Statistical Analysis Plan
(SAP)

for the trial

A multi-centre randomized, placebo-controlled trial of mirabegron, a new beta3-adrenergic receptor agonist on the progression of left ventricular mass and diastolic function in patients with structural heart disease

(BETA3_LVH)

NCT01067703

Coordinating investigator
Prof. Jean-Luc Balligand
Full professor, Head of Department
Pharmacology and Therapeutics (FATH)
Avenue Mounier, 52, B1.53.09
1200 Brussels, Belgium
Phone: +32-2-7645260/+32-2-7645262
Fax: +32-2-7645269

Biometry
Dr. Oana Brosteanu, Dr. Dirk Hasenclever
Clinical Trial Centre Leipzig
Universität Leipzig
Härtelstraße 16-18,
04107 Leipzig, Germany
Phone: +49-341-97 16 250
Fax: +49-341-97 16 189

Sponsor
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
PLACE DE L'UNIVERSITE 1, 1348 LOUVAIN LA NEUVE, Belgium

Date of the document: 2022-08-10
Version of the document: Final 1.0

We hereby approve the Statistical Analysis Plan in its final version:

Coordinating Investigator		
	Date	Signature
Biometrician		
	Date	Signature
Biometrician		
	Date	Signature

Table of contents

1 Introduction 5

2 Changes in the planned analysis 5

2.1 Fine-tuning the basic analytic model ... 5
2.2 Sample size revisited in the light of the COVID-19 pandemic ... 5
2.3 Additional secondary endpoints ... 6

3 Endpoints and further variables 6

3.1 Primary Endpoints ... 6
3.1.1 Change in left ventricular mass index (LVMI) ... 6
3.1.2 Change in diastolic function (E/e') ... 7
3.1.3 Dealing with co-primary endpoints ... 7
3.2 Secondary Endpoints ... 7
3.3 Safety endpoints ... 9
3.3.1 Adverse and serious adverse events ... 9
3.3.2 Safety events of special interest ... 10
3.4 Covariates for confirmatory analyses ... 10
3.5 Further Variables ... 11
3.5.1 Baseline characteristics ... 11
3.5.2 Trial intervention ... 13

4 General Analysis Definitions 13

4.1 Study periods ... 13
4.2 Analysis Populations ... 13
4.3 Handling of centre effects ... 14

5 Planned analysis 14

5.1 Demographic and other baseline parameter ... 14
5.2 Concomitant Diseases and Medication ... 14
5.3 Compliance with regard to the Study Intervention ... 14
5.4 Primary and secondary efficacy endpoints ... 15
5.4.1 Choice of scale of analysis ... 15
5.4.2 Basic analytic model ... 15
5.4.3 Sensitivity analyses ... 17
5.5 Subgroup analysis ... 17
5.6 Safety aspects ... 18
5.6.1 Adverse events ... 18
5.6.2 Serious adverse events ... 19
5.6.3 Events of special interest ... 19

6 References 20

7 Appendix 20

7.1 Abbreviations 20

1 Introduction

The purpose of this document is to provide a detailed elaboration of the statistical analysis described in the protocol, including detailed procedures for the confirmatory analysis of the primary and secondary endpoints and other variables.

The Statistical Analysis Plan (SAP) assumes familiarity with the Study Protocol (Version Final 2.0 dated 13.11.2015), including Protocol Amendments.

The SAP is based on the planned analysis specification as written in the study protocol Section 9 "Biometry". SAP readers may consult the study protocol for more background information on the study, e.g., on study objectives, study design and population, trial intervention, definition of measurements and variables, planning of sample size, and randomization.

R 4.2.0 (R packages named below) will be used for all statistical analyses.

2 Changes in the planned analysis

2.1 Fine-tuning the basic analytic model

There were two challenges in choosing the structure of the basic analytic model:

- Dealing with missing of failed measurements at study visits and
- Adjusting for chance baseline imbalances

Because failed measurements were already expected to occur at baseline, in the study protocol we chose a linear mixed model with a 3-dimentional random effect (Baseline, V3, V5) with a general variance covariance structure. The random effect models the intra-individual dependence in the vector of residuals from the mean structure for each patient. This random effect includes the baseline measurement. This allows the model to accommodate all patients with at least one valid measurement.

However, as the baseline measurement is included in the random effect and can be missing, it cannot be used as a fixed effect adjustment covariate, without restricting the analysis set to patients with valid baseline only.

Fitzmaurice et al. 2004 (Chapter 5.7) propose a solution to this short-coming: Because the study is randomised, it is justified to force the model to estimate a common mean at baseline without separating the arms. This change in the mean structure of the model performs an indirect baseline adjustment via its influence on the random effect. Simulations of Fitzmaurice et al. 2004 (and as well as our own) confirm that results with this indirect baseline adjustment are nearly identical to using baseline as fixed effect covariate. In addition, the gain in power compared to unadjusted analysis is the same.

Therefore, indirect baseline adjustment was added to the basic analysis (details see 5.4 below).

2.2 Sample size revisited in the light of the COVID-19 pandemic

The COVID-19 pandemic affected the study in slowing down accrual and increasing the risk for dropouts due to missed visits. Therefore, the steering committee of BETA3_LVH discussed on 2021-01-19 whether to end the study as planned or to increase the sample size due to a higher than expected dropout rate.

In a statistical appraisal (NoReasonsForOverrecuiting 2021-03-01.pdf), the eventual dropout rates were projected and the assumptions underlying the sample size calculations revisited based on blinded data.

- With conservative assumptions, the projected number of fully + partially informative patients at the time was at least 75% for both primary endpoints.

- The original sample size calculation was based on a very conservative assumption on the standard deviation of the changes in E/e'. The study data at the time suggested sd < 2.5 instead of sd = 3.
- Thus, BETA3_LVH is sufficiently powered for drop-out rates up to 30%, although initially a dropout rate of only 8% was assumed.

Accrual into the study was stopped as planned when the target sample size was achieved (decision of the Steering Committee 2021-04-06).

2.3 Additional secondary endpoints

While we were completing our data collection, the group of H. Bundgaard in Denmark published the results of an independent study on mirabegron in patients with severe heart failure with reduced ejection fraction (<35%); while mirabegron was used at 300 mg daily (under full beta1-blocker therapy) for only 1 week, invasive hemodynamic measurements showed a significantly larger increase in cardiac index in the verum compared to placebo (Bundgaard H et al. 2022). Therefore, we decided to add a secondary key endpoint reflecting the LV systolic function in our analysis, i.e. stroke volume index (SVI) measured by high precision cMRI. In addition, we add left ventricular ejection fraction (LVEF) as a further non-key endpoint.

Likewise, we are aware of the protocol of the SPHERE study (Garcia-Lunar I. et al 2020), testing the effect of mirabegron (vs. placebo) in patients with both pre- and post-capillary pulmonary hypertension from both heart failure with preserved or reduced ejection fraction, where right ventricle ejection fraction, reflecting the RV function, is a key secondary endpoint (and pulmonary vascular resistance is the primary endpoint). As we postulate a favourable myocardial remodelling and improved diastolic function with mirabegron as our main hypothesis, if this is verified, we expect an improvement in filling pressures that may have an impact on pulmonary vascular pressures and RV function. For this reason, we decided to add an index of RV systolic function as a secondary key endpoint, i.e. right ventricular ejection fraction (RVEF) measured by high-precision cMRI.

3 Endpoints and further variables

3.1 Primary Endpoints

Two equally ranked primary endpoints are defined, in order to assess both structural and functional aspects of left ventricular remodelling.

The measurements at baseline, 6 months and 12 months will be analysed together by a linear mixed model for repeated measurements (see 5.4). The treatment effect at 12 months is the primary contrast of interest. The treatment effect at 6 months is a secondary endpoint.

3.1.1 Change in left ventricular mass index (LVMI)

Change in left ventricular mass index (**LVMI** in g/m^2 is defined as left ventricular mass (LVM) divided by body surface area, BSA) measured at baseline, 6 months (V3 = visit 3), and 12 months (V5 = visit 5) after randomisation. Cardiac MRI is performed locally according to a standardized protocol, and LVM is measured in the central MRI core lab.

LVMI will be computed using

- LVM as measured by the core lab
- body surface area (BSA) derived from height and weight as documented in the electronic CRF for the respective visit, using the DuBois formula

$$BSA\ [m^2] = 0.007184\ m^2 * height\ [cm]^{0.725} * weight\ [kg]^{0.425}$$

The MRI core lab has classified the quality of the LVM measurement with a quality score from 0 to 3, meaning

- 0 = OK

- 1 = minor problems (values probably ok)
- 2 = major problems (affecting values)
- 3 = data missing

MRI scans classified as having “major problems (affecting values)” will be counted as missing.

3.1.2 Change in diastolic function (E/e’)

Change in diastolic function, assessed as the ratio of peak early transmitral ventricular filling velocity to early diastolic tissue Doppler velocity (**E/e’**) measured at baseline, 6 and 12 months after randomisation. This parameter is assessed by echocardiography, performed locally according to a standardized protocol, and is measured in the central echocardiography core lab.

E/e’ will be measured by the echo core lab according to the echo core lab echo image analysis plan. Only average E/e’ values will be used for the statistical analysis of this co-primary endpoint. E/e’ average values are calculated as the ratio of E divided by the average of e' septal and e' lateral. In cases either e’ septal or e’ lateral cannot be reliably measured by the echo core lab, the respective E/e’ average value will be set as missing value due to low image quality in the echo core lab report.

3.1.3 Dealing with co-primary endpoints

This trial wants to demonstrate that mirabegron as add-on to standard treatment compared to standard treatment alone improves at least one of the two primary endpoints over 12 months.

The null hypotheses are that the 12 month mean changes from baseline mean are identical in both arms. Two-sided tests are used, i.e. the alternative is that mean changes from baseline mean differ by arm.

The Hochberg method will be used to adjust for endpoint multiplicity: If both p-values are below 0.05 we will claim efficacy in both primary endpoints; if otherwise the smallest p-value is below 0.025 we will claim efficacy in the respective primary endpoint. This procedure controls the family-wise error rate (FWER) in the strong sense at a two-sided significance level of 5%. (Hochberg 1988). The Hochberg procedure seems appropriate since we expect non-negatively correlated test-statistics.

3.2 Secondary Endpoints

- Further cMRI endpoints (all measured in the central MRI core lab)
  - Cardiac fibrosis at baseline and at 12 months (V5). Fibrosis is a key pathogenic mechanism of diastolic dysfunction, which is at the origin of HFpEF. Cardiac fibrosis has been shown to be reduced by beta3AR in pre-clinical studies.

    The following separate markers of cardiac fibrosis are analyzed:

    <u>Late Gadolinium Enhancement</u>
    - LGEvf: Late Gadolinium Enhancement volume fraction

    <u>Native T1</u>
    - T1_native: T1 relaxation time

    <u>Extracellular Volume Fraction</u>
    - ECV_AllDataPoints as summary (based on ECV_5m, ECV_15m, ECV_30m as raw data)
    - The four parameters are non-independent

    The MRI core lab has classified the quality of measurements of cardiac fibrosis with a quality score from 0 to 3. As suggested by the core lab, only measurements classified as 0 or 1 will be used for analysis. All other measurements will be counted as missing.

  - Left atrial volume index (LAVOLI) at baseline and at 12 months (V5). This parameter

determines diastolic filling (and was shown to predict treatment efficacy in HFpEF in the J-DHF trial (Yamamoto et al. 2013).

LAVOLI will be computed using

- Left atrial volume (LAVOL) as measured by the core lab
- body surface area (BSA) derived from height and weight as documented in the electronic CRF for the respective visit, using the DuBois formula

$$BSA\ [m^2] = 0.007184\ m^2 * height\ [cm]^{0.725} * weight\ [kg]^{0.425}$$

The MRI core lab has classified the quality of the LAVOL measurement with a quality score from 0 to 3. As suggested by the core lab, only measurements classified as 0 or 1 will be used for analysis. All other measurements will be counted as missing.

- Stroke volume index (SVI), defined as
  ([Left Ventricular End-Diastolic Volume (LVEDV) - Left Ventricular End-Systolic Volume (LVESV)], divided by Body Surface Area (BSA)

  measured by cardiac MRI at Baseline and 12 months.

  The MRI core lab has classified the quality of the left ventricular measurements with a quality score from 0 to 3. As suggested by the core lab, only measurements classified as 0 or 1 will be used for analysis. All other measurements will be counted as missing.

- Left Ventricular Ejection fraction (LVEF) by cardiac MRI at Baseline and 12 months

  The MRI core lab has classified the quality of the left ventricular measurements with a quality score from 0 to 3. As suggested by the core lab, only measurements classified as 0 or 1 will be used for analysis. All other measurements will be counted as missing.

- Right Ventricular Ejection Fraction (RVEF) by cardiac MRI at Baseline and 12 months

  The MRI core lab has classified the quality of the right ventricular measurements with a quality score from 0 to 3. As suggested by the core lab, only measurements classified as 0 or 1 will be used for analysis. All other measurements will be counted as missing.

- LV mass index (by cardiac MRI measured by the MRI core lab) at 6 months,

- Diastolic function (E/e'), measured by the echo core lab, at 6 months;

- Laboratory parameters at baseline and at 3 (V2), 6 (V3) and 12 months (V5), measured by the central core lab
  - serum biomarkers
    - NT-proBNP
    - hsTnT
    - Galectin3
    - GDF15
  - metabolic parameters
    - Glycemic profile
    - HbA1c
    - Total cholesterol
    - Calculated LDL cholesterol
    - HDL cholesterol
    - Triglycerides
    - modified HOMA test (HOMA test /%S, HOMA test / % B)
    - Insulin

- Maximal exercise capacity (peak VO_2) at baseline and 12 months (V5).

  Up to four different parameters are provided depending on the device used and are recorded in the eCRF:

  - VO_2max in [L/min]
  - VO_2max in [mL/min/kg]
  - VO_2max in [L/min] in % predicted values
  - VO_2max in [mL/min/kg] in % predicted values

  In principle, all parameters should be interconvertible:

  - There is a formula provided by Wonisch et al for the predicted values which could be used for transformation of absolute values into % predicted values
  - Weight is recorded in the eCRF, allowing for transformation between the two units used for absolute measurements.

  The preferred secondary endpoint for analysis is peak VO_2max in [mL/min/kg] according to the core lab.

  In rare cases where the preferred secondary endpoint for analysis is VO_2max in [mL/min/kg] is not documented directly, but any of the three other endpoints, this formula will be used to derive it.

  Because of the large number of secondary endpoints and in order to limit multiplicity problems, we prioritize those secondary efficacy endpoints in which we particularly expect a potential treatment effect:

  Key secondary endpoints are:

  - NT-proBNP (as continuous variable)
  - Left atrial volume index
  - myocardial fibrosis: ECV_AllDataPoints
  - metabolic measurement: HOMA test: HOMA /%S
  - peak VO_2 max (ml/min/kg)
  - Stroke volume index (SVI)
  - RVEF

  All other secondary endpoints are of lower ranking interest.

3.3 Safety endpoints

3.3.1 Adverse and serious adverse events

Adverse and serious adverse events are recorded on corresponding forms, which contain the following information:

- Type of event
- Start and end date
- Maximal intensity (mild / moderate / severe / life-threatening / death related to AE)
- Therapy of event (yes / no)
- Outcome of event (recovered / recovering / not recovered / recovered with sequelae / fatal / unknown)
- Causal relationship with trial drug (reasonable possibility / no reasonable possibility)
- Seriousness (yes / no)
- Seriousness criterion (death / life risk / hospitalization / disability / birth defect / Intervention required)

- Action taken with trial drug due to event (drug withdrawn / dose reduced / dose increased / dose not changed / unknown / not applicable)

Adverse Events (AE) and Serious Adverse Events (SAE) will be coded according to Medical Dictionary for Regulatory Activities (MedDRA) Version 24.1. For analysis, classifications by body system and by preferred term will be used.

3.3.2 Safety events of special interest

In addition to the adverse event documentation, the following safety events of special interest were closely monitored.

Blood pressure

Three measurements of blood pressure (systolic and diastolic), taken after 5 min of rest in sitting position are documented at every visit. The average of these three measurements is used for further analyses.

Elevated blood pressure (yes / no) is defined as systolic blood pressure > 160 mm Hg or diastolic blood pressure > 100 mm Hg after start of study treatment.

Highly elevated blood pressure (yes / no) is defined as systolic blood pressure > 180 mm Hg or diastolic blood pressure > 110 mm Hg after start of study treatment.

Hepatic impairment

Alanine aminotransferase (ALT) and aspartate aminotransaminase (AST) are measured locally at baseline and at visits V1, V2, V3 and V5.

Hepatic impairment (yes / no) is defined as ALT > 2 x upper level normal or AST > 2 x upper level normal after start of study treatment, where the upper level of normal is defined by the local laboratory.

Renal impairment

Serum Creatinine (SCr) is measured locally at baseline and at visits V1, V2, V3 and V5. The estimated glomerular filtration rate eGFR is derived from this measurement using the MDRD formula:

$$eGFR = 175 \times (SCr \text{ in mg/dl})^{-1.154} \times (\text{age in years})^{-0.203} \times 0.742 \text{ [if female]} \times 1.212 \text{ [if Black]}$$

Renal impairment (yes / no) is defined as eGFR < 30 ml/min after start of study treatment.

Anaemia

Haemoglobin (Hb) is measured locally at baseline, V3 and V5.

Anaemia (yes / no) is defined as Hb < 13 g/dl in males and Hb < 12 g/dl in females after start of study treatment.

Atrial fibrillation /atrial flutter

Any event of atrial fibrillation (yes / no) or atrial flutter (yes / no) after start of study treatment.

Mortality

Death of any cause (yes / no) after start of study treatment.

3.4 Covariates for confirmatory analyses

According to the protocol, atrial fibrillation (yes / no) and diabetes mellitus (yes / no) as documented at randomisation will be incorporated as covariates in the primary analysis because they were used to stratify randomisation.

Unfortunately, the documentation at time of randomisation did not distinguish between permanent or paroxysmal atrial fibrillation.

3.5 Further Variables

3.5.1 Baseline characteristics

Age [years] is computed as difference between year of registration and birth year as recorded on the patient registration CRF.

Sex is recorded on the patient registration CRF.

Ethnic origin (Caucasian / Asian / African / other) is recorded on the patient registration CRF.

Baseline Assessments

The following baseline assessments are recorded on the baseline CRF:

- **Body mass index** [kg/m^2] is derived from weight [kg] and height [m] as recorded on the baseline CRF
- **Waist / hip ratio** is derived from waist circumference [cm] and hip circumference [cm] as recorded on the baseline CRF
- **Blood pressure** [mm Hg]: as defined in 3.3.2
- **Heart frequency** [bpm] after 5 min rest
- **Left ventricular mass index** [g/m^2] as determined by local echocardiography and recorded at registration.
- **End diastolic wall thickness** [mm] as determined by local echocardiography and recorded at registration.
- **Ejection fraction** [%] as determined by local echocardiography and recorded at registration.
- **QT-interval [ms]** as determined at screening ECG and recorded at registration**.**
- **Atrial fibrillation (yes / no)** is recorded thrice:
    - at patient registration
    - on the baseline CRF (result of the baseline electrocardiogram: rhythm coded as sinus rhythm / atrial fibrillation / atrial flutter)
    - as assessed by the central echocardiography core lab from the baseline echo

  For atrial fibrillation in general (without distinction between permanent and paroxysmal), the status reported on the registration CRF will be used. However, this item is crosschecked with the ECG resp. the echo documentation triggering queries in case of inconsistency.

  It may be possible to derive the distinction permanent versus paroxysmal atrial fibrillation based on additional information from ECG resp. baseline echocardiography.

Heart Failure classification

- **NYHA class** (no heart failure / I / II / III / IV)
- If NYHA class I-IV: **classification of heart failure** (right sided / left-sided / global / unknown)

Cardiovascular risk factors

The following risk factors are recorded on the baseline CRF. Documentation as “unknown” is counted as missing.

- **Diabetes mellitus** (yes / no) and type of diabetes mellitus

  Diabetes mellitus (yes / no) is recorded twice: at patient registration and on the baseline CRF.

  The status reported on the baseline CRF will be used

  We crosschecked with the documentation on the registration form as well as the reported use of anti-diabetes drugs triggering queries in case of inconsistency.
- **Hyperlipidaemia** (yes/no/unknown)
- **Hyperuricemia** (yes/no/unknown)
- **Hypertension** (yes/no/unknown)

- **Sleep apnoea syndrome** (yes/no/unknown)
- **Myocardial infarction in relatives of first degree** (parents, siblings, children) before 60 years of age (yes/no/unknown)
- **Smoking habits** (never / yes / former smoker) as packyears (=twenty cigarettes smoked every day for one year) in active or former smoker
- **Drinks per week:** 1 drink = 0.25 l beer or 0.1 l wine or 0.02 l spirits

Current medication

On the baseline CRF, a list of relevant current medication is specified and current application recorded as (yes / no)

Concomitant diseases

On the baseline CRF, a list of relevant concomitant disease is specified and current diseases are recorded as (yes / no /unknown). Documentation as "unknown" is counted as missing.

Concomitant diseases documented in addition to the pre-specified list will be coded according MedDRA preferred terms.

3.5.2 Trial intervention

Compliance with the randomised intervention (mirabegron/placebo) will be assessed by:

Total dose taken

- **Days on treatment**, i.e. total number of capsules handed out to the patient and not documented as returned

Days on treatment will be computed using the following documentation on the CRF:

- Number of complete boxes (à 4 blisters = 28 capsules in total) handed out
- Number of capsules returned

Intended total dose

Days on study, i.e. from randomisation to individual visit 5 after 12 months or to the day of premature study termination (if before visit 5)

Relative dose

Days on treatment divided by days on study

4 General Analysis Definitions

4.1 Study periods

The following study periods are distinguished for analysis

- Screening period
- Baseline Assessment
- Treatment period, including regular assessment of end points
- Follow-up at 1 month after end of treatment

4.2 Analysis Populations

Analysis populations are defined according to the protocol of the BETA3_LVH-Study (see protocol chapter 9.5.1).

The **Full Analysis Set** (FAS, also called intention-to-treat (ITT) population) includes all randomised patients with valid informed consent and at least one valid measurement in at least one of the primary endpoints (baseline, 6 months (V3) or 12 months (V5)).

The confirmatory analysis is performed on the FAS and is based on the randomized arm.

The **Per-Protocol Set (PPS)** is a subset of the FAS consisting of **patients without major protocol deviations.**

The PPS approximates an ideal (contra-factual) study in which conduct was perfect and patients compliant and selected such that they tolerate the treatment. In a study to demonstrate a treatment effect, the PPS analysis is particularly useful in case of a non-significant result, in order to assess whether the ideal biological efficacy was compromised by protocol deviations.

As specified in the protocol, the following **protocol deviations are classified as major**:

- Violation of an eligibility criterion
- Patients who received less than 50% of the total dose taken of study medication (mirabegron resp. placebo)
- Patients who received both mirabegron and placebo due to a mix-up of the study drug boxes
- No valid measurement of any of the primary endpoints at the 12 month (V5) visit.

However, patients remain in the PPS if:

- Study medication had to be interrupted because of medical reasons, e.g. (S)AE, and therefore the patient received lower than 50% of the intended cumulative dose,
- they deceased during the treatment phase

The **Safety Analysis Set** consists of all randomised patients belonging to the FAS who received at least one dose of study medication. Patients will be analysed according to treatment taken.

4.3 Handling of centre effects

Random centre effects are not included in the primary analytical model since the number of patients per trial site is small in most centres in order to avoid convergence problems in over-complex models. A subgroup analysis of the treatment effect by centre will be provided as sensitivity analysis.

5 Planned analysis

A flowchart according to the CONSORT statement will describe the disposition of all patients registered to the trial detailing screening failure before randomization, withdrawals, drop-outs and inclusion in the analyses sets defined above. Respective listings will be provided.

In addition, patients with major protocol violations excluded from the PPS will be listed.

Standard methods of descriptive statistics will be used always indicating the number of valid and missing values. Summary statistics will be reasonably rounded to avoid pseudo-precision.

5.1 Demographic and other baseline parameter

Demographic and other baseline parameter will be described by randomization arm.

5.2 Concomitant Diseases and Medication

Frequencies of concomitant diseases and medication will be described by randomization arm.

5.3 Compliance with regard to the Study Intervention

Patients will be listed on whom no intervention or an intervention not corresponding to the randomization arm was performed.

Percentage of received dose of intended dose will be described by randomization arm and plotted as an empirical cumulative distribution function.

5.4 Primary and secondary efficacy endpoints

Confirmatory analysis follows the intention to treat principle as close as possible and will be based on the full analysis set.

5.4.1 Choice of scale of analysis

All time courses of metric measurements will be analysed on an appropriate scale such that linear models are adequate.

- The two primary endpoints will be analysed without scale transformation.
- The laboratory parameters
    - GDF15
    - NT-proBNP
    - hs TnT
    - Glycemic Profile
    - Insulin
    - HOMA test / % B
    - HOMA test / % S
    - Triglycerides

  are analysed on the log10 scale.
- All other parameters are analysed without scale transformation.

5.4.2 Basic analytic model

Analyses of both primary endpoints are identically structured. For secondary endpoints the same model is used, possibly adapting the dimension of the random effect to the number of visits.

We use a linear mixed model for repeated measurements. This choice of a repeated measurements linear mixed model as the primary analytic model is in line with the recommendations of (Mallincrodt et al. 2008). There is a certain rate of patients with missing measurements on the endpoints, also at baseline.

A 3-dimensional random effect with a general unstructured variance covariance matrix is used to model the dependence of measurements within-patients.

The mean structure is modelled using (0/1)-coded variables and include the following fixed effects:

- Intercept corresponding to baseline
- Timepoints: V3 and V5
- Treatment: Verum ,
- Treatment by Timepoint interaction at 6 months, 12 months: V3:Verum, V5:Verum
- Atrial fibrillation,
- Diabetes mellitus

The model implicitly uses the estimated covariance matrix to deal with missed visits. We expect that missing endpoints will be missing at random (MAR) given the specified model structure above. The above model can deal with patients with incomplete data in the endpoints as long as at least one valid measurement is documented.

The Treatment by baseline interaction is not included in the model as patients are randomised at baseline. This indirectly adjusts for possible random chance fluctuation at baseline and generally increases the power (Fitzmaurice et al. 2004, Chapter 5.7). For further discussion compare 2.1 above.

The model is fitted using the "nlme" R-package using the formula:

```
nlme::lme(Endpoint ~  V3 + V5 + V3:Verum + V5:Verum  +  AF +  DB ,
                random = ~ BASE + V3 + V5 - 1 | PATNO,
                     data = Beta3, method="REML")
```

As specified in the protocol, analysis will be based on restricted maximum likelihood (REML), which is a method to remove bias from the estimation of variance components and was shown to be advantageous in simulation studies (e.g. McNeish D 2017).

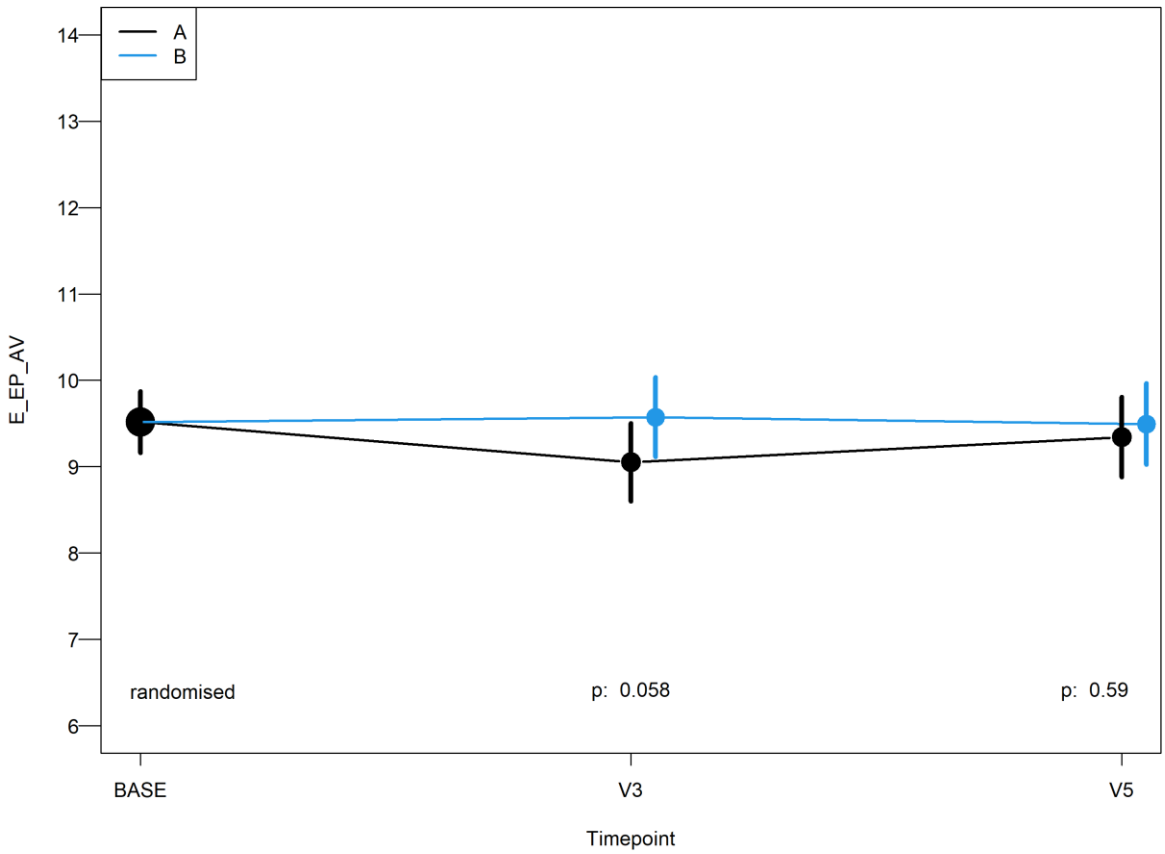
The contrast of interest for the primary endpoints is the treatment by visit interaction V5:Verum at 12 months. Respective inference will be based on Wald type confidence intervals and p-values.

The secondary endpoints concerning treatment effects at 6 months will be assessed **within the same model** by looking at the treatment by visit interaction at 6 months.

The model results will be illustrated as in this prototype figure using fake example data for e/e' depicting

- the 95% confidence intervals for the estimates of the mean time courses by treatment arm as well as
- p-values for the null hypotheses of no treatment effect at the respective time points and

The intervention effect estimate will be quantified the 95% confidence interval for the main treatment effect at 12 months.



5.4.3 Sensitivity analyses

There will be two sensitivity analyses for both primary endpoints:

1. Analysis restricted to the per protocol population. The per-protocol analysis tries to answer the study question in a hypothetical world in which the treatment strategy was implemented optimally. The per-protocol analysis becomes important, when superiority is not shown: Is there simply no effect or did we miss to demonstrate the treatment effect due to protocol deviations?
2. An ANCOVA with Baseline values, Atrial fibrillation, Diabetes mellitus and randomised arm as covariates in all randomized patients with baseline and 12 months measurements. This is a complete case analysis using a robust simple linear model without random effects.

The protocol also mentions an ANCOVA “with imputation of missing values by last information carried forward (LOCF)”. We drop this, since last information carried forward (LOCF) is an obsolete, no more state-of-the-art method.

5.5 Subgroup analysis

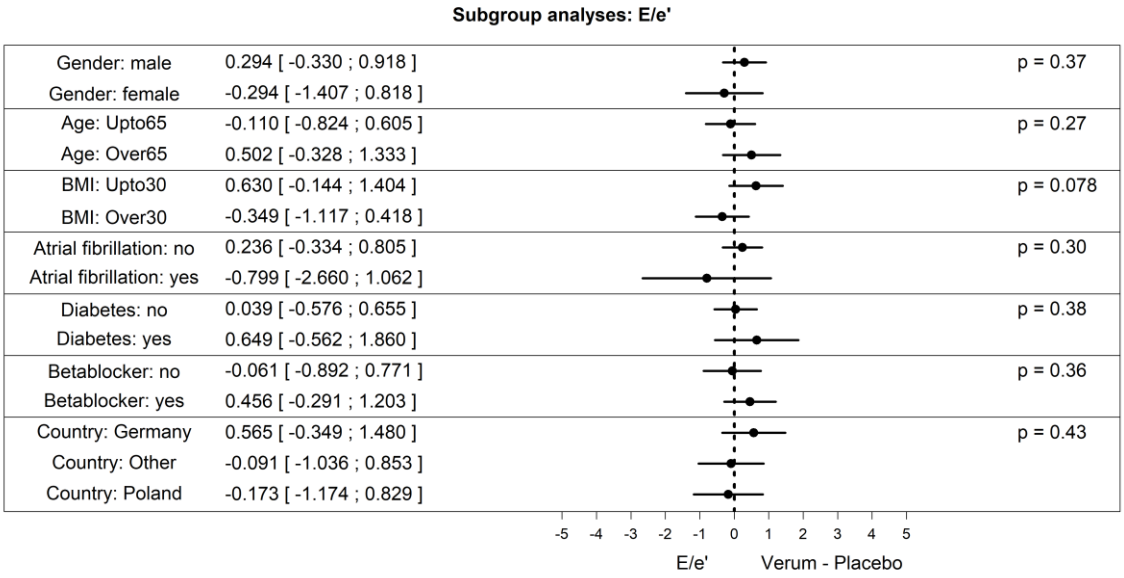
No confirmatory subgroup analyses are planned.

Strictly exploratory subgroup analyses for the two primary endpoints will be performed to assess whether the **overall treatment effect is homogenous across relevant subgroups**.

Planned explorative subgroups by baseline characteristics:

- Gender
- Use of a beta-blocker in their standard treatment (yes/no)
- Diabetes mellitus (yes / no)
- Atrial fibrillation (yes / no) at registration
- Age > 65 years at baseline
- BMI > 30 [kg/m2] at baseline
- Region (Poland / Germany / Other countries)

Results will be displayed as a forest plot with interaction tests as illustrated below with a dummy arm.



Subgroup analyses will be performed by looking at a threefold interaction:

Example for sex with an (0/1) indicator variable Male:

```
nlme::lme(Endpoint ~ Male + V3 + V3:Male + V3:Verum + V3:Verum:Male
      + V5 + V5:Male + V5:Verum + V5:Verum:Male +  AF +  DB
            random = ~ BASE + V3 + V5 - 1 | PATNO ,
                  data = Beta3, method="REML")
```

The term ***V5:Verum:Male*** assesses the relevant interaction of sex with the treatment effect at 12 months and is used to test for the presence of a treatment by subgroup interaction.

In addition, subgroup specific estimates of the treatment effect will be provided with 95% confidence intervals.

5.6 Safety aspects

5.6.1 Adverse events

The following descriptive analyses will be provided:

- Patients with at least one AE (number and %) total and by arm
- AEs per patient – descriptive statistics (number, mean, SD, minimum, maximum) total and by arm

- AEs per month of drug exposure, total and by arm
- Characteristics of AEs (seriousness, relatedness, severity, outcome, measures regarding study drug) number and % of all adverse events, total and by arm
- AEs by body system resp. preferred term: number of adverse events total and by arm
- AEs by body system: number of patients with at least one event, total and by arm,
- AEs by preferred term: number of patients with at least one event, total and by arm

5.6.2 Serious adverse events

The following descriptive analyses will be provided:

- SAEs by body system resp. preferred term: number of adverse events total and by arm
- SARs (serious adverse reactions) by body system resp. preferred term: number of adverse events total and by arm
- Listings of all serious adverse events

5.6.3 Events of special interest

The following descriptive analyses will be provided:

Blood pressure

- Patients with at least one event of elevated blood pressure, total and by arm,
- Patients with at least one event of highly elevated blood pressure, total and by arm,
- Listing of all patients with elevated blood pressure
- Individual time course of systolic and diastolic blood pressure, by arm (in particular for patients with elevated blood pressure)

Hepatic impairment

- Patients with at least one event of hepatic impairment, total and by arm,
- Listing of all patients with hepatic impairment
- Individual time course of ALT [ULN] and AST [ULN], by arm (in particular for patients with hepatic impairment), displayed on the log10 scale

Renal impairment

- Patients with at least one event of renal impairment, total and by arm,
- Listing of all patients with renal impairment
- Individual time course of eGFR [ml/min] by arm (in particular for patients with renal impairment)

Anaemia

- Patients with at least one event of anaemia, total and by arm,
- Listing of all patients with anaemia
- Individual time course of Hb [g/dl] by arm (in particular for patients with anaemia)

Atrial fibrillation / flutter

- Patients with at least one event of atrial fibrillation, total and by arm,
- Listing of all patients with atrial fibrillation

Deaths

- Listing of all deceased patients

6 References

Bundgaard H.; Axelsson Raja, A.; Iversen, K.; Valeur, N; Tønder, N.; Schou, M.; Hørby Christensen, A.; Eske Bruun, N.; Søholm, H.; Ghanizada, M.: Fry, N.A.S.; Hamilton, E.J.; Boesgaard, S.; Møller, M.B.; Wolsk, E.; Rossing, K.; Køber, L.; Rasmussen, H.H.; Vissing, C.R.: Hemodynamic Effects of Cyclic Guanosine Monophosphate-Dependent Signaling Through β3 Adrenoceptor Stimulation in Patients With Advanced Heart Failure: A Randomized Invasive Clinical Trial. Circ Heart Fail 2022 Jul;15(7):e009120.

DuBois D, DuBois EF: A formula to estimate the approximate surface area if height and weight be known (1916). Archives of Internal Medicine. 17, 1916, S. 863–871.

Fitzmaurice, Garrett M.; Laird, Nan M.; Ware, James H. (2004): Applied Longitudinal Analysis, Book: Wiley-Interscience, Hoboken, New Jersey.

Garcia-Lunar L.; Blanco, I. ; Fernández-Friera, L. ; Gonzàlez, S. ; Jordà, P.; Sánchez, P.; Pereda, D.; Pujadas S.; Rivas M.; Solé-Gonzalez, E.; Vázquez J.; Blázquez, Z.; García-Picart J.; Caravaca, P.; Escalera, N.; Garcia-Pavia, P.; Delgado, J.; Segovia-Cubero J.; Fuster V.; Roig, E.; Albert Barberá, J.; Ibanez, B.; García-Álvarez, A.: Design of the β3-Adrenergic Agonist Treatment in Chronic Pulmonary Hypertension Secondary to Heart Failure Trial. JACC Basic Transl Sci 2020 Mar 11;5(4):317-327.

Hochberg, Yosef (1988): A sharper Bonferroni procedure for multiple tests of significance. In: Biometrika 75 (4), S. 800–802.

McNeish, Daniel (2017) Small Sample Methods for Multilevel Modeling: A Colloquial Elucidation of REML and the Kenward-Roger Correction, Multivariate Behavioral Research, 52:5, 661-670.

Mallincrodt, Craig H.; Lane, Peter W.; Schnell, Dan; Peng, Yahong; Mancuso, James P. (2008): Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. In: Drug Information Journal (42), S. 303–319

Wonisch M, Hofmann P, Pokan R, Kraxner W, Hödl R, Maier R, Watzinger N, Smekal G, Klein W, Fruhwald F. (2003). Spiroergometrie in der Kardiologie-Grundlagen der Physiologie und Terminologie. Journal für Kardiologie-Austrian Journal of Cardiology, 10(9), 383-390.

Yamamoto, Kazuhiro; Origasa, Hideki; Hori, Masatsugu (2013): Effects of carvedilol on heart failure with preserved ejection fraction: the Japanese Diastolic Heart Failure Study (J-DHF). In: European journal of heart failure 15 (1), S. 110–118.

7 Appendix

7.1 Abbreviations

ABPM	Ambulatory blood pressure monitoring
AE	Adverse Event
AF	Atrial Fibrillation
ALT	Alanine aminotransferase
AR	Adrenergic Receptor
AST	Aspartate aminotransaminase
BMI	Body Mass Index
BPM	Beats per Minute
BSA	Body Surface Area
CA	Competent Authority
cMRI	cardiac MRI
CRF	Case Report Form
CTC	Clinical Trial Centre

DSMB	Data Safety and Monitoring Board
EAC	Event Adjudication Committee
EC	Ethics Committee
eCRF	Electronic Case Report Form
EF	Ejection Fraction
eGFR	Estimated glomerular filtration rate
GCP	Good Clinical Practice
Hb	Hemoglobin
HCT	Hematocrit
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction
HR	Heart Rate
ICH	International Conference on Harmonisation
ICMJE	International Committee of Medical Journal Editors
ISF	Investigator Site File
LAVOLI	Left atrial volume index
LGE	Left Gadolinium Enhancement
LV	Left Ventricle
LVM	Left Ventricular Mass
LVMI	Left Ventricular Mass Index
MHRD	Maximum recommended human dose
OAB	Overactive Bladder
REML	Restricted maximum likelihood estimation
SAE	Serious Adverse Event
SAR	Serious Adverse Reaction
SUSAR	Suspected Unexpected Serious Adverse Reaction
TMF	Trial Master File
V1-5	Visits 1-6, at 1, 3, 6, 9, and 12 months after randomisation
WOCB	Women of child bearing potential