Statistical Analysis Plan (SAP) for “FDG-PET/CT versus conventional CT for response monitoring in metastatic breast cancer: A multicenter randomized clinical trial (MONITOR-RCT)”

Vach, Werner; Gerke, Oke; Hildebrandt, Malene Grubbe

DOI:
10.21996/kazm-cx40

Publication date:
2024

Document version:
Final published version

Document license:
Unspecified

Citation for pulished version (APA):
Vach, W., Gerke, O., & Hildebrandt, M. G. (2024, Nov 8). Statistical Analysis Plan (SAP) for “FDG-PET/CT versus conventional CT for response monitoring in metastatic breast cancer: A multicenter randomized clinical trial (MONITOR-RCT)”. https://doi.org/10.21996/kazm-cx40

Funded by the European Union

UK Research and Innovation

Statistical Analysis Plan (SAP) for "FDG-PET/CT versus conventional CT for response monitoring in metastatic breast cancer: A multicenter randomized clinical trial (MONITOR-RCT)"

Table of contents

1. Administrative Information

Trial registration number: EU CT 2023-508591-11-00
SAP version: 1.0, 08 November 2024
Protocol version: 3.0, 22 September 2024
SAP version history:

Date	Version	Change
22/10/2024	0.1	First draft of the document
08/11/2024	1.0	First version of the SAP

Authors: Werner Vach[1], Oke Gerke[2,3], Malene Grubbe Hildebrandt[2,3]
[1]Basel Academy for Quality and Research in Medicine, Basel, Switzerland
[2]Department of Nuclear Medicine, Odense University Hospital, Odense, Denmark
[3]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

2. Introduction

This statistical analysis plan follows closely the recommendations by Gamble and colleagues (1). This SAP is supposed to be used in conjunction with the study protocol (SP). Parts already described in the SP are kept to a minimum and referenced to the SP.
The plan describes all analyses to be included in the primary statistical report of the MONITOR-RCT. It also clarifies which analyses are intended to be included in the main publication. Partially, we also mention analyses to be included in secondary publications to avoid the impression that they are regarded as non-substantial.

2.1 Background and rationale

FDG-PET/CT is increasingly used in cancer staging. Several studies have shown improved sensitivity of FDG-PET/CT compared with conventional imaging for diagnosing metastatic breast cancer in retrospective and smaller prospective studies. We expect FDG-PET/CT to detect disease progression earlier than CT in patients treated for metastatic breast cancer, enabling earlier start of second-line therapies. This has the potential to increase the beneficial effect of second-line therapies at the individual level and result in a delayed need for third-line therapies, prolonged overall survival, and improved quality of life compared with patients monitored with conventional CT.
SP, 2a), 2b)

2.2 Objectives

Primary: To demonstrate superiority in overall survival of response monitoring with FDG-PET/CT in patients with metastatic breast cancer over response monitoring based on CT. Appropriately adapted PERCIST criteria for FDG-PET/CT and the RECIST1.1 criteria for CT will be used.

Secondary: To demonstrate superiority in quality of life and exposure to oncologic treatment with FDG-PET/CT and to investigate the cost-effectiveness.
SP, 3a)

3. Study Methods

3.1 Trial design

International multicenter, parallel group, comparative, randomized trial comparing an experimental monitoring strategy based on FDG-PET/CT with a standard monitoring strategy based on CT.
SP, 3b)

3.2 Randomization

1:1 randomization, stratified by site and a prognostic index based on blocks of length 4 or 6. The prognostic index will be based on the same variables as used for adjustment in the primary analysis (SP, 8a); see also Section 6.2.1 below). The prognostic index will describe the estimated 42-months-survival-probability when monitored by PET/CT and will be dichotomized at a threshold of 0.5.
SP, 3c)

3.2.1 Selection of prognostic factors

Patients with metastatic breast cancer are a rather heterogeneous population with respect to overall survival. Already the two markers hormone receptor (HR) and HER2 status allow to define subgroups differing in median survival between 15 months and 50 months (2). Based on several patient characteristics, Plichta et al. (3) could constructed four groups of de novo metastatic breast cancer patients with 3-year survival rates varying from about 20% to about 75%.

Due to this huge variation, it was decided to adjust the main analyses for strong prognostic factors and to include a prognostic index in the stratified randomization.

The selection of prognostic factors was mainly based on the publication of Deluche et al. (2) reporting the results of a multivariable analysis based on 22,109 metastatic breast cancer patients from a French Cohort, supplemented by discussions among members of the research teams of the participating centers.

Deluche et al. (2) identified five prognostic factors in the overall population and six prognostic factors in a subpopulation of 8,595 patients with known performance status according to the WHO scale. Since performance status turned out as strong prognostic factor and will be available for all patients in MONITOR-RCT, we refer in the following to the latter analysis.

1) Age: Age was associated with an HR of 1.05 per decade (after adjustment for performance status) and was hence not considered as strong prognostic factor.

2) HR/HER2: Two third of patients were characterized as HR+/HER2- patients. HR-/HER2- patients covered 14% of the population and showed worse survival with a HR of 2.681. HER2+ patients covered 20% and showed better survival with a HR of 0.663. Here, we define HER2+ and HER2- as follows:
    a. HER2+ : HER2 3+ or 2+ with gene amplification.
    b. HER2- : HER2 0, HER2 1+, and HER2 2+ without gene amplification.

3) Relapse- and metastases-free interval: 2/3 of the patients had a relapse. Compared to de novo patients, patients with a metastases-free interval >24 months had a worse survival with an HR of 1.259 and patients with a metastases-free interval between 0 and 24 months had a worse survival with a HR of 2.425.

4) Metastases: Presence of visceral metastases in 57% of the patients (including brain metastases) was associated with worse survival with an HR of 1.437. It remains, however, unclear, whether also other metastatic sites were (to a lower degree) associated with survival or whether the lack of other metastatic sites in the final model was due to an inadequate modelling strategy.

5) Number of metastatic sites: Presence of three or more metastatic sites in 21% of the patients was associated with worse survival with an HR of 1.453. Here, we consider the following sites: (a) Bone/bone marrow; (b) liver; (c) lung; (d) subcutis/skin; (e) brain; (f) lymph node/axilla lymph node; (g) peritoneal karcinomatose; (h) breast; (i) pleural fluid/pleural metastases; (j) ascites; (k) gastrointestinal; (l) ovaries; (m) muscle; (n) other.

6) Performance status: Increasing the performance index from 0 to 1 point was associated with worse survival and a hazard ratio of about 1.477. Only about 40% of the patients had a performance status of 0, and 20% of 2 or above.

The final decision was hence to consider the five prognostic factors

- **HR/HER2,**
- **relapse/metastases-free-interval,**
- **visceral metastases (including brain metastases),**
- **number of metastatic sites, and**
- **performance status.**

The handling of these factors in constructing the prognostic index and as adjustment in the analyses is described separately in Section 3.2.2 and Section 6.2.1, respectively.

3.2.2 Construction of a prognostic index

Although the model of Deluche et al. (2) includes one additional factor (age), we decided to use the hazard ratios reported for Deluche's model to compute the prognostic index.

Hence the prognostic index is based on the following binary indicators and HRs:

Indicator	*HR*	*log HR*	*Probability*
HER2+ subtype[a]	*0.663*	*-0.411*	*0.195*
HR-/HER2- subtype[a]	*2.681*	*0.986*	*0.143*
Relapse & Metastasis-free interval < 24 month[b]	*2.425*	*0.886*	*0.133*
Relapse & Metastasis-free interval >24 months[b]	*1.259*	*0.230*	*0.542*
Presence of visceral metastases[c]	*1.437*	*0.363*	*0.573*
Number of metastatic sites >=3	*1.453*	*0.374*	*0.207*
Performance status = 1	*1.477*	*0.390*	*0.360*
Performance status = 2	*2.836*	*1.042*	*0.188*
Performance status = 3	*3.703*	*1.309*	*0.02*
Performance status = 4	*7.378*	*2.0*	*0.00*

[a]*Reference: HR+/HER2-.*
[b]*Reference: de novo MBC.*
[c]*Reference: absence of visceral metastases.*

The prognostic index is derived by summing up the log hazard ratio values for those indicators that apply for a respective patient.

To determine the threshold, we applied the index in the "observational data set" from Naghavi-Behzad et al. (4) described in Appendix 1 (omitting patients with missing HER2 status or missing performance status and stretching all survival times by a factor of 1.1) and chose the median value. This suggested splitting the population in two groups of equal size when performing the randomized stratification. The threshold determined this way was **0.593**. Based on a Cox regression using the index as only covariate, this value corresponds to a survival probability after 42 months close to 0.5.

Missing values will be handled by probability imputation, i.e. the missing value is replaced by the relative frequency of the variable as observed in the study by Deluche et al. (2). These values are also shown in the table above. The study reported only the aggregated probability of having a performance status of 2 or above. This probability of 0.208 was distributed to the three possible values, taking into account that according to the inclusion criteria it is unlikely that patients with performance status 3 or 4 will enter the trial.

3.3 Sample size

420 patients.
SP, 8b)

3.4 Framework

Superiority of FDG-PET/CT over conventional CT.
SP, 3a)

3.5 Statistical interim analyses and stopping guidance

No interim analyses with respect to the primary and secondary outcome are planned. (SP, 8d)

However, the trial requires a continuous monitoring of the scanning and treatment activities in order to be able to discuss these activities – in particular regarding differences between the participating centers – with the participating centers. We will export and analyze corresponding data accordingly.

3.6 Timing of final analysis

First Patient First Visit in mid-November 2024 and a maximal follow-up time of 54 months (as per SP) imply Last Patient Last Visit in mid-May 2029. The final analysis will take place after data cleaning and freezing the database, meaning presumably after 01 July 2029.

3.7 Timing of outcome assessments

Table 2 in SP.

4. Statistical Principles

4.1 Confidence intervals and P values

Level of significance is 5% (two-sided). Adjustment for multiple testing does not apply due to one primary outcome. 95% confidence intervals will supplement point estimates to the widest possible extent. As long as not stated otherwise, confidence intervals and p-values will be based on the Wald test principle. An alternative approach is, for instance, profile likelihood-based estimation and 95% confidence intervals (5).

4.2 Adherence and protocol deviations (description and how to be summarized)

Non-compliance at the scan level can happen due to omission of scheduled scan visits or failure to apply the intended modality as first modality at a visit. Reasons for non-compliance can be participation in clinical trials requiring a specific modality, indication of alternative scans for medical reasons, organizational reasons, patient preferences, patient absence and so on (SP, 3d). The amount and types of non-compliance is rather unpredictable. Formal non-compliance may not be equal to clinical non-compliance if an alternative scan provides a similar amount of information.

Hence, we do not provide an exact definition of non-compliance in this plan but will develop one or several definitions in the light of the observed non-compliance types and patterns.

4.3 Analysis populations

The intention-to-treat population comprises all patients according to the applied randomization. This will be the population for the primary analysis.

As non-compliance as described above may imply an underestimation of the potential intervention effect, additional, secondary analyses are planned.

Based on the compliance definition(s) mentioned, three types of analyses are planned:

- Per-protocol analysis: Exclusion of patients with a compliance rate of maximally 80%. The compliance rate is here defined as the fraction of scans performed with the modality a patient is randomized to.

- As-treated analysis: Shift of patients to the opposite modality in case of a higher compliance rate for the opposite modality (i.e. omitted scans are not counted here).

- Per-protocol analysis at the scan level: Censoring patients 12 months after shift to the alternative modality if they stay in this modality until the end of the study (with at least a compliance rate of 80%).

These analyses will be reported in the main publication only if they imply analysis sets that differ by at least 5% in the number of patients compared to the intention-to-treat population.

5. Trial Population

5.1 Screening data

Screening data will not be collected.

5.2 Eligibility

SP, 4)

5.3 Recruitment

SP, Figure 1

5.4 Withdrawal/follow-up (level, timing, reasons)

Patients who withdraw or are withdrawn by the sponsor will be offered to continue monitoring at the participating center according to local standards. Withdrawn patients contribute with data to the study until the time point of withdrawal.
SP, 4c)

5.5 Baseline patient characteristics (list and how to be descriptively summarized)

A list of baseline characteristics extracted from hospital records is shown in the SP, Table 3. Descriptive summaries will be done by modality as per randomization. Categorical data will be summarized by frequencies and percentages. Continuous data will be summarized by mean, standard deviation, median, minimum, 5%, 10%, 25%, 75%, 90%, 95% percentiles, maximum, and skewness. Stratified histograms, stacked bar charts and violin plots will be used for visualization of the distributions. Exploratory hypothesis testing for baseline characteristics will be done with (generalizations) of Fisher's exact test and Wilcoxon rank sum tests.
Selected descriptive characteristics will be reported in the main part of the main publication. Selection depends on the shape of distributions observed and the aim of a uniform presentation of results. Visualizations will be presented in the supplementary material. P-values will not be reported in the main publication.

6. Analysis

6.1 Outcome definitions (specification, timing, units, transformations)

6.1.1 Primary outcome

Overall survival (in years) will be addressed by "time from randomization until death or censoring". Minimal (planned) follow-up time of the patients will be 36 months, maximal (planned) follow-up time 54 months.
SP, 6a)

6.1.2 Secondary outcomes

"Quality of life" will be addressed by two outcome variables: (1) overall summary score of the FACT-B, (2) complaints related to the conduct of scans reported by the patients.

"Exposure to oncologic treatment" will be addressed by the following outcome variables describing different aspects of oncological treatment:

1) Experience of progression
2) Start of a new treatment line because of progression
3) Time to first progression
4) Time from first to second progression
5) Time from second to third progression
6) Experiencing other diagnostic procedures
7) Hospitalization

These seven secondary outcomes do not only aim to point to further, patient-relevant differences between the two monitoring strategies, but serve also as a check and demonstration of how monitoring by FDG-PET/CT implies a survival (and potentially also quality of life) benefit. An earlier detection of first progression implies earlier shift to effective therapies, longer time until subsequent progression, and less need for diagnostic procedures and hospitalization.

"Cost-effectiveness" assessment relates overall survival to costs. Costs will be assessed by two different approaches: (1) Difference in costs related to the scans and subsequent oncological treatment, (2) Difference in cost of using of health care as documented in registry data.
SP, 6a), 8a)

6.2 Analysis methods (method and effect description, covariate adjustment, assumptions, alternative methods)

6.2.1 Covariate adjustment

Several of the analyses described in the following include an adjustment for the five prognostic factors defined above. This will be approached by using regression models. The factors will enter the models using the following covariates

- HR/HER2: indicator variables for HER2+ as well as HR-/HER2-,
- relapse/metastases-free-interval: indicator variable for relapse; continuous variable for metastases-free interval, set to 0 for patients with de novo metastatic breast cancer,
- visceral metastases (including brain metastases): indicator variable,
- Number of metastatic sites and
- performance status: categorical variable.

The coefficients from the models will be reported in the supplemental material of the main publication.

Models with additional baseline characteristics (age, center, de novo or recurrent metastatic breast cancer, histological type, location of confirmatory biopsy, burden of disease, metastatic sites, first-line therapy and so on) will be fitted to check whether strong prognostic factors have been overlooked. If so, an additional selection of factors will be considered in a sensitivity analysis.

The value of covariate adjustment could be confirmed in simulating the clinical trial (Appendix 1). Adjustment for the four factors implied a smaller p-value with a probability of 0.76.

6.2.2 Primary outcome

We will visualize the difference in survival by model-based Kaplan-Meier-curves. To this end, a Cox model stratified by the interventions and adjusted for all prognostic factors (Section 6.2.1) will be fitted, and an estimate of the baseline survival function will be used to obtain estimates of the

survival function in each intervention group. The two estimates will be shown graphically. As some readers may not be familiar with model-based estimation of survival curves, ordinary Kaplan-Meier estimates will be presented in the supplementary material of the main paper.

The model-based Kaplan-Meier estimates will also be used to compute the difference in survival rates at 36, 42, and 48 months together with respective 95% bootstrapped confidence intervals (6).

The statistical significance of the difference in survival will be assessed using a statistical test procedure not directly relying on the assumption of proportional hazards. We will make use of a Weibull accelerated failure time model. The model will include the prognostic factors (Section 6.2.1) and the intervention as covariates. In general, the use of an accelerated failure time model seems to be reasonable in this context, as it can be interpreted as prolongation of the survival time by a factor due to the intervention.

A recent investigation by Klinglmüller et al. (7) has shown that this approach has a higher power than many alternative approaches when a delayed intervention effect has to be expected. This result could also be confirmed when simulating the trial (Appendix 1). The approach had a higher power than the log rank test, a Cox model, or an alternative accelerated failure time models.

The Weibull model provides also an estimate for the acceleration factor. This estimate will also be reported together with a 95% confidence interval.

6.2.3 Secondary outcomes

Quality of life

The quality of life of a patient will vary over time. It has to be assumed that the quality of life will often deteriorate prior to a patient's death. Similarly, we cannot exclude that at least in some patients we will observe a slow decrease over time simply due to aging and living with disease burden. However, we may observe substantial individual deviations from such "typical courses".

Detailed analyses of the observed trajectories and their typical and atypical courses are planned as part of further analyses subsequent to the analyses presented in the main paper.

With respect to the analysis presented in the main paper, the idea is to focus on a summary measure over time and patients in a manner allowing a simple interpretation for all stakeholders. The average quality of life (averaged first over all time points and then over patients) provides such a measure. If it indicates an advantage for one modality, this simply means that patients can expect on average a higher quality of life for the rest of their life.

However, it may happen that this simple approach has to be regarded as misleading due to ignoring the variation in individual courses and a potential confounding with the survival difference between the two groups. In particular, if patients live longer in the FDG-PET/CT arm and quality of life

decreases on average over time, this may imply a lower quality of life on average in the FDG-PET/CT arm.

Hence, as part of preparing the statistical report, we will perform additional analyses based on inspecting the individual courses of the quality of life scores and time-point specific comparisons of quality of life – with time points also defined relative to end of life. In case the simple approach outlined above has to be regarded as potentially misleading, alternative approaches will be considered, in particular the approach outlined in Appendix 2. This approach is based on first averaging over patients at each time point and then averaging over time points.

Patients' complaints related to the conduct of scans will be categorized and the frequencies reported for each arm. Differences in frequencies will be described by odds ratios.

Exposure to oncologic treatment

Differences in the risk of progression, the risk of starting a new treatment line because of progression, the risk of experiencing diagnostic procedures, and the risk of hospitalization (secondary outcomes #1,2,6,7) will be quantified by incidence rate ratios based on a Poisson regression with the risk time as offset. The first two analyses will also be adjusted for the prognostic factors (Section 6.2.1). The raw frequencies as well as the raw rates will be reported, too.

Differences in the time until progression and time between progressions (secondary outcomes #3–5) will be quantified by hazard ratios based on a Cox model with adjustment for the prognostic factors (Section 6.2.1). Median and quartiles of the times based on model-based Kaplan-Meier estimates will be reported, too.

Two different definitions of progression will be considered. The first is progression according to the reports by the radiological or nuclear medicine specialists. The second is clinical progression, i.e. a treatment shift following an indication for progression. The time until progression is always defined from the start of the last treatment shift.

Note: Additional treatment characteristics such as choice of first line, second line, and third line therapies, treatment for oligometastatic progression (SBRT, ablation, surgery, and others), and participation in experimental treatment offers (new treatment testing studies) will be reported descriptively by arm. These are, however, not regarded as outcomes.

Cost-effectiveness

Cost-effectiveness analyses will be based on computing an incremental cost effectiveness ratio (ICER). Effectiveness will be assessed by the gain in quality-adjusted life years (QUALY) based on the data from the EQ-5D-5L and survival.

6.2.4 Planned sensitivity analyses

See the section on missing values (6.3) and on covariate adjustment (6.2.1).

6.2.5 Planned subgroup analyses (incl. definition of subgroups)

Subgroup analyses are planned for different reasons. They all have in common that they try to check certain expectations we have about how monitoring by FDG-PET/CT may benefit the patients. It is not the aim to "prove" that there are differences in survival between specific patient groups or that some patient groups have a higher benefit than other groups. Due to the exploratory character of the subgroup analyses, they are restricted to showing the empirical intervention-group specific survival curves estimates and reporting an estimate of the acceleration factor with a 95% confidence interval.

Heterogeneity of intervention effects due to differences in prognosis

As pointed out in the study protocol, we have to assume that not all patients will benefit to the same degree from FDG-PET/CT. Patients can be expected to benefit from FDG-PET/CT if they experience at least one progression and if they can still benefit at that time point from an earlier shift to an (effective) second line therapy.

This may imply that patients with a poor prognosis may benefit less from FDG-PET/CT than patients with a moderate prognosis. Further, it may be even the case that patients with a very good prognosis do not benefit as they will not experience a progression during the study period.

It has to be expected that the results of the trial are driven by patients with a "moderate" time until first progression, i.e. close to the median time of 15 months observed by Vogsen et al. (8). This is in line with the result of Naghavi-Behzad et al. (4), indicating that survival curves do not differ substantially until about 24 months after diagnoses – a pattern expected and taken into account in the sample size considerations of this study. However, this carries the risk of misinterpretation in the sense that no patient can benefit from FDG-PET/CT earlier. This needs not to be the case, as a minority of patients with limited survival may still benefit already within the first two years.

Consequently, subgroup analyses focusing on subgroups with poor or good prognosis are useful to contribute to a better understanding of how FDG-PET/CT may provide a survival benefit.

In a first step, we will construct a prognostic index based on the five selected prognostic factors in the study population of the MONITOR-RCT and will use this to stratify the population in those with the 25% worst prognosis, the 25% best prognosis, and those in between. (The model may be extended as pointed out above in Section 6.2.1.)

In addition, we consider three – relatively small – groups with (expected) poor prognosis based on single clinical characteristics

- HR-/HER2- (15%)

- Performance status >=2 (20%)
- Relapse and metastases-free interval less than 24 months (13%)

as well as one group with good prognosis based on combining favorable characteristics:

- Performance status 0 AND no visceral metastases AND HR+ or HER2+ AND (de novo metastatic breast cancer OR metastases-free interval of at least 48 months).

Heterogeneity of intervention effects due to metastatic site-specific differences in sensitivity between FDG-PET/CT and CT

It is well known that the sensitivity of FDG-PET/CT varies across different metastatic sites and the same holds true for CT. Consequently, for some metastatic sites we expect a higher benefit from FDG-PET/CT than from other sites. Checking these expectations may again contribute to a better understanding of the benefit from FDG-PET/CT. Metastatic sites with an assumed distinct benefit from FDG-PET/CT are bone and liver. We will perform analyses on subgroups of patients based on observed bone and liver metastases (i.e. bone and liver metastases both present, only bone metastases, only liver metastases, neither bone or liver metastases (reference group)).

These analyses will be conducted twice, once only considering metastases at baseline and once considering any metastasis detected during follow up.

Delayed intervention effect

The expectation of a delayed intervention affect has been already taken into account in designing the analysis of the primary outcome. It will be also visible in the model-based Kaplan-Meier curves. However, this aspect can be further depicted by landmark analyses (also known as conditional survival analyses): The population is restricted to those patients surviving until a specific landmark time point and the survival time thereafter is analyzed.

We will perform landmark analyses with respect to the following landmarks: 12, 18, and 24 months; first progression, first clinical progression.

6.3 Missing data

The only outcome for which we expect missing data is quality of life.

In the primary analysis, we will ignore these missing values, i.e. we will base the computation on the available values.

Formally, this approach may be biased even if we assume that the missing at random assumption (MAR) holds. As a sensitivity analysis, we will perform an analysis that should provide unbiased estimates under the MAR assumption. This analysis will be based on multiple imputation using a

multivariate normal regression with arm-specific parameters and including age, gender and the five prognostic factors mentioned in Section 6.2.1 as additional predictors.

We do not expect missing values in the prognostic factors as these are required for randomization.

6.4 Additional analyses

We will pursue the following additional analyses for the primary outcome.

6.4.1 Bone scintigraphy as add-on to CT

In the USA, guidelines recommend the combination of CT with bone scintigraphy as standard monitoring. European guidelines do only recommend CT, but in some centers bone scintigraphy is used as an add-on – in all patients or selected patients and regularly at all monitoring visits or irregularly/on demand. To inform also patients from the USA, we plan to analyze the primary outcome including in the CT arm only patients for whom bone scintigraphy has been used regularly.

MRI scans will typically be performed in some Danish centers to improve the sensitivity of progression in bone metastases. This is the same argument as in the USA, where an additional bone scintigraphy is recommended. MRI and bone scintigraphy are both performed to have a better evaluation of bone metastases. However, there may be differences in the added value of either of these scans as the techniques are different. The bone scintigraphy detects mainly osteosclerotic metastases and the MRI often detects in addition bone marrow metastases.

As both MRI and bone scintigraphy (bone scan) can provide information on bone metastases in addition to the CT, we will perform an analysis including all patients for whom either bone scintigraphy or MRI was applied regularly.

Regularly is defined at the patient level as having obtained on average the corresponding scan biannually, ignoring a period without such a scan up to 6 months prior to end of follow-up/death but requiring at least one scan.

6.4.2 Subgroups of centers with specific strategies with respect to CT or PET/CT

There may be further differences across centers with respect to performing CT or PET/CT. We will inspect the practice of the centers and may define subgroups of centers following a specific strategy or exclude centers with specific strategies.

6.5 Harms

Adverse events, adverse reactions, serious adverse events, serious adverse reactions, and suspected unexpected serious adverse reactions will be tabulated by intervention (see SP, 2c), 7a)). However, these numbers will not be reported completely in the main publication; instead, we will focus on events which go beyond the usual and well known side effects.

6.6 Statistical software

We will use the statistical analysis package Stata and the programming language R.

Appendix 1: Simulation of the clinical trial

As pointed out in the study protocol, the observational study published by Naghavi-Behzad et al. (4) mimics closely the scenario we have to expect in the MONITOR-RCT, as the choice of the monitoring modality was made by the treating clinician reflecting mainly personal preferences.

We did not only make use of this data set to inform the power calculations presented in the study protocol, but also some of the considerations presented in this statistical analysis plan. Patients treated prior to 2009 were excluded in order to be closer to the situation we can expect today.

This left 83 patients in the FDG-PET/CT group, 137 patients in the CE-CT group and 65 patients in the combined group. We refer to this in the sequel as “the observational data set”.

We used the observational data set as starting point to simulate the MONITOR-RCT, but excluded the patients from the combined group. To inform the simulation of the trial, a stratified Cox proportional hazard model with the prognostic factors described in Section 3.2.1 and intervention-group-specific baseline hazard functions was fitted to the data. This enables the computation of the individual survival function for any subject with a specific intervention and specific values of the prognostic factors. To take the progress in oncological treatment over the last decade into account, the baseline survivor functions were stretched by the factor 1.1.

A single trial was simulated by generating a bootstrap sample of size 420 from the 196 patients. Only the information on the prognostic factors was used from this sample. The 420 patients were then randomized to one of the two interventions. Individual survival curves were computed as described above and a survival time drawn based on this curve for each individual. Censoring was simulated by drawing from a uniform distribution on the interval from 36 to 54 months.

The results refer always to 1000 simulations of the trial.

Appendix 2: The time averaged mean difference in quality of life

This alternative measure is defined in two steps. (1) At each follow up time-point, we consider the expected difference in quality of life among those who survived until this time point. (2) The time point-specific differences are then summed up with weights proportional to the fraction of patients reaching the time point but summing up to 1.

Estimation of this measure is straightforward in the sense that the expected difference can be estimated by the corresponding mean values at each time point and the weights can be obtained from the empirical estimates of the proportions. We will base 95% confidence intervals and p-values on setting up the computations formally as a regression model and use robust standard errors. (The stochastic nature of the weights is ignored here.)

However, it is well known that adjustment for baseline values can provide a substantial increase in power in such analyses, and it is to be expected that this holds true also here. As we cannot exclude that some baseline assessments are missing, we prefer the alternative to include baseline values as outcomes and to inform the model explicitly about the randomization at baseline. This is typically approached within the framework of mixed models. However, the estimate defined above is actually based on marginal parameters. We will make use of a generalized estimating equations (GEE) approach with an exchangeable working correlation matrix, and we will include the baseline assessments as outcome and simultaneously assuming no systematic difference in the baseline assessment. As the GEE approach is biased even if the non-MAR assumption holds, the analysis will make use of multiple imputation as described in Section 6.3.

References

1. 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.
2. Deluche E, Antoine A, Bachelot T, Lardy-Cleaud A, Dieras V, Brain E, et al. Contemporary outcomes of metastatic breast cancer among 22,000 women from the multicentre ESME cohort 2008-2016. Eur J Cancer. 2020;129:60-70.
3. Plichta JK, Thomas SM, Hayes DF, Chavez-MacGregor M, Allison K, de Los Santos J, et al. Novel Prognostic Staging System for Patients With De Novo Metastatic Breast Cancer. J Clin Oncol. 2023;41(14):2546-60.
4. Naghavi-Behzad M, Vogsen M, Vester RM, Olsen MMB, Oltmann H, Braad PE, et al. Response monitoring in metastatic breast cancer: a comparison of survival times between FDG-PET/CT and CE-CT. Br J Cancer. 2022;126(9):1271-9.
5. Royston P. Profile likelihood for estimation and confidence intervals. Stata Journal. 2007;7(3):376-87.
6. Ruhe C. Bootstrap pointwise confidence intervals for covariate-adjusted survivor functions in the Cox model. Stata Journal. 2019;19(1):185-99.
7. Klinglmüller F, Fellinger T, König F, Friede T, Hooker AC, Heinzl H, et al. A neutral comparison of statistical methods for time-to-event analyses under non-proportional hazards. ArXiv; 2023.
8. Vogsen M, Harbo F, Jakobsen NM, Nissen HJ, Dahlsgaard-Wallenius SE, Gerke O, et al. Response Monitoring in Metastatic Breast Cancer: A Prospective Study Comparing (18)F-FDG PET/CT with Conventional CT. J Nucl Med. 2023;64(3):355-61.

Other references

Data Management Plan. Version 1.0, dated 08 November 2024.