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

Lene Kongsgaard Nielsen,[1*] Claudia Stege,[2*] Birgit Lissenberg-Witte,[3] Bronno van der Holt,[4] Ulf-Henrik Mellqvist,[5] Morten Salomo,[6] Gerard Bos,[7] Mark-David Levin,[8] Heleen Visser-Wisselaar,[4] Markus Hansson,[9] Annette van der Velden,[10] Wendy Deenik,[11] Juleon Coenen,[12] Maja Hinge,[13] Saskia Klein,[14] Bea Tanis,[15] Damian Szatkowski,[16] Rolf Brouwer,[17] Matthijs Westerman,[18] Rineke Leys,[19] Harm Sinnige,[20] Einar Haukås,[21] Klaas van der Hem,[22] Marc Durian,[23] Peter Gimsing,[24] Niels van de Donk,[2] Pieter Sonneveld,[25] Anders Waage,[26] Niels Abildgaard[1] and Sonja Zweegman[2]

[1]Quality of Life Research Center, Department of Haematology, Odense University Hospital, Odense, Denmark; [2]Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Hematology, Cancer Center Amsterdam, Amsterdam, the Netherlands; [3]Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam, the Netherlands; [4]HOVON Data Center, Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands; [5]Section of Hematology and Coagulation, Department of Medicine, Sahlgrenska University Hospital, Gotheborg, Sweden; [6]Department of Haematology, Rigshospitalet, Copenhagen, Denmark; [7]Department of Haematology, Maastricht University Medical Center, Maastricht, the Netherlands; [8]Department of Internal Medicine, Albert Schweitzer Hospital, Dordrecht, the Netherlands; [9]Department of Haematology and Wallenberg Center for Molecular Medicine, Skåne University Hospital, Lund University, Lund, Sweden; [10]Department of Internal Medicine, Martini Ziekenhuis, Groningen, the Netherlands; [11]Department of Internal Medicine, Tergooi Ziekenhuis, Hilversum, the Netherlands; [12]Department of Internal Medicine, Isala, Zwolle, the Netherlands; [13]Department of Internal Medicine, Division of Hematology, Vejle Hospital, Vejle, Denmark; [14]Department of Internal Medicine, Meander Medisch Centrum, Amersfoort, the Netherlands; [15]Department of Internal Medicine, Groene Hart Ziekenhuis, Gouda, the Netherlands; [16]Department of Oncology, Haematology and Palliative Care, Førde Central Hospital, Førde, Norway; [17]Department of Internal Medicine, Reinier de Graaf Ziekenhuis, Delft, the Netherlands; [18]Department of Internal Medicine, Northwest Clinics, Alkmaar, the Netherlands; [19]Department of Internal Medicine, Maasstad Ziekenhuis, Rotterdam, the Netherlands; [20]Department of Internal Medicine, Jeroen Bosch Ziekenhuis, Den Bosch, the Netherlands; [21]Department of Haematology, Stavanger University Hospital, Stavanger, Norway, [22]Department of Internal Medicine, Zaans Medisch Centrum, Zaandam, the Netherlands; [23]Department of Internal Medicine, Tweesteden Ziekenhuis, Tilburg, the Netherlands; [24]Department of Haematology, Rigshospitalet, Copenhagen, Denmark; [25]Department of Haematology, Erasmus Medical Center Cancer Center, Rotterdam, the Netherlands and [26]Department of Haematology, St Olavs Hospital and Norwegian University of Science and Technology, Trondheim, Norway

**LKN and CS contributed equally as co-first authors.*

 doi:10.3324/haematol.2019.222299

Received: April 19, 2019.

Accepted: September 11, 2019.

Pre-published: September 12, 2019.

Correspondence: *LENE KONGSGAARD NIELSEN* -lene.kongsgaard.nielsen@rsyd.dk

Index of the Online Supplementary Appendix

A. HRQoL data collection and categorization of questionnaires

Health related quality of life (HRQoL) was assessed by using two questionnaires: the cancer specific European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 and the multiple myeloma (MM) specific EORTC QLQ-MY20 module. The QLQ-C30 consists of 15 scales: one global quality of life (QoL) scale, five functional scales (physical, role, emotional, cognitive and social functioning) and nine symptoms scales (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties). The 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 in item 13 of the QLQ-MY20 *"Did you have tingling hands or feet?"*. The peripheral neuropathy scale was an additional single-item scale, transformed to a 0 to 100 score, according to the EORTC manual.[1] In this HRQoL study, all the above mentioned scales were evaluated.

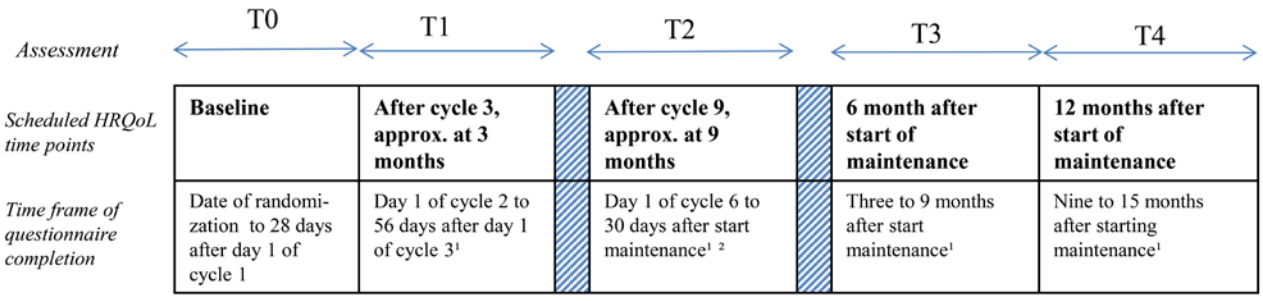
Patients received a paper version of the questionnaires. For patients recruited by an investigator from HOVON the local QoL coordinators e-mailed the answered questionnaires to the HOVON Data Center, Amsterdam. For patients recruited by an investigator from Nordic Myeloma Study Group (NMSG) the local 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 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 these time frames were excluded from the analysis. The timing of questionnaires is illustrated in Figure A1. All questionnaires had to be completed at most 1 month after going off protocol. HRQoL assessment was terminated from the time a patient went off protocol.

Supplemental Figure A1. Time of the collected HRQoL questionnaires



HRQoL; health-related quality of life

[1]A questionnaire answered by the patient within the first month off protocol was categorized into the fitting category of T0-T4, [2]Earlier in case of early start of maintenance.

B. Statistical analysis

Change in HRQoL over time was assessed by linear mixed models. For differences *within* each arm separately, the linear mixed model only included a fixed effect for time and random intercept for subject. The null hypothesis for within arm analysis was "no change in HRQoL over time". For differences *between* arms, the model included fixed effects for time, treatment arm and their two-way interaction and a random intercept for subject. The null hypothesis for between arm analysis was "no difference between arms". Model estimates were used for post-hoc comparisons of change from baseline, within and between arms.
Patient-reported peripheral neuropathy, using the results of the QLQ-C30 question "*Did you have tingling hands or feet?*", was compared with investigator-reported peripheral neuropathy, using the grading assessed by National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) version 3.0.[2] All available questionnaires were used, and the individual CTCAE score at a similar time point was matched to the results of these questionnaires. No/mild peripheral neuropathy (defined by either patient reported "not at all" or "a little tingling" and investigator reported CTCAE 0-1) was compared to moderate/severe neuropathy (defined by either patient reported "quite a bit" or "very much tingling" and investigator-reported CTCAE 2-3) with a kappa score.

B1. Minimal important difference

Clinical relevance of the differences and clinically meaningful changes in HRQoL were assessed by minimal important difference (MID). The MID threshold for clinically relevant differences *between* treatment arms at one specific time point was defined as > 5 points, which is a generally accepted threshold.[3] Clinically meaningful change from baseline *within* arms was based on the calculated standard error of measurements (SEM) for multi-item scales[4] and Cohen´s criteria for medium effect size for single-item scales.[5,6] 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. Calculated MID thresholds for each subscale can be found in *Online Supplementary Table B1*.

Currently, anchor based methods for the assessment of clinically meaningful changes within arms [7] and clinically meaningful differences between arms [3] are increasingly being used in the QoL field. These methods have the advantage of defining what change/difference is minimally important, or, what is the effect size of the HRQoL change. Unfortunately, anchor based guidelines for clinically relevant HRQoL changes are only available for the QLQ-C30 and not the QLQ-MY20 subscales. Therefore, it was decided to use these anchor based MID-values for *within arm* changes *and between arm* differences (according to Cocks et al) only as a supplement for the calculated distribution-based method.[3,7]

For each arm separately, we calculated the percentage of patients who improved and deteriorated by more than MID in HRQoL from their baseline and from the start of maintenance (T2).[8] Percentages of improvement or deterioration were compared between arms using the chi-square test.

Supplemental Table B1. MID thresholds for clinically meaningful HRQoL change from baseline/start maintenance

Health-related quality of life subscales	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
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.

B2. Impact of early treatment discontinuation

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.[9,10] Therefore, HRQoL data could be missing not at random (MNAR) and be informative since the mechanism might be due to the missing HRQoL result.[11,12]

Fortunately, several patients returned a questionnaire after study treatment discontinuation. This allowed us to compare HRQoL between patients who discontinued treatment and patient who were still on treatment at that time. Therefore, patients who discontinued treatment and completed a questionnaire after the off protocol date were matched to patients who were on protocol at a similar treatment phase. Patients were matched on treatment arm (equal), treatment response (±1, for example PR vs VGPR), and baseline WHO-status (±1). Patients who did not start maintenance were also matched to the number of induction cycles (±1 cycle), while patients who started maintenance were also matched on time on maintenance (±3 months), and number of induction cycles (±2 cycles). The time periods

in which the off protocol questionnaires were returned and the reason for treatment discontinuation are tabulated in *Online Supplementary Table G1*. HRQoL of patients on and off protocol was compared via the independent samples *t*-test (p-value <0.005 was considered statistically significant to account for multiple testing).

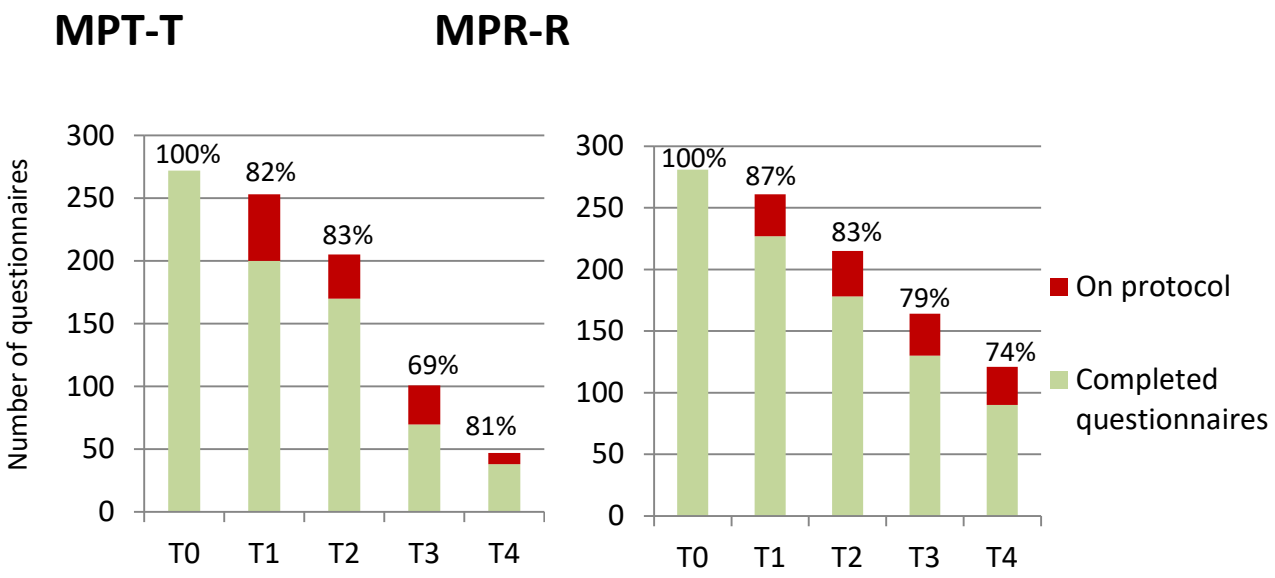
As recommended by Bell and colleagues[9], we explored the impact of missing data by comparing the HRQoL course over time between patients who discontinued treatment "early" (e.g. during induction therapy) and patients discontinuing treatment "late" (e.g. after starting maintenance therapy) or never. 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).

Finally, differences in course of HRQoL were compared between patients discontinuing treatment due to peripheral neuropathy and patients completing at least 12 months of maintenance therapy. Again, a linear mixed model was used, including fixed effect for time, group (off protocol due to peripheral neuropathy vs on protocol until 12 months maintenance) and their two-way interaction and a random intercept for subject. A p-value <0.005 was considered statistically significant to account for multiple testing.

C. Compliance

Supplemental Figure C1. Number of patients on protocol and questionnaires completed at each time point

The bars represent the number of patients on protocol at each time point (T0 to T4). The red bars represent the number of patients on protocol, but without having returned a questionnaire at that specific time point. The percentages above the bars indicate the compliance rate at that specific time point (also reflected by the green bars). Of note, since only patients with at least a baseline (T0) questionnaire were included, compliance at baseline is 100%.



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

D. Baseline scores

Supplemental Table D1. Mean baseline (T0) scores for all EORTC QLQ-C30 and EORTC QLQ-MY20 subscales for each treatment arm.

Values range from 0 to 100. A higher score for a functional scale or global health status/QOL represents a higher/healthier level of functioning. A higher score for a symptom scale represents a high level of symptoms/problems.
Note that there were no statistically significant differences in baseline scores between arms (all p>0.005).

Health-related quality of life subscales	MPT-T Baseline mean scores (SD) (N=272)	MPR-R Baseline mean scores (SD) (N=281)	p-value between arm difference
EORTC QLQ-C30			
Global health status/QoL	52.9 (24.1)	55.5 (25.3)	0.22
Physical functioning	57.0 (27.7)	58.6 (28.2)	0.51
Role functioning	45.8 (35.4)	49.2 (37.1)	0.27
Emotional functioning	70.8 (21.7)	71.2 (21.7)	0.84
Cognitive functioning	80.2 (23.1)	80.4 (23.2)	0.94
Social functioning	68.3 (30.4)	68.2 (31.4)	0.97
Fatigue	48.2 (29.1)	43.8 (28.5)	0.071
Nausea and vomiting	10.5 (19.2)	8.7 (17.7)	0.23
Pain	51.1 (36.3)	47.2 (34.7)	0.19
Dyspnoea	32.5 (32.1)	25.7 (28.6)	0.009
Insomnia	30.4 (31.6)	28.1 (31.4)	0.40
Appetite loss	26.0 (33.6)	23.0 (32.1)	0.28
Constipation	22.4 (30.4)	22.1 (30.8)	0.89
Diarrhoea	10.6 (22.6)	8.8 (21.2)	0.34
Financial difficulties	5.2 (15.4)	4.5 (14.5)	0.61
EORTC QLQ-MY20			
Disease symptoms	31.2 (22.6)	31.1 (22.4)	0.93
Side effects of treatment	19.4 (13.9)	18.8 (13.6)	0.61
Future perspective	52.3 (25.2)	52.3 (25.5)	0.99
Body image	80.8 (29.3)	81.4 (28.1)	0.80
Peripheral neuropathy	10.9 (20.7)	11.7 (21.4)	0.69

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

Supplemental Table D2. Mean scores at start maintenance (T2) for all EORTC QLQ-C30 and EORTC QLQ-MY20 subscales for each treatment arm

Values range from 0 to 100. A higher score for a functional scale or global health status/QOL represents a higher/healthier level of functioning. A higher score for a symptom scale represents a high level of symptoms/problems.
For some scales a statistically significant difference in baseline T2 scores between arms was present (in **bold**).

Health-related quality of life subscales	**MPT-T** **Baseline mean scores (SD)** (N=272)	**MPR-R** **Baseline mean scores (SD)** (N=281)	**p-value between arm difference**
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
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, 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

E. HRQoL induction and maintenance phase together

Supplemental Table E1. HRQoL course over time within each arm separately.

The **bold** p-values represent a significant change over time as well as a change from baseline that is above the threshold for calculated minimal important difference (MID) at least at one time point.

Health-related quality of life scales	Change over time *within* MPT-T*	Largest HRQoL change	Effect size of change[$]	Change over time *within* MPR-R[*]	Largest HRQoL change	Effect size of change[$]
EORTC QLQ-C30						
Global health status/QoL	**<0.001**	13.4	medium	**<0.001**	14.8	medium
Physical functioning	<0.001	8.9	medium	**<0.001**	12.4	medium
Role functioning	**<0.001**	14.7	medium	**<0.001**	19.5	medium
Emotional functioning	**<0.001**	10.6	medium	**<0.001**	13.1	medium
Cognitive functioning	0.34	0.91	trivial	0.003	5.7	small
Social functioning	**<0.001**	14.6	medium	<0.001	12.2	medium
Fatigue	**<0.001**	-13.8	medium	**<0.001**	-11.3	medium
Nausea and vomiting	<0.001	-5.9	small	0.001	-4.6	small
Pain	**<0.001**	-23.7	large	**<0.001**	-24.5	large
Dyspnoea	0.18	-5.4	small	<0.001	-4.7	small
Insomnia	**<0.001**	-16.5	medium	<0.001	-11.1	medium
Appetite loss	**<0.001**	-17.3	medium	<0.001	-13.4	medium
Constipation	0.003[#]	9.1	small	0.002	-8.5	small
Diarrhoea	0.012	-5.0	small	**<0.001**[#]	11.1	small
Financial difficulties	0.57	-2.6	trivial	0.77	1.2	trivial
EORTC QLQ-MY20						
Disease symptoms	<0.001	-11.1	NA	<0.001	-10.0	NA
Side effects of treatment	<0.001[#]	4.5	NA	<0.001	-3.7	NA
Future perspective	**<0.001**	15.8	NA	**<0.001**	19.3	NA
Body image	0.50	-2.4	NA	0.13	3.5	NA
Peripheral neuropathy	**<0.001**[#]	39	NA	0.001[#]	6.9	NA

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, NA; not applicable

[*] *for visualisation of clinical meaningful (MID) changes in HRQoL within arms, please also be referred to manuscript Figure 3 and Online Supplementary Figure D1. MID thresholds for within arm changes are provided in Online Supplementary Table B1.*

[$] *Effect size of the change within arms for the QLQ-C30 subscales, according to Cocks*[7]

[#] *represents a statistically significant worsening in HRQoL over time.*

Supplemental Table E2. HRQoL course over time between arms.

The **bold** p-values represent a significant between-arm difference over time as well as a clinically meaningful difference (MID) of > 5 points at least at one time point.

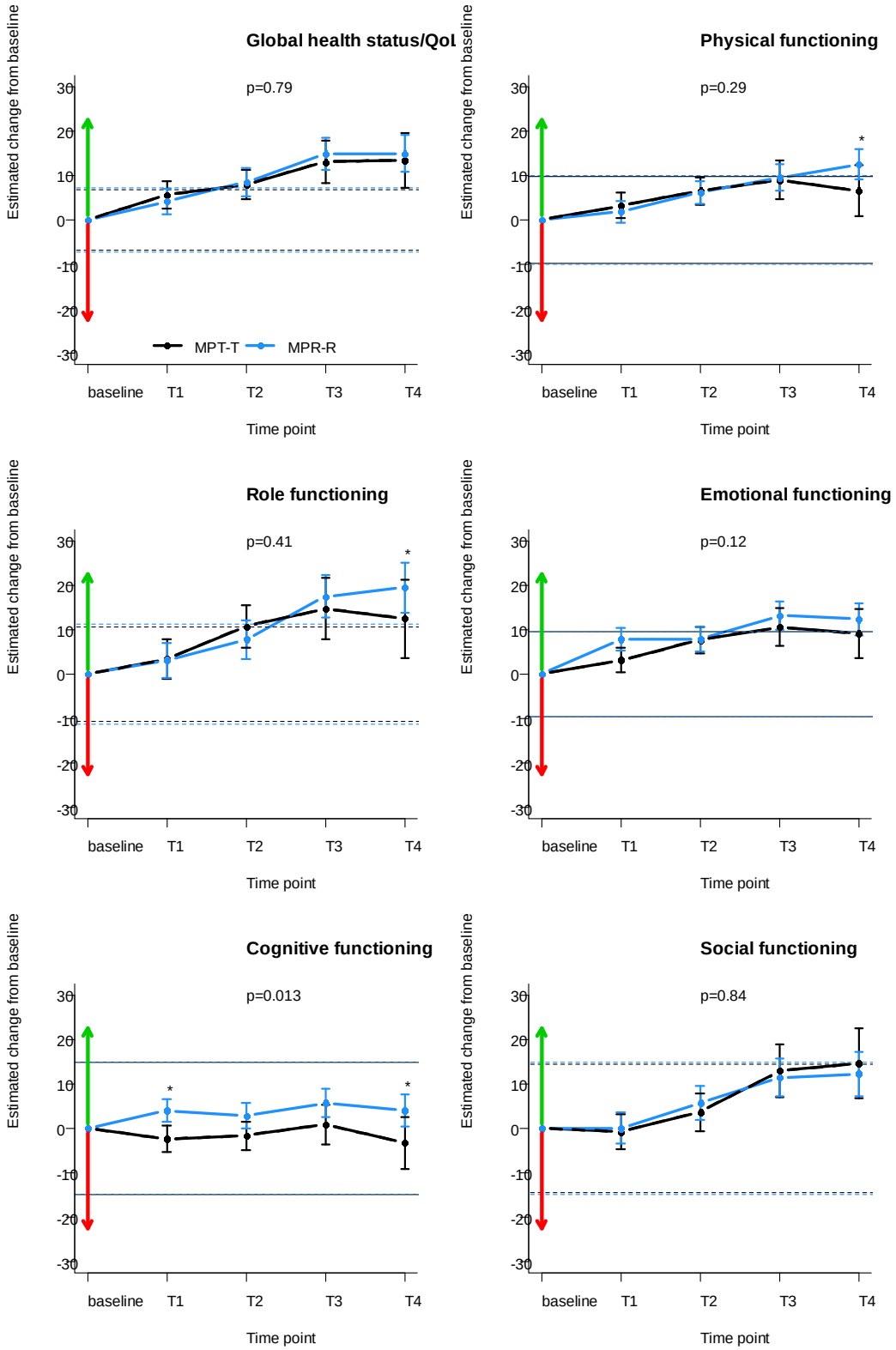
Health-related quality of life scales	Change over time *between* arms*	Largest between-arm difference	Favouring arm	Effect size of difference[$]
EORTC QLQ-C30				
Global health status/QoL	0.79	2.1	MPR-R	trivial
Physical functioning	0.29	6.2	MPR-R	small
Role functioning	0.41	7.4	MPR-R	small
Emotional functioning	0.12	4.7	MPR-R	NA
Cognitive functioning	0.013	7.3	MPR-R	small
Social functioning	0.84	2.3	MPR-R	trivial
Fatigue	0.17	5.6	MPT-T	small
Nausea and vomiting	0.86	1.8	MPT-T	trivial
Pain	**0.004**	9.6	MPT-T	small
Dyspnoea	0.56	4.2	MPT-T	small
Insomnia	**0.004**	10.0	MPT-T	small
Appetite loss	0.29	7.4	MPT-T	small
Constipation	0.008	12.3	MPR-R	small
Diarrhoea	**<0.001**	14.3	MPT-T	medium
Financial difficulties	0.84	2.3	MPT-T	trivial
EORTC QLQ-MY20				
Disease symptoms	0.38	3.1	MPT-T	NA
Side effects of treatment	**0.003**	5.5	MPR-R	NA
Future perspective	0.32	5.5	MPR-R	NA
Body image	0.11	5.1	MPR-R	NA
Peripheral neuropathy	**<0.001**	32.9	MPR-R	NA

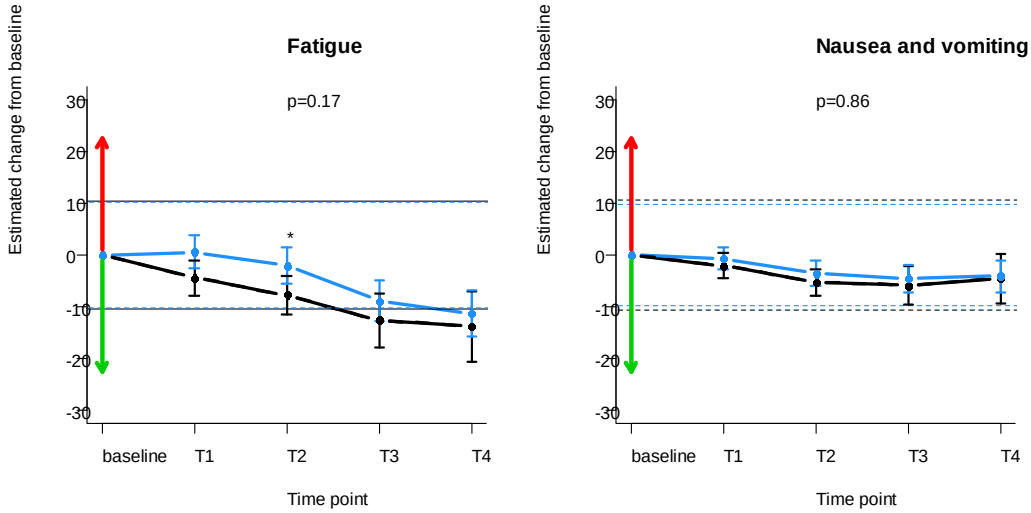
EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20, MPT-T; melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy, MPR-R; melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy, NA; not applicable

** for visualisation of differences in HRQoL course between arms, please also be referred to manuscript Figure 3 and Online Supplementary Figure D1*

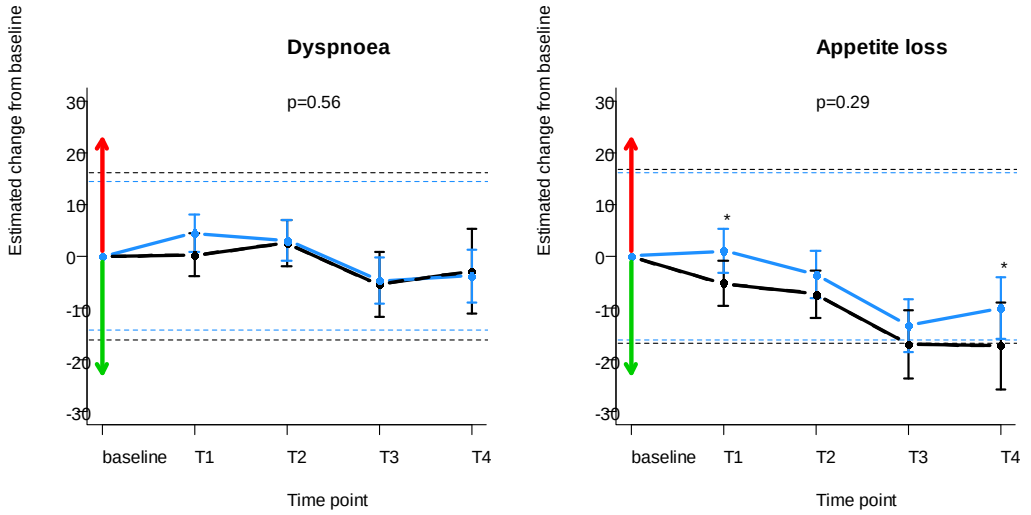
[$] *Effect size of the difference between arms for the QLQ-C30 subscales, according to Cocks[3]*

Supplemental Figure E1. HRQoL over time between and within treatment arms

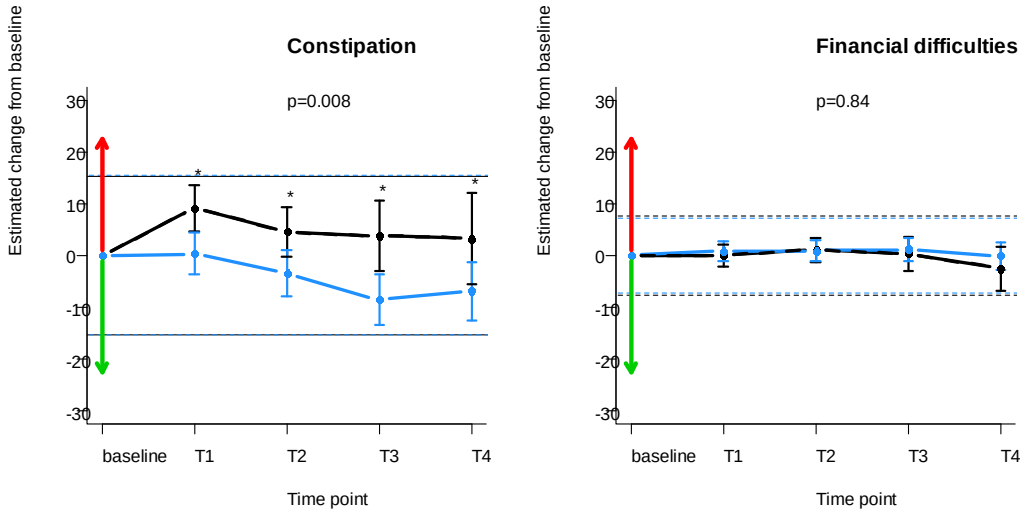



Fatigue
p=0.17
Estimated change from baseline
30
20
10
0
-10
-20
-30
*
baseline
T1
T2
T3
T4
Time point

Nausea and vomiting
p=0.86
Estimated change from baseline
30
20
10
0
-10
-20
-30
baseline
T1
T2
T3
T4
Time point


Dyspnoea
p=0.56
Estimated change from baseline
30
20
10
0
-10
-20
-30
baseline
T1
T2
T3
T4
Time point

Appetite loss
p=0.29
Estimated change from baseline
30
20
10
0
-10
-20
-30
*
*
baseline
T1
T2
T3
T4
Time point


Constipation
p=0.008
Estimated change from baseline
30
20
10
0
-10
-20
-30
*
*
*
*
baseline
T1
T2
T3
T4
Time point

Financial difficulties
p=0.84
Estimated change from baseline
30
20
10
0
-10
-20
-30
baseline
T1
T2
T3
T4
Time point

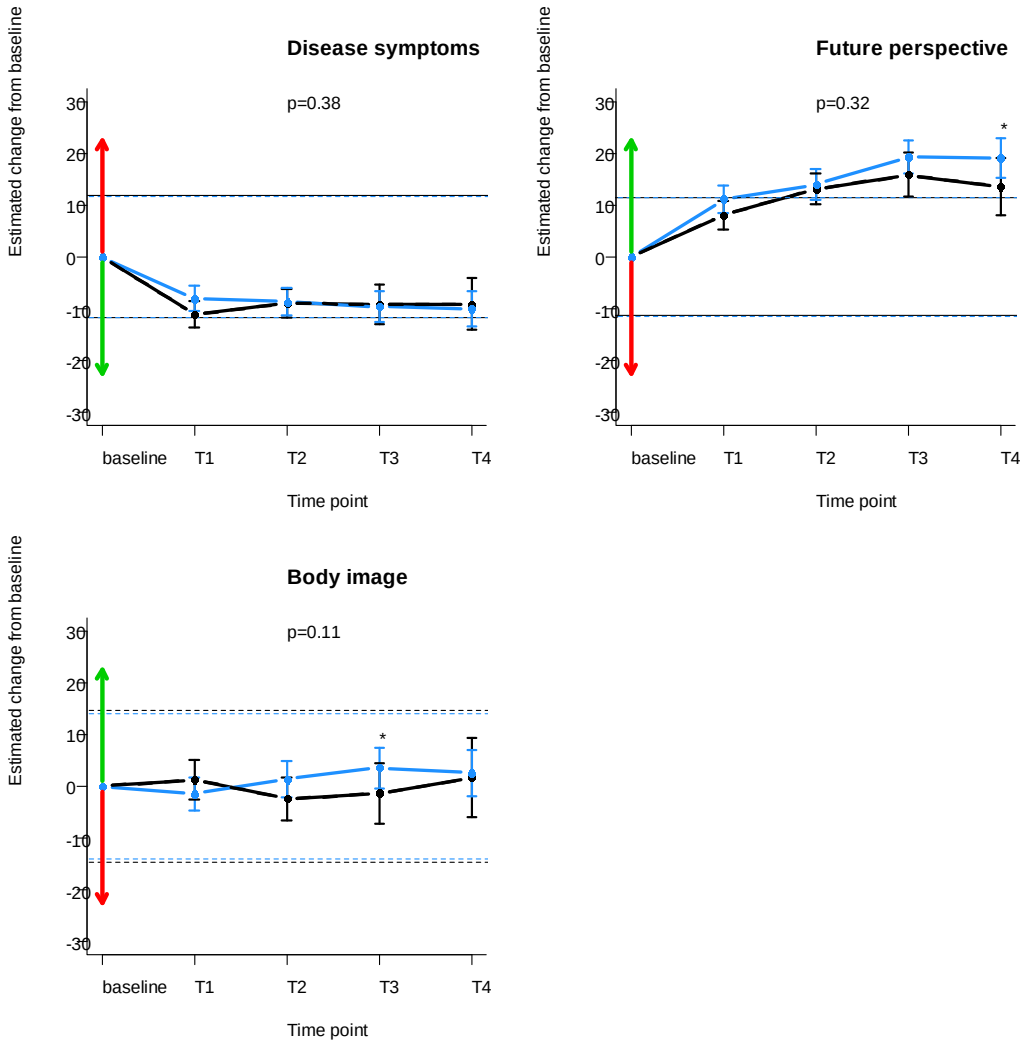


Figure E1. Estimated change in HRQoL score from baseline with corresponding 95% confidence intervals for the 15 scales with no statistically significant difference in change over time between treatment arms.

The p-values correspond to the difference in change over time *between* arms. Time points with clinical relevant difference between arms are marked with *. The dashed horizontal line represents the calculated threshold for minimal important difference, black for MPT-T and blue for the MPR-R treatment. The green arrows indicate the direction of improvement in functioning for functional scales, or reduction in symptoms for symptom scales. The red arrows indicate the direction of deterioration in functioning or worsening in symptoms.Scales with a statistically significant difference in HRQoL change over time between arms are presented in Figure 3 in the main article.

MPT-T; melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy, MPR-R; melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy

Supplemental Figure E2. Pain symptoms over time within treatment arms

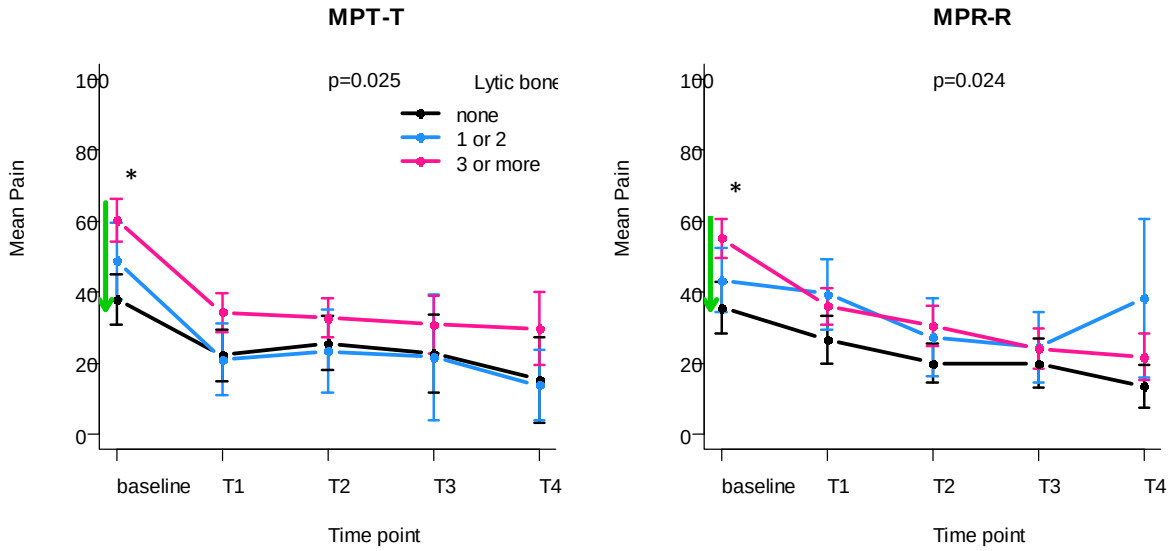


Figure E2. Estimated mean pain per time point with corresponding 95% confidence intervals between patients with no, 1-2 or 3 or more bone lesions, respectively.

The p-values correspond to the difference in change over time *between* patients with no versus 1-2 versus ≥3 lesions. The black line indicates 'no', the blue line '1 or 2' and the pink line '3 or more' bone lesions. The green arrows indicate the direction of reduction in pain symptoms.

* Statistically significant difference in baseline pain score between patients with no, 1-2 or ≥3 bone lesions (for both MPT-T and MPR-R p<0.001).

MPT-T; melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy, MPR-R; melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy

F. HRQoL maintenance phase separately

Supplemental Table F1. HRQoL course over time within each arm separately.

The **bold** p-values represent a significant change over time as well as a change from baseline that is above the threshold for calculated minimal important difference (MID) at least at one time point.

Health-related quality of li scales	Change over time *within* Thalidomide*	Largest HRQoL change	Effect size of change$	Change over time *within* lenalidomide*	Largest HRQoL change	Effect size of change$
EORTC QLQ-C30						
Global health status/QoL	0.042	4.9	trivial	**0.003**	5.1	small
Physical functioning	0.64	1.7	trivial	<0.001	5.9	Small
Role functioning	0.58	3.3	trivial	**<0.001**	10.0	small
Emotional functioning	0.20	3.4	trivial	0.032	3.7	trivial
Cognitive functioning	0.053	4.4	small	0.066	3.2	small
Social functioning	0.003	9.5	medium	0. 024	5.0	small
Fatigue	0.095	-5.0	trivial	<0.001	-8.4	small
Nausea and vomiting	0.53	1.8	trivial	0.59	-1.1	trivial
Pain	0.95	-0.57	trivial	0.028	-5.4	small
Dyspnoea	0.051	-6.9	trivial	0.004	-6.8	small
Insomnia	0.82	-0.99	trivial	0.12	-2.5	trivial
Appetite loss	0.003	-8.9	small	<0.001	-7.4	small
Constipation	0.036	-8.1	small	0.15	-3.3	trivial
Diarrhoea	0.52	3.0	trivial	0.12	4.8	trivial
Financial difficulties	0.22	-3.7	small	0.76	-0.88	trivial
EORTC QLQ-MY20						
Disease symptoms	0.71	1.2	NA	0.98	-0.3	NA
Side effects of treatment	0.34	-1.7	NA	0.019	-2.5	NA
Future perspective	0.28	3.1	NA	0.011	4.6	NA
Body image	0.74	3.0	NA	0.80	0.8	NA
Peripheral neuropathy	**<0.001**#	20.4	NA	0.89	1.3	NA

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, NA; not applicable

* *for visualisation of clinically meaningful (MID) changes in HRQoL within arms, please also be referred to manuscript Figure 3 and Online Supplementary Figure F1. MID thresholds for within arm changes are provided in Online Supplementary Table B1.*

$ *Effect size of the change within arms for the QLQ-C30 subscales, according to Cocks*[7]

\# *represents a statistically significant worsening in HRQoL over time.*

Supplemental Table F2. HRQOL course over time during maintenance between arms.

The **bold** p-values represent a significant between-arm difference over time as well as a clinically meaningful difference (MID) of > 5 points at least at one time point.

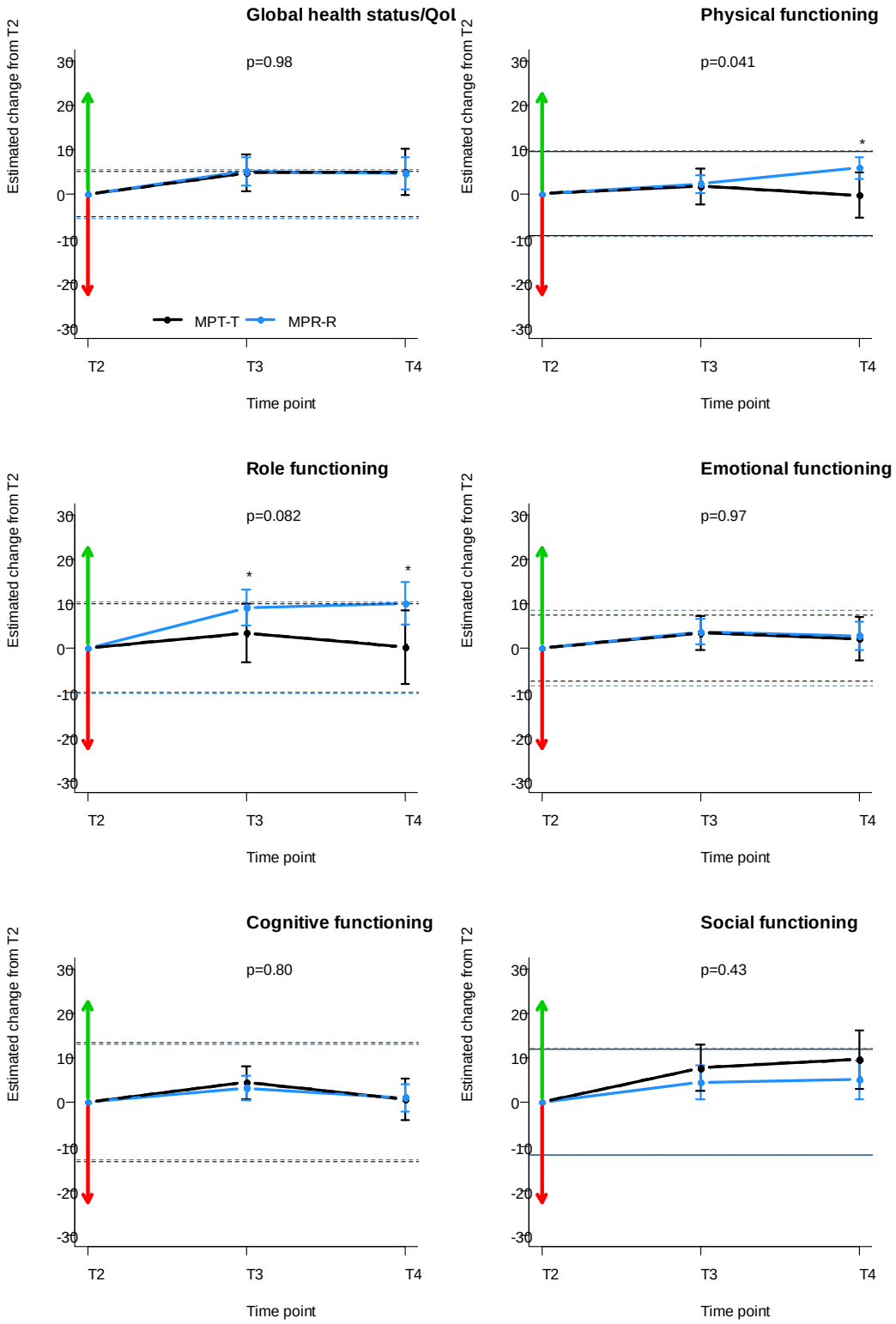
Health-related quality of life scales	Change over time *between* arms*	Largest between-arm difference	Favouring arm	Effect size of difference[$]
EORTC QLQ-C30				
Global health status/QoL	0.98	0.3	lenalidomide	trivial
Physical functioning	0.041	6.1	lenalidomide	small
Role functioning	0.082	9.6	lenalidomide	small
Emotional functioning	0.97	0.7	lenalidomide	NA
Cognitive functioning	0.80	1.2	thalidomide	trivial
Social functioning	0.43	4.5	thalidomide	trivial
Fatigue	0.58	3.4	lenalidomide	trivial
Nausea and vomiting	0.39	2.8	lenalidomide	trivial
Pain	0.26	4.8	lenalidomide	trivial
Dyspnoea	0.92	1.6	lenalidomide	trivial
Insomnia	0.31	4.3	lenalidomide	small
Appetite loss	0.28	6.1	thalidomide	small
Constipation	0.49	4.6	thalidomide	trivial
Diarrhoea	0.85	1.5	thalidomide	trivial
Financial difficulties	0.48	2.7	thalidomide	trivial
EORTC QLQ-MY20				
Disease symptoms	0.77	1.4	lenalidomide	NA
Side effects of treatment	0.86	0.8	lenalidomide	NA
Future perspective	0.84	1.5	lenalidomide	NA
Body image	0.70	3.2	thalidomide	NA
Peripheral neuropathy	**<0.001**	19.4	lenalidomide	NA

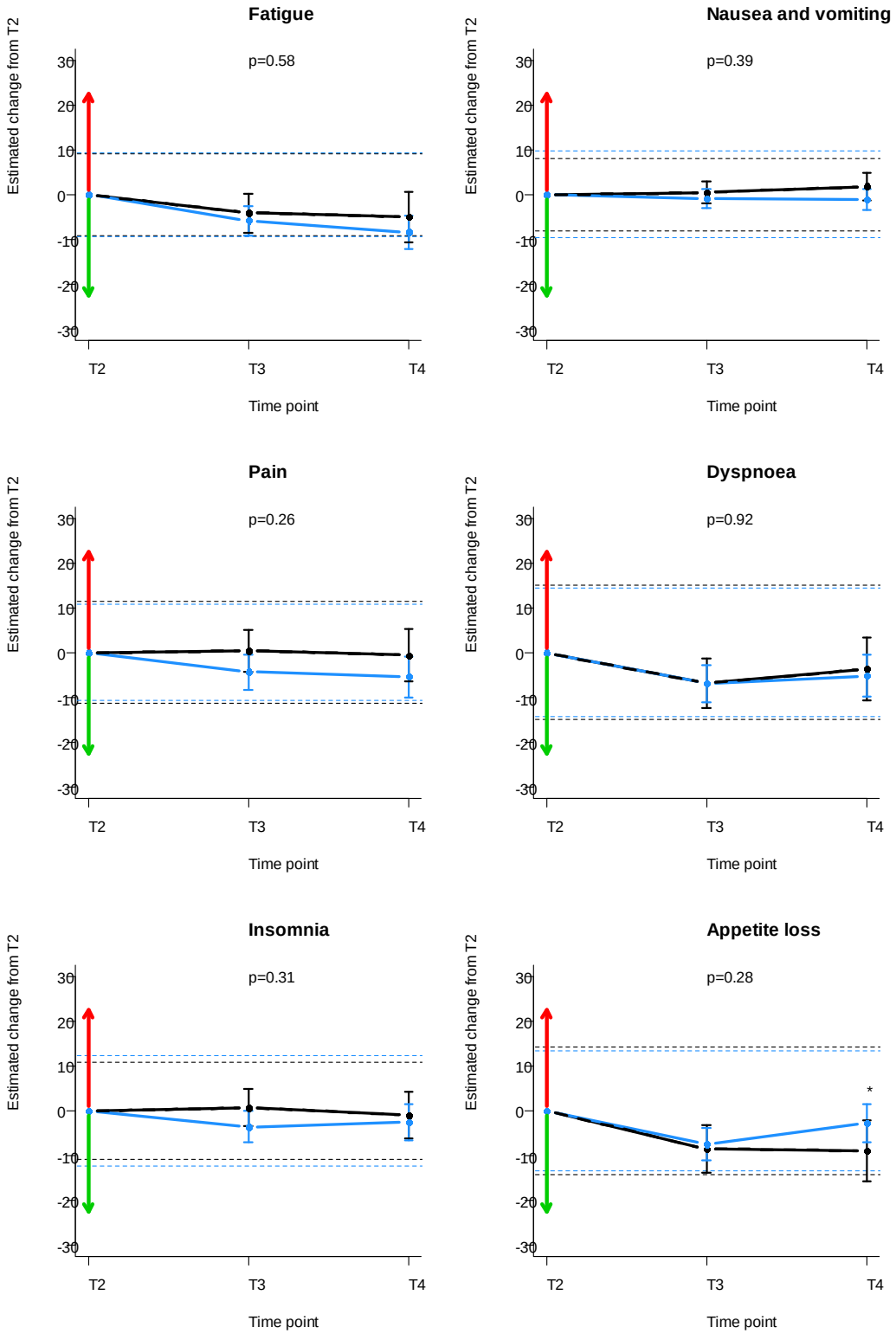
EORTC QLQ-C30; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire C30, EORTC QLQ-MY20; European Organisation for Research and Treatment of Cancer Quality of life Questionnaire MY20, NA; not applicable

** for visualisation of differences in HRQoL course between arms, please also be referred to manuscript Figure 3 and Online Supplementary Figure F1*

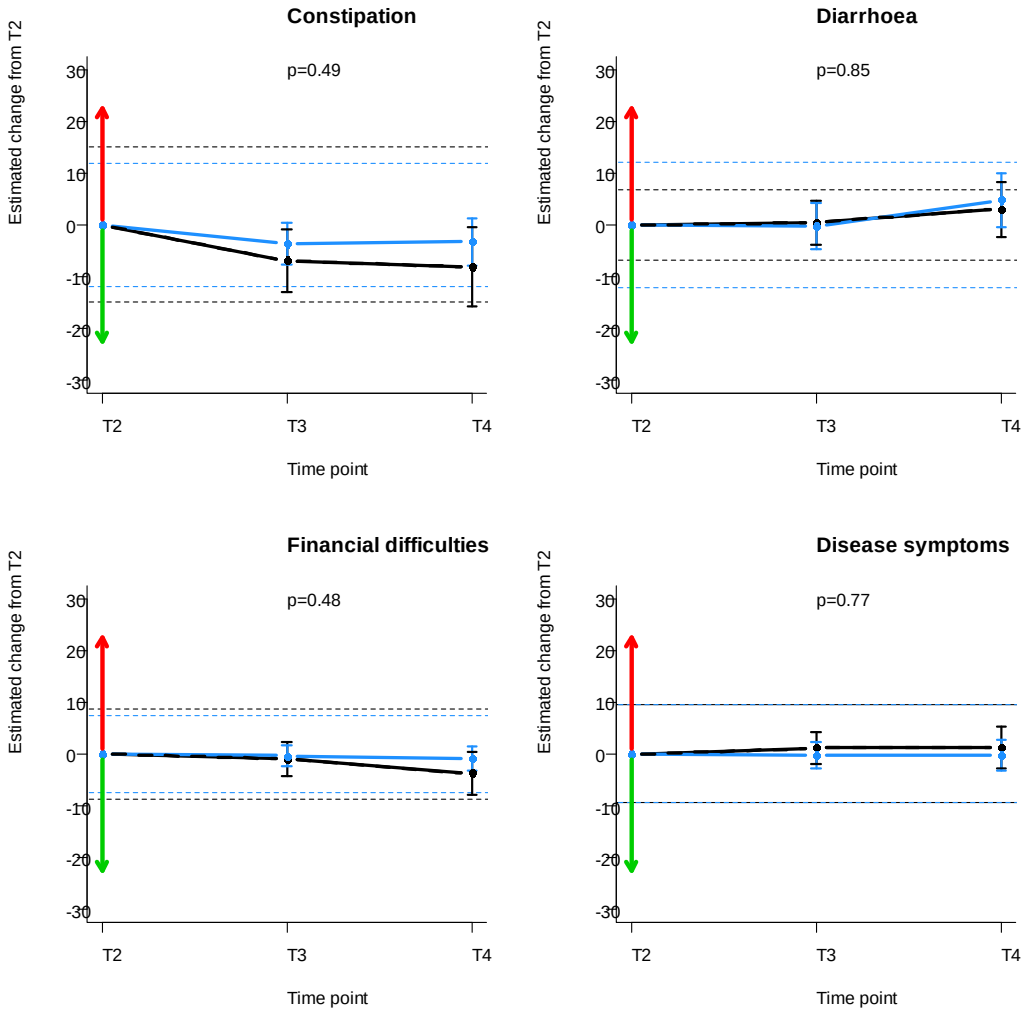
[$] *Effect size of the difference between arms for the QLQ-C30 subscales, according to Cocks*[3]

Supplemental Figure F1. HRQoL over time between and within treatment arms from start maintenance




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
*
Estimated change from T2
30
20
10
0
-10
-20
-30
T2
T3
T4
Time point

\


Constipation
p=0.49
Diarrhoea
p=0.85
Financial difficulties
p=0.48
Disease symptoms
p=0.77
Estimated change from T2
30
20
10
0
-10
-20
-30
T2
T3
T4
Time point

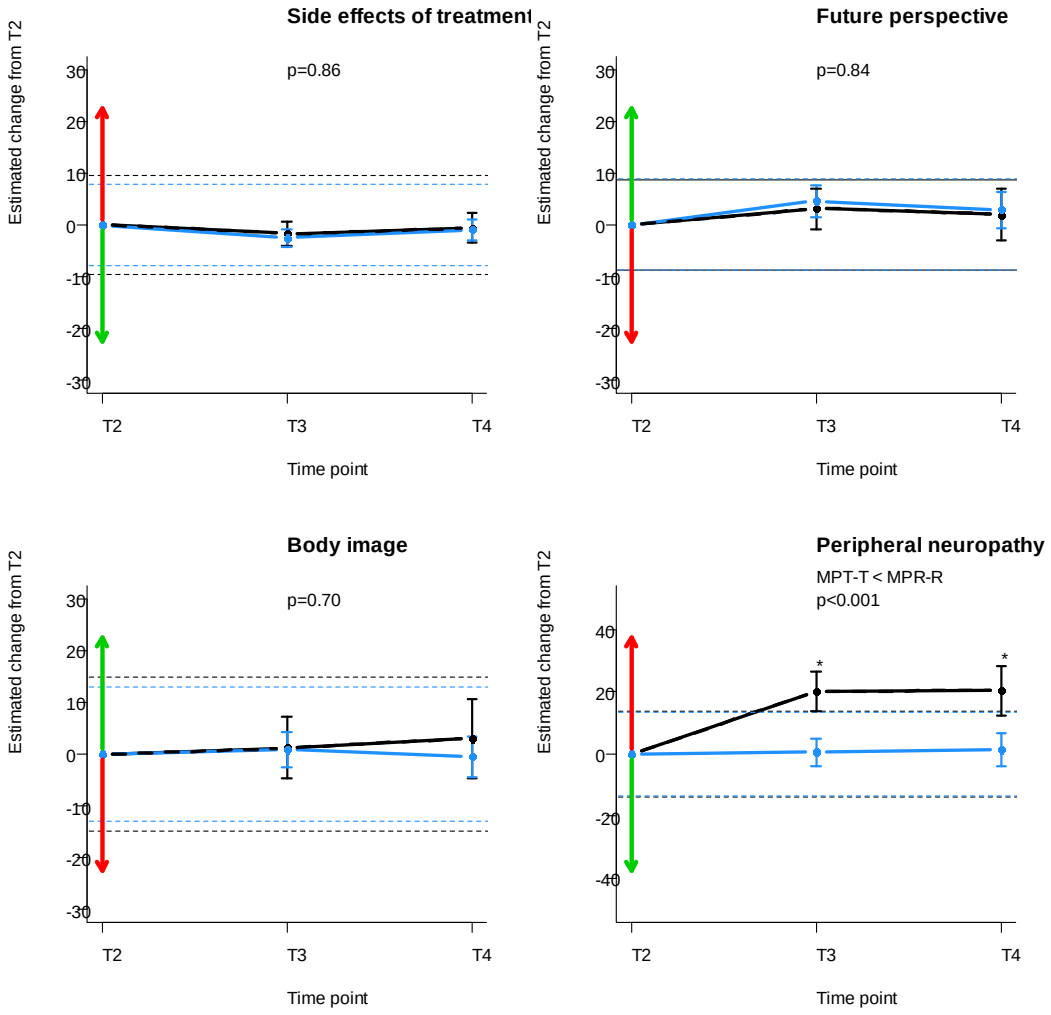


Figure F1. Graphs of the estimated mean score HRQoL change <u>from start of maintenance</u> with corresponding 95% confidence intervals and p-value for change over time between treatment arms.

Time points with clinically relevant difference in mean change from start of maintenance between arms are marked with *. 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

G. Impact of patient drop-out

Many patients discontinued treatment, due to for example progressive disease, toxicity or non-compliance. Post-protocol HRQoL assessment was not part of the protocol. As this might induce a bias, we performed several analyses to investigate whether potential bias might be present, as described in *Online Supplementary Appendix B2*.

G1. Comparison of patients on and off protocol

Results of the comparison of the course of HRQoL between patients who discontinued treatment but from whom a questionnaire was available after going off protocol, matched with patients who were still on protocol.

Supplemental Table G1A. Timing and reason for treatment discontinuation

Timing and reason for treatment discontinuation for the patients who returned a questionnaire after going off protocol.

	Reason for treatment discontinuation (TD)			
Timing of TD	**PD (n)**	**Toxicity (n)**	**Other (n)**	**Total (n)**
Between T0 and T1	0	2	3	5
Between T1 and T2	5	6	4	15
Between T2 and T3	9	22	2	33
Between T3 and T4	12	14	2	28
After T4	2	1	0	3
Total	28	45	11	84

T0; baseline, T1; after 3 induction cycles, T2; after 9 induction cycles, T3; after 6 months maintenance treatment, T4; after 12 months maintenance treatment, PD; progressive disease, n; number

Supplemental Table G1B. Mean HRQoL scores of patients off protocol matched to patients on protocol

Health-related quality of life subscales	On protocol n	On protocol Mean score	On protocol SD	Off protocol n	Off protocol Mean score	Off protocol SD	p-value
EORTC QLQ-C30							
Global health status/QoL	84	59.9	20.8	84	66.3	19.5	0.043
Physical functioning	84	62.0	24.3	84	68.0	23.6	0.10
Role functioning	83	56.6	32.2	84	58.9	32.2	0.64
Emotional functioning	84	76.6	21.0	84	78.7	21.2	0.51
Cognitive functioning	84	83.5	19.3	84	86.1	19.0	0.38
Social functioning	84	71.6	30.9	84	77.0	25.5	0.22
Fatigue	84	42.6	27.1	84	35.1	26.7	0.071
Nausea and vomiting	84	6.0	13.6	84	4.2	10.3	0.34
Pain	84	33.7	31.7	84	27.6	29.0	0.19
Dyspnoea	84	31.3	32.5	81	18.5	27.4	0.007
Insomnia	84	17.1	26.1	84	17.5	25.6	0.92
Appetite loss	83	13.7	25.0	83	11.2	24.0	0.53
Constipation	84	20.2	26.9	84	15.9	27.1	0.30
Diarrhoea	81	12.8	23.3	84	7.5	19.6	0.12
Financial difficulties	83	6.0	14.9	84	4.4	13.5	0.45
EORTC QLQ-MY20							
Disease symptoms	84	25.4	19.8	82	19.7	17.0	0.050
Side effects of treatment	84	20.3	13.5	82	16.6	13.9	0.082
Future perspective	84	39.7	23.5	81	37.3	20.1	0.48
Body image	84	80.6	26.5	81	81.1	26.3	0.90
Peripheral neuropathy	81	30.5	30.8	80	31.7	33.5	0.81

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, SD; standard deviation

G2. Timing off protocol

The results of the comparison of the HRQoL course between patients who discontinued treatment early (before or at T2) versus late/never (after T2 or never) are presented in Table G2 and Figure G2. The **bold** p-values represent a significant difference (p<0.005).

Supplemental Table G2. P-values of two-way interaction for the comparison of HRQoL course over time between patients discontinuing treatment early versus late

Health related quality of life subscale	Change over time within arm MPT-T	Change over time within arm MPR-R
EORTC QLQ-C30		
Global health status/QoL	0.49	0.31
Physical functioning	0.11	0.65
Role functioning	0.14	0.36
Emotional functioning	0.027	**0.002**
Cognitive functioning	0.12	0.97
Social functioning	0.072	0.18
Fatigue	0.27	0.077
Nausea and vomiting	0.48	0.53
Pain	0.14	0.85
Dyspnoea	0.91	0.93
Insomnia	**<0.001**	0.67
Appetite loss	0.13	0.015
Constipation	0.011	0.31
Diarrhoea	0.20	0.65
Financial difficulties	1.00	0.25
EORTC QLQ-MY20		
Disease symptoms	0.19	0.54
Side effects of treatment	0.51	0.23
Future perspective	0.60	0.23
Body image	0.61	0.24
Peripheral neuropathy	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

Supplemental Figure G2. HRQoL over time between patients discontinuing early versus late/never

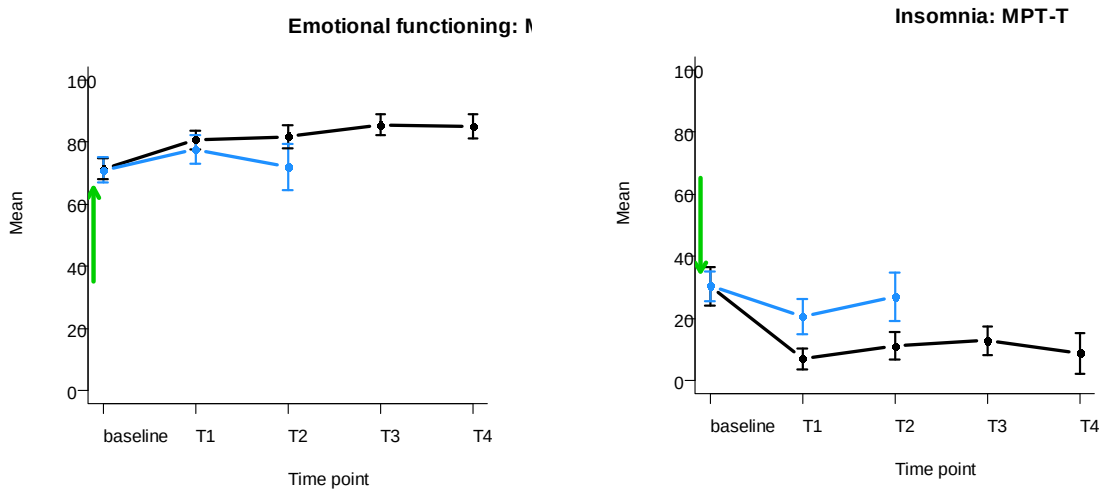


Figure G2. HRQoL course for scales with a statistically significant difference between patients who discontinued treatment early (before or at T2, blue line) versus late (after T2 or never, black line)

The green arrows indicate 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

G3. Impact of treatment discontinuation due to peripheral neuropathy

The results of the comparison of the course of HRQoL between patients who discontinued treatment due to investigator-reported peripheral neuropathy versus continuing treatment for at least 12 months of maintenance treatment, are presented in table G3 and Figure G3. The **bold** p-values represent a significant difference (p<0.005).

Supplemental Table G3. P-values of two-way interaction for the comparison of HRQoL course over time between patients discontinuing treatment due to peripheral neuropathy versus patients continuing treatment for at least 12 months of maintenance therapy

Health related quality of life subscale	Change over time within arm MPT-T	Change over time within arm MPR-R
EORTC QLQ-C30		
Global health status/QoL	0.87	0.49
Physical functioning	0.63	0.32
Role functioning	0.75	0.21
Emotional functioning	0.74	0.032
Cognitive functioning	0.20	0.23
Social functioning	0.85	0.83
Fatigue	0.90	0.20
Nausea and vomiting	0.021	0.38
Pain	0.52	0.69
Dyspnoea	0.43	0.23
Insomnia	0.22	0.20
Appetite loss	0.18	0.32
Constipation	0.009	0.18
Diarrhoea	**<0.001**	0.25
Financial difficulties	0.85	0.052
EORTC QLQ-MY20		
Disease symptoms	0.92	0.012
Side effects of treatment	0.52	0.009
Future perspective	0.99	0.050
Body image	0.39	0.98
Peripheral neuropathy	**<0.001**	0.17

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

Supplemental Figure G3. HRQoL over time between patients discontinuing due to neuropathy versus continuing treatment

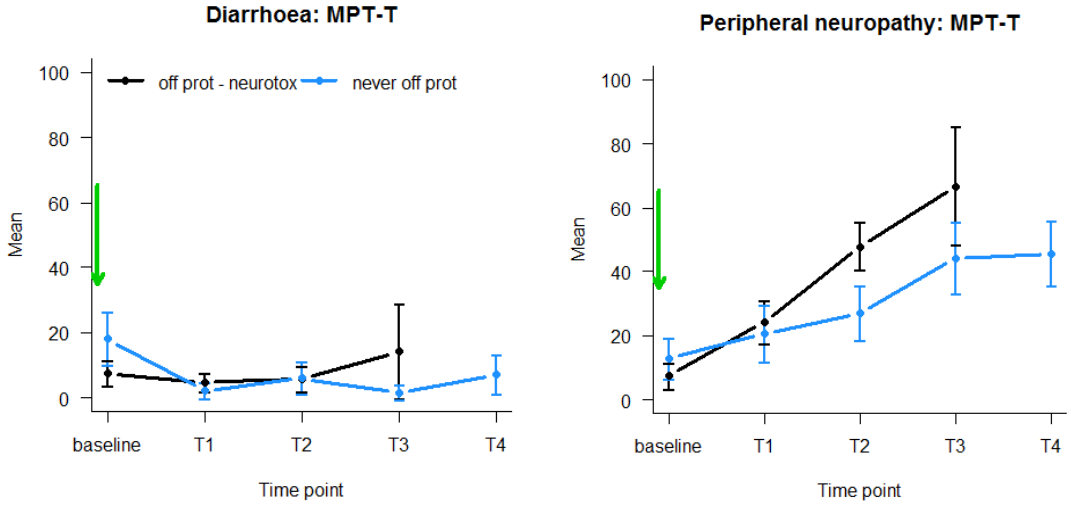


Figure G3. HRQoL course for scales with a statistically significant difference between patients who discontinued treatment due to investigator reported peripheral neuropathy (black line) versus continuation of treatment for at least 12 months of maintenance treatment (blue line)

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

MPT-T; melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy

H. Effect modification

H1. Effect modification by World Health Organization status

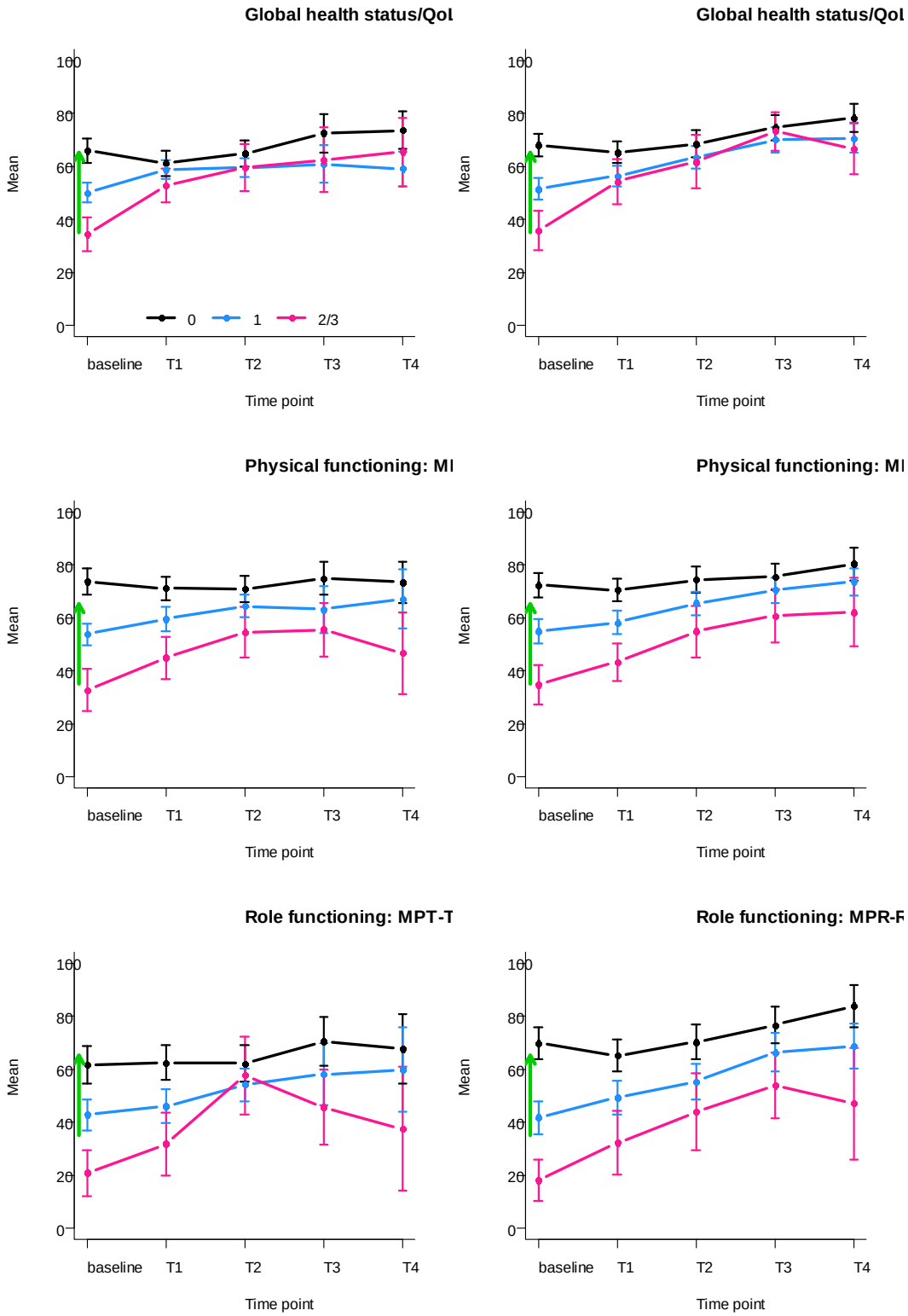
Supplemental Table H1. Comparison of HRQoL course over time between patients with a baseline World Health Organization Performance (WHO) status 0 vs 1 vs 2/3

P-values of two-way interaction for the comparison of HRQoL over time. The **bold** p-values represent a significant (p<0.005) difference in HRQoL course between patients with a WHO of 0 vs 1 vs 2/3.

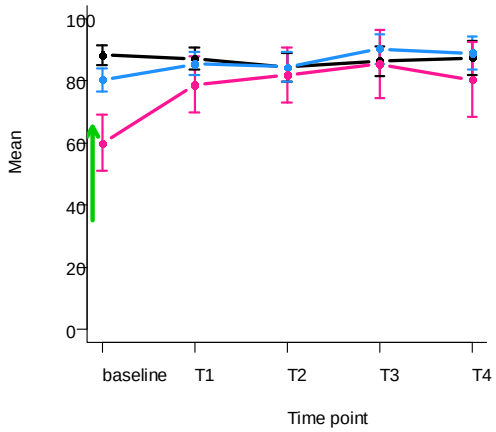
	Patients treated with MPT-T p-value	**Patients treated with MPR-R** p-value
Global health status/QoL	**<0.001**	**<0.001**
Physical functioning	**<0.001**	**<0.001**
Role functioning	**0.004**	**<0.001**
Emotional functioning	0.26	0.012
Cognitive functioning	0.015	**<0.001**
Social functioning	0.030	**<0.001**
Fatigue	**<0.001**	**<0.001**
Nausea and vomiting	0.066	**<0.001**
Pain	**0.001**	**<0.001**
Dyspnoea	0.51	0.12
Insomnia	0.88	0.18
Appetite loss	**0.001**	**<0.001**
Constipation	0.015	**<0.001**
Diarrhoea	0.17	0.74
Financial difficulties	0.55	0.59
Disease symptoms	0.12	0.010
Side effects of treatment	0.22	**<0.001**
Future perspective	0.15	0.64
Body image	0.64	0.078
Peripheral neuropathy	0.79	0.18

MPT-T; melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R; melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy

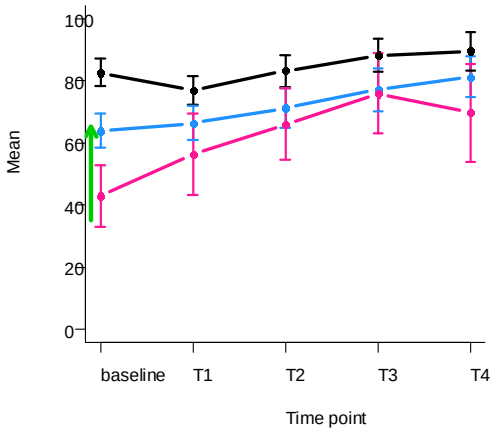
Supplemental Figure H1. Effect modification by WHO status within arms



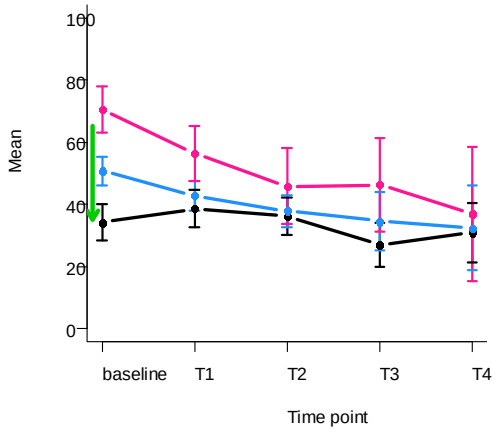
Cognitive functioning: M



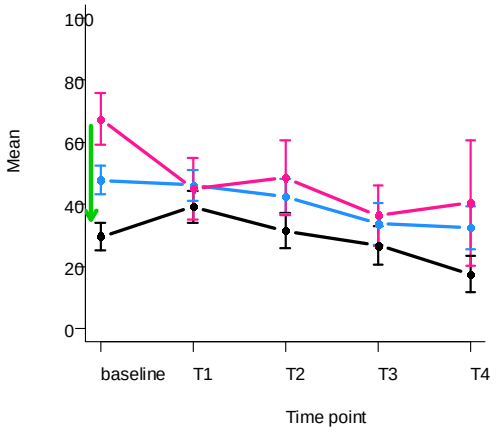
Social functioning: MPR



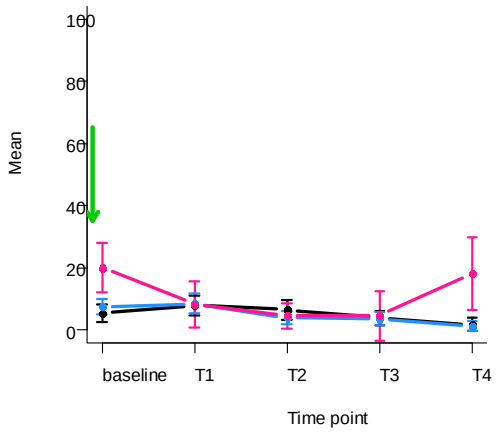
Fatigue: MPT-T



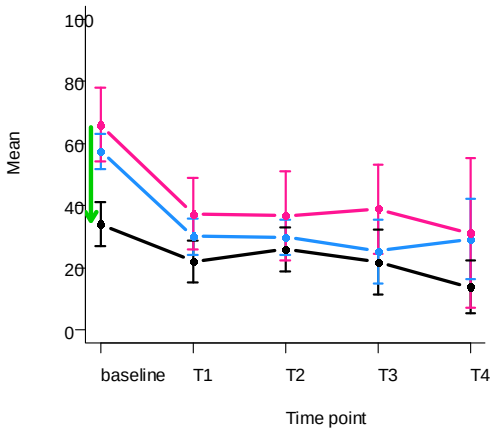
Fatigue: MPR-R



Nausea and vomiting: M



Pain: MPT-T



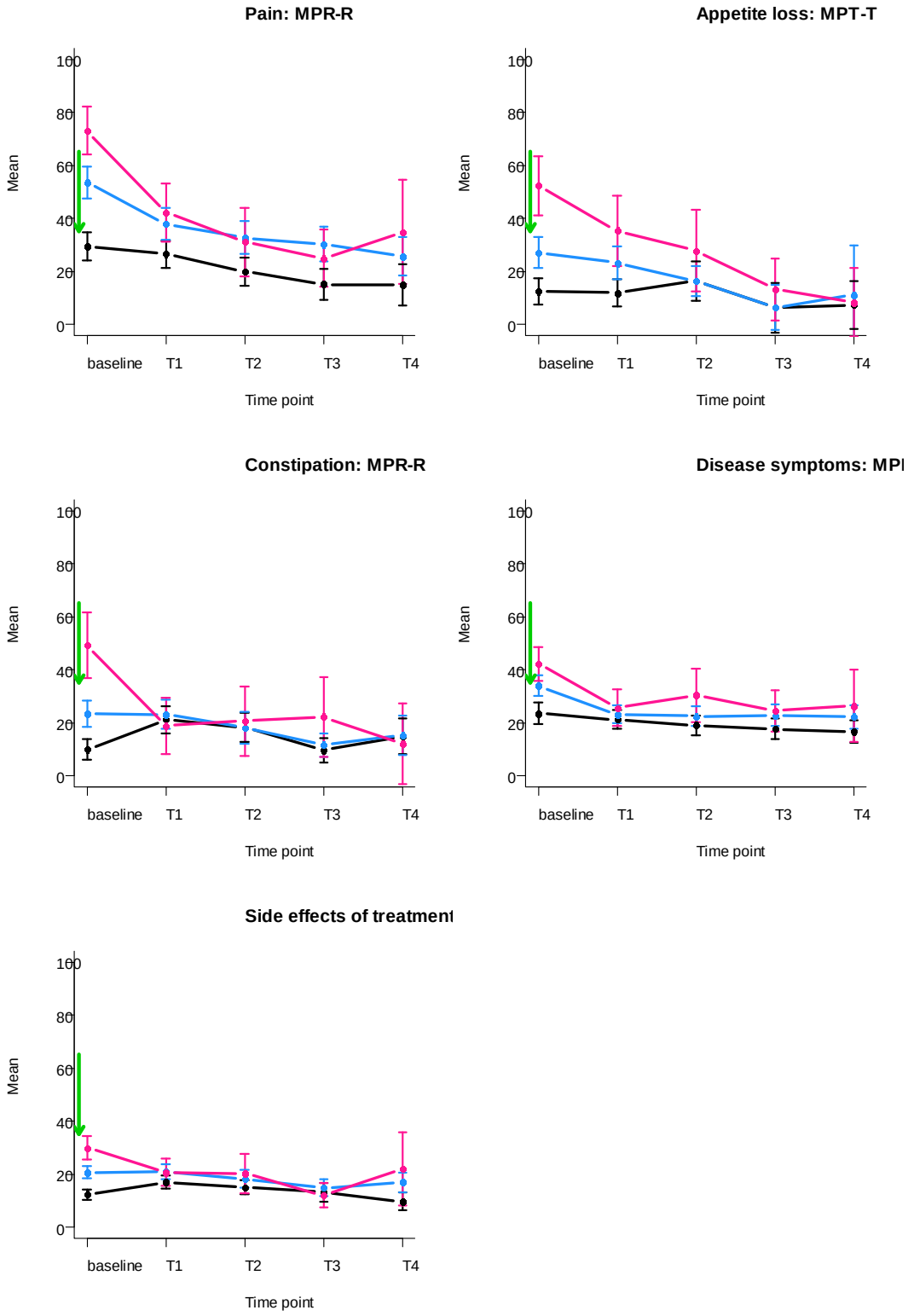


Figure H1. Effect modification by WHO status

Mean HRQoL scores for each time point with corresponding 95% confidence intervals for subscales with effect modification by World Health Organization (WHO) performance status 0 vs 1 vs 2/3 evaluated at baseline. The black curves represent the patients with WHO status 0, the blue curves the patients with WHO status 1 and the pink lines

the patients with WHO status 2/3 at baseline. The green arrows indicate the direction of improvement in functional scales and symptom reduction for symptom scales.
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

H2. Effect modification by age

Supplemental Table H2. P-values of two-way interaction for the comparison of HRQoL course over time between patients aged ≤75 years versus >75 years

The **bold** p-values represent a significant difference in HRQoL course between patients aged ≤75 versus >75 years

	Patients treated with MPT-T p-value	**Patients treated with MPR-R** p-value
Global health status/QoL	0.26	0.91
Physical functioning	0.54	0.94
Role functioning	0.006	0.13
Emotional functioning	0.23	0.92
Cognitive functioning	0.34	0.50
Social functioning	0.11	0.58
Fatigue	0.76	0.36
Nausea and vomiting	0.65	0.57
Pain	0.28	0.79
Dyspnoea	0.56	0.47
Insomnia	0.014	0.71
Appetite loss	0.61	0.23
Constipation	0.72	1.00
Diarrhoea	0.27	0.76
Financial difficulties	0.97	0.42
Disease symptoms	0.40	0.28
Side effects of treatment	0.71	0.75
Future perspective	0.41	0.51
Body image	0.031	0.98
Peripheral neuropathy	**<0.001**	0.06

MPT-T; melphalan-prednisone-thalidomide induction and thalidomide maintenance therapy; MPR-R; melphalan-prednisone-lenalidomide induction and lenalidomide maintenance therapy

Supplemental Figure H2. Effect modification by age

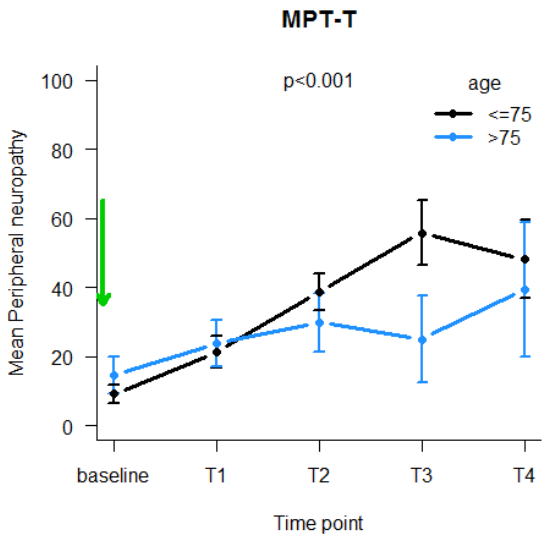


Figure H2. Effect modification by age Mean HRQoL scores for each time point with corresponding 95% confidence intervals for subscales with effect modification by age ≤75 vs >75 years evaluated at baseline. The black curves represent the patients aged ≤75 years and the pink lines the patients aged >75 years at baseline. The green arrows indicate the direction of improvement in functional scales and symptom reduction for symptom scales.
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

I. References

1. Fayers P, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A. The EORTC QLQ-C30 Scoring Manual (3rd edition). 2001.
2. National Cancer Institute NIoH. Common Terminnology Criteria for Adverse Events (CTCAE) Version 3.0, revised December 12, 2003. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm: US Department of Health and Human Services; 2003.
3. 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.
4. 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.
5. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale N.J.: Lawrence Erlbaum Associates; 1988.
6. 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.
7. 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): 1713-21.
8. Osoba D, Bezjak A, Brundage M, et al. 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.
9. 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.
10. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. Bmj 2013; 346: e8668.
11. Rubin DB. Inference and Missing Data. Biometrika 1976; 63(3): 581-92.
12. Fairclough DL. Design and analysis of quality of life studies in clinical trials. 2. ed. ed. Boca Raton: Chapman & Hall/CRC; 2010.