Statistical Analysis Plan (SAP) for the CARFI trial
-
Analysis of secondary endpoints: patient-reported outcomes

SAP contributors: Henrik Rode Eshoj[1,5], Madeleine T King[1, 2], Niels Abildgaard[1], Henrik Gregersen[3], Lene Kongsgaard Nielsen[1,4], Sören Möller[5]

Affiliations:
[1]Quality of Life Research Center, Department of Haematology, Odense University Hospital, Denmark
[2]School of Psychology, University of Sydney, Australia
[3]Department of Haematology, Aalborg University Hospital, Denmark[4]
[4]Department of Internal Medicine and Cardiology, Regional Hospital Viborg, Denmark
[5]OPEN - Open Patient data Explorative Network, Odense University Hospital and University of Southern Denmark, Denmark

Publication date: 08.09.2020

Citation for published version: Henrik Rode Eshoj, Madeleine T King, Niels Abildgaard, Henrik Gregersen, Lene Kongsgaard Nielsen, Sören Möller
Statistical analysis plan for patient-reported outcomes of the CARFI trial: a multi-centre randomized controlled trial of Carfilzomib-cyclophosphamide-dexamethasone and High-dose Melphalan (HDT) followed by Randomization between Observation or Maintenance with Carfilzomib and Dexamethasone in Patients with Relapsed Multiple Myeloma after HDT

Abbreviations

ASCT autologous stem cell transplantation
C30 Core 30 items
CI confidence interval
EORTC European Organisation for Research and Treatment of Cancer
FACT/GOG-Ntx Functional Assessment of Cancer Therapy/Gynaecologic Oncology Group/Neurotoxicity
GHS/QoL Global health status/Quality of life
HDT High-dose Melphalan
HR hazard ratio
HRQL health-related quality of life
ITT Intention to Treat
Kd carfilzomib-dexamethasone
KRd carfilzomib-lenalidomide-dexamethasone
MAR missing at random
MID minimally important difference
MM multiple myeloma
MMRM mixed effects model for repeated measures
MNAR missing not at random
MY20 Multiple Myeloma Module
NTC non-transplantation consolidation
OS overall survival
PFS progression free survival
PRO patient-reported outcome
QAPFP quality-adjusted progression-free-period
QLQ Quality of Life Questionnaire
QLU-C10D Quality of Life Utility Measure-Core 10 dimensions
RCT randomized controlled trial
Sum Sc Summary Score
TTP time-to-progression
Vd bortezomib-dexamethasone

Table of contents

ADMINISTRATIVE INFORMATION

Title & trial registration

- Statistical analysis plan for patient-reported outcomes of the CARFI trial: a multi-centre randomized controlled trial of Carfilzomib-cyclophosphamide-dexamethasone and High-dose Melphalan (HDT) followed by Randomization between Observation or Maintenance with Carfilzomib and Dexamethasone in Patients with Relapsed Multiple Myeloma after HDT
- Trial registration: *ClinicalTrials.gov ID: NCT02572492*

Version

- 1.0. Date: 04.09.2020

Protocol version

- This document has been written based on the study protocol version of the CARFI trial, dated 15.07.2015

Roles and responsibility

- Chief trial investigator: Henrik Gregersen, MD, PhD
  Chief investigator of the patient-reported outcomes (PRO): Henrik Rode Eshoj, PT, PhD
  Senior HRQL advisor: Madeleine T King
  Senior statistical advisor: Sören Möller, OPEN, OUH
  Contributors to this document: Lene Kongsgaard Nielsen, Niels Abildgaard
  Database manager: Lars Søgaard, OPEN, OUH

Signatures

- Chief trial investigator signature:
  Date: 7 Sep 2020

Statistician signature:
Date: 07.09.2020

Chief PRO investigator:
Date: 7.9.2020

1 INTRODUCTION

1.1 Background

Patients with multiple myeloma (MM) report high symptom burden and poor health-related quality of life (HRQL) compared with those with other hematologic malignancies (1, 2). More than half of MM patients report fatigue and pain due to bone lesions and about a third experience **shortness of breath**, **appetite loss**, **insomnia**, and **constipation** (2). With the introduction of high-dose chemotherapy and subsequent autologous stem cell transplantation (ASCT) therapy, both the survival and HRQL in younger MM patients has improved (3, 4). However, since MM is an incurable disease, patients with MM will experience disease relapse sooner or later. For fit and younger MM patients salvage ASCT has therefore been introduced. However, the outcome of salvage ASCT has shown contradicting results in terms of survival (5). The risk of disease relapse therefore remains high following salvage ASCT and there is a need for improvement of the response of salvage ASCT.

The overarching objective of the CARFI trial was therefore to determine whether salvage ASCT and subsequent carfilzomib-dexamethasone (Kd) maintenance therapy prolonged the time to progression, compared to salvage ASCT and observation only. The primary trial endpoint was time-to-progression (TTP) with patients being followed until disease progression, unacceptable toxicities, withdrawal or end of study. The main study hypothesis was that Kd maintenance therapy would prolong the TTP. In addition, it was also assumed that Kd maintenance therapy would be well tolerated and not adversely impact on HRQL. HRQL was therefore a predefined secondary endpoint in the maintenance phase of the CARFI trial, and was evaluated by means of patient-reported outcome (PRO) measures appropriate to this patient group and the trial interventions.

The date for the data cut-off (equal to the last patient's last visit, i.e. nine months after randomization) was the 1 September 2019. The primary analysis showed that Kd maintenance therapy post salvage ASCT significantly prolonged the TTP by approximately eight months compared to observation only post salvage ASCT.

This document specifies the statistical analysis plan for the secondary endpoint of HRQL. These analyses are intended to support the interpretation of the primary endpoint. Specifically, the positive effect of Kd maintenance therapy on TTP should be balanced against any drug-related side effects as experienced by patient and reflected in the PRO data.

1.2 Rationale for PRO objectives and hypotheses in the CARFI trial

1.2.1 Impact of autologous stem cell transplantation (ASCT) on HRQL of MM patients

A recent systematic review found that MM patients who receive ASCT experience an immediate deterioration in HRQL, specifically worse functioning scores and increased symptom burden (6). The review also revealed that the adverse impact of ASCT was only short-lived and patients

recovered their HRQL functioning scores and returned to pre-ASCT symptom levels as early as one to two months after undergoing ASCT therapy.

1.2.2 Impact of salvage autologous stem cell transplantation (ASCT) on HRQL of MM patients

To the best of our knowledge, only one study has described the impact of salvage ASCT on HRQL using PRO measures in relapsed MM patients.

In a phase III randomized controlled trial (n=288, the UK Myeloma X trial), published in 2019, relapsed MM patients were randomized between salvage ASCT after a prior ASCT or non-transplantation consolidation (NTC) with oral cyclophosphamide once weekly (7-9). HRQL was evaluated with two PRO measures, the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire – Core 30 items (QLQ-C30) and the Quality of Life Multiple Myeloma questionnaire (QLQ-MY20). Data was collected at randomization (to salvage ASCT or NTC), at day 100, six and 12 months after random assignment, and then annually until last randomized patient reached two years post randomization.

Overall, the Myeloma X trial showed that salvage ASCT was superior to NTC in terms of TTP. Within-groups analysis revealed that the group randomized to salvage ASCT experienced an immediate deterioration in HRQL (functioning and symptom scores), but also that most of the HRQL scores (functioning and symptom) were restored to baseline health status (pre-salvage ASCT therapy) two to three months post salvage ASCT therapy. Specifically, **emotional functioning** and symptoms of **insomnia** and **constipation** recovered the fastest (within the first three months post salvage ASCT) while Global health status/Quality of life (**GHS/QoL)**, **role functioning** and **body image** in addition to **symptoms of nausea/vomiting** began to recover three months post ASCT. For **social functioning** and **symptoms of fatigue**, **appetite loss** and **dyspnea** a steady recovery throughout the first six months post salvage ASCT was seen.

1.2.3 Impact of maintenance therapy following autologous stem cell transplantation (ASCT) on HRQL of MM patients

To the best of our knowledge, no studies have investigated the effect of maintenance therapy with a Kd-based treatment regimen after salvage ASCT. A recent study compared HRQL after first-line ASCT between patients at baseline (30-100 days post ASCT and no maintenance therapy) and patients >100 days post ASCT receiving maintenance therapy (monotherapy or combination therapy with the drugs lenalidomide, bortezomib, dexamethasone) or no maintenance therapy. The study found that HRQL (various aspects of functioning and symptoms assessed by the PRO measure(s) EORTC QLQ-C30 and QLQ-MY20 and the generic EuroQoL five dimension (EQ-5D)) in the patients receiving maintenance therapy was not associated with a notable detriment, compared to no maintenance, except for clinically relevant worse scores of **diarrhea** and **future perspectives** (10).

1.2.4 Impact of Carfilzomib-dexamethasone (Kd)-based regimens on HRQL of MM patients

A recent, large randomized controlled trial (RCT) (n=929, ENDEAVOR trial) comparing Kd with bortezomib-dexamethasone (Vd) in relapsed/refractory MM patients reported that Kd, relative to Vd, significantly prolonged progression free survival (PFS) and improved overall survival (OS) (11, 12).

PROs were included as exploratory endpoints, evaluated with the EORTC QLQ-C30, QLQ-MY20 and the Functional Assessment of Cancer Therapy/Gynaecologic Oncology Group/Neurotoxicity (FACT/GOG-Ntx) sub-scale instrument, and collected at randomization and at day 1 of each 28-day treatment cycle until disease progression, withdrawal of consent or if patients received another anticancer treatment (13) .

The longitudinal within-group changes for the Kd group in the ENDEAVOR trial is of special interest, since this is the only study that describes the impact of Kd on PROs in relapsed MM patients. The ENDEAVOR trial found statistically significant and clinically meaningful within-Kd-group **worsening** in the PRO domains **fatigue** from week 48 (=11 months), **role functioning** from week 60 (=15 months) and **physical functioning** at week 72 (=17 months). No other of the PRO subscales reached the thresholds for clinically meaningful changes from baseline.
It should be noted, that the Kd therapy in the ENDEAVOR trial was different to the current CARFI trial. Whereas the dosages for carfilzomib (56 mg/m^2) and dexamethasone (20 mg) were identical, the drugs were given more extensively in the ENDEAVOR trial compared to the CARFI trial. In addition, the starting point for treatment with Kd may not be the same either, since the MM patients in the ENDEAVOR trial did not undergo ASCT, which is in contrast to the current CARFI trial, where all patients receive salvage ASCT prior to Kd therapy.

In a recent RCT (n=792; the ASPIRE trial), published in 2015, carfilzomib-lenalidomide-dexamethasone (KRd) were compared to lenalidomide-dexamethasone in relapsed/refractory MM patients (14). Besides the primary endpoints, PFS and OS, PROs were evaluated with the EORTC QLQ-C30 and QLQ-MY20 instruments (15), collected at randomization, and subsequently at day 1 of each 28-day treatment cycle. Carfilzomib (27 mg/m^2) was given six times and dexamethasone (40 mg) four times during each cycle.

The ASPIRE study found that median PFS and OS was significantly prolonged in the KRd group compared to the control group (P<0.001) (14). With regards to the PRO data, within-group changes in the KRd-arm revealed statistically, but not clinically meaningful improvements from baseline in the following PRO domains: **GHS/QoL**, **physical functioning**, **role functioning**, **fatigue**, **nausea/vomiting**, **pain**, **adverse effects of treatment** and **disease symptoms** (15).

Of note, besides variations in treatment intensity between the ASPIRE and CARFI trial (different drug dosages), the MM patients in the ASPIRE trial nor had received ASCT, which may limit the basis for direct comparison between the two trials.

1.2.5 Impact of carfilzomib-dexamethasone-based regimen on toxicity and peripheral neuropathy of MM patients

A phase I-II trial, investigating the effect of carfilzomib, lenalidomide and dexamethasone in newly diagnosed MM patients, found clinician-reported peripheral neuropathy (any grade) in 23-39% of the patients. The peripheral neuropathy was, however, mostly of only grade 1 to 2 according to common terminology criteria for adverse events (CTCAE) evaluation (16, 17).
Furthermore, a recent systematic review, summarizing the effect of carfilzomib-based regimens in MM patients, demonstrated that Kd-based regimens has a favorable toxicity profile (18), with the most frequent symptomatic adverse events of any grade to carfilzomib being **fatigue**, **nausea**, **dyspnoe** and **diarrhea** (19).

1.2.6 Impact of dexamethasone-based regimens on HRQL

Corticosteroids, like dexamethasone, are frequently used in treatment combinations for the myeloma disease (20). However, the downside of corticosteroids is a number of side effects, i.e. **psychological disturbances**, **depression**, **anxiety**, **agitation** and **irritation** and **swelling tendencies** due to fluid retention, especially following long-term use (21).

1.3 PRO and HRQL summary

In conclusion, the evidence indicates that treatment with ASCT initially impacts core aspects of HRQL (GHS/QoL, role, social, physical functioning, fatigue, dyspnea, body image, appetite loss), but that most of these recover to pre-ASCT levels of HRQL by three to six months. In addition, although data are scarce, maintenance therapy post ASCT seems not to adversely impact HRQL, expect for worse scores for **diarrhea** and **future perspectives**. Of note, the latter is most likely drug-dependent and may be controlled through supportive care post-transplant. Specifically for carfilzomib, this drug seems to have a favorable toxicity profile and does not appear to negatively impact core aspects of HRQL. However, if provided with higher intensity (higher dosages per injection), carfilzomib may induce toxicity symptoms of **fatigue**, **nausea**, **dyspnoe** and **diarrhea.** Finally, dexamethasone alone may too give rise to a number of side effects, which may impact core aspects of HRQL such as **emotional function** and **body image** as well as symptoms of **agitation and irritation**.

2 OBJECTIVES and HYPOTHESES

Pre-specified HRQL objectives and/or hypotheses were not described a priori in the CARFI study protocol (version 1, 2015). Hence, the stated PRO objectives and hypothesis described below were developed post-hoc after study initiation. Further, for transparency and to avoid data-driven analysis, the current SAP was finalized while the database was locked and before that PRO data was accessed. Furthermore, it was made publicly available prior to analysis.

2.1 PRO objectives and hypothesis

Primary objective and outcome:
To compare the effects of Kd maintenance therapy, relative to observation only, on changes in HRQL as assessed by the EORTC-QLQ-C30 Summary score from randomisation to eight months follow-up.

The primary endpoint for the HRQL analysis will thus be changes in the EORTC QLQ-C30 Summary Score. The primary analysis time point, eight months from randomization, is selected since the date of data cut-off was when the last patient had been followed for nine months from randomization (which was finalized in September 2019). The last PRO assessment occurred at eight months. This time point is further anticipated to be the time point that most patients will reach before experiencing progressive disease. Hence, we expect the eight months´ time point to be the time point that includes the largest proportion of recruited patients still on study in both groups, but also the time point least affected by non-ignorable missing PRO data, thus minimizing statistical bias. Finally, this time point allows us to compare HRQL when subjects had had long enough time to recover from the effects of salvage ASCT. Note that all available data (from randomisation and until progression/death/drug discontinuation/end of study progressive disease) will be presented/graphed for all outcomes, as described below.

> Hypothesis:
> *There will be no difference between the two groups in change from randomisation to eight months follow-up of the EORTC-QLQ-C30 Summary score.*

Secondary objectives and outcomes:
Secondary objectives of the study include exploration of the effects on other important PRO domains. For each PRO, the effect of Kd maintenance therapy will be compared to observation alone. The analysis metrics vary among domains, as described below.

1. Change in HRQL as assessed by the individual domains of the instruments EORTC QLQ-C30, EORTC MY20 and FACT/GOG-ntx subscale from randomisation to eight months follow-up.

> Hypothesis:
> *There will be no difference in HRQL, except for the domains potentially impacted by Kd-related side effects: "Nausea/vomiting","Fatigue","Dyspnea" and "Diarrhea"*

and a poorer "Body Image", compared to patients randomised to observation only. Even for these five domains, the differences will not exceed the threshold for clinically relevant differences.

2. The proportion of patients who have improved, remained stable and worsened in HRQL, as assessed by the functional domains (except cognitive functioning) and GHS/QoL scale of the EORTC QLQ-C30 and body image domain of the EORTC QLQ-MY20, from randomisation to eight months follow-up.

Hypothesis:
The proportion of patients that improved/remained stable/worsened will be similar between the two groups

3. The time to first recorded improvement in HRQL, as assessed by the functional domains (except cognitive functioning) and GHS/QoL scale of the EORTC QLQ-C30 and the EORTC MY20 body image domain, from randomisation to eight months follow-up.

Hypothesis:
Patients randomised to Kd-therapy will take a longer to time to achieve significant clinically relevant improvement in emotional functioning, role functioning and social functioning as well as in the body image domain, compared to observation only, but will not differ in time to improve in physical functioning and GHS/QoL.

4. The proportion of patients with Kd-related patient-reported symptoms, and the proportion of patients with Kd-related moderate to severe symptoms of "fatigue", "dyspnea", "nausea/vomiting", "insomnia" and "agitation and irritation", from randomisation to eight months follow-up.

Hypothesis:
A larger proportion of patients randomized to Kd therapy will report more severe levels of symptoms compared to observation only

5. The time in quality-adjusted progression-free-period (QAPFP) (time in remission) from randomisation to progression/death/drug discontinuation/end of study using two quality adjusted utility scores: **1)** The EORTC QLQ-C30 Summary Score and, **2)** The EORTC Quality of Life Utility Measure-Core 10 dimensions (QLU-C10D), a preference-based multi-attribute health state classification based on EORTC QLQ-C30. At the time of writing this SAP, the Danish QLU-C10D value set was not available; the German value set will be used if the Danish QLU-C10D value set is not available at the time of analysis.

Hypothesis:
QAPFP will not differ substantially from the difference observed in PFS since we do not expect Kd to impact significantly on HRQL

3 STUDY DESIGN

3.1 Design

The CARFI trial is a phase II, parallel group RCT, multi-centre study investigating the TTP in relapsed MM patients treated with salvage ASCT and subsequent maintenance therapy or observation only.

3.2 Randomization and blinding

Two months after salvage ASCT patients were randomized (1:1) to maintenance therapy with intravenous Carfilzomib 27 mg/m^2 and dexamethasone 20 mg every second week or observation only. The maintenance dose of Carfilzomib was escalated to 56 mg/m^2 after 4 weeks provided acceptable side effects. The randomization procedure was stratified according to relapse 1-2 year or >2 years after the primary ASCT, multiple myeloma international staging system and standard versus high-risk cytogenetics. Due to the open-label design, blinding of clinicians and patients was not possible.

4 PATIENT-REPORTED OUTCOMES

4.1 Administration and timing of PRO assessments

PRO data were collected at randomization (two months post salvage ASCT) and then every second month until progression/death/drug discontinuation/end of study (last patient, last visit, which was nine months from randomization). If an otherwise scheduled visit was postponed (due to vacation, public holidays, acute toxicities beyond grade 2), so was the PRO data completion. Thus the PRO data completion followed the clinic visit, and the time windows between the PRO completions could therefore potentially be longer than two months if visits were postponed.

The PRO instruments were completed electronically or in hard copy when patients were at the outpatient clinic for study protocol visit. Study nurses had the opportunity to report reasons for non-completion of questionnaires, but this was not mandatory. Therefore, complete recording of reasons for non-compliance is not expected.

4.2 PRO measures

The rationale for including the PROs (the EORTC QLQ-C30, QLQ-MY20 and the FACT/GOG-ntx subscale) was that the EORTC QLQ-C30 and QLQ-MY20 instruments were the most frequently used instruments for HRQL measurement in patients with MM (22, 23). They therefore provide the greatest opportunity for comparison with previous MM studies. The FACT/GOG-ntx was included to investigate peripheral neuropathy, which is a frequent side effect of anti-myeloma treatment and not covered by the two other instruments. Translated instruments were available for use in all of the participating countries.

4.2.1 EORTC QLQ-C30

The EORTC QLQ-C30 instrument is a cancer-generic questionnaire that consists of 30 items, combined into 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). Each item is scored on a four-point categorical scale. The domain scores will be calculated according to the EORTC manual and translated into a scale of 0-100 (24) . A high score represents good HRQL for functional domains, and a low score presents low degree of symptoms for the symptom domains. The EORTC QLQ-C30 has been validated in patients with MM (25). The questionnaire uses a recall period of 7 days.

Besides the individual domains of the EORTC QLQ-C30, two summary scores from the QLQ-C30 instrument will also be derived.

The EORTC QLQ-C30 Summary Score (QLQ-C30 Sum Sc) is an established summary score calculated from the mean of 13 (27 items in total) out of 15 QLQ-C30 domains (GHS/Global QoL and Financial impact are excluded). Prior to calculating the mean, the symptom scales will be reversed to obtain a uniform direction of all scales. After reversing the appropriate items, scores will be summarised and divided by the number of domains (=13). The QLQ-C30 Sum Sc ranges from 0-100, with 100 being best (26).

The EORTC QLU-C10D is a multi-attribute utility instrument, which allows a preference-based HRQL summary score to be calculated from data collected with the EORTC QLQ-C30 instrument. The QLU-C10D contains 13 items of the QLQ-C30´s 30 items, and 10 of its 15 domains: physical functioning (mobility), role functioning, social functioning, emotional functioning, pain, fatigue, sleep, appetite, nausea, bowel problems. The severity of impairment in each dimension is expressed by the four categories used in the QLQ-C30: “not at all”, “a little”, “quite a bit”, “very much”. An algorithm is used for the calculation of health-state utilities according to the 10 dimensions (27). The QLU-C10D will be used as a second approach to a HRQL summary score because this is arguably more appropriate than the QLQ-C30 Sum Sc for quality adjusting progression-free-period. Also, it is methodologically interesting to include both types of summary measure, and compare them (e.g. correlation) as an exploratory descriptive analysis, as they are both relatively new.

4.2.2 EORTC QLQ-MY20

EORTC QLQ-MY20 consists of four domains: two functional domains (future perspective and body image) and two symptom domains (disease symptoms and side effects of treatment) (28, 29). The module is administered in addition to the EORTC QLQ-C30 questionnaire and all domains, except for body image, are multi-item domains. The domain scores will be calculated according to the EORTC manual and translated into a scale of 0-100 (24) . A high score represents good HRQL

for functional domains, and a low score represents a low degree of symptoms for symptom domains. The questionnaire uses a recall period of 7 days.

4.2.3 FACT/GOG-ntx subscale

The FACT/GOG-ntx subscale is a single domain 11-item questionnaire (30). The questionnaire has been used in MM patients for evaluation of treatment-related PN (31). The instrument is scored from 0-44 and a higher score means a lower degree of PN. The questionnaire has a recall period of 7 days. The score will be calculated in accordance with the FACT scoring guidelines including instructions on how to handle missing data (30). For the FACT/GOG-ntx subscale a between-group MID that exceeds 11.8 point will be interpreted as a clinically meaningful between-group difference (32). However, others have estimated the MID to be no more than 4.4 points (33). Thus, a conservative strategy for interpreting clinical meaningful difference in neuropathy will be used. Since the Kd treatment is expected to be a low toxicity treatment regimen a MID threshold of 4.4 points will be used as the primary MID. However, a sensitivity analysis with the more stringent MID threshold of 11.8 points will also be conducted.

4.3 Primary endpoint

The primary endpoint for the HRQL analysis will be the EORTC QLQ-C30 Sum Sc. The reason for choosing the QLQ-C30 Sum Sc as the primary endpoint is as follows: 1) there is currently no standard HRQL assessment for use in MM maintenance trials; 2) the QLQ-C30 Sum Sc captures the most generic HRQL concepts of importance; and 3) the use of an overall score avoids issues with multiplicity. However, since this is a fairly new score and has not yet been subjected to extensive psychometric validation, the individual QLQ-C30 domains must be analysed as secondary outcomes to allow in depth clinical interpretation of the contributions of each of the subscales to the overall QLQ-C30 Sum Sc. Each PRO subscale will be interpreted according to clinically relevant minimal important differences (MID) using the evidence-based guidelines for between-group differences (34) and change over time (35). The smallest thresholds for MIDs between groups will be used to interpret the clinical importance of the results for each of the key PRO subscales of the EORTC QLQ-C30 instrument. However, as a minimally important difference (MID) has not yet been established for the QLQ-C30 Sum Sc, a 10 point change/difference will be used as an indicator of a between-group MID on this endpoint (36, 37). Thus, a change from baseline of $\geq +10$ points will be considered a clinically relevant improvement, a change of ≤ -10 points will be considered a clinically relevant worsening, and changes below +10 and -10 points will be considered no clinically relevant changes.

4.4 Secondary endpoints

1) The 15 domains of the EORTC QLQ-C30

   a. Five functional domains
      i. Physical
      ii. Role
      iii. Social

iv. Emotional
v. Cognitive

b. Nine symptom domains
i. Fatigue
ii. Nausea and vomiting
iii. Pain
iv. Dyspnoea
v. Insomnia
vi. Appetite loss
vii. Constipation
viii. Diarrhoea
ix. Financial difficulties

c. Global health state/QoL

2) The four domains of the EORTC MY20

a. Two functional domains
i. Future perspective
ii. Body image

b. Two Symptom domains
i. Disease symptoms
ii. Side effects of treatment

3) The FACT/GOG-ntx subscale

4) Kd-related single-item patient-reported symptomatic side effects

a. Single items from the EORTC QLQ-C30 instrument assessing
i. Dyspnea; *"Where you short of breath" (item 8)*
ii. Insomnia; *"Have you had trouble sleeping"(item 11)*
iii. Fatigue; *"Did you need to rest" (item 10); "Have you felt weak" (item 12); "Were you tired" (item 18)*
iv. Nausea/Vomiting; *"Have you felt nauseated" (*item 14), "*Have you vomited"(*item 15)
v. Diarrhea; *"Have you had diarrhea"(item 17)*

b. Single item from the EORTC MY20 instrument assessing

i. Restlessness and agitation; *"Did you feel restless or agitated" (item 44)*

5) The Multi-attribute utility instrument (EORTC QLU-C10D)

4.5 PRO compliance

Compliance will be calculated using the proportion of randomized patients with completed questionnaires (EORTC QLQ-C30, MY-20 and FACT/GOG-ntx) and the proportion of patients expected to complete questionnaires (alive and still on study). This will be tabulated per trial arm at baseline and then subsequently for all scheduled visits until the date of data cut-off (last patient, last visit) (Table xx).

5 STUDY POPULATION

The primary analyses will be based on the Intention to Treat (ITT) population, including all randomised participants with available data at baseline and at least one completed set of PROs at the subsequent assessment time points. The ITT principles assert the effect of a treatment policy (that is, the planned treatment regimen) rather than the actual treatment given (i.e. it is independent of adherence). This has the consequence that participants allocated to a treatment group should be followed up, assessed and analysed as members of that group irrespective of their compliance to the planned course of treatment (i.e. independent of withdrawals and cross-over phenomena). Thus, all randomized subjects will be included in the analysis population according to the treatment they were randomized to receive.

6 STATISTICAL METHODS

6.1 Sample size and power calculation

Sample size in the maintenance phase of the CARFI trial was determined according to the primary trial endpoint (TTP). According to the trial protocol, it was expected that 150 patients would continue to the maintenance phase, thus constituting the study population for this secondary analyses of patient-reported HRQL. No formal power analysis for the PROs was performed.

6.2 Framework

The main PRO objective is to determine whether Kd-therapy adversely impacts HRQL. Thus, the primary endpoint is evaluated in a non-inferiority setting with the agreement-limit for non-inferiority being the MID (as described below). The secondary PRO outcomes are evaluated for non-inferiority for the individual domains of scales and superiority for remaining outcomes.

6.3 Statistical analysis

6.3.1 Descriptive statistics

Graphs and summary statistics will be reported for each of the PROs , including subscale domains and items, by visit as well as change in scores from randomization to the end of study for all patients (PRO trajectories) still on trial at the date of data cut-off (last patient, last visit) (Table 1).

6.3.2 Primary endpoint

The impact of Kd maintenance therapy, relative to observation only, post salvage ASCT on HRQL as assessed by the EORTC QLQ-C30 Summary Score

6.3.2.1 Primary analysis

The change from baseline in QLQ-C30 Sum Sc will be the primary response variable and analyzed using a linear mixed effects model for repeated measures (MMRM) of the change from baseline in QLQ-C30 Sum Sc for each visit until eight months follow-up allowing us to compare the average effect of Kd maintenance therapy, relative to observation only. The model will include patients as a random variable and time (baseline, 2, 4, 6 and 8 months), treatment arm (Kd maintenance; observation only) and country (Denmark, Sweden, Norway, Finland, Lithuania) as fixed factors.

Coefficient for mean differences with 95% confidence intervals from the model will be presented in a table and means will be presented in a figure. Normality assumptions on residuals and random effects will be checked by quantile-quantile plots. In case of non-normality non-parametric confidence intervals will be estimated by bootstrapping with 1000 repetitions. The main analysis will include all answers from baseline to the eight months follow-up. Mean change in score from baseline and mean differences between treatment arms will be estimated with 95% confidence interval (CI) from the mixed effects models. The results will be presented in a table, or as text.

6.3.2.2 Supportive secondary analysis of the primary endpoint

A summary measure approach applied to the QLQ-C30 Sum Sc will be used as a supportive analysis of the primary endpoint. This method considers the individual as the basic unit and uses the responses for each individual subject to construct a simple number, which summarizes (some

aspects of) that subject´s response curve. Hence, calculation of mean QLQ-C30 Sum Sc. per patient using all available assessment time points will be conducted.

I.e., the average QLQ-C30 Sum Sc. of all PRO assessment per patient, averaged across **all available** time points from randomization to progressive disease/death/drug discontinuation/end of study, will be calculated. Hence, if a patient has any intermittent missing data, these will not be imputed (e.g. if the 6 month PRO data are missing the sum score will be derived as follows: baseline + 2 month + 4 month + 8 month divided by four (=number of completed assessments)). Subsequently, the per-patient average QLQ-C30 Sum Sc´s will be used as the analysis variable to compare differences between the two treatment groups with appropriate statistical testing, applying bootstrapping in case of non-fulfilled distributional assumptions. The results will be presented in a table, or as text.

6.3.3 Secondary endpoints

6.3.3.1 Individual domains of the EORTC QLQ-C30, MY20 and FACT/GOG-ntx subscale

Within and between groups comparison of change from baseline to progression/death/drug discontinuation/end of study in the secondary endpoints (all individual PRO domains) will be analysed using similar methodology to the primary analysis of the primary endpoint (Table 2).

6.3.3.2 The proportion of patients, that has improved, remained stable or worsened

The proportion of patients that have perceived improvement, remained stable or worsening from randomisation to each assessment time point (2, 4, 6, 8 months follow-up) will be evaluated using PRO data at a patient-level (38). Thus, each time point will be compared to the baseline level (as is obtained at randomisation). The results will be presented in a table, or as text.

For the EORTC QLQ-C30 multi-item domains and GHS/QoL, patients will be categorized as having perceived either an improvement or worsening from baseline if they experience a change score (in the direction of improvement respectively worsening) that exceeds at least 20 points. Hence, a stable patient will be defined as a patient with neither improvement nor worsening from baseline of more than 20 points (39). For the EORTC MY20 single-item domain, Body Image, patients will be categorized as having perceived either an improvement or worsening from baseline if they experience at least a one unit change (in the direction of improvement respectively worsening) from baseline corresponding to a score that exceeds at least 33 points. A stable patient will therefore be defined as a patient with neither improvements nor worsening from baseline of more than 33 points (40).

Between treatment groups, this endpoint will be analyzed as a binary endpoint, comparing the proportion of patients with improvements using chi-square test of homogeneity. Data will be presented as percentage of patients who experienced improvement and odds ratios with 95% CI.

6.3.3.3 Time to first recorded improvement in the functional domains and GHS/QoL scale of the EORTC QLQ-C30 and the EORTC MY20 body image domain

In this analysis, data at a patient-level is used once again (38). An event (first recorded improvement) will be calculated as time from randomisation to the first time the patient records a change (in the direction of improvement) of at least 20 points for the EORTC multi-item domains (39), and 33 points (in the direction of improvement) for the Body Image domain of the EORTC MY20 instrument (40). Patients that do not record an improvement at any time point will be censored at progression/death/drug discontinuation/end of analysis time point (eight months follow-up) (whichever comes first).

Patients with no baseline score and patients with high functioning as well as good GHS/QoL and Body Image at randomization will be excluded from this analysis since this leaves "no room for improvement". "No room for improvement" will be defined as follows: a baseline score greater than 80 for the functional domains and the GHS/QoL scale of the EORTC QLQ-C30 instrument. For the single-item domain of the EORTC MY20, Body Image, a good baseline score will be less than 33 points. Analysis will only be performed if there are a sufficient number of patients available.

The mean time to improve will be compared between the treatment groups using a proportional hazards cox regression model with treatment and baseline scores as well as country as covariates and presented as a hazard ratio (HR) with corresponding 95% CI and p-value and displayed using Kaplan-Meier curves.

6.3.3.4 The proportion of patients with Kd-related symptoms and proportion of patients with moderate to severe patient-reported symptomatic side effects

For all items, an individual raw score will be calculated in accordance with the EORTC QLQ-C30 manual for calculating raw scores for single-item symptom scales (24).

The proportion of patients reporting Kd-related symptoms will be calculated according to the following definition: a patient is defined to have a Kd-related symptom if the scale score corresponds to at least "a little" (≥33 points).

The proportion of patients reporting moderate to severe Kd-related symptoms will be calculated according to the following definition: a patient is defined to have moderate to severe Kd-related symptoms if the scale score corresponds to at least "quite a bit" (≥67 points). This dichotomized grading has previously been used in patients with hematological malignancies (2).

Proportions between treatment groups will be compared using similar statistical methodology to the analysis in section 7.4.2. The degree of symptoms at each time point (x-axis) will be presented graphically as proportions (y-axis) and will be presented in a table, or in a figure.

6.3.3.5 Time in quality-adjusted progression-free-period

If Kd maintenance therapy significantly improves the time from randomisation to progression/death/drug discontinuation/end of study, but at the same time causes substantial symptoms and/or toxicity, a measure of the net clinical benefit is indicated. This can be demonstrated by incorporating time in quality-adjusted progression-free-period (QAPFP) and HRQL into a single measure. This is defined as time in QAPFP and describes the duration of good quality of life.

A. EORTC QLQ-C30 Summary Score

In the calculation of time in QAPFP, the mean QLQ-C30 Sum Sc from randomisation to progression/death/drop-out/end of study will be derived using a mixed model for repeated measures separately for each treatment group. This model includes all QLQ-C30 Sum Sc up to the measurement closest to progression/death/drop-out/end of study. To estimate the QAPFP for each treatment arm, the mean EORTC QLQ-C30 Sum Sc will be divided by 100 so that it is on a 0-1 scale and then multiplied with the estimated mean time to progression. From these estimates restricted mean PFS time up to 9 months follow-up will be estimated by applying the Kaplan-Meier method. The difference in mean QAPFP between treatment groups will be presented with corresponding 95% CI and p-value, which will be calculated using bootstrap methods. As a sensitivity analysis QAPFP will be estimated for the full follow-up, without artificial censoring at 9 months and including all QLQ-C30 measurements. The results will be presented in text or in a table (Table xx).

B. EORTC QLU-CD10

Calculation of time in QAPFP using the EORTC QLU-CD10 and difference in mean QAPFP between the two treatment groups will be analysed using similar statistical methodology to the analysis in section 6.3.3.5 A. The results will be presented in text or a table (Table xx).

6.4.1 Missing data patterns

The primary analysis, using MMRM, is valid only under the assumption that missing data are missing at random (MAR). With an assumed faster attrition rate for the patients randomized to observation only, especially beyond the eight months follow-up, the numbers of missing data are expected to be larger for the observation only group. It cannot be rejected that these data are missing not at random (MNAR), thus potentially inflicting bias to the MMRM analysis.

6.4.1.1 Dropout (yes vs no)

An MMRM analysis will be conducted adjusting for dropout using a simple dropout variable (yes vs no) (41).

6.4.1.2 Pattern mixture models

To validate the primary analysis we will explore whether the missing data patterns are informative. We will do this by stratifying the patients randomised to Kd or observation only into groups based on their last completed PRO (for example, group 1: drop out before PRO visit 2 (baseline to month 2), group 2: drop out after PRO visit 2, but before PRO visit 3 (between month 2 and 4), group 3: drop out after PRO visit 3, but before PRO visit 4 (between month 4 and 6), group 4: full

responders (participants who completed questionnaires at PRO visit 5 (baseline to eight months) regardless of intermittently missing PRO assessments).

An MMRM analysis estimating the average score for the QLQ-C30 Sum Sc scale for each treatment group per dropout time will be conducted (42). Subsequently, the average PRO data for group 4 (full responders) will be compared to each of the drop-out groups, stratified by time of dropout (for example group 4 vs. group 2). Of note, comparison of mean PRO scores between the various dropout groups will only be conducted within treatment arms. Differences will be examined with appropriate statistical testing and deemed significant if $p < 0.05$. Furthermore, MIDs will be determined using evidence-based guidelines for interpretation of the QLQ-C30 (35). Hence, if the mean PRO score of subgroup 2 is statistically and clinically significantly different compared to group 4 (full responders), the non-responses of subgroup 2 will be considered as missing not at random (e.g. PRO scores deteriorate in the lead-up to drop-out) and the MMRM results need careful interpretation. HRQL trajectories grouped by the timing of dropout will be plotted by treatment group and graphed to assess the trends over time by missing data pattern (41).

In addition, we will also explore the missing data patterns for the whole dataset by stratifying the patients randomised to Kd or observation only into groups based on their dropout history. Early dropout will be defined as patients who drop out before the 10 months follow-up, late dropout will be defined as patients who drop out at or after 10 months follow-up and patients still in study at the date of cut-off will be defined as no drop-out. These groups will be explored with the same techniques as above. Of note, to ensure reasonable precision in mean PRO scores for each group stratified by time of dropout, only subgroups with more than 10 patients will be analyzed. The number of patients contributing to each mean will be shown in a table at the bottom of the graph.

6.6 Statistical principles

6.6.1 Level of statistical significance

All applicable statistical tests will be two-sided and will be performed using a 5% significance level for primary analyses and 1% for secondary analyses. All confidence intervals will be 95% and two-sided.

6.6.2 Adjusting for multiplicity

This SAP clearly defines one primary endpoint (average effect of Kd relative to observation only on HRQL using the mean QLQ-C30 Summary Score) and adjustment for multiplicity is therefore not needed for the primary analysis. As all other outcomes serve as supportive and/or explanatory outcomes, any adjustment for multiplicity is, by nature, not needed (43). However, p values are provided to aid interpretation, but must be interpreted with caution to account for the multiple testing. To decrease the risk of false positive interpretation, we only will interpret p-values < 0.01 as indicating evidence for differences for secondary outcomes.

6.7 Methods used for assumptions to be checked for statistical methods

Normality assumptions will be checked by normal quantile-quantile plots. In case of problems with the normality assumptions we will estimate confidence intervals and p-values by bootstrapping with 1000 repetitions as a sensitivity analysis. The proportional hazards assumption in Cox regression will be checked by Schoenfeldt residuals.

7 OPERATING PROCEDURES

7.1 Database

The PRO data are registered in the eCRF developed by the Quality of Life Research Center, Department of Hematology, Odense University Hospital. The project is approved for admission in the research infrastructure group of Odense Patient Explorative data Network (OPEN) and uses the electronic data capture platform REDCap (44). The REDCap database meets the safety requirement set by the Danish Data Protection Agency for storage of person sensitive data.

7.2 Blinded interpretation

The fact that the result of the primary endpoint in the CARFI trial (TTP) is known at the time of interpreting the secondary PRO endpoints, these interpretations may be vulnerable to prior convictions, wishful thinking and conflict of interest (45). Thus, to avoid misleading interpretation of the current PRO data a blinded interpretation process will be conducted as follows:

1) An independent database manager will code each treatment arm into “treatment A” and “treatment B” and thus leaving all others blinded from treatment during analysis.

2) Blinded data will be delivered to the statistician

3) Primary and secondary endpoint analysis will be made blinded from treatment

4) Results will be presented to the writing committee of the secondary PRO data from the CARFI trial (identical to the study chair in this SAP) where any uncertainties will be clarified. Two blinded interpretations of the primary endpoint results will be conducted prior to unblinding of data. One interpretation assumes that A is the experimental intervention (Kd) and another assumes that A is the control (observation only). After agreeing that there will be no further changes to the two interpretations, the writing committee records their decisions and signs the resulting document. Subsequently, the randomization code is broken, the correct interpretation is chosen, and the manuscript finalized (46).

8 STATISTICAL SOFTWARE

Analyses will be carried out using Stata 16.

9 REFERENCES

1. Priscilla D, Hamidin A, Azhar MZ, Noorjan KO, Salmiah MS, Bahariah K. Quality of life among patients with hematological cancer in a Malaysian hospital. The Medical journal of Malaysia. 2011;66(2):117-20.
2. 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.
3. 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. Med Oncol. 2001;18(1):65-77.
4. 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.
5. Morris C, Iacobelli S, Brand R, Bjorkstrand B, Drake M, Niederwieser D, et al. Benefit and timing of second transplantations in multiple myeloma: clinical findings and methodological limitations in a European Group for Blood and Marrow Transplantation registry study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2004;22(9):1674-81.
6. Chakraborty R, Hamilton BK, Hashmi SK, Kumar SK, Majhail NS. Health-Related Quality of Life after Autologous Stem Cell Transplantation for Multiple Myeloma. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2018;24(8):1546-53.
7. Cook G, Ashcroft AJ, Cairns DA, Williams CD, Brown JM, Cavenagh JD, et al. The effect of salvage autologous stem-cell transplantation on overall survival in patients with relapsed multiple myeloma (final results from BSBMT/UKMF Myeloma X Relapse [Intensive]): a randomised, open-label, phase 3 trial. Lancet Haematol. 2016;3(7):e340-51.
8. Cook G, Williams C, Brown JM, Cairns DA, Cavenagh J, Snowden JA, et al. High-dose chemotherapy plus autologous stem-cell transplantation as consolidation therapy in patients with relapsed multiple myeloma after previous autologous stem-cell transplantation (NCRI Myeloma X Relapse [Intensive trial]): a randomised, open-label, phase 3 trial. Lancet Oncol. 2014;15(8):874-85.
9. Ahmedzai SH, Snowden JA, Ashcroft AJ, Cairns DA, Williams C, Hockaday A, et al. Patient-Reported Outcome Results From the Open-Label, Randomized Phase III Myeloma X Trial Evaluating Salvage Autologous Stem-Cell Transplantation in Relapsed Multiple Myeloma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2019;37(19):1617-28.
10. Tay J, Vij R, Norkin M, Buadi F, Kindwall-Keller TL, Roberts JS, et al. Health related quality of life for multiple myeloma patients according to treatment strategy after autologous stem cell transplant: a cross-sectional study using EORTC, EQ-5D and MY-20 scales. Leukemia & lymphoma. 2019;60(5):1275-82.
11. Dimopoulos MA, Goldschmidt H, Niesvizky R, Joshua D, Chng WJ, Oriol A, et al. Carfilzomib or bortezomib in relapsed or refractory multiple myeloma (ENDEAVOR): an interim overall survival analysis of an open-label, randomised, phase 3 trial. The Lancet Oncology. 2017;18(10):1327-37.
12. Dimopoulos MA, Moreau P, Palumbo A, Joshua D, Pour L, Hajek R, et al. Carfilzomib and dexamethasone versus bortezomib and dexamethasone for patients with relapsed or

refractory multiple myeloma (ENDEAVOR): a randomised, phase 3, open-label, multicentre study. Lancet Oncol. 2016;17(1):27-38.
13. Ludwig H, Moreau P, Dimopoulos MA, Mateos MV, Kaiser M, Hajek R, et al. Health-related quality of life in the ENDEAVOR study: carfilzomib-dexamethasone vs bortezomib-dexamethasone in relapsed/refractory multiple myeloma. Blood cancer journal. 2019;9(3):23.
14. Stewart AK, Rajkumar SV, Dimopoulos MA, Masszi T, Spicka I, Oriol A, et al. Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. N Engl J Med. 2015;372(2):142-52.
15. Stewart AK, Dimopoulos MA, Masszi T, Špička I, Oriol A, Hájek 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 : official journal of the American Society of Clinical Oncology. 2016;34(32):3921-30.
16. Jakubowiak AJ, Dytfeld D, Griffith KA, Lebovic D, Vesole DH, Jagannath S, et al. A phase 1/2 study of carfilzomib in combination with lenalidomide and low-dose dexamethasone as a frontline treatment for multiple myeloma. Blood. 2012;120(9):1801-9.
17. Zimmerman T, Raje NS, Vij R, Reece D, Berdeja JG, Stephens LA, et al. Final Results of a Phase 2 Trial of Extended Treatment (tx) with Carfilzomib (CFZ), Lenalidomide (LEN), and Dexamethasone (KRd) Plus Autologous Stem Cell Transplantation (ASCT) in Newly Diagnosed Multiple Myeloma (NDMM). 2016;128(22):675-.
18. Mushtaq A, Kapoor V, Latif A, Iftikhar A, Zahid U, McBride A, et al. Efficacy and toxicity profile of carfilzomib based regimens for treatment of multiple myeloma: A systematic review. Critical reviews in oncology/hematology. 2018;125:1-11.
19. Siegel D, Martin T, Nooka A, Harvey RD, Vij R, Niesvizky R, et al. Integrated safety profile of single-agent carfilzomib: experience from 526 patients enrolled in 4 phase II clinical studies. Haematologica. 2013;98(11):1753-61.
20. King T, Faiman B. Steroid-Associated Side Effects: A Symptom Management Update on Multiple Myeloma Treatment. Clinical journal of oncology nursing. 2017;21(2):240-9.
21. Nathan Cherny MF, Stein Kaasa, Russell K. Portenoy, and David C. Currow. Oxford Textbook of Palliative Medicine (5 ed.): Oxford University Press; 2015.
22. 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.
23. 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. Eur J Cancer. 2007;43(11):1670-8.
24. Fayers P, Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., & Bottomley, A. . The EORTC QLQ-C30 Scoring Manual (3rd Edition). Brussels: European Organisation for Research and Treatment of Cancer. 2001.
25. 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.
26. Giesinger JM, Kieffer JM, Fayers PM, Groenvold M, Petersen MA, Scott NW, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. J Clin Epidemiol. 2016;69:79-88.
27. King MT, Costa DS, Aaronson NK, Brazier JE, Cella DF, Fayers PM, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC

QLQ-C30. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2016;25(3):625-36.
28. 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.
29. Cocks K, Cohen D, Wisløff 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.
30. 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.
31. 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.
32. Calhoun EA, Welshman EE, Chang CH, Lurain JR, Fishman DA, Hunt TL, et al. Psychometric evaluation of the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (Fact/GOG-Ntx) questionnaire for patients receiving systemic chemotherapy. International journal of gynecological cancer : official journal of the International Gynecological Cancer Society. 2003;13(6):741-8.
33. Yost KJ, Cella D, Chawla A, Holmgren E, Eton DT, Ayanian JZ, et al. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. Journal of clinical epidemiology. 2005;58(12):1241-51.
34. 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.
35. 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. European journal of cancer (Oxford, England : 1990). 2012;48(11):1713-21.
36. Osoba D, Bezjak A, Brundage M, Zee B, Tu D, Pater J, 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.
37. Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA. Interpreting clinically significant changes in patient-reported outcomes. Cancer. 2007;110(1):196-202.
38. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert review of pharmacoeconomics & outcomes research. 2011;11(2):163-9.
39. Kemmler G, Zabernigg A, Gattringer K, Rumpold G, Giesinger J, Sperner-Unterweger B, et al. A new approach to combining clinical relevance and statistical significance for evaluation of quality of life changes in the individual patient. Journal of clinical epidemiology. 2010;63(2):171-9.

40. Sully K, Trigg A, Bonner N, Moreno-Koehler A, Trennery C, Shah N, et al. Estimation of minimally important differences and responder definitions for EORTC QLQ-MY20 scores in multiple myeloma patients. European journal of haematology. 2019;103(5):500-9.
41. 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.
42. Ramsey I, de Rooij BH, Mols F, Corsini N, Horevoorts NJE, Eckert M, et al. Cancer survivors who fully participate in the PROFILES registry have better health-related quality of life than those who drop out. Journal of cancer survivorship : research and practice. 2019;13(6):829-39.
43. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. Jama. 2017;318(23):2337-43.
44. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009;42(2):377-81.
45. Gotzsche PC. Blinding during data analysis and writing of manuscripts. Control Clin Trials. 1996;17(4):285-90; discussion 90-3.
46. Järvinen TL, Sihvonen R, Bhandari M, Sprague S, Malmivaara A, Paavola M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. Journal of clinical epidemiology. 2014;67(7):769-72.

10 TABLES and FIGURES

10.1 Table xx. Baseline characteristics

	Kd (n=xx)	**Observation (n=xx)**
Age (years (median), range))		
Sex Female Male		
Time from MM diagnosis (months (median), range))		
WHO performance status 0 1 2 3 or 4 Missing		
Geographical region Denmark Sweden Norway Finland Lithuania		
Patient-reported Outcomes		
-EORTC QLQ-C30 functional domain scores, mean (SD) (0 to 100, 100=best) Global Health Status/Quality of Life Physical functioning Role functioning Emotional functioning Cognitive functioning Social functioning		
-EORTC QLQ-C30 symptom domain scores, mean (SD); (0 to 100, 100=worst) Fatigue Nausea/vomiting Pain Dyspnea Insomnia Appetite loss Constipation Diarrhea Financial difficulties		
-EORTC QLQ-MY20 scores, mean (SD) Disease symptoms (0 to 100, 100=worst) Side effects of treatment (0 to 100, 100=worst) Future perspectives worries (0 to 100, 100=best) Body image loss (0 to 100, 100=best)		
-FACT-GOG/Ntx score, mean (SD) (0 to 44, 44=worst)		
Kd, carfilzomib-dexamethasone; **EORTC QLQ-C30,** European Organization for the Research and Treatment of Cancer QoL Questionnaire C30**; SD,** standard deviation; **MY20,** Multiple Myeloma Module**; FACT-GOG/Ntx,** Functional Assessment of Cancer Therapy/Gynaecologic Oncology Group/Neurotoxicity		

10.2 Table xx. Completed and missing patient-reported outcome data

Timing of assessment	Kd (n=)			Observation (n=)			Total (n=xxx)		
	Number of forms, n (%)			Number of forms, n (%)			Number of forms, n (%)		
	Expected	Received	Missing	Expected	Received	Missing	Expected	Received	Missing
Randomi-sation (Baseline)									
Month 2									
Month 4									
Month 6									
Month 8									
…									
…									
Total									
Kd, carfilzomib-dexamethasone; Expected, still alive and still on study at each specific assessment time point									

10.3. Table xx. Quality-adjusted progression-free survival (QAPFS)

	Kd	Observation	Difference (estimated 95% CI)	P value
Mean (SD) PFS				
Mean utility score (SD) -EORTC QLQ-C30 Sum Sc. -QLQ U10				
QAPFS (SD)				

SD, standard deviation; **EORTC QLQ-C30,** European Organization for the Research and Treatment of Cancer QoL Questionnaire C30; **QLQ U10**, Quality of Life Utility Measure-Core 10 dimensions

10.4. Figure x. Flow chart