health-related quality of life in patients with multiple myeloma: results of a European, multicenter cohort study. *Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer, 22*(2), 417-426, doi:10.1007/s00520-013-1991-4 [doi].

21. Johnsen, A. T., Tholstrup, D., Petersen, M. A., Pedersen, L., & Groenvold, M. (2009). Health related quality of life in a nationally representative sample of haematological patients. *European journal of haematology, 83*(2), 139-148, doi:10.1111/j.1600-0609.2009.01250.x [doi].

22. Priscilla, D., Hamidin, A., Azhar, M. Z., Noorjan, K. O. N., Salmiah, M. S., & Bahariah, K. (2011). Quality of life among patients with hematological cancer in a Malaysian hospital. *Medical Journal of Malaysia, 66*(2), 117.

23. Nielsen, L. K., Jarden, M., Andersen, C. L., Frederiksen, H., & Abildgaard, N. (2017). A systematic review of health-related quality of life in longitudinal studies of myeloma patients. *European journal of haematology, 99*(1), 3-17, doi:10.1111/ejh.12882.

24. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute, 85*(5), 365-376.

25. Bell, M. L., & Fairclough, D. L. (2014). Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res, 23*(5), 440-459, doi:10.1177/0962280213476378.

26. Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials* (2. ed. ed., Chapman & Hall/CRC interdisciplinary statistics series). Boca Raton: Chapman & Hall/CRC.

27. Mercieca-Bebber, R., Palmer, M. J., Brundage, M., Calvert, M., Stockler, M. R., & King, M. T. (2016). Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *The BMJ Open, 6*(6), e010938, doi:10.1136/bmjopen-2015-010938.

28. Mercieca-Bebber, R. L., Price, M. A., Bell, M. L., King, M. T., Webb, P. M., & Butow, P. N. (2017). Ovarian cancer study dropouts had worse health-related quality of life and psychosocial symptoms at baseline and over time. *Asia-Pacific Journal of Clinical Oncology, 13*(5), e381-e388, doi:10.1111/ajco.12580.

29. Mercieca-Bebber, R., Calvert, M., Kyte, D., Stockler, M., & King, M. T. (2018). The administration of patient-reported outcome questionnaires in cancer trials: Interviews with trial coordinators regarding their roles, experiences, challenges and training. *Contemporary Clinical Trials Communications, 9*, 23-32.

30. Mercieca-Bebber, R., Kyte, D. G., Calvert, M. J., Stockler, M., & King, M. T. (2017). Challenges and training needs of study coordinators responsible for administration of patient-reported outcome questionnaires in cancer trials. [Conference Abstract]. *Quality of Life Research, 26 (1 Supplement 1)*, 102.

31. Fairclough, D. L., Peterson, H. F., Cella, D., & Bonomi, P. (1998). Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Stat Med, 17*(5-7), 781-796.
32. Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., et al. (1998). Missing quality of life data in cancer clinical trials: serious problems and challenges. *Stat Med, 17*(5-7), 517-532.
33. Bottomley, A., Pe, M., Sloan, J., Basch, E., Bonnetain, F., Calvert, M., et al. (2016). Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *The Lancet Oncology, 17*(11), e510-e514, doi:10.1016/s1470-2045(16)30510-1.
34. Basch, E. (2018). Patient-reported outcomes: an essential component of oncology drug development and regulatory review. *Lancet Oncol, 19*(5), 595-597, doi:10.1016/s1470-2045(18)30141-4.
35. Palmer, M. J., Mercieca-Bebber, R., King, M., Calvert, M., Richardson, H., & Brundage, M. (2018). A systematic review and development of a classification framework for factors associated with missing patient-reported outcome data. *Clinical trials, 15*(1), 95-106, doi:10.1177/1740774517741113.
36. Basch, E., Abernethy, A. P., Mullins, C. D., Reeve, B. B., Smith, M. L., Coons, S. J., et al. (2012). Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *J Clin Oncol, 30*(34), 4249-4255, doi:10.1200/jco.2012.42.5967.
37. Kumar, S., Paiva, B., Anderson, K. C., Durie, B., Landgren, O., Moreau, P., et al. (2016). International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncol, 17*(8), e328-e346, doi:10.1016/s1470-2045(16)30206-6.
38. Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases, 40*(5), 373-383.
39. Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of Illness in the Aged. the Index of Adl: a Standardized Measure of Biological and Psychosocial Function. *Jama, 185*, 914-919.
40. Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *The Gerontologist, 9*(3), 179-186.
41. Kleber, M., Ihorst, G., Terhorst, M., Koch, B., Deschler, B., Wasch, R., et al. (2011). Comorbidity as a prognostic variable in multiple myeloma: comparative evaluation of common comorbidity scores and use of a novel MM-comorbidity score. *Blood Cancer J, 1*(9), e35, doi:10.1038/bcj.2011.34.
42. Palumbo, A., Bringhen, S., Mateos, M. V., Larocca, A., Facon, T., Kumar, S. K., et al. (2015). Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. *Blood, 125*(13), 2068-2074, doi:10.1182/blood-2014-12-615187 [doi].
43. Mor, V., Laliberte, L., Morris, J. N., & Wiemann, M. (1984). The Karnofsky Performance Status Scale. An examination of its reliability and validity in a research setting. *Cancer, 53*(9), 2002-2007.
44. Gimsing, P., Holmstrom, M. O., Klausen, T. W., Andersen, N. F., Gregersen, H., Pedersen, R. S., et al. (2016). The Danish National Multiple Myeloma Registry. *Clin Epidemiol, 8*, 583-587, doi:10.2147/clep.s99463.
45. Folmer Andersen, T., Madsen, M., Jørgensen, J., Mellemkjær, L., & Olsen, J. H. (1999). The Danish National Hospital Register : a valuable source of data for modern health sciences. *Danish medical bulletin, Vol. 46, nr. 3 (1999)*, 263-268.
46. Cocks, K., Cohen, D., Wisloff, F., Sezer, O., Lee, S., Hippe, E., et al. (2007). An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. *European journal of cancer (Oxford, England : 1990), 43*(11), 1670-1678, doi:S0959-8049(07)00351-6 [pii].

47. Postma, T. J., Aaronson, N. K., Heimans, J. J., Muller, M. J., Hildebrand, J. G., Delattre, J. Y., et al. (2005). The development of an EORTC quality of life questionnaire to assess chemotherapy-induced peripheral neuropathy: the QLQ-CIPN20. *European journal of cancer (Oxford, England : 1990), 41*(8), 1135-1139, doi:S0959-8049(05)00160-7 [pii].

48. Ware, J., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care, 34*(3), 220-233.

49. Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform, 42*(2), 377-381, doi:10.1016/j.jbi.2008.08.010.

50. Nielsen, L., Abildgaard, N., Jarden, M., & Klausen, T. (Accepted for publication 2018). Methodological aspects of health-related quality of life measurement and analysis in patients with multiple myeloma. *British journal of haematology*.

51. Gulbrandsen, N., Wisloff, F., Brinch, L., Carlson, K., Dahl, I. M., Gimsing, P., et al. (2001). Health-related quality of life in multiple myeloma patients receiving high-dose chemotherapy with autologous blood stem-cell support. *Medical oncology (Northwood, London, England), 18*(1), 65-77, doi:MO:18:1:65 [pii].

52. Wisloff, F., Eika, S., Hippe, E., Hjorth, M., Holmberg, E., Kaasa, S., et al. (1996). Measurement of health-related quality of life in multiple myeloma. Nordic Myeloma Study Group. *British journal of haematology, 92*(3), 604-613.

53. Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., et al. (2012). The prevention and treatment of missing data in clinical trials. *The New England journal of medicine, 367*(14), 1355-1360, doi:10.1056/NEJMsr1203730.

54. Nielsen, L., Jarden, M., Andersen, C., Frederiksen, H., Gregersen, H., Klostergaard, A., et al. (2018). Quality of life in newly diagnosed and relapsed Danish multiple myeloma patients - a national longitudinal study of Danish Myeloma Study Group. [Conference abstract]. *HemaSphere, PF687*.

55. Giesinger, J. M., Wintner, L. M., Zabernigg, A., Gamper, E. M., Oberguggenberger, A. S., Sztankay, M. J., et al. (2014). Assessing quality of life on the day of chemotherapy administration underestimates patients' true symptom burden. *BMC cancer, 14*, 758, doi:10.1186/1471-2407-14-758.

56. Klee, M. C., King, M. T., Machin, D., & Hansen, H. H. (2000). A clinical model for quality of life assessment in cancer patients receiving chemotherapy. *Ann Oncol, 11*(1), 23-30.

57. Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A. W., King, M. T., et al. (2018). Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *The Journal of the American Medical Association, 319*(5), 483-494, doi:10.1001/jama.2017.21903.

Table 1. Demographic and disease characteristics at entry.

Characteristics	Patients P=271
Mean age (SD)	67.5 (9.1)
Median age, years (IQR)	69 (62; 74)
Age ≤ 65 / 66-75 /≥ 76 years, N (%)	90 (33%)/118 (44%)/63 (23%)
Sex, female/male, N (%)	106 (39%)/165 (61%)
Marital status, Married or cohabiting/single[a], N (%)	213 (79%)/58 (21%)
Weekly alcohol intake, no alcohol intake/1-7/>8 items, N (%)	60 (22%)/142 (52%)/69 (25%)
Daily smoking, yes/former smoker/never smoker, N (%)	31 (11%)/119 (44%)/121 (45%)
Charlson Comorbidity Index, 0/1/2/≥3, N (%)	154 (57%)/45 (17%)/45 (17%)/27 (10%)
Freiburg Comorbidity Index, 0/1/2 or 3, N (%)	223 (82%)/47 (17%)/<5 (0%)
MWG myeloma frailty score, Fit/Intermediate Fitness/Frail, N (%)	148 (55%)/87 (32%)/36 (13%)
Karnofsky Performance Status Scale, 100/90/80/≥70%, N (%)	79 (29%)/116 (43%)/49 (18%)/27 (10%)
Mean time from diagnosis to inclusion (years)(SD)	2.21 (3.12)
Number of lines of therapy	
First line	150 (55%)
Second line	49 (18%)
3-4 line	42 (15%)
5 or more lines	30 (11%)
Anti-myeloma treatment starting	
Induction therapy and HDT	85 (31%)
Melphalan-prednisolon-bortezomib	48 (18%)
Containing daratumumab	69 (25%)
Containing elotuzumab	7 (3%)
Lenalidomid	13 (5%)
Containing Ixazomib	7 (3%)
Containing carfilzomib	13 (5%)
Containing pomalidomid	7 (3%)
M-component subtype, IgG/IgA/light chain/> 1 M-component/ non-secretory/missing[b]	85 (31%)/31 (11%)/10 (4%)/<5 (1%)/6 (2%)/137 (51%)
International Staging System, ISS I/ISS II/ISS III/missing[b], N (%)	24 (9%)/51 (19%)/22 (8%)/174 (64%)

SD; standard deviation, IQR; interquartile range, IMWG; International Myeloma Working Group, HDT; high dose therapy with stem cell support, ISS; International Staging System, [a] separated, divorced, widow or unmarried, [b]Missings are due to time delay in entering disease data into The Danish National Multiple Myeloma Registry or unknown.
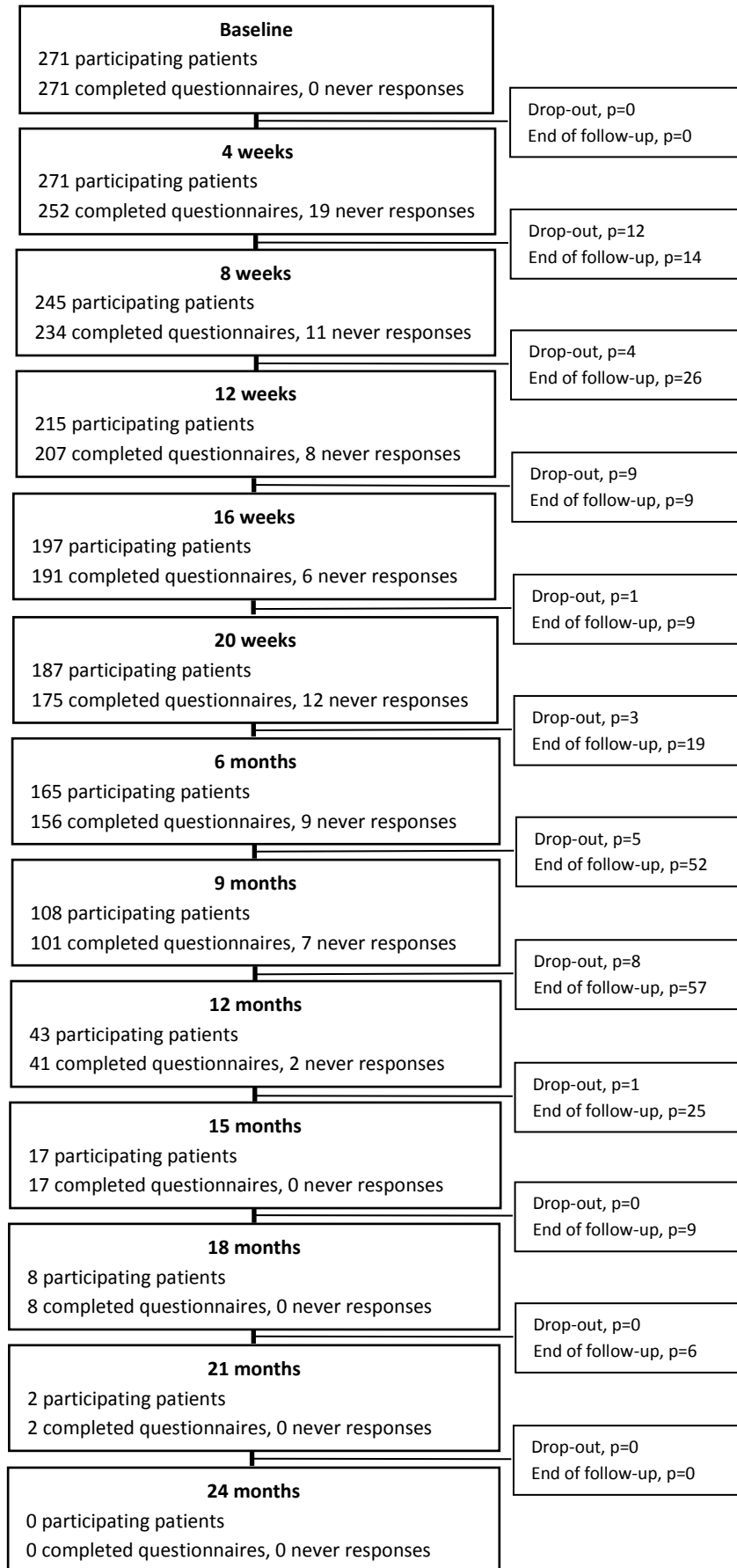
Baseline

271 participating patients

271 completed questionnaires, 0 never responses

Drop-out, p=0
End of follow-up, p=0

4 weeks

271 participating patients

252 completed questionnaires, 19 never responses

Drop-out, p=12
End of follow-up, p=14

8 weeks

245 participating patients

234 completed questionnaires, 11 never responses

Drop-out, p=4
End of follow-up, p=26

12 weeks

215 participating patients

207 completed questionnaires, 8 never responses

Drop-out, p=9
End of follow-up, p=9

16 weeks

197 participating patients

191 completed questionnaires, 6 never responses

Drop-out, p=1
End of follow-up, p=9

20 weeks

187 participating patients

175 completed questionnaires, 12 never responses

Drop-out, p=3
End of follow-up, p=19

6 months

165 participating patients

156 completed questionnaires, 9 never responses

Drop-out, p=5
End of follow-up, p=52

9 months

108 participating patients

101 completed questionnaires, 7 never responses

Drop-out, p=8
End of follow-up, p=57

12 months

43 participating patients

41 completed questionnaires, 2 never responses

Drop-out, p=1
End of follow-up, p=25

15 months

17 participating patients

17 completed questionnaires, 0 never responses

Drop-out, p=0
End of follow-up, p=9

18 months

8 participating patients

8 completed questionnaires, 0 never responses

Drop-out, p=0
End of follow-up, p=6

21 months

2 participating patients

2 completed questionnaires, 0 never responses

Drop-out, p=0
End of follow-up, p=0

24 months

0 participating patients

0 completed questionnaires, 0 never responses

Figure 1. Flow diagram of the patients in follow-up. The reduced number of patients in follow-up was due to drop-out or end of follow-up.
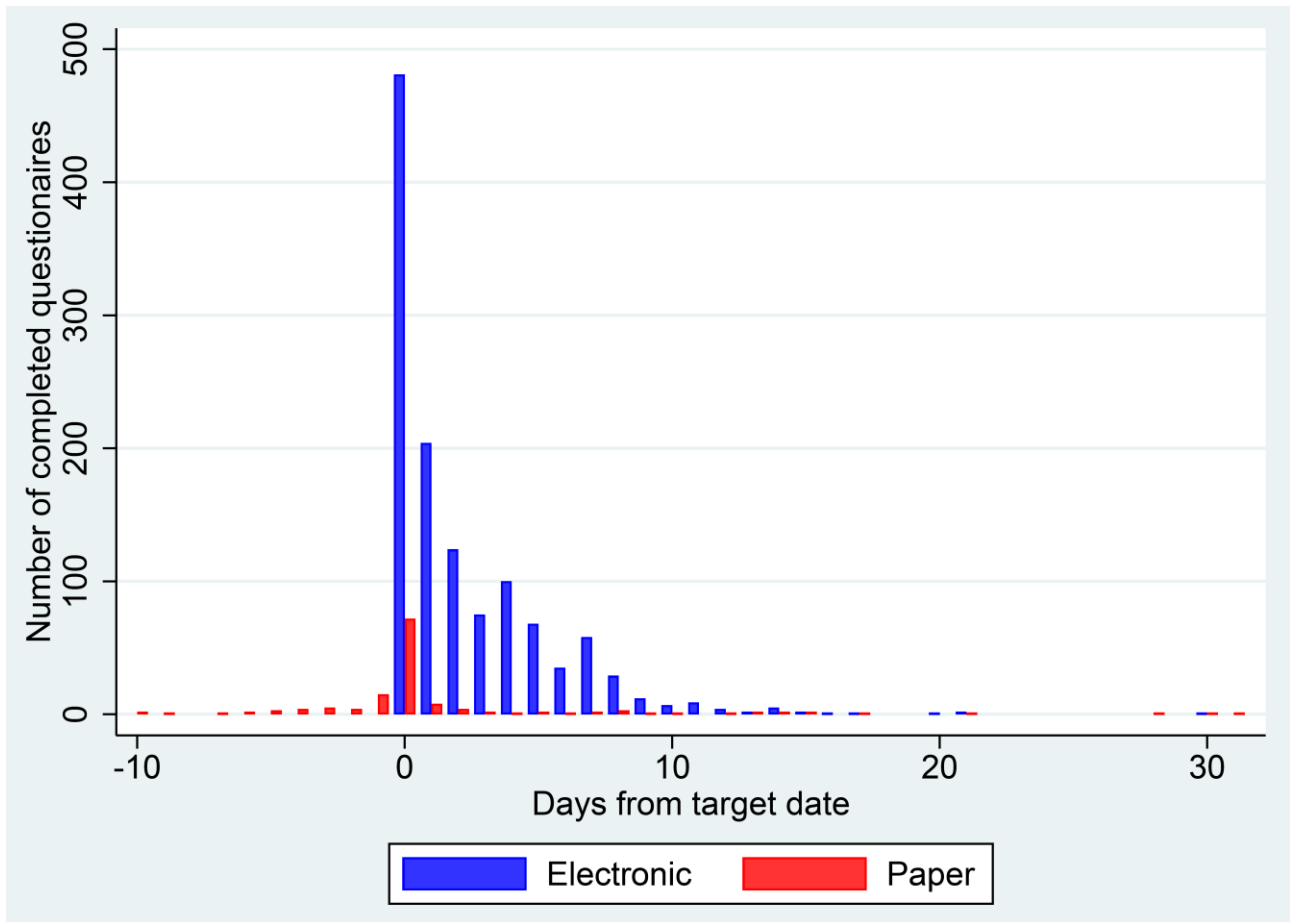
Figure 2. Day of response to paper and electronic questionnaires.

Day 0 is the target day, when the patients were instructed to complete the questionnaires. The patients completing the questionnaires electronically received an email with a link to the questions on day 0. The local nurses provided the paper questionnaires with inscribed target dates for patients completing on paper at home. If the patient had completed the EORTC QLQ-C30, which was the first health-related quality of life instrument in each set of questionnaires, the set of questionnaires was defined as completed.

Table 2. Pattern of response.
Questionnaires completed before or within the seven day window are termed "on-time responses".
Questionnaires completed after the seven day window are termed "salvage responses". 5% of the
scheduled questionnaires were never completed.

Proportion of scheduled questionnaires	Time of response	Completed follow-up questionnaires (Q=1367)
On-time response 84%	**Before day 0**	**37 (3%)**
	Electronic	0
	Paper	37
	Day 0 – the target day	**553 (40%)**
	Electronic	*481*
	Paper	*72*
	Day 1-3	**417 (31%)**
	Day 1	212
	Day 2	128
	Day 3	77
	Electronic	*403*
	Paper	*14*
	Day 4*-6	**207 (15%)**
	Day 4	101
	Day 5	70
	Day 6	36
	Electronic	*203*
	Paper	*4*
Salvage response 11%	**Day 7 or later**	**153 (11%)**
	Day 7	60
	Day 8	32
	Day 9	13
	After day 9	48
	Electronic	*134*
	Paper	*19*

*The patients completing questionnaires electronically received a reminder on day 4 if the answer
was not provided.