Document Type:	Statistical Analysis Plan
Official Title:	A multicenter, randomized, placebo-controlled, parallel group, double blind, dose-finding Phase II trial to study the efficacy, safety, pharmacokinetic and pharmacodynamic effects of the oral partial adenosine A1 receptor agonist neladenoson bialanate over 20 weeks in patients with chronic heart failure with reduced ejection fraction
NCT Number:	NCT02992288
Document Date:	29 MAY 2018

Title page

A multicenter, randomized, placebo-controlled, parallel group, double blind, dose-finding Phase II trial to study the efficacy, safety, pharmacokinetic and pharmacodynamic effects of the oral partial adenosine A1 receptor agonist neladenoson bialanate over 20 weeks in patients with chronic heart failure with reduced ejection fraction

short title: PANTHEON

Bayer study drug	Neladenoson bialanate / BAY 1067197		
Study purpose:	dose finding		
Clinical study phase:	IIb	**Date:**	29MAY2018
Study No.:	BAY 1067197 / 15128	**Version:**	3.0
Author:	PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Abbreviations

A1R	A1 receptor
ACE	angiotensin-converting enzyme
ACEi	ACE inhibitor
AE(s)	adverse event(s)
AF	atrial fibrillation
ALT	alanine aminotransferase
ARB(s)	angiotensin receptor blocker(s)
ARNI(s)	angiotensin receptor-neprilysin inhibitor(s)
AST	aspartate aminotransferase
AUC	area under the time-concentration curve
AV	atrioventricular
BMI	body mass index
BNP	b-type natriuretic peptide
BSA	body surface area
CHF	chronic heart failure
C_{max}	maximum drug concentration in plasma
CO	cardiac output
CRF	case report form
CV	cardiovascular
DMC	Data Monitoring Committee
e.g.	for example (*exempli gratia*)
ECG	electrocardiogram
eCRF	electronic CRF
EF	ejection fraction
eGFR	estimated glomerular filtration rate
EQ-5D-5L	EuroQol Group 5-dimensional, 5-level questionnaire
EU	European Union
EWDT	E-wave deceleration time
FAS	full analysis set
GFR	glomerular filtration rate
GGT	gamma glutamyl transpeptidase
HDL	high-density lipoprotein
HF	heart failure
HFpEF	heart failure with preserved ejection fraction
HFrEF	heart failure with reduced ejection fraction
HR	heart rate
hs-TNT	high sensitivity troponin T
i.e.	that is (*id est*)
ICF	informed consent form

ICH	International Conference on Harmonization
INR	international normalized ratio
ITT	intent-to-treat
IVSD	interventricular septum diameter
KCCQ	Kansas City cardiomyopathy questionnaire
LA	left atrial
LAV	LA volume
LAVI	LA volume index
LDH	lactate dehydrogenase
LDL	low-density lipoprotein
LV	left ventricular
LVEDV	LV end-diastolic volume
LVEDVI	LVEDV index
LVEF	left ventricular ejection fraction
LVESV	left ventricular end-systolic volume
LVESVI	LVESV index
MAP	mean arterial pressure
MCH	mean corpuscular hemoglobin
MCHC	MCH concentration
MCP	multiple comparison procedures
MCV	mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
mL	milliliter
MRA	mineralocorticoid receptor antagonist
NGAL	neutrophil gelatinase-associated lipocalin
NONMEM	non-linear mixed effect modeling
NT-proBNP	N-terminal pro-hormone b-type natriuretic peptide
NYHA	New York Heart Association
PASP	pulmonary artery systolic pressure
PDV	premature discontinuation visit
Pes	end-systolic pressure
pg	picogram
PK	pharmacokinetic
PKS	PK analysis set
PP	pulse pressure
PPS	per-protocol set
PTT	partial thromboplastin time
PWT	posterior wall thickness
RV	right ventricular
SAC	systemic arterial compliance

SAE(s)	serious adverse event(s)
SAF	safety analysis set
SAP	statistical analysis plan
SBP	systolic blood pressure
SV	stroke volume
SVI	SV index
TAPSE	tricuspid annular plane systolic excursion
TD	tissue Doppler
TPR	total peripheral resistance
UACR	urine albumin-to-creatinine ratio
US	United States (of America)

1. Introduction

Chronic heart failure (CHF) is a major public health problem characterized by significant mortality, frequent hospitalizations, and poor quality of life, with an overall prevalence that is increasing throughout the world. HF carries a prognosis comparable to many forms of cancer with a 5-year survival rate of approximately 50% (1), which exceeds that of many cancers (2). Frequent hospitalizations, along with other direct and indirect costs, also place an enormous financial burden on healthcare systems; more is spent annually in the US for the treatment of HF by Medicare than on any other Medicare-covered condition (3).Upon hospitalization for HFrEF mortality and re-hospitalization for HF remain high despite treatment with stable chronic HFrEF neurohumoral blockers, indicating that this does not fully explore the complexity of HF pathophysiology. Additionally, the repeated stepwise addition of hemodynamically active medications raises tolerability and safety concerns (e.g. hypotension and bradycardia) (4).

Therefore, addressing the failing heart directly and getting viable but not functional myocardium back to work might be a new option for the development of the next generation of hemodynamically silent HF drugs. In this context, neladenoson bialanate holds promise as a potentially hemodynamically neutral therapy for HF that could simultaneously improve cardiomyocyte energetics, calcium homeostasis, cardiac structure and function, and long-term clinical outcomes when added to background therapies.

Neladenoson bialanate (BAY 1067197, which is the free base of the hydrochloride BAY 86-8901) is the pro-drug of the pharmacologically active compound BAY 84-3174, a highly potent and selective non-adenosine like partial adenosine A1 receptor agonist suitable for once daily oral use.

The main limitations of using full A1R agonists in cardiovascular indications such as HF are undesired cardiac effects, such as bradycardia and higher degree AV block as well as negative cardiac inotropy and dromotropy. In contrast, preclinical data show that the partial adenosine A1R agonist neladenoson bialanate can be used to modulate and trigger primarily favorable pharmacological responses for HF therapy and avoid undesired effects such as AV conduction abnormalities and higher degree AV Block. Nevertheless, based on the mode of action, there are theoretical concerns particularly with regard to undesired effects, such as bradycardia and higher degree AV block for an A1R agonist, which might be aggravated by concomitant use of standard heart failure therapy.

The purpose of this clinical trial is to assess the safety, tolerability and the pharmacokinetic and pharmacodynamic response of 20 weeks' treatment with neladenoson bialanate compared to placebo in patients with chronic HFrEF on standard therapy for HF and to find the optimal dose for a further Phase III trial.

This statistical analysis plan (SAP) is based on the Global Integrated Clinical Study Protocol BAY 106-7197/15128 version 2.0, dated 17 FEB 2017.

2. Study Objectives

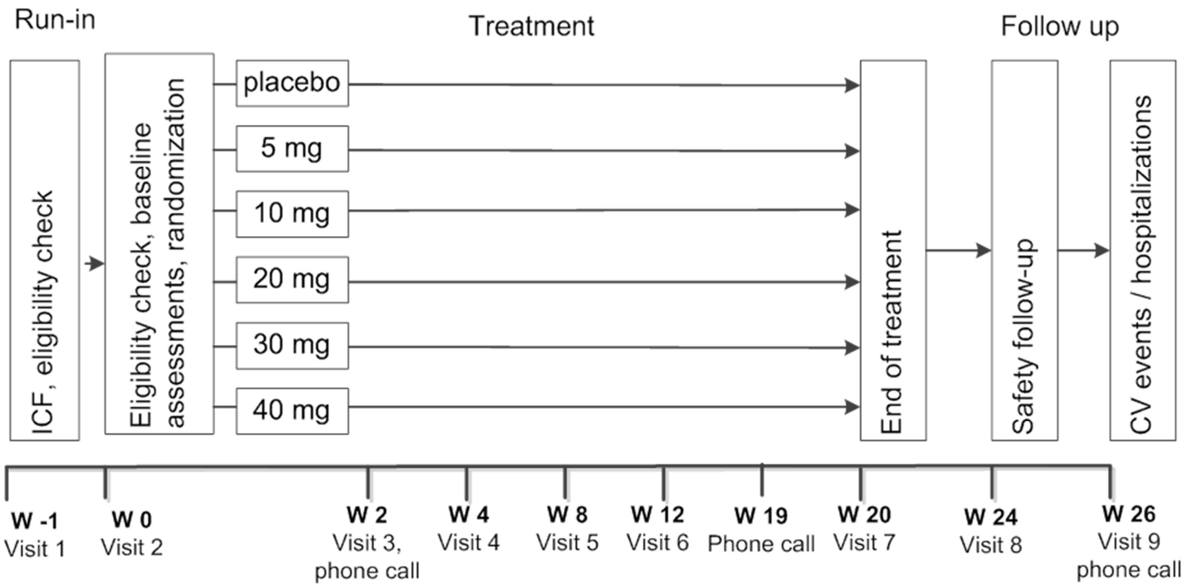
The objective of the study is to find the optimal dose of neladenoson bialanate for the Phase III trial by detecting and characterizing a significant dose-response relationship in the two primary efficacy endpoints, absolute change from baseline in LVEF and log-transformed NT-proBNP at 20 weeks, in patients with chronic heart failure with reduced ejection fraction (HFrEF), and by characterizing the safety, tolerability and pharmacodynamic effects of the compound when given in addition to standard therapy for HFrEF.

An exploratory objective is to further assess pharmacokinetic parameters and blood and urine biomarkers.

3. Study Design

Study 15128 is a randomized, double-blind, placebo-controlled, parallel group, multicenter, Phase II study. Figure 3–1 displays the overall study design.

Figure 3–1: Study design overview



Abbreviations: CV = cardiovascular; ICF = informed consent form; W = week

Approximately 384 patients from approximately 90 study centers worldwide will be randomized to one of the active treatment dose arms or placebo, in addition to standard of care heart failure therapy (for details see Section 7).

The study will comprise a 1-week run-in period, 20-week treatment period, and a 6-week follow-up period (27 weeks total).

Patients will have site visits at weeks -1, 0 (baseline), 4, 8, 12, 20 (end-of-treatment visit) and 24 (safety follow-up visit). In addition, 2 phone calls at weeks 2 and 26 will be made to assess patients' safety, and one additional phone call - to remind the patients of AVIVO self-application at Week 19.

LVEF will be measured by echocardiography at baseline and end of treatment, NT-proBNP will be measured at baseline, Week 4 and end of treatment. Safety will be monitored throughout the study. PK samples will be taken from all patients at dedicated time points.

Additional biomarkers reflecting the pharmacodynamic activity of the drug will be examined, as well as candidate biomarkers that may predict drug response.

For detailed visit descriptions and rules for patients who discontinue study treatment earlier, please see protocol Sections 9.1 and 9.2.

The anticipated duration of the study as a whole is approximately 19 months: this includes an anticipated recruitment period of 13 months followed by a run-in period of 1 week, a treatment period of 20 weeks and a follow-up period of 6 weeks after enrollment of the last patient into the trial.

A parallel group design was chosen to compare five different once-daily dose regimens and one placebo arm to find the best dose for Phase III. Placebo control is used to control for observer and subject bias, and randomization - to control for assignment bias. Based on Phase I data in healthy volunteers and Phase II data on heart failure patients, a sequential dose escalation design was not deemed necessary for neladenoson bialanate since the safety profile of the compound could be verified in dose ranges up to 40 mg. Safety of the subjects in this parallel study design will be closely monitored by an Data Monitoring Committee (DMC).

The end of the study as a whole will be reached as soon as the last visit of the last patient has occurred in all centers in all participating countries (EU and non-EU).

4. General Statistical Considerations

4.1 General Principles

The statistical evaluation will be performed by using the software package SAS version 9.2 or higher (SAS Institute Inc., Cary, NC, USA). Unless otherwise noted, data will be analyzed by descriptive statistical methods: The number of data available and missing data, mean, standard deviation, minimum, quartiles, median, and maximum will be calculated for metric data. Frequency tables will be generated for categorical data.

Definition of efficacy and safety endpoints, analysis strategies, structure of analysis datasets and layout of analysis data displays are following Bayer AG standards: Neladenoson bialanate Project Standards, the Therapeutic Area Cardiovascular Standards (TAS) and the Global Medical Standards (GMS), respectively. The given ordering reflects the priority of the different standards, i.e. specifications of the latter ones have to be followed only if not specified in standards mentioned before. Study-specific specifications may be included in addition to the Project Standards, if needed.

4.2 Handling of Dropouts

A “dropout” is defined as a subject who has been randomized and discontinues study participation prematurely for any reason. Subjects withdrawn from study treatment will not be replaced. Refer to Section 6.4 in the Clinical Study Protocol for withdrawal of subjects from study.

See following chapters for more details on deriving efficacy endpoints in case of missing data.

4.3 Handling of Missing Data

All missing or partial data will be presented in the patient data listing as they are recorded in the eCRF.

A number of descriptive analyses will be performed to better understand missing data patterns. The frequency, proportion and the reasons for premature discontinuation of both the study and study treatment will be reported. Kaplan-Meier plots for "time to end of study treatment (calculated as days from first dose to the earliest date of stop medication, including premature stop of study medication and death, for the calculation all the subjects will be considered to have an event-stop of study medication)" and "time to end of study"(calculated from randomization to the earliest date of visit 9, death, and the last visit if subject drops off from study prematurely, for the calculation all the subjects will be considered to have an event-stop of study) will be provided, by treatment group and overall.

The number of patients who prematurely discontinue study participation or intake of study medication will be summarized with respect to the key subgroups, see section 6.4.1. and the reasons for premature discontinuation of study and / or study treatment. If the proportion of patients who withdraw across the dose groups is not fairly balanced, the impact on the primary variables will be further explored. To further explore the missingness pattern with regards to the "missing at random" assumption, the mean of the baseline values of the efficacy variable will be summarized for patients with and without post-baseline observations, by treatment group and overall.

For the analysis of the primary and secondary variables, it cannot necessarily be assumed that data are missing at random. As the choice of primary analysis will be based on assumptions that cannot be verified, the robustness of the results of the primary analysis will be investigated through appropriate sensitivity analyses making different assumptions, in accordance with the EMA "Guideline on missing data in confirmatory clinical trials". Detail missing data handling are specified in Section 6.2.1.4.

When appropriate, the following rules will be implemented so as not to exclude subjects from statistical analyses due to missing or incomplete data:

- **Date of CHF diagnosis**
  For cases where start month and year are reported but day is missing, impute it with the 1st day of month. If the month is not available, this date will not be imputed.
- **Clinical outcomes**
  For cases where start month and year are reported but day is missing, impute the maximum of (date of randomization, first date of study medication, 15.month.year). For cases where only start year is reported or completely missing, impute the maximum of (date of randomization, first date of study medication, 15.01.year), but not later than death date if the subject died.
- **Heart failure related concomitant medication start date**
  For case where start month and year are reported but day is missing, impute it with 15th day of month. For cases where only start year is reported or completely missing, impute it as maximum of (15.01.year, randomization).
- **Heart failure related concomitant medication stop date**

For case where stop month and year are reported but day is missing, impute it as minimum of [(15, month, year) and (last visit date) and (death date)].
If the stop day and month are missing, then the stop date will be imputed as minimum of [(15.12.year) and (last visit date) and (death date)].
If the date is completely missing then the stop date will be imputed as minimum of [the last visit date and death date]. If the concomitant medication is "Ongoing at subject's last visit", for the respective stop date variable the 'last visit date' from the corresponding domain is merged in the concomitant medication database by data management programming.

- **Study medication start date for subjects took study medication**
  If the start date and time is missing it will be imputed with the randomization date and time. If start date and time is recorded as earlier than randomization and cannot be clarified, date and time of randomization will be used for the statistical analysis.
- **Study medication stop date for subjects took study medication**
  If the stop day is missing, but the stop month and stop year are available then the stop date will be imputed as minimum of [(15, month, year) and (last on treatment visit date before the observational visit) and (death date)].
  If the stop day and month are missing or the date is completely missing then the stop date will be imputed as minimum of [the last on treatment visit date before the observational visit and death date].

4.4 Interim Analyses and Data Monitoring

A formal interim analysis is not planned. Periodic data review by a DMC will be performed to monitor safety and tolerability. An Independent Data Monitoring Committee (IDMC) will be applied to this study. An external statistical analysis center will provide results to the DMC

4.5 Data Rules

Generally, for each date stored in database a set of organizational variables will be derived in order to describe the temporal context of that date in the specific study: Phase of treatment (pre, during or post study treatment), day relative to the start of study treatment, day relative to the end of study treatment will be provided.

4.5.1 Baseline and Change from Baseline

For efficacy endpoints, the efficacy baseline is defined as the last available value prior to or on the date of randomization. Safety baseline is defined as the last available value before the first study medication administration. If values are missing at the baseline (visit 2, week 0), data recorded at run-in (Visit 1) will be considered as safety baseline value. If run-in record is also missing, the baseline value will be left as missing.

Change from baseline for vital signs or laboratory parameters will in general be displayed as the difference to baseline defined as:

Change = Post baseline value – baseline value.

In addition, for some parameters the relative change will be analyzed defined as

*Relative change = 100% * [(post baseline value – baseline value) / baseline value].*

4.5.2 Repeated Measurements

At all visits post-randomization and if not stated otherwise, only the values at scheduled time points will be used for analysis, although unscheduled results will be included in tables reporting any abnormalities, e.g. incidences of high laboratory abnormalities.

For the derived visit "Any time post baseline" this will include any measurement after initiation of study drug, including unscheduled assessments.

4.5.3 Laboratory Data Handling

The data of hematology, clinical chemistry, and coagulation will be provided by central laboratories. Additional re-tests for liver monitoring will be done locally.

For values which are below the lower limit of quantification (LLOQ), half the value of the LLOQ will be used for analysis. Differences between two values of below the LLOQ will be assigned values of 0.

In case of measurements above the upper limit of quantification (ULOQ), the following rules will be applied:

- The ULOQ will be used for calculations.
- Corresponding tables and figures will get a footnote indicating that "Values above the upper limit of quantification of ULOQ were replaced by ULOQ."
- Tables displaying maximum values will show up ">ULOQ" as maximum.

If more than one assessment occurred at any post-baseline visit (repeated measures at same visit), the last valid (non-missing) value will be used in the summaries. Unscheduled laboratory data will be listed and included in the summary tables.

4.5.4 Subgroup analyses

In order to assess the homogeneity of the dose response across the most important prognostic and predictive factors, subgroup analyses will be performed.

'Key' subgroups include:

- LVEF(%) at baseline: ≤25 vs. >25
- LVEF(%) at baseline: ≤40 vs. >40
- LVEF(%) at baseline: ≤45 vs. >45
- NT-proBNP (pg/ml) at baseline: ≤ median vs. > median
- NYHA class at baseline: II vs. III / IV
- Prior ß-blocker: yes vs. no

All other exploratory subgroups comprise demographic and baseline characteristics specified in section 6.1.2.

If the total number of subjects in a subgroup category is too small, the respective subgroup category will be either omitted from the analysis or combined with other categories, if a logical combination to another subgroup category is possible.

Prior hospitalization for heart failure is defined based on one of inclusion criterions.

CRF item (subject group) selected	Prior hospitalization (outpatients visit)
CHF Hospitalization/ unscheduled outpatient clinic visit with BNP/NTproBNP measurement in the last 3 months	Yes
Only BNP/NTproBNP measurement in the last 4 weeks	No
Combination of CHF Hospitalization/ unscheduled outpatient clinic visit with BNP/NTproBNP measurement in the last 3 months and additional measurement in the last 4 weeks	Yes

4.6 Blind Review

The results of the final data assessment will be documented in the final list of important deviations, validity findings and assignment to analysis set(s). Any changes to the statistical analysis prompted by the results of the review of study data will be documented in an amendment and, if applicable, in a supplement to this SAP.

5. Analysis Sets

Documentation of protocol deviations and assignment of patients to analysis sets will be performed according to the sponsor's applicable Standard Operating Procedures and ocor Operation Instructions.

5.1 Assignment of analysis sets

Final decisions regarding the assignment of subjects to analysis sets will be made during the review of study data and documented in the final list of important deviations, validity findings and assignment to analysis set(s) (see section 4.6).

Data for all patients who signed informed consent but were not randomized will not be included in any statistical analyses except standard disposition tables and listings provided in the clinical study report (Screening failures and discontinued patients).

The statistical analysis sets are defined as follows:

Full analysis set (FAS)

The FAS population consists of all randomized unique patients. According to the ICH E9 guideline, this analysis set is as complete as possible and as close as possible to the intent-to-treat (ITT) ideal. Patients will be analyzed as randomized. The FAS will be used to display baseline characteristics and to display efficacy analyses. Sensitivity analyses are based on the FAS population. For the analyses conducted in FAS, patients will be analyzed as randomized per IxRs.

Safety analysis set (SAF)

The SAF population consists of all randomized patients who received at least one dose of study medication after randomization. The SAF will be used to display safety analyses. For analyses conducted in SAF, patients will be analyzed as treated.

Per-protocol set (PPS)

The protocol defines a PPS. Since there are two primary efficacy variables, we will define PPS LVEF and PPS BNP, specific to the presence of non-missing data for each variable separately.

Per-protocol set for LVEF (PPS LVEF): The PPS LVEF population consists of all FAS population patients without validity findings. Validity findings may include adherence and compliance issues and the violation of inclusion / exclusion criteria affecting efficacy evaluation. If the left ventricular ejection fraction of this subject is measured after first dose of study medication but within a specified time frame this subject will not be excluded from PPS LVEF. The patients with invalid/missing baseline value or missing post-baseline value not due to CV death or HF hospitalization will be excluded from the PPS LVEF. The primary analysis of LVEF will be conducted in PPS LVEF.

Per-Protocol set for NT-proBNP (PPS BNP): The PPS BNP population consists of all FAS population patients without validity findings. Validity findings may include adherence and compliance issues and the violation of inclusion / exclusion criteria affecting efficacy evaluation. If the NT-proBNP (pg/mL) of this subject is measured after first dose of study medication but within a specified time frame this subject will not be excluded from PPS BNP. The patients with invalid/missing baseline value or missing post-baseline value not due to CV death or HF hospitalization will be excluded from the PPS BNP. The primary analysis of NT-proBNP will be conducted in PPS BNP.

A list of potential validity findings will be provided in a separate important deviations and validity specifications document which will be finalized before database lock. The detailed definitions and the assignment of patients to this analysis set will be based on the validity review meeting. Patients will be analyzed as treated.

Pharmacokinetic analysis set (PKS)

The PKS population consists of all patients treated with neladenoson bialanate with at least 1 valid BAY 84-3174 plasma concentration and without protocol deviation that would interfere with the evaluation of the PK data.

6. Statistical Methodology

The formal statistical analyses will be both descriptive and inferential. Summaries will be provided for each of the treatment group. All of analyses planned in this SAP will be repeated in Japanese subjects only.

6.1 Population characteristics

6.1.1 Disposition of Subjects

The following will be tabulated overall and/or by treatment group:

- Study sample sizes (FAS, PPS LVEF, PPS BNP, SAF and PKS)
- Study sample sizes by region, country, and site
- Subject disposition
- Number of subjects and primary reasons for screening failures (only overall)
- Number of subjects and primary reasons for permanent discontinuation of study medication (by treatment group and overall for FAS and SAF)
- Number of subjects and primary reasons for discontinuation from study (by treatment group and overall for FAS and SAF)

6.1.2 Demographic and Baseline Characteristics

Descriptive summaries of demographics and baseline characteristics will be presented by treatment group and overall for the PPS LVEF, PPS BNP, FAS and SAF populations. Comparability of the treatment groups with respect to demographics and baseline characteristics will be assessed using the descriptive summaries. Same analyses will also be performed for subjects who prematurely discontinue study participation or intake of study medication.

The following demographic data will be summarized:

- Age at baseline (years);
- Age category: <65, 65-75, >75 years;
- Gender (male vs. female)
- Race and ethnicity;
- Height (cm);
- Weight (kg) at baseline;
- BMI (kg/m^2);
- BMI category: $\leq$30 vs. >30 kg/m^2;
- Tobacco smoking history
- Alcohol consumption history
- Recent caffeine-containing beverage and chocolate consumption history

The following baseline characteristics will be summarized:

- LVEF(%) at baseline: ≤25 vs. >25
- LVEF(%) at baseline: ≤40 vs. >40
- LVEF(%) at baseline: ≤45 vs. >45
- LVEF(%) at baseline: ≤mean vs. >mean
- NT-proBNP (pg/ml) at baseline: ≤ median vs. > median
- NYHA class at baseline: II vs. III / IV
- Prior ß-blocker: yes vs. no
- Region: North America /West Europe / East Europe / Asia
- Region: Japan vs. rest-of-the-world (analysis to be included in reports specific for Japan only)
- Time of CHF diagnosis to randomization (months): ≤3 vs. >3
- Prior hospitalization for heart failure: yes vs. no
- Etiology of CHF: ischemic vs. non-ischemic
- Diabetes: yes vs. no
- Atrial fibrillation: yes vs. no
- Hypertension: yes vs. no
- Prior CHF medication:
    - Prior ß-blocker in max tolerated dose: yes vs. no
    - Prior use of ivabradine: yes vs. no
    - Prior use of MRA: yes vs. no
    - Prior use of ACE inhibitor: yes vs. no
    - Prior use of ARB: yes vs. no
    - Prior use of angiotensin receptor-neprilysin inhibitor (ARNI): yes vs. no
- Estimated GFR (ml/min/1.73 m^2): ≤60 vs. >60

6.1.3 Medical history

Medical history findings will be summarized using medical dictionary for regulatory activities (MedDRA, version refers to TS domain) terms for the FAS population by treatment group.

6.1.4 Prior and Concomitant Medications

All non-study medications taken during the study will be coded using the World Health Organization Drug Dictionary (WHO-DD) in its latest version which is specified in the TS domain and the Anatomical Therapeutic Chemical (ATC) classification system. Coding will include the drug class and preferred drug name.

Non-study medications taken during the study will be categorized as prior medications, concomitant medications during the treatment period, and post treatment medications during the safety follow-up.

Prior medications will be defined as a non-study medication with a stop date prior to the first dose of study treatment.

Concomitant medications will be defined as:

- Non-study medications with a start or stop date on or after the date of the first dose of study treatment;
- Non-study medications that started prior to the first dose of study treatment and are ongoing during the treatment period;
- Non-study medications with partial start dates that indicate that the medication could be concomitant in relation to the date of the first dose of study treatment;
- Non-study medications with completely missing start dates, unless their stop dates confirm otherwise (i.e. the stop date is before the first dose of study treatment).

Post treatment medications are defined as non-study medications taken up to 6 weeks after the last study medication intake.

All concomitant medications will be listed, including verbatim descriptions and coded terms, and flags for prior medications. Prior, concomitant, and post treatment medications will be summarized using frequencies of subjects reporting each drug category and perferred drug name. Additionally

- concomitant medications of interest required by the HF guideline, i.e. ACEI/ARB/ARNI, MRA, beta blockers and/or Ivabradine and their combinations
- other medications of interest relevant to the cardiovascular system, i.e.digitalis glycosides,loop and thiazide diuretics, Potassium sparing agents (excluding MRAs), Statins, anticoagulants, antiplatelets, natriuretic peptides, soluble guanylate cyclase stimulators, GLP-1 antagonists, insulins, and SGLT-2 inhibitors,

will be summarized using frequencies of subjects reporting each preferred drug name at baseline and post-baseline. Frequency of subjects who experienced dose increase, dose decrease or no dose change of concomitant medication of interested required in the HF guideline will also be summarized.

For each subject, multiple records of the same concomitant medication will be counted once within a drug class and preferred name.

6.2 Efficacy

6.2.1 Primary efficacy variables and analyses

6.2.1.1 Primary efficacy variables

- Absolute change from baseline in left ventricular ejection fraction (LVEF; %) after 20 weeks measured by echocardiography, i.e., LVEF at week 20 minus LVEF at baseline.
- Absolute change from baseline in log-transformed NT-proBNP (pg/mL) after 20 weeks, i.e., log-transformed NT-proBNP at week 20 minus log-transformed NT-proBNP at baseline.

6.2.1.2 Primary analysis of primary efficacy variables

Primary analysis will be performed in PPS LVEF and PPS BNP respectively, a subset of the FAS comprising “compliant and adherent” patients (as much as possible, defined via validity criteria). To avoid bias, the PPS LVEF and PPS BNP will also include those “compliant and adherent” patients who are “censored” due to CV death or a hospitalization for HF preventing the assessment of the relevant efficacy endpoints 20 weeks after randomization to take place as planned. For missing post-baseline value due to CV death or study drug/study discontinuation due to HF, a worst case approach (WOV) will be applied. The worst change by treatment group will be calculated. If the worst change in that treatment group is positive for LVEF / negative for NT-proBNP, the worst case in that treatment group is 0. The missing change from baseline value will be imputed with a multiple 1.0 of the median of worst changes from baseline value among all treatment groups. The post-baseline value at week 20 will be calculated accordingly as baseline value + imputed change from baseline. If this imputed post-baseline value for LVEF is < 0% in case of CV death preventing the measurement, then the post-baseline value will be replaced with 0 and the imputed change from baseline will be modified to (- baseline value). Similarly, if the imputed post-baseline value is < 5% in case of a hospitalization for HF and no CV death preventing the measurement, the post-baseline value will be set to 5% and the corresponding imputed change from baseline will be reset to (5 - baseline value).

For both primary efficacy variables, it is planned to perform a test for a dose-response signal, respectively, under the assumption of a nearly monotone dose-response relationship in the dose range considered. The MCP-Mod method (5) combining multiple comparison procedures (MCP) principles with modeling techniques will be used for the primary statistical analysis of the primary efficacy variables. This method allows the flexibility of modeling for dose estimation, while preserving the robustness to model misspecification associated with MCP procedures. The MCP-Mod method will be used based on SAS programs provided (9) and the results may be validated within R (10) with the actual DoseFinding package (11).

Two primary variables have been defined to better evaluate multiple dimensions of the treatment effect, thereby adding a source of multiplicity. The overall outcome of the primary efficacy analysis will be considered positive if at least one of the two null hypotheses related to the primary efficacy variables can be rejected. The Hochberg (step-up) procedure will be applied to control the family-wise error rate (FWER):

Consider the null hypotheses for LVEF, $H_{0,LVEF}$, and for NT-proBNP, $H_{0,NT\text{-}proBNP}$, and their associated raw p-values p_{LVEF} and $p_{NT\text{-}proBNP}$. The Hochberg test is performed using the following steps:

1. If $\max(p_{LVEF}, p_{NT\text{-}proBNP}) \leq \alpha = 0.05$, both null hypotheses $H_{0,LVEF}$ and $H_{0,NT\text{-}proBNP}$ will be rejected and the procedure stops.
2. If $\max(p_{LVEF}, p_{NT\text{-}proBNP}) > \alpha = 0.05$ and $\min(p_{LVEF}, p_{NT\text{-}proBNP}) \leq \alpha/2 = 0.025$, we will fail to reject the null hypothesis related to the larger p-value $\max(p_{LVEF}, p_{NT\text{-}proBNP})$ and the null hypothesis related to the smaller p-value $\min(p_{LVEF}, p_{NT\text{-}proBNP})$ will be rejected.
3. Otherwise, we will fail to reject both null hypotheses $H_{0,LVEF}$ and $H_{0,NT\text{-}proBNP}$.

Assumptions

Five active doses of neladenoson bialanate will be used in this study: 5 mg, 10 mg, 20 mg, 30 mg, and 40 mg, as well as a placebo arm corresponding to a 0 mg dose.

The measurements of the primary efficacy variables are assumed to be normally distributed with the same standard deviation σ and independent between patients, respectively.

LVEF: The following assumptions were made for the absolute change from baseline in LVEF over 20 weeks:

- the expected mean effect under the placebo dose is conservatively assumed as an absolute increase from baseline of up to $\Delta = 2\%$ with a standard deviation of $\sigma = 7\%$
- while the maximum observable mean effect under neladenoson bialanate within the dose range considered is assumed as an absolute increase of $\Delta = 5\%$ with a standard deviation of $\sigma = 7\%$.

This results in an expected maximum effect size of (5-2)/7 = 3/7 = 0.43 under conservative assumptions. Note that the threshold of clinical relevance is an absolute difference of 3% in the change from baseline over the placebo control group.

NT-proBNP: The following assumptions were made for the absolute change from baseline in the log-transformed NT-proBNP over 20 weeks:

- the expected mean effect under the placebo dose is assumed as an absolute decrease from baseline of -0.105 log(pg/mL) with a standard deviation of $\sigma = 0.85$ log(pg/mL), which is equivalent to a 10% relative reduction from a mean of 2,500 pg/mL at baseline.
- while the maximum observable mean effect under neladenoson bialanate within the dose range considered is assumed as an absolute decrease from baseline of -0.357 log(pg/mL) with a standard deviation of $\sigma = 0.85$ log(pg/mL), which is equivalent to a 30% relative reduction from a mean of 2,500 pg/mL at baseline.

It is assumed that the respective primary response efficacy variable, denoted as Y, is observed for the 6 parallel groups corresponding to doses levels: (placebo =) $d_1 < d_2 < \ldots < d_k$, where $k = 6$.

For patient j within treatment group i the response can then be described by the following model:

$$Y_{ij} = f(d, \boldsymbol{\theta}) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2), \quad i = 1, \dots, k, \; j = 1, \; \dots, \; n_i,$$

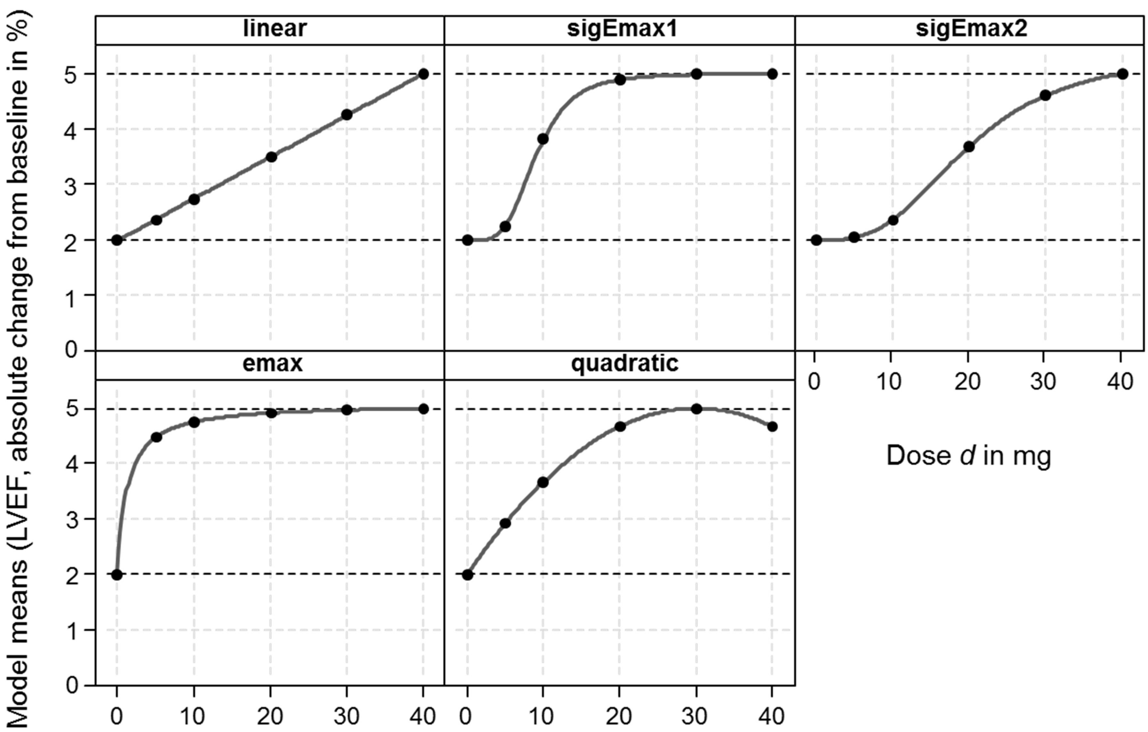
where $f(.)$ is parameterized by a vector of parameters $\boldsymbol{\theta}$ and ε_{ij} is the error term.

A candidate set with M=5 different dose response shapes $f(.)$ based on four models was chosen for the MCP-Mod method. Table 6–1 displays the response expressions for the shapes in the candidate sets for both, change in LVEF and change in log-transformed NT-proBNP. Figure 6–1 shows the corresponding dose-response shapes for change in LVEF. The model parameters were obtained through discussions with experts in the clinical team, taking prior beliefs and uncertainty into account. The dose-response shapes for NT-proBNP are equivalent to those for LVEF but with decreasing trends.

Table 6–1: Dose-response shapes used in the candidate set

Model	Response as function of dose *d*	
	LVEF	**NT-proBNP**
Linear	$2 + 0.075d$	$-0.105 - 0.0063d$
Sigmoidal E_{max} 1	$2 + 3.007689d^4 / (9^4 + d^4)$	$-0.105 - 0.2526d^4 / (9^4 + d^4)$
Sigmoidal E_{max} 2	$2 + 3.375d^3 / (20^3 + d^3)$	$-0.105 - 0.2835d^3 / (20^3 + d^3)$
E_{max}	$2 + 3.09375d / (1.25 + d)$	$-0.105 - 0.2599d / (1.25 + d)$
Quadratic	$2 + 0.2d - 0.00333d^2$	$-0.105 - 0.0168d + 0.00028d^2$

Figure 6–1: Dose-response shapes used in the candidate set for change in LVEF



Based on the standardized versions of the models in the candidate set and the sample size allocation planned for this study, the optimum contrast coefficients for the 5 contrast tests on the dose-response shapes can be derived for both primary variables (Table 6–2).

These contrast coefficients will be updated based on the actual number of subjects per treatment group which are finally available for the analysis.

Table 6–2: Contrast coefficients

Model	Contrast coefficients for dose					
	0.0	5.0	10.0	20.00	30.00	40.00
Linear	-0.674	-0.159	-0.186	0.077	0.340	0.603
Sigmoidal E_{max} 1	-0.784	-0.223	0.004	0.315	0.342	0.347
Sigmoidal E_{max} 2	-0.615	-0.197	-0.290	0.130	0.423	0.550
E_{max}	-0.901	0.046	0.170	0.215	0.231	0.240
Quadratic	-0.815	-0.129	-0.024	0.288	0.392	0.288

Analysis

Step 1: Detection of dose-response signal

For detecting an overall trend, or a dose-response signal, each of the M=5 dose-response shapes in the candidate set will be tested, using a single contrast test based on the updated version of contrast coefficients taking the actual sample sizes per treatment group into account.

For each model m, m = 1, …, 5, in the candidate set

the null hypothesis $H_{0m}: c_m\mu_m^0 = 0$

will be tested against

the respective 1-sided alternative hypothesis $H_{1m}: c_m\mu_m^0 > 0$,

where $\mu_m^0 = (\mu_{m1}^0, \dots, \mu_{m6}^0)' = (f_m^0(d_1, \boldsymbol{\theta}_m^0), \dots, f_m^0(d_6, \boldsymbol{\theta}_m^0))'$ and

f^0 is the standardized version of the dose-response model $f(d, \boldsymbol{\theta}) = \theta_0 + \theta_1 f^0(d, \boldsymbol{\theta}^0)$. In this parameterization, θ_0 is a location parameter and θ_1 is a scale parameter such that only $\boldsymbol{\theta}^0$ determines the shape of the model function.

The contrast coefficients c_{m1} … c_{mi} for the m-th model are chosen such that they maximize the power to detect the underlying model. These optimal contrast coefficients depend only on the parameters in the standardized model function $\boldsymbol{\theta}^0$, which determine the model shape (5) and the actual group sample sizes (which is known after unblinding of the study). The *i*th member of the optimal contrast vector $\boldsymbol{c}_{opt,m}$ for testing the shape model m is proportional to

$n_i(\mu_{mi}^0, \dots, \bar{\mu}), i = 1, \dots, 6,$

where $\bar{\mu} = N^{-1}\sum_{i=1}^{6}\mu_{mi}^0 n_i$. In case of unequal sample sizes per treatment arm, $\boldsymbol{c}_{opt,m}$ cannot be expressed in closed form and numerical optimization techniques are required (9, 11). The $\boldsymbol{c}_{opt,m}$ is derived by fulfilling the condition $\sum_{i=1}^{6} c_{mi}^2 = 1$.

The single contrast test for detecting the m-th model shape is defined by

$T_m = \frac{\sum_{i=1}^{6} c_{mi}\bar{Y}_i}{S\sqrt{\sum_{i=1}^{6} c_{mi}^2/n_i}}, m = 1, \ldots, 5,$ where $S^2 = \frac{\sum_{i=1}^{6}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_i\right)^2}{N-6}$.

Under the null hypothesis of no dose-response effect, i.e. $\mu_{d_1} = \cdots = \mu_{d_6}$, the test statistic $T = (T_1, \ldots, T_5)'$ follows a central multivariate t distribution with N-6 degrees of freedom and correlation matrix $R = (\vartheta_{ij})$, where $\vartheta_{ij} = \frac{\sum_{l=1}^{6} c_{il}c_{jl}/n_l}{\sqrt{\sum_{l=1}^{6} c_{il}^2/n_l \sum_{l=1}^{6} c_{jl}^2/n_l}}$.

The final test statistic T_{max} is based on the maximum contrast test and a "proof-of-concept" dose-response relationship is detected if this maximum statistic T_{max}, and thus at least one single contrast test, is statistically significant, while controlling the familywise error rate at level α. If $q_{1-\alpha}$ denotes the multiplicity adjusted critical value, a dose-response signal is established if $T_{max} \geq q_{1-\alpha}$.The smallest of the corresponding adjusted p-values will be used for the Hochberg test.

This analysis will be performed for the FAS and PPS LVEF, PPS BNP populations respectively, where the PPS LVEF and PPS BNP analysis is the primary analysis.

If no candidate model is statistically significant, the procedure stops, indicating that a dose-response relationship cannot be established from the observed data.

Out of the statistically significant models in the candidate set a best model can be selected for the next step: modeling and estimation.

Step 2: Modeling and estimation of target doses

If a dose-response signal is established, the selected dose-response model(s) will be fitted to the observed data to estimate the model parameters.

The estimated dose-response model will be plotted against the doses including 90% confidence bands. Once the dose-response model has been successfully fitted to the data, target dose(s) of interest are estimated. Given a clinically relevant effect Δf, a minimum effective dose (MED_Δ) associated with model $f(d, \boldsymbol{\theta})$ is defined as

$MED_\Delta = \text{argmin}_{d\in(d_1,d_6]}\{f(d, \boldsymbol{\theta}) \geq f(d_1, \boldsymbol{\theta}) + \Delta\}$.

Estimates of MED_Δ will be calculated for a clinically relevant change in LVEF assumed as Δwil% and potentially a plausible range of Δ values which will be defined based on the observed data. Estimation for change in log-transformed NT-proBNP will be performed analogously, ensuring an adequate back-transformation to allow for better interpretability of the results. In addition, estimates considering confidence bounds for the predicted value at a certain dose may be used. The final choice of the target dose depends on the evaluation of the primary efficacy variables and other efficacy variables, as well as safety considerations.

Additionally LVEF and increasing change from baseline, NT-pro-BNP and decreasing change from baseline, and log-transformed NT-proBNP and decreasing change from baseline will be descriptively summarized by treatment and overall, and visit in PPS LVEF and PPS BNP respectively.

6.2.1.3 Secondary analysis of primary efficacy variables

As a secondary analysis pairwise comparisons of the active neladenoson bialanate dose groups with the placebo group will be performed without controlling the family-wise error rate, by calculating the 90% confidence interval for the difference in primary efficacy variable between each active dose of neladenoson bialanate and placebo.

6.2.1.4 Sensitivity analyses of primary efficacy variables

6.2.1.4.1 Sensitivity analysis of primary efficacy variables in PPS LVEF and PPS BNP

As a sensitivity analysis, primary analysis of primary efficacy variables on the "completers and treatment adherers" excluding the censored patients in the per-protocol analysis set will be repeated. This strategy leads to unbiased estimates only if missing values are "missing completely at random" (MCAR), i.e. the missingness – including missing data due to death – is independent of both observed and unobserved outcomes. This condition is unlikely to hold exactly but rather approximately.

Further sensitivity analyses on the PPS LVEF and PPS BNP may be performed if the missing data patterns suggest further exploration.

6.2.1.4.2 Sensitivity analysis of primary efficacy variables in FAS

Additional efficacy analyses in the FAS will include the following:

- Primary analyses of primary efficacy variables specified in section 6.2.1.2 will be performed in FAS without any imputation.
- Generally, it will be assumed that missing observations for the respective efficacy variables are missing at random. This implies that the behavior of the post dropout observations can be predicted from the observed variables using appropriate imputation models. Likely exceptions to the missing at random assumption are observations which are missing due to a patient's CV death or HF hospitalization prior to the visit in Week 20. These observations can be assumed to be missing not at random (MNAR), i.e., that missingness depends both on observed and unobserved outcomes and that an explicit model for the patient's statistical behavior after drop-out (or death) is required. Therefore, an analysis based on a pattern mixture framework (6) with different imputation rules depending on the reason for missingness will be used using a multiple imputation model, followed by a modification of the imputed data applying penalties:
    1. First, multiple imputation will be applied to draw sets of completed data, using an appropriate imputation model. Baseline characteristics which should be considered in the imputation model include but are not restricted to the baseline values of the respective efficacy variable, the treatment group, and sex.
    2. The imputed data will be modified by applying penalties. The penalty is chosen as the median of the worst changes from baseline across treatment groups (or 0 if the median worst change should be positive for LVEF or negative for LVESV/LVEDV or log-transformed NT-proBNP/hs-TNT).

3. After modifying the completed data sets, the primary analysis specified in section 6.2.1.2 will be applied to the multiple imputed datasets and the point estimate and variance of the contrast from multiple imputed datasets will be combined based on Rubin's rule (8). For more details see Appendix 9.4.

- Primary analyses where for each patient without an observation at the visit in Week 20 the missing value will be imputed according to a last observation carried forward approach, including the baseline value.

For reproducibility, the SAS seed number for creating the random numbers for the multiple imputation will be set to the study number.

6.2.1.5 Additional analysis of primary efficacy variables

Adjusted primary analysis of primary efficacy variables specified in section 6.2.1.2 (dose-response test) will performed in PPS LVEF and PPS BNP respectively. Baseline values of atrial fibrillation, eGFR ($\leq$60 mL/min/1.73m^2 , >60 mL/min/1.73m^2), gender , age group (<50, 50-75 and >75), and BMI > 30 kg/m^2 will be used as covariates. In exploratory manner, the absolute change in LVEF (%) from baseline to week 20 will be categorized into four groups ($\leq$ -5, > -5 to <5, 5 to <10, $\geq$10). The number (%) of patients in each category of LVEF will be displayed by treatment group and overall. For further details including further categorization see the TLF Specification document.

6.2.2 Secondary efficacy variables and analyses

6.2.2.1 Secondary efficacy variables

Secondary efficacy variables across different domains are:

- Key echocardiographic parameters, measured values and absolute change from baseline at 20 weeks:
    - Left ventricular end-systolic volume (LVESV; mL)
    - Left ventricular end-diastolic volume (LVEDV; mL)
- High sensitivity troponin T (hs-TNT; ng/L), measured values (log transformed) and absolute / relative change from baseline at 20 weeks as a biomarker of myocardial injury
- Time from randomization to clinical outcomes: CV mortality, HF hospitalization and urgent visits for HF (separate and composite)

6.2.2.2 Primary analyses of secondary efficacy variables

The primary analysis of key echocardiographic parameters will be performed in PPS LVEF and other secondary efficacy variables will be performed in PPS BNP. The missing values of post baseline of LVESV and LVEDV will be imputed by WOV if the baseline values are not missing and the subjects have CV death or HF hospitalization, otherwise remains missing. Similarly The missing values of post baseline of hs-TNT will be imputed by WOV if the baseline values are not missing and the subjects have CV death or HF hospitalization, otherwise remains missing. The secondary efficacy variables, except for the clinical outcomes, will be analyzed using similar statistical methods as for the primary efficacy variables, i.e. the MCP-Mod method with the same standardized candidate dose-response shapes (with the proper direction: decreasing trend for change of LVESV, LVEDV and log-transformed hs-TNT) and corresponding coefficients as for the primary variables.

Time to adjudicated clinical outcome event since randomization (using both separate and composite outcomes) will be described by means of Kaplan-Meier (KM) estimates by treatment group. The subjects who do not have the corresponding clinical outcomes until week 26 visit (upper time limit, i.e. 182+7days) will be considered as right-censored at the minimum of date of last visit, date of week 26 visit, study day 189, and date of death (in case death is non CV death). KM estimates for each visit will be tabulated.

Additionally time to adjudicated on-treatment clinical outcome events since randomization (using both separate and composite outcomes) will also described up to 6 weeks after last dose in each treatment arm respectively. The subjects who do not have the corresponding clinical outcomes until 6 weeks after last dose will be considered right censored at minimum of 6 weeks after last dose, date of last visit and date of death (in case death is non CV death).

6.2.2.3 Sensitivity analyses of secondary efficacy variables

Sensitivity analyses of secondary variables except for clinical events will be performed in the FAS using the same methods as for the primary efficacy variables. If the baseline values of secondary endpoints are missing, then only multiple imputation will be applied for those subjects. Please see the details in section 6.2.1.4. In addition clinical events will be analyzed by using life table statistics and Kaplan-Meier estimates in FAS (only time scope up to week 26) and SAF.

6.2.2.4 Additional analyses of secondary efficacy variables

Percentage change in hs-TNT from baseline to week 20 will be summarized by treatment group and overall in PPS BNP. In addition to analyses comparing population means in the different dose groups, the number of patients in whom the individual change from baseline value crossed clinically meaningful thresholds will be reported in post-hoc analysis separately.

6.2.3 Exploratory efficacy variables and analyses

Exploratory efficacy variables include:

- Activity (duration, intensity) reported values and absolute change from baseline at 20 weeks. Variables as specified in appendices 9.1
- Echocardiographic parameters, as described in Section 9.4.1 of Clinical Study Protocol, measured values and absolute / relative change from baseline at 20 weeks
- Left ventricular ejection fraction (LVEF; %), measured values, number of patients in whom absolute change from baseline after 20 weeks is > 5%
- Mandatory biomarkers, measured values and absolute / relative change from baseline at 20 weeks, including
    - NT-proBNP (pg/mL), measured values and relative change from baseline at 20 weeks to assess elevated filling pressures
    - UACR, cystatin-C, NGAL for the evaluation of kidney function

- KCCQ and EQ-5D-5L, as described in Section 9.4.4.1 and 9.4.4.2 of Clinical Study Protocol. Frequencies of answers to individual questions will be displayed by treatment group and overall by visit. In addition, changes from baseline for the single questions will be categorized into the categories improvement / no change / worsening. Frequencies of the different categories will be displayed by treatment group and overall by visit.
  The 7 individual domain scores and 3 summary scores of the KCCQ and their changes to baseline will be summarized by treatment group and overall by visit as continuous variables. The symptom stability domain (a measure of change in symptoms over the previous two weeks) and self-efficacy domain (a measure of patient knowledge how to prevent heart failure exacerbations and management of complications when they arise), will be reported and summarized but will not be considered measures of treatment efficiency. Change from baseline to week 20 of KCCQ domains of Physical Limitation Score, Symptom Frequency Score, Symptom Burden Score, Total Symptom Score, Clinical Summary Score and Overall Summary Score will be categorized into groups by 5 points and the number (%) of patients in each category of KCCQ domains respectively will be displayed by treatment group and overall. For scoring see Appendix, Section 9.3. The EQ visual analogue scale (VAS) values of the EQ-5D-5L and their changes to baseline will be summarized by treatment group and overall by visit as continuous variable.
- Time from randomization to all-cause mortality, non-fatal myocardial infarction, non-fatal stroke
- Change in NYHA class

Time to adjudicated clinical outcome events since randomization (using both separate and composite outcomes) will be described by means of life table statistics and Kaplan-Meier estimates in FAS. Other than time to event variables will be analyzed descriptively in FAS without missing data. In addition, percentage change in NT-proBNP from baseline to week 20 will be summarized by treatment group and overall. The percentage change of NT-proBNP from baseline to week 20 will be categorized into four groups (≤ -20, > -20 to ≤+20, >20 and ≤+30, > +30) . The number (%) of patients in each category of NT-proBNP respectively will be displayed by treatment group and overall.

6.3 Safety

The summaries of the safety data will be completed for the safety analysis population (SAF). No formal statistical test will be performed for the safety variables.

6.3.1 Extent of exposure

Study medication will be summarized for the safety population by treatment group, using descriptive statistics such as frequency and proportion (for categorical variables), mean, median, and standard deviation (for continuous variables).

All summaries related to intake of study medication will be by treatment group based on SAF:

- The treatment duration (date of last study medication- date of first study medication+1 day) will be summarized descriptively. Additionally the number of subjects by treatment duration category will be given (≤28 days, >28-≤56 days, >56-≤84 days, >84-≤140 days).

- The time on study medication (treatment duration excluding days off study medication) will be calculated and summarized descriptively.
- The number of tablets taken will be summarized descriptively, as well as corresponding extent of exposure (total amount of intake in mg).

6.3.2 Treatment Compliance

Compliance is defined as 100*number of tablets taken / number of tablets planned.

The compliance will be summarized descriptively by treatment group and overall. In addition, compliance will be categorized into three groups (<80%, 80-120%, >120%) and summarized by treatment group and overall.

6.3.3 Adverse events

All adverse events (AEs) will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) in its latest version which is specified in the TS domain.

A treatment-emergent AE is defined as any event arising or worsening after the start of study drug administration until 6 weeks after the last study medication intake.

Summary statistics (frequency and percentage of subjects) will be presented by treatment group using MedDRA for the following:

- Incidence rate of treatment-emergent AEs.
- Incidence rate of drug-related treatment-emergent AEs.
- Incidence rate of treatment-emergent AEs leading to death.
- Incidence rate of treatment-emergent AEs leading to permanent withdrawal of medication.
- Incidence rate of treatment-emergent AEs leading to interruption of medication.
- Incidence rate of treatment-emergent serious adverse events.
- Incidence rate of treatment-emergent drug-related serious adverse events.
- Incidence rate of adverse events of special interest:
    - Symptomatic bradycardia (HR < 50 bpm)
    - Findings in ECG and / or AVIVO device as follows:
        - Mobitz type I AV-block leading to withdrawal or interruption of study drug
        - Mobitz type II AV block leading to withdrawal or interruption of study drug or leading to any change in therapy
        - Third degree AV blocks

- Listing of treatment-emergent AEs leading to withdrawal: subject ID, investigator AE term, SOC, preferred term, start and stop date of study drug administration, start and stop date(relative days) of AE, drug related (yes/no), serious (yes/no), intensity, outcome.
- Listing of treatment-emergent serious AEs: subject ID, investigator AE term, SOC, preferred term, start and stop dates of study treatment, start and stop date of AE, drug related (yes/no), outcome, action taken.

6.3.4 Deaths

Deaths reported during the study period will be tabulated by treatment group.

- Summary table of deaths (all deaths, all deaths during treatment and up to 6 weeks after last dose of study drug, all deaths later than 6 weeks after last dose of study medication)
- Listing of subjects who died during treatment and up to 6 weeks after last dose: subject ID, start and stop date of study medication, date of death, relative day and cause of death.

6.3.5 Clinical laboratory evaluations

All laboratory evaluations will be done by central laboratory.

Hematology: erythrocytes, hemoglobin, hematocrit, MCV, MCH, MCHC, reticulocytes, leukocytes, differential blood count, platelets

Clinical chemistry: aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (AP), gamma glutamyl transpeptidase (GGT), lactate dehydrogenase (LDH), creatine kinase (CK), amylase, lipase, glucose, cholesterol (HDL, LDL, total), triglycerides, creatinine, urea, uric acid, bilirubin, total protein, albumin, sodium, potassium, calcium, chloride, magnesium, anorganic phosphate

Coagulation: partial thromboplastin time (PTT), international normalized ratio (INR)

The safety evaluation of laboratory data will include:

- Descriptive analysis of continuous laboratory parameters, and their changes from baseline by visit and treatment group.
- Incidence rates of treatment-emergent laboratory values outside of normal range by treatment group.
- Listings of laboratory data out of normal range.

Laboratory abnormalities will be summarized in table of change from baseline by visit and treatment:

- Change in renal function measured by eGFR change from baseline
- Change in liver function measured by bilirubin (total and fractions), AST and ALT from baseline

6.3.6 AVIVO/HealthPatch monitoring

AVIVO Mobile Patient Management System is intended to continuously measure, record and periodically transmit ECG data.

Notifiable ECG-findings (like AV-conduction abnormalities) triggered by system/patients based, as well as reportable ECG finding according to AVIVO will be summarized in frequency tables by treatment group. ECG-findings according to HealthPatch will also be summarized in frequency tables by treatment group . A table displaying the number of patients with AV-block > I° according to AVIVO/HealthPatch will be provided by treatment group. All patients with significant ECG-finding will be listed. The definition of the findings that trigger a notifiable report is in Appendices 9.1.

6.3.7 Other safety measures

The last pre-treatment safety measurement, i.e., SBP(systolic blood pressure), DBP(diastolic blood pressure), weight, body temperature, heart rate, respiration rate and electrocardiogram (12 lead ECG) will be used as “baseline value.”

When more than one value is collected at the same post-baseline visit, the value retained at that particular visit for summary statistics will be the average of the different measures reported for that visit.

For each treatment group, vital signs will be tabulated and summarized by visit for observed values and changes from baseline using descriptive statistics, as appropriate.

The incidence rates of treatment-emergent 12-lead ECG abnormalities will be tabulated by treatment group. A descriptive analysis of continuous ECG parameters and their changes from baseline by visit and treatment group will also be presented. PR interval will be summarized by visit and treatment group.

6.4 Subgroup Analysis

6.4.1 Subgroups

Subgroups are specified in Section 4.5.4.

6.4.2 Subgroup analysis of efficacy variables

Primary, secondary and sensitivity analyses (without imputed values) of primary efficacy variables will be performed based on key subgroups. Additionally primary efficacy variables will be descriptively summarized based on exploratory subgroups.

Secondary efficacy variables will be descriptively summarized based on key subgroups and exploratory efficacy variables will be descriptively summarized based on LVEF(%) at baseline: $\leq$40 vs. >40.

6.4.3 Subgroup analysis of safety variables

Incidence rate of treatment-emergent AEs and treatment-emergent ECG abnormalities not presented at baseline (by AVIVO and HealthPatch, Japan only) will be summarized based on key and exploratory subgroups.

6.5 Pharmacokinetics/pharmacodynamics

Pharmacokinetic analyses will be performed on the population valid for pharmacokinetics.

For the investigation of systemic exposure to BAY 84-3174 and its relationship with treatment effects, the plasma concentrations of BAY 84-3174 will be determined at different time points using a sparse sampling approach in all participating patients (see Clinical Study Protocol Section 9.5). The plasma concentration vs. time data will be evaluated descriptively separated by dose and visit. Plots will be prepared pooling all individual plasma concentrations (naive pooling) vs. actual relative study times (time of sample collection after time of study drug administration).

Furthermore, the pharmacokinetic data and the relationship of markers of BAY 84-3174 exposure (e.g. C_{max}, AUC) with treatment effects will be evaluated using non-linear mixed effect modeling (NONMEM). The latter evaluation will be described in a separate analysis plan and will be reported under separate cover.

The PK bioanalysis will be performed under the responsibility of the Sponsor's Bioanalytics Laboratory.

6.6 Biomarker analyses

Biomarker data will be described by the following summary statistics: arithmetic mean, standard deviation, median, quantiles, minimum, and maximum.

Box plots and line plots of means of biomarkers over visits, by treatment group will be provided. Additional analyses of safety and efficacy biomarkers and their results will be provided in a separate report.

7. Document history and changes in the planned statistical analysis

Version	Date	Action
Version 1.0	13Nov2017	
Version 2.0	19 April 2018	• Change the wording in section 6.3.6 • Add more detailed description of plots we plan to provide in section 6.6
Version 3.0	29 May 2018	• Add one more key subgroup: LVEF(%) at baseline: ≤45 vs. >45 • Add more details of worst case imputation and censoring rule of time to event variables in section 6.2 • Some minor changes in subgroup

		analyses in section 6.4 • Remove some unused variables from Appendix 9.1

8. References

1. McCullough PA, Philbin EF, Spertus JA, Kaatz S, Sandberg KR, Weaver WD, et al. Confirmation of a heart failure epidemic: findings from the Resource Utilization Among Congestive Heart Failure (REACH) study. Journal of the American College of Cardiology. 2002;39(1):60-9.
2. Stewart S, MacIntyre K, Hole DJ, Capewell S, McMurray JJ. More 'malignant' than cancer? Five-year survival following a first admission for heart failure. European journal of heart failure. 2001;3(3):315-22.
3. Executive summary: HFSA 2006 Comprehensive Heart Failure Practice Guideline. J Card Fail. 2006;12(1):10-38.
4. McMurray JJ, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med. 2014;371(11):993-1004.
5. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics. 2005;61(3):738-48.
6. Carpenter JR, Kenward MG. Missing data in randomized controlled trials - a practical guide. Chapter 6.2008.
7. Kelly RP, Ting CT, Yang TM, Liu CP, Maughan WL, Chang MS, et al. Effective arterial elastance as index of arterial vascular load in humans. Circulation. 1992;86(2):513-21.
8. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
9. Menon S, Zink RC. Modern Approaches to Clinical Trials Using SAS: Classical, Adaptive, and Bayesian Methods. SAS Institute; December 2015
10. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
11. Bjoern Bornkamp, Jose Pinheiro and Frank Bretz (2016). DoseFinding: Planning and Analyzing Dose Finding Experiments. R package version 0.9-15. URL https://cran.r-project.org/web/packages/DoseFinding
12. The Kansas City Cardiomyopathy Questionnaire Scoring Instructions (2000-2003), John Spertus MD MPH [http://cvoutcomes.org/ (6th September 2013)].
13. C. Patrick G, et al. Development and Evaluation of the Kansas City Cardiomyopathy Questionnaire: A New Health Status Measure for Heart Failure, Journal of the American College of Cardiology, 2000; 35(5):1245-55

14. Barnard J, Rubin DB. Small-Sample Degrees of Freedom with Multiple Imputation. Biometrika 1999; 86: 948–955.
15. Rubin, D. B. (1996), "Multiple Imputation after 18+ Years," Journal of the American Statistical Association, 91, 473–489.

9. Appendices

9.1 AVIVO device variable specification

In the following, we describe the variables which will be analyzed

- **Activity**

Variable	Summary measure(s)	Unit	Length of intervals
activity duration	duration	seconds	daily
activity intensity	mean	mG	1 hour
activity intensity	mean, max	%	daily

- **Notifiable findings**

These should be the findings that trigger **a notifiable report,** definitions as below:

Finding	Notification criteria
Ventricular fibrillation	always notified
ICD discharge	always notified
Ventricular Tachycardia	any rate and ≥10 beats
Wide complex Tachycardia	any rate and ≥10 beats
PVCs	never notified
Sinus Bradycardia	≤30bpm
Sinus Tachycardia	≥ 180bpm
Supraventricular tachycardia	≥ 150 bpm AND ≥ 30 sec
A. Fibrillation or A. flutter	≥ 150 bpm AND ≥ 30 sec
A. Fibrillation or A. flutter	≤30bpm and ≥ 30 sec
A. Fibrillation or A. flutter	when notification criteria are met
pause	≥ 3.0 sec
AV block 2nd (Mobitz I)	≤ 50bpm
AV block 2nd (Mobitz II)	always notified
Isolated 2nd degree AV block (2:1)	≤ 50bpm
High degree AV block	always notified

3rd Degree AV block	always notified
other	
Patient triggered Events	when notification criteria are met
Technicians discretion	any

9.2 Echocardiography parameters

The list of parameters is

- LV ejection fraction (LVEF, %)
- LV end-diastolic volume (LVEDV), LVEDV index (LVEDVI, calculated as LVEDV/BSA)
- LV end-systolic volume (LVESV), LVESV index (LVESVI, calculated as LVESV/BSA)
- LA size (LA diameter, area, volume index [LAVI, calculated as LAV/BSA])
- Lateral e' (early diastolic mitral annular relaxation velocity at the lateral mitral annulus by Tissue Doppler, TD)
- Septal e' (early diastolic mitral annular relaxation velocity at septal mitral annulus by TD), including calculation of average e'
- Global longitudinal strain (%)
- Pulmonary artery systolic pressure (PASP), estimated by tricuspid regurgitation velocity and inferior vena cava diameter, including its change with respiration, and hepatic vein flow in patients with tricuspid regurgitation
- Tricuspid annular plane systolic excursion (TAPSE), right ventricular (RV) s' (velocity of the tricuspid annular systolic excursion at the RV free wall by TD)
- Mitral regurgitation
- LV mass, LV mass index (calculated as LV mass/BSA)
- Wall thicknesses, incl. interventricular septum diameter (IVSD), posterior wall thickness (PWT), anteroseptal wall thickness (ASWT)
- E, A (if in sinus rhythm), calculation of E/A and E/e' (using lateral, septal, average e') ratios
- E-wave deceleration time (EWDT)
- Stroke volume (SV, calculated by LVEDV - LVESV) and derived parameters, including SV index (SVI, calculated as SV/BSA), cardiac output (CO, calculated as SV*HR), cardiac index (CI, calculated as CO/BSA), systemic arterial compliance (SAC, calculated as SV/PP), total peripheral resistance (TPR, calculated as MAP/CO*80)

- Effective arterial elastance (Ea), estimated as end-systolic pressure (Pes) [Pes calculated as SBP times 0.9 (7)] divided by SV (SBP*0.9/SV)

Final details of all echocardiography parameters to be measured and analyzed will be included in a separate echocardiography manual.

9.3 KCCQ Scoring

As described in the KCCQ Scoring instruction [12][13], the following derivations will be used.

Generally only questions actually answered are used for derivation of the scores in the following way:

If there are n questions in a scale, and the subject must answer m to score the scale, but the subject answers only n-i, where $n\text{-}i \geq m$, calculate the mean of those questions as

(sum of the responses to those n-i questions) / (n-i)

not

(sum of the responses to those n-i questions) / n

The 7 individual domain scores and 3 summary scores will be calculated as follows:

9.3.1 Physical Limitation

Code responses to each of Questions 1a-f as follows:

Extremely limited = 1
Quite a bit limited = 2
Moderately limited = 3
Slightly limited = 4
Not at all limited = 5
Limited for other reasons or did not do = <*missing value*>

If at least three of Questions 1a-f are not missing, then compute

Physical Limitation Score = 100*[(mean of Questions 1a-f actually answered) – 1]/4

9.3.2 Symptom Stability

Code the response to Question 2 as follows:

Much worse = 1
Slightly worse = 2
Not changed = 3
Slightly better = 4
Much better = 5
I've had no symptoms over the last 2 weeks = 3

If Question 2 is not missing, then compute

Symptom Stability Score = 100*[(Question 2) – 1]/4

9.3.3 Symptom Frequency

Code responses to Questions 3, 5, 7 and 9 as follows:

Question 3

Every morning = 1
3 or more times a week but not every day = 2
1-2 times a week = 3
Less than once a week = 4
Never over the past 2 weeks = 5

Questions 5 and 7

All of the time = 1
Several times a day = 2
At least once a day = 3
3 or more times a week but not every day = 4
1-2 times a week = 5
Less than once a week = 6
Never over the past 2 weeks = 7

Question 9

Every night = 1
3 or more times a week but not every day = 2
1-2 times a week = 3
Less than once a week = 4
Never over the past 2 weeks = 5

If at least two of Questions 3, 5, 7 and 9 are not missing, then compute:

S3 = [(Question 3) – 1]/4
S5 = [(Question 5) – 1]/6
S7 = [(Question 7) – 1]/6
S9 = [(Question 9) – 1]/4

Symptom Frequency Score = 100*(mean of S3, S5, S7 and S9)

9.3.4 Symptom Burden

Code responses to each of Questions 4, 6 and 8 as follows:

Extremely bothersome = 1
Quite a bit bothersome = 2
Moderately bothersome = 3
Slightly bothersome = 4
Not at all bothersome = 5
I've had no swelling/fatigue/shortness of breath = 5

If at least one of Questions 4, 6 and 8 is not missing, then compute

Symptom Burden Score = 100*[(mean of Questions 4, 6 and 8 actually answered) – 1]/4

9.3.5 Self-Efficacy

Code responses to Questions 10 and 11 as follows:

Question 10

Not at all sure = 1
Not very sure = 2
Somewhat sure = 3
Mostly sure = 4
Completely sure = 5

Question 11

Do not understand at all = 1
Do not understand very well = 2
Somewhat understand = 3
Mostly understand = 4
Completely understand = 5

If at least one of Questions 10 and 11 is not missing, then compute

Self-Efficacy Score = 100*[(mean of Questions 10 and 11 actually answered) – 1]/4

9.3.6 Quality of Life

Code responses to Questions 12, 13 and 14 as follows:

Question 12

It has extremely limited my enjoyment of life = 1
It has limited my enjoyment of life quite a bit = 2
It has moderately limited my enjoyment of life = 3
It has slightly limited my enjoyment of life = 4
It has not limited my enjoyment of life at all = 5

Question 13

Not at all satisfied = 1
Mostly dissatisfied = 2
Somewhat satisfied = 3
Mostly satisfied = 4
Completely satisfied = 5

Question 14

I felt that way all of the time = 1
I felt that way most of the time = 2
I occasionally felt that way = 3

I rarely felt that way = 4
I never felt that way = 5

If at least one of Questions 12, 13 and 14 is not missing, then compute

Quality of Life Score = 100*[(mean of Questions 12, 13 and 14 actually answered) – 1]/4

9.3.7 Social Limitation

Code responses to each of Questions 15a-d as follows:

Severely limited = 1
Limited quite a bit = 2
Moderately limited = 3
Slightly limited = 4
Did not limit at all = 5
Does not apply or did not do for other reasons = <missing value>

If at least two of Questions 15a-d are not missing, then compute

Social Limitation Score = 100*[(mean of Questions 15a-d actually answered) – 1]/4

9.3.8 Total Symptom Score

= mean of the following available summary scores:

Symptom Frequency Score
Symptom Burden Score

9.3.9 Overall Summary Score

= mean of the following available summary scores:

Physical Limitation Score
Total Symptom Score
Quality of Life Score
Social Limitation Score

9.3.10 Clinical Summary Score

= mean of the following available summary scores:

Physical Limitation Score
Total Symptom Score

9.4 Combining inferences from multiple imputed data sets

With *m* imputations, *m* different sets of the point and variance estimates for a parameter Q (in our case the contrast estimate) can be computed (see SAS version 9.2 Help for Proc Mianalyze). Suppose that $\hat{Q}_i$ and $\hat{W}_i$ are the point and variance estimates, respectively, from the *i*th imputed data set, *i*= 1, 2, ..., *m*. Then the combined point estimate for Q from multiple imputation is the average of the *m* complete-data estimates:

$$\overline{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i$$

Suppose that $\overline{W}$ is the within-imputation variance, which is the average of the *m* complete-data estimates:

$$\overline{W} = \frac{1}{m}\sum_{i=1}^{m}\hat{W}_i$$

And suppose that B is the between-imputation variance:

$$B = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_i - \overline{Q})^2$$

Then the variance estimate associated with $\overline{Q}$ is the total variance [8]

$$T = \overline{W} + (1 + \frac{1}{m})B$$

The statistic $(Q - \overline{Q})T^{-(1/2)}$ is approximately distributed as *t* with v_m degrees of freedom [14], where

$$v_m = (m-1)\left[1 + \frac{\overline{W}}{(1+m^{-1})B}\right]^2$$

The degrees of freedom v_m depend on *m* and the ratio

$$r = \frac{(1+m^{-1})B}{\overline{W}}$$

The ratio *r* is called the relative increase in variance due to nonresponse [8].When there is no missing information about Q , the values of *r* and *B* are both zero. With a large value of *m* or a small value of *r*, the degrees of freedom v_m will be large and the distribution of $(Q - \overline{Q})T^{-(1/2)}$ will be approximately normal.
Another useful statistic is the fraction of missing information about Q:

$$\hat{\lambda} = \frac{r + 2/(v_m+3)}{r+1}$$

Both statistics *r* and λ are helpful diagnostics for assessing how the missing data contribute to the uncertainty about Q.
When the complete-data degrees of freedom v_0 are small, and there is only a modest proportion of missing data, the computed degrees of freedom, v_m, can be much larger than v_0, which is inappropriate. For example, with *m*=5 and *r*=10%, the computed degrees of freedom

v_m =484, which is inappropriate for data sets with complete-data degrees of freedom less than 484.
[14] recommend the use of adjusted degrees of freedom

$$v_m^* = \left[\frac{1}{v_m} + \frac{1}{\hat{v}_{obs}}\right]^{-1}$$

where $\hat{v}_{obs} = (1-\gamma)v_0(v_0+1)/(v_0+3)$ and $\gamma = (1+m^{-1})B/T$.

We will specify the complete-data degrees of freedom v_0 with the EDF= option, the MIANALYZE procedure uses the adjusted degrees of freedom, v_m^*, for inference.