Statistical Analysis Plan (SAP)
for the
scheduled interim analysis
in the GRAFT-Trial

Granulocyte colony stimulating factor (G-CSF) to treat acute-on-chronic liver failure:
A multicentre randomized trial

Coordinating investigator

Prof. Thomas Berg, MD

University Hospital of Leipzig;

Section Hepatology;

Department of Internal Medicine, Neurology, Dermatology

Biometry

Annegret Franke, PhD

University of Leipzig;

Clinical Trial Centre Leipzig

Sponsor

University of Leipzig;

Ritterstr. 26

04109 Leipzig

Date: 2018-01-14

Version: final 1.0

EudraCT-No.: 2015-002212-32

ClinicalTrials.gov - no.: NCT02669680

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

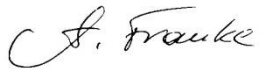
Coordinating Investigator		
	Date	Signature
Biometrician	14.01.2019	
	Date	Signature

Table of contents

1 Introduction

The purpose of this document is to provide a detailed elaboration of the scheduled interim analysis regarding the primary endpoint "**T**ransplant-**F**ree **S**urvival after **90** days of observation" (TFS90) and the major secondary endpoint "**O**verall **S**ur**v**ival" (OSV), including detailed procedures for analyses of futility.

The Statistical Analysis Plan (SAP) assumes familiarity with the Study Protocol (Version final 4.0, dated 2018-07-26) which included all protocol amendments. Protocol version 1.0 was from 2015-10-09. If in doubt the study protocol formulation takes precedence.

The SAP is based on the planned analysis specification as written in the protocol section 8 "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.

At the time of writing the SAP, the results of the Annual Safety Reports (ASR) of 2017 and 2018 were known. Since deaths occurring during the course of the trial represent a major aspect of patients' safety and the GRAFT trial was unblinded due to the nature of the intervention investigated, we are aware of relevant information regarding the major endpoints of the clinical trial at this time.

The programming of the statistical analysis was performed in parallel and after finalization of the SAP. Analysis programmes/routines of the ASRs are re-used.

Finalization of the SAP is done before freezing the study data base for the respective (interim or final) analysis without detailed knowledge on the complete data to be included. However, in preparation of the analyses complex data checks were performed to ensure data completeness and consistency.

IBM® SPSS® statistics, version 25 [1], and PASS sample size®, version 14 [2] are used for statistical analyses.

2 Planned analysis and changes required

2.1 Sequence of Planned Analyses

Anticipating an absolute difference of 20 % (from 42 % to 62 %, hazard ratio 1.815) in transplant-free survival at day 90 and assuming a drop-out rate not exceeding 10%, we originally determined (acc. to protocol) that the enrolment of 292 patients would be required to show beneficial effect of G-CSF.

A single interim analysis was scheduled after 50% of envisaged randomized sample size. As patient accrual was lower than expected, this number was available only after about 30 months of recruitment. On October 09, 2018, half of the patients was included. Given the maximally 3 months of follow up period for the assessment of the primary endpoint and the time frame for completion of documentation and data management, we agreed to conduct the interim analysis and the analyses for the 3rd Annual Safety Report almost simultaneous using the same database (DB) freezing. This is possible because of random but favourable recruitment timing.

No further interim analyses of the clinical trial are planned.

The final analyses are conducted after completion of recruitment and documentation of all planned visits of the last patient included and will follow a more comprehensive SAP including all secondary and various exploratory analyses if necessary and suitable.

2.2 Changes required

During the previous safety analyses, the last one in March 2018 for Annual Safety Reports and DMSB information, it became aware that a higher number of early deaths occurred albeit balanced between treatment arms and without hints on any causality regarding treatment/ trial

procedures. After receiving consent from the Independent Data Monitoring and Safety Board (DMSB), we decided that recruitment of patients should continue until the sample size of the scheduled interim analysis was reached. However, together with the scheduled interim analysis we agreed to perform a futility analysis (using estimations of the conditional power) of the trial. Based on the results of the interim analysis we will determine a premature termination of trial because of futility after consideration of DMSB's recommendations.

Details regarding the futility analysis are described in sec. 7.

If a premature termination is required, further recruitment of patients might be finished but all patients within the trial are followed up until their regular end of observation or the primary endpoint is reached. After LPO and finalization of all data management procedures, the final analysis will take place.

2.3 Purpose of the Statistical Analysis Plan

The purpose of this SAP is to outline the procedures and definitions of the scheduled interim analysis. All results of analyses are strictly confidential and are solely presented to the independent DMSB) of this trial but not to participating centres unless decisions regarding the trial's continuation were to make.

In addition, exploratory analyses not necessarily part of this SAP may be performed to support the decision on the further conduction of the clinical trial. Any post-hoc or unplanned analyses are clearly labelled as post hoc in the respective Clinical Study Report (CSR) of the interim analysis.

The final analyses will follow a separate and more comprehensive SAP including all secondary and further exploratory analyses.

3 Trial centres

3.1 Description of Trial Centres

A table containing the following information for all participating trial sites are provided:

- date of first patient in,
- number of patients screened and randomized until November 16th, 2011, and
- number of patients in the analysis populations.

3.2 Impact on Statistical Analysis

Due to the possible heterogeneity of the trial centres, randomisation was stratified by trial centres so that no bias regarding arm balances should have occurred. No further exploratory analyses are performed in case no apparent deviations of one or few centres attract attention in data descriptions.

4 Endpoints and further variables

4.1 Primary Endpoint

The primary endpoint – transplant-free survival (TFS90) - is defined as a composite of

- Death from any cause or
- Orthotopic liver transplantation (OLT) performed before,

until day 90 after randomization (D90) censoring patients alive without OLT for primary analysis. Operational definitions for the primary endpoint are detailed in the following sub-sections.

4.1.1 Time period

- Individual documentation of a patient lasts until 360 days after randomization (visit 8=V8 with allowed deviation of +/- 14 days) ***or*** until the End of Study (ES)-CRF page indicates an individual's premature termination.
- In case of earlier death during the course of trial *or* premature termination of trial, the ES-CRF page has to be filled which contains the date of death *or* the last date of patients' contact to identify the time of observation per patient.
- In cases in which patients received a liver transplant, the observation was continued to provide data regarding the major secondary endpoint overall survival (OSV). Therefore, an OLT-CRF page at every regular visit has to be filled which contains the date of transplantation as well as survival (SV) status, any reaction regarding organ repulsion or development of malignancies.
- A one-day tolerance policy was applied calculating the duration until the first of the either or other event as difference between both dates plus one day since not detailed information was available whether an event occurred in the morning or night of a day. The date of death was defined as last date per patient within the trial documentation. A duration of >90 days leads to a censored status regarding the primary endpoint.

4.1.2 Status

- The status "deceased" is derived from the ES-CRF page depending on whether or not the reason for premature trial termination is given by patient's death and/or a date of death is documented.
- In case, no ES form is available for a patient at the time of database (DB) snapshot/freezing the (still missing) end-of-study information is kept in another item, but patients' survival status is censored with the latest date available in case documentation.
- If an OLT was necessary and done, and whether or not a patient died thereafter is aggregated in two different items via the OLT forms from all regular visits.
- Or-combination of "deceased" and OLT status serves as status variable for primary analysis and analysis of the transplant-free survival at D360 after randomization (see secondary endpoints).

4.2 Major Secondary and Safety Endpoints

The only secondary endpoint which will be considered within the interim analysis is

- Overall Survival (OSV) at the end of study: Status and time see 4.1

Analyses of all further secondary endpoints named within the trial protocol are not part of the interim analysis but only of the final analysis.

Safety aspects of the treatment are presented to investigate any possible association between treatment and occurrence of adverse events:

- Occurrence of (Serious) Adverse Events [(S)AE] as per MedDRA-coded Preferred Terms (PT), High Level Group Terms (HLGT) or System Organ Class (SOC)
- reported reasons of death if deaths occurred during the predefined SAE reporting period (from randomization, in G-CSF arm from first injection of G-CSF until visit 5(=D28); for exceptions see protocol); reasons of death were MedDRA as described for SAE

4.3 Covariates for confirmatory interim analyses

According to the trial protocol, Cox regression models with/without adjusting for grade of Acute on Chronic Liver Failure (ACLF) are used to estimate the treatment effect. Hazard ratios for

treatment effect and ACLF grade and their confidence intervals are presented within the full analysis set (FAS) of patient.

According to Moreau et al [Moreau 2013] grades of ACLF present a major predictor of mortality and therefore the ***baseline ACLF score*** was originally considered as stratification criterion in the planning stage of the trial to account for prognostic heterogeneity.

However, it became apparent that most of the trial centres were unfamiliar with its scoring scheme and the existing ACLF calculators in the Web showed deviations in some criteria of classification used. So we decided to document all sub-criteria essential to calculate/report ACLF grades but not to use it as stratum in a first step.

Together with first ASR/DMSB analyses, the ACLF grades within the study population were calculated from CRF-reported sub-criteria and according to published calculation schemes. Agreement between centre-reported and DB-calculated grades was compared and revealed a relevant proportion of patients with deviating grades (~20%). Therefore, we decided to use the program-based and validated ACLF grades for analyses and abstain from stratification by ACLF grade in general.

The ACLF calculation scheme according to Moreau (2013) follows the rules described below:

- ACLF grade 0
  a) no organ failure
  b) single non-kidney failure (of liver, coagulation, circulculation, or respiration) and no hepato-encephalopathy and low creatinine (<1.5 mg/dL)
  c) single cerebral failure and low creatinine.
- ACLF grade 1: single organ failure
  a) single kidney failure
  b) single non-kidney failure (of liver, coagulation, circulculation, or respiration) and moderate creatinine and/or mild to moderate hepatoencephalopathy
  c) single cerebral failure and moderate creatinine (>=1.5 to 1.9 mg/dL)
- ACLF grade 2: two organ failures
- ACLF grade 3: three or more organ failures

The organ failure definitions came from Jalan (2014). They refer to single/few laboratory or clinical measures/ assessments per organ system (see Table 1 and sec. 4.4.3).

4.4 Further Variables

4.4.1 Baseline characteristics

Gender and ACLF grade at registration is recorded on patient registration PR-CRF.

Age is derived (as difference of years) from year of birth and date of informed consent (both on patient registration=PR-CRF), **body mass index** from height and weight (on patient baseline BL-CRF).

Vital signs (systolic and diastolic blood pressure, heart and respiratory rate, body temperature) and existing ascites are collected at BL-CRF.

Mean arterial blood pressure (MAP) is derived from [2* diastolic pressure+ systolic pressure]/3 (both on patient baseline BL-CRF) since the BL-CRF documented MAP showed various deviations from the expected values.

Pre-existing conditions

Details von supported ventriculation are recorded on baseline CRF (and all visits performed later) as “no / non-invasive/ invasive”. In addition, reasons for ventilator support are collected (“respiratory insufficiency / hepatic encephalopathy (HE)/ others”) and measures in detail:

- fraction of inspired oxygen (FIO2); in case of missing values (MV) and no supportive measures a rate of 0,21 is substituted in database.

- partial pressure of arterial oxygen (PAO2) or peripheral oxygen saturation (measured via pulse oximeter; SPO2

The ratios of PAO2/FIO2 and SPO2/FIO2 are derived (per visit) and a respiratory organs failure identified according to Table XX from either or the other ratio per visit.

Causes of ACLF development are recorded on baseline CRF as “known/ unknown”. In addition, most frequent causes (yes/no) plus an “others” option are documented on the baseline CRF. A combination of causes is derived with missings identifiable by separate codes.

Complications of ACLF are recorded on baseline CRF in various (most frequent) different categories) as “yes / no / unknown” (from patient history) with further details as types and/or grades A combination of complications is derived with “unknown” assessments identifiable by separate codes. Regarding bacterial infections, various manifestations may be documented. A combination of multiple nominations is derived with items not selected or information generally missing identifiable by separate codes.

Occurrence of **concomitant diseases/medical history (MH)** is recorded on baseline CRF as “yes / no” and by terms of diseases and required medications on a separate MH CRF if necessary. For documentation of **concomitant medication (CM)** serves a separate CM CRF if necessary. The reported terms of diseases/indications are MedDRA-coded to ensure analyses on the most suitable aggregation level, e.g. MedDRA preferred term (PT) and/or systems organ class (SOC).
In addition, for both MH and CM the individual numbers (counted MH/CM datasets per patient) and strings comprising all patients’ information (using MEdDRA low level terms) were derived from the MH/CM CRF pages, with separate CM strings of named medications and indications due to medical history [MH] or due to adverse events [AE].

Standard therapy of the ACLF per centre (i.e. the control intervention in the GRAFT trial) is documented regarding various frequently used measures in regular care on the baseline CRF as “yes / no”. As described for concomitant medications, individual number of measures applied to improve/control ACLF and a string comprising all measures per patient and visit was derived. This was also done for all visits performed per patient.

Laboratory data is recorded on baseline CRF regarding a variety of relevant measures (as well as at every visit). Furthermore, a clinical assessment is documented which classifies the observed values in “normal / abnormal without clinical relevance / abnormal with clinical relevance”.
In addition, individual numbers of clinically relevant abnormal measures per patient (and visit) and a strings referring to which entitites were concerned were derived from the respective CRF pages per visit.

4.4.2 Derivation of disease describing dates, times and durations

First date in patients’ documentation is derived as minimum date from randomisation CRF and all BL CRF pages.

Last visit date in patients’ documentation is derived as maximum date from visit dates performed (CRF pages BL, V1 to V8).

Last date in patients’ documentation is derived as maximum date from randomisation CRF, BL CRF page, <last visit date> (derived item), all OLT CRF pages from V1 to V8, and ES CRF.

Duration of observation is derived from date of randomisation (Rando CRF) to <last date>. As described in sec. 4.1.1 a one-day tolerance policy was applied if duration in [days] is calculated.

Duration of hospitalisation is derived from date of randomisation (Rando CRF) to date of discharge from hospital as requested at the visit’s regular CRF page. As described in sec. 4.1.1 a one-day tolerance policy was applied if duration in [days] is calculated.

4.4.3 Derivation of scores describing the liver function

All the scores/items named below are derived per visit based on detailed data documented at the respective CRF pages.

MELD score is derived from serum bilirubin and creatinine, both in [mg/dl], and international normalized ratio for prothrombin time (INR), recorded on the visit's CRFs.

Lowest values should be 1, highest values of creatinine should be 4 as in usual scoring schemes see Hofmann WP et al. 2008 [3], and further references [4, 5] so that the score's range is [6; 40]. If any compound is missing in spite of queries, the score p. visit cannot be computed.

Child-Pugh score is derived using serum albumin [g/dl], bilirubin [mg/dl] and INR, grades of ascites and HE recorded on the visit's CRFs according to [3]. All these entities are classified into 1 to 3 points with higher numbers referring to higher deviations from normal. At last, all points are added to classify a patient as "A" to "C" with "C" indicating highest severity of disease.

Within the scheduled interim analysis, both scores will only be used in description of treatment groups at baseline.

4.4.4 Derivation of organ failures in other organ systems

Organ failures are according to Jalan R et al. [6] (yes / no) is derived per patient an visit using the laboratory, and clinical data as presented and classified in Table XX: all necessary information is collected on baseline CRF and at every visit's CRFs until V8.

Organ/system	Subscore = 1	Subscore = 2	Subscore = 3
Liver	Bilirubin <6 mg/dl	Bilirubin ≥6 mg/dl and <12 mg/dl	Bilirubin ≥12 mg/dl
Kidney	Creatinine <2 mg/dl	Creatinine ≥2 mg/dl and <3.5 mg/dl	Creatinine ≥3.5 mg/dl or renal replacement
Brain (West-Haven grade for HE*)	Grade 0	Grade 1-2	Grade 3-4**
Coagulation	INR <2.0	INR ≥2.0 and <2.5	INR ≥2.5
Circulatory	MAP ≥70 mmHg	MAP <70 mmHg	Use of vasopressors
Respiratory PaO_2/FiO_2 or SpO_2/FiO_2	 >300 or >357	 ≤300 and >200 or >214 and ≤357	 ≤200# or ≤214#

The shaded area describes criteria for diagnosing organ failures.
*HE, hepatic encephalopathy; FiO_2, fraction of inspired oxygen; PaO_2, partial pressure of arterial oxygen; SpO_2, pulse oximetric saturation.
**Patients submitted to Mechanical Ventilation (MV) due to HE and not due to a respiratory failure were considered as presenting a cerebral failure (cerebral subscore = 3).
#Other patients enroled in the study with MV were considered as presenting a respiratory failure (respiratory subscore = 3).

Table 1 : Organ failures defined according to the CLIF-C OFs (Jalan R et al. 2014)

Calculation of **ACLF grade** per patient and visit is based on the presence of organ failures acc. to Jalan (see above) and additional clinical and/or laboratory values observed for a patient. Details are already describe in sec. 4.3.

Within the scheduled interim analysis, frequency of organ failures and (calculated) ACLF grade will only be used in description of treatment groups at baseline.

4.4.5 Trial intervention

While in the control arm treatment as usual (TAU) is applied, in the interventional arm repeated G-CSF injections are given in parallel to TAU. The following variables are used to describe the study intervention and assess compliance.

G-CSF intervention

- **G-CSF given** (yes/no) as recorded on the STHER CRF per visit;
- **Cumulative G-CSF dose:** A visit-independent separate GCSF-CRF page contains the data regarding all G-CSF applications with dates, times, doses (in agreement with the initial weight of patient, acc. to protocol without changes over the whole period of intervention), numbers of charges and ambulatory or stationary application used.

Although 12 injections are regularly planned as trial intervention, the possibility exists to document more applications if done (e.g. due to misunderstanding between various departments involved in a patient's care). From this data, the total dose applied and possibly deviation from/ violations of protocol are derived.

- **Number of injections:** see above
- **Total duration of intervention:** interventional period [in days], defined as time differences from start date to end date of all injections documented. As described in sec. 4.1.1 a one-day tolerance policy was applied since duration in [days] are calculated.

Intervention as randomized (yes/no) is derived from identical data regarding treatment allocation (as recorded on the randomization CRF) and aggregated documentation (beside further information from on-site-monitoring or direct communication with the centres): from all STHER CRF per visit if performed via the GCSF item (see above) as well as an existing GCSF-CRF (with documented applications). If no discrepancy became known it is assumed that a patient is treated according to the randomized arm allocation as intended to treat.

Deviation from/ violations of interventional procedures regarding G-CSF application are derived with following items (from GCSF CRF page):

- Numbers of injections following the scheduled regimen (1d/ 3d differences between subsequent injections acc. to protocol);
- Shortened resp. *relevantly* shortened series of injections *(<9 from 12 injections planned);*
- Applied dose per injection acc. to protocol in all applications;

Description of treatment as usual (TAU; within both arms)

- **Combination of regular interventions** applied per visit and patient are aggregated from predefined treatments options (from all STHER CRF per visit if performed) to described the standard intervention/s of ACLF within the centres.
- The total **number of TAU** reported as well as mean number per visit is derived from pre-defined options on STHER CRF page (yes/no) to describe comparable intensity of basic treatment.

4.4.6 Assessment of complications of ACLF

Observed complications of ACLF are documented on the EP CRF-pages and refer to

- hepatorenal syndrome (HRS),
- variceal or gastrointestinal bleedings,
- ascites, and
- hepatic encephalopathy (yes/no) including the respective durations (if possible).

Furthermore, proven bacterial infections (yes/no) necessitating systemic use of antibiotics including their manifestation:

- urinary tract infection,
- peritonitis
- pneumonitis
- cellulitis
- sepsis
- Clostridium d. infection, or others,

are reported.

Within the scheduled interim analysis, baseline characteristics will only be part of description of treatment groups. More details will be derived in detail will be described within the SAP of the final analysis.

4.4.7 Ruptures of the spleen

Beside the ACL complications, **size of spleen** and **ruptures** if occurred (N/Y) will be part of the SAP for the final analysis

4.4.8 Adverse Events

Other adverse events which may or may not be associate with trial procedures/intervention are reported on AE CRF-pages and MedDRA-coded.

Data processing for analyses of AEs (with possibly multiple data sets per patient) is performed in vertical structure of data sets. It comprises all items of the 3 AE-related DB-Tables (AE.AE.AE_Hdr/ ~.AE/ ~.AE_Comm) ordered by Patients'-ID (PATNR) and AE-sequential number (AESeq).

For every AE documented following items are processed:

- Sequence-no.: ~ AESEQ,
- Event: ~.AETerm, ~AELLT, ~AEDECOD, ~AEBodSys,
- Start and Enddate/ ongoing AE: ~.AEStDt/ ~AEEnDt, ~AEOng
- AE relationship to trial intervention ~.AERel,
- Seriousness: ~.AESer.

At first, AEs without an assessed possible causality to G-CSF are identified and listed separately. AEs with assessed possible causal relationship to G-CSF are analyzed in more detailed manner and separately described.

Data aggregation per patient summarized following derived variables:

- Number of AEs per patient in total (AnzAE);
- Number and kind of AEs per patient unrelated to any study procedure;
- Number and kind of AEs per patient related to G-CSF.

All AEs are MedDRA-coded so that tabulations of frequencies according to Preferred Terms (PT) und System Organ Classes (SOC) are possible.

See sec. 6.5.3 for analysis.

4.4.9 Serious Adverse Events

Data processing SAEs is performed based on the two DB-Views (from the pharmacovigilance database):

- Vw_saes_no death_data, and
- Vw_saes_death_data.

Beside the reason(s) of death, a derivation of all SAE relevant data comes from the first view. However, a separation was suitable because in some deceased patients more than one SAE might be ongoing until death and more than one might be regarded the assumed reason for patients' death. Therefore, MedDRA-coded terms of death are reported per patient based on the 2nd DB view.

In general, serious adverse events are present in vertical structure of data sets (with multiple data sets per patient) and ordered by Patients'-ID and case-ID (consecutive number of event/case p.pat.). In all SAE the most recent SAE-(Update-)report is used for further processing. All SAEs are coded by the recent MedDRA version.

A SAE-AE reconciliation will be done and discrepancies clarified with the centre to ensure data consistency before database freezing/closure.

See sec. 6.5.3 for analysis.

4.4.10 Post-interventional/ follow-up assessments

Documentation for visits 6 (after 90 days), 7 (after 180 days), and V8 (after 360 days) are comparable to those of V1 to V5 with the same complications documented. During the FU period however, it is possible to suffer from more than one episode of the respective complication named in sec. 4.4.6. Therefore, these numbers reported on EP-CRF pages are included in any counting and detailed analysis of re-occurrence within the final analysis.

Additionally, occurrence of a potential relapse of ACLF (N/Y) is reported (EP-FUP CRF page). Information on a relapse of ACLF it derived from aggregation over those visits V6 to V8. Detailed analysis will be part of the final analysis only.

5 General issues for Statistical Analysis

5.1 Study periods

The following study periods are distinguished for analysis:

- Registration
- Baseline Assessment
- Randomization and Intervention (G-CSF & TAU or TAU alone; until Day 26 respective visit 5/ D28)
- Follow-up at D90, D180, D360

5.2 Disposition of Subjects and Withdrawals

The number of patients

- screened,
- randomized to the treatment arms

will be provided within the interim analysis.

5.3 Analysis Populations

Analysis populations are defined according to the protocol of the GRAFT-Study (see protocol chapter 8.5).

The **Full Analysis Set (FAS)** for the intent to treat analysis consists of all randomized patients with valid informed consent. The confirmatory analysis based on the FAS utilizes the randomized arm independent of the true intervention.

The **Per-Protocol Set (PPS)** is a subset of the FAS consisting of patients without major protocol violations (see sec. 5.4).

The **Safety Analysis Set (SAS)** consists of all patients of the FAS, but patients are classified whether or not they received at least a single G-CSF injection, irrespective of the randomized group allocation.

Additional sub-populations are defined to explore possible benefits of G-CSF treatment in patients with possibly best prognosis:

1. **3d Survivor-Set (3dSV)**: all patients who are alive at least 3 days after randomisation,
2. **7d Survivor-Set (7dSV)**: all patients who are alive at least 7 days after randomisation: → both conditions are derived from the patients' documentation on the ES-CRF and Rando-CRF pages, see also sec. 4.1.1
3. **Patients with alcoholic hepatitis (AlcHep)**: derived from the ACLF page at Baseline, Item RALC="ja"

4. **Patients without bacterial infections (NoInf)**: at randomisation: derived from the ACLF page at Baseline, Item RINF="ja"
5. **Patients** fulfilling the APASL criteria of ACLF according to Sarin et al. [7] (**APASL**):
   „…an acute hepatic insult manifesting as jaundice [serum bilirubin level >5 mg/dL] and coagulopathy [international normalized ratio >1.5), complicated within 4 weeks by ascites and/or encephalopathy in a patient with previously diagnosed or undiagnosed chronic liver disease"

   Present ascites and/or encephalopathy are derived from the randomisation page of CRF (E01A, E01B="ja") and will be verified at baseline from the VIT resp. the ACLF modules Bilirubin and INR will come from the Lab CRF-module at baseline as well.

All these sub-sets of patients (1)… (5) are defined for the interim/futility analyses of TFS90 and OSV.

The results found and the depending conditional power estimated within these sub-sets may allow investigating potential benefits of G-CSF treatment in sub-populations with better prognosis than that defined by the present inclusion/exclusion criteria for the trial population since - based on the recent evidence in literature - we assume these sub-populations may benefit most from the trial intervention.

5.4 Major protocol violations

The following protocol violations are considered major according to protocol:

- Violation of inclusion or exclusion criteria
- Intervention (G-CSF+TAU/TAU alone) not done or not performed according to randomization result: see sec. 4.4.5
- Patients who did receive less than 75 % of the intended total dose of G-CSF or who have received G-CSF on less than 9 treatment days for reasons others than early death, early OLT, or (serious) adverse events *or relevant clinical reasons;* see also sec. 4.4.5 for details regarding variables of G-CSF interventions;

5.4.1 Violation of inclusion or exclusion criteria

Inclusion and exclusion criteria were recorded as indicator variables on the patient registration CRF. Patients were only registered if all inclusion criteria were documented as "yes" and all exclusion criteria were documented as "no". However, in few cases monitoring on site detected, that an exclusion criterion was already met at registration. In these cases, the respective variable was corrected by the local investigators.

In addition, some of the exclusion criteria could be cross-checked with data recorded in the baseline CRF.

A violation of inclusion or exclusion criteria was considered present if:

- Any of the inclusion criteria was recorded as "no" or
- Any of the exclusion criteria was recorded as "yes" or
- *ACLF grade at baseline did not fulfill the inclusion criteria (because of timely changes observed between registration and baseline assessment (calculated from detailed patient data)*

5.4.2 Intervention not per protocol (without medical reason)

A major protocol violation with respect to the trial intervention is considered present if:

- **Intervention as randomized** is derived as "no" regarding the allocated treatment arm
- **applied G-CSF dose** is derived as too low or
- **applied number of G-CSF injections** is derived as too low,

both without a relevant medical reason given; from the **End of Treatment documentation** (TE-GCSF CRF page) are derived whether or not any of the named deviations occurred due to medical reason or has to be regarded as protocol violation.

Patients' inclusion in the per-protocol population depends on occurrence of any of the named criteria above (until OLT and/or death occurred).

5.5 Monitoring of data quality and protocol compliance

The number of outstanding documentations of visits, the number of open queries and the contents of queries with frequencies >10 times are presented.

6 Planned analysis

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

General on the planned confirmatory analyses

All statistical methods for analyses as predefined within the trial protocol are adhered to, without changes (see also sec. 6.4). However, additional analyses are planned (as described in laters sections) to

Significance level within the scheduled interim analysis

The scheduled interim analysis after outcomes of 50 % of the patients allows to detect marked superiority of the experimental treatment. Multiplicity adjustment followed O'Brien [8] with nominal significance levels p=0.005 at interim analysis and p=0.048 at final analysis. According to protocol the interim analysis was designed to have a power of at least 80% to detect a difference in transplant-free survival in the order of 33 % or higher.

Premature termination of the trial due to futility

During the previous meeting in March 2018 the Data Monitoring and Safety board (DMSB) recommended to include an option to prematurely terminate the trial due to futility in the statistical analysis plan for the scheduled interim analysis given the rather high numbers of early deaths observed in both arms without any indication of safety problems.

Therefore, sec. 7 details the procedures and definitions on which a decision on a premature termination of the trial are based.

6.1 Demographic and other baseline parameter

Demographic and other baseline parameter are described for the whole FAS and by randomization arm.

6.2 Concomitant Diseases and Medication

Lists of diseases as well as medications and underlying indications per patient are provided to allow a clinical assessment of associations between those and possible adverse events.

6.3 Compliance with regard to the study intervention

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

6.4 Primary endpoint

6.4.1 Confirmatory analysis by Cox regression

The study protocol specifies: The **primary efficacy analysis** of a treatment effect on transplant-free survival within 90 days are performed by Cox regression adjusting for ACLF grade to gain power… Confirmatory analysis follows the intention to treat principle and are based on the full analysis set (FAS). Hazard ratios for treatment effect and ACLF grade and their confidence intervals are presented." This will be done without changes in both interim and final analysis.

Previous data on TFS suggest that the proportional hazard assumption is justified [Garg 2012, Duan 2013]. The time-to-event data is further described by Kaplan-Meier curves per arm.

6.4.2 Further specification of the analyses

Because of the frequent occurrences of early deaths, for exploratory reasons, and to perform futility analyses we also will analyse TFS90 within predefined five sub-populations, see sec. 5.3.

Although the criteria defined there are not used in a stratified randomisation, we are convinced that it might be valuable to perform futility analysis (see sec. 7) within these sub-populations, too. If a premature termination of trial might be suggested by the conditional power the TFS90 may reach at the present level of information within the FAS, it seems reasonable to analyse sub-populations with possibly better prognosis in whom potential benefits of G-CSF therapy might be enhanced.

Analysis populations: FAS, 3dSV, 7dSV, AlcHep, NoInf, APASL

6.5 Secondary endpoints

6.5.1 Analysis of time to death and Overall survival at 360 days

Time to death is described using Kaplan-Meier product limit estimator. The same Cox regression model as for the primary endpoint is used in the analysis of the treatment effect including the same covariate (ACLF grade) as in the analysis of the primary endpoint.
Impact of further covariates can be explored in supplementary analyses.

Analysis populations: FAS, 3dSV, 7dSV, AlcHep, NoInf, APASL

6.5.2 Analysis of further secondary endpoints and other characteristics

No further secondary endpoints and further characteristics will be considered within the interim analysis but all will be part of the final SAP and analysis.

6.5.3 Analysis of Safety characteristics

Adverse Event (AE) and Serious Adverse Event (SAE) reporting was limited by the trial protocol to a time horizon of 28 days (V5) **except** for newly developed malignancies starting from the 1st injection of G-CSF in the intervention and from randomisation in the control arm. If an OLT occurred before day 28, SAEs have to documented **up to 3 days** after the last injection of G-CSF in the experimental arm; OLT itself must **not** be reported as SAE; a special OLT CRF page is to be filled in those cases on every visit performed later on.

Adverse events and serious adverse events are listed per patient and study arm. Tables acc. to aggregated Med-DRA levels (preferred terms or higher aggregation levels up to system organ classes) will allow comparing both arms descriptively. In cases of relevantly different frequencies exploratory Fisher- or Chi-squared test are provided.

Analysis populations: SAS, PPS, BCS

6.5.3.1 Analysis of SAR

Acc. to Amendment 4 and the previously acknowledged changes of the SmPC (Summary of medicinal Product Characteristics) any diagnostically confirmed aortitis causes a SAE reporting even if it does not meet the SAE criteria.

Serious adverse reactions (SAR) and suspected unexpected SAR (SUSAR) are provided by patient listings – if applicable – including the information whether or not the respective patient belongs to the PPS. In summary tables, the frequencies of events will also be presented per arm by MedDRA terms like done in all annual safety reports during the course of trial using the most adequate level of aggregation to allow identification of descriptive group differences.

6.5.3.2 Analysis of AR

For adverse reactions (AR: AE with possible causal relationship to G-CSF) absolute and relative frequencies per treatment arm are provided with regard to outcome and severity.

Analysis populations: SAS, PPS, BCS

6.6 Sub-group analyses

Detailed sub-group analyses are part of the final SAP and analysis.

7 Analysis of Futility

7.1.1 Background

According to Sully et al. [12] publicly funded trials often have poor recruitment rates - as also observed in the GRAFT trial – and numbers of them are neither successful in recruiting their initial target sample size nor resulting in statistically significant and clinically important results. From 1993 to 2008 only 19% of the publicly funded superiority trials (from the two major UK funding bodies: the HTA programme and the Medical Research Council) which Sully et al. considered in a systematic review provided significant results. Furthermore, the authors simulated futility analyses in these UK trials in-between 2002 and 2008. Their results revealed that inclusion of a single futility analysis using conditional power after 75% of patients had been recruited with a stopping boundary of 30%, would have correctly stopped about a third of all trials for futility which went on to have negative results.

Although the recruitment in the GRAFT trial improved within the previous (3rd) year, the initially planned period of recruitment is already exceeded even though only half of the sample size is recruited yet. The Deutsche Forschungsgemeinschaft (DFG), which financed the trial, allowed the trial's continuation and extended the duration of the trial until September 2019 so far but without further financial support.

In parallel, during both periods in which annual safety reports (ASR) for the GRAFT trial previously had to be provided we observed relevant proportions of early deaths in the randomized patients. These deaths occurred that near after randomization that - due to the nature of the G-CSF treatment under investigation, with somewhat delayed beneficial effects expected - it seems implausible to expect an informative contribution to the study question although formally the primary endpoint was reached and albeit apparently no safety concerns are present.

Therefore, at the previous DMSB meeting we decided to combine the scheduled interim analysis after 50% of planned recruitment with an analysis of futility although not explicitly stated within the trial protocol.

7.1.2 Technical solution to achieve conditional power

The PASS software [2] contains a procedure, which allows to calculate the conditional power – the probability that the final result of a trial are significant, given the data obtained up to the interim look. This procedure uses the z-statistic of a logrank test (which compares the hazard

rate of a treatment group to that of a control group) and is adequate for time-to event endpoints like TFS90, and OSV within the GRAFT trial.

Indeed the logrank test provides nearly the same results (p values) as the Cox regression model, which is defined for the confirmatory analysis in the GRAFT trial (see sec. 6.4) protocol regarding the binary treatment variable.

Beside the *logrank z score statistic* and the *number of event observed (total of both groups)* at the time of the scheduled look at the data, which are only available from the interim analysis for the primary endpoint, following further information is necessary:
regarding the interim look

1. the information level (corresponding to the fraction of information at the time of the look compared to that at the designed end of trial), and

regarding the trial's design data from the planning stage of the trial:

2. the target number of events (in total),
3. the randomisation ratio and
4. the assumed hazard ratio (treatment/control) under the alternative hypothesis (for which the trial was initially powered)

Based on these details it is possible to estimate the conditional power at the time of the interim analysis. Before any data is collected at the beginning of the trial, the conditional power is equal to the unconditional one. Therefore, it is only useful to calculate the conditional power not too early within a trial: say if already for 50 % or 75 % of patients' endpoint data is available.

Steps of data preparation

To derive the level of information at the time of the interim analysis (point 1) the total number of *observed* events is needed. It comes from the latest data based on the DB snapshot for the interim analysis and calculations/analyses as described before. The information level I_k will then be calculated from the observed number and the randomisation ratio.

Regarding the design information mentioned above: in the GRAFT trial time-to-event data presents primary endpoint. Since only a part of the sample will experience the event of interest, not on the number of patients but the number of initially expected events (2) are relevant for further estimation. This target number of events expected in GRAFT (2) is derived from:

a) the hazard rates assumed for both treatment and

b) the sample size planned (n=131 patients per group, with randomisation ratio 1:1 (3)),

given the defined nominal significance level of α=0.05 (with adjustment to a single interim analysis as described in the study protocol), and a power of 1-β=0.9.

Based on the CANONIC Study, a transplant-free survival rate of 42 % at day 90 in the control group was assumed, therefore an event rate of 58%. Furthermore, we aimed to detect an absolute between-groups difference of 20 % (from 42 % to 62 %) in transplant-survival at day 90, therefore a reduced event rate in the G-CSF arm of 38%. The expected numbers of events per group are then:

$n_{G\text{-}CSF}$=131*0,38=49,78≈50 and

$n_{Control}$=131*0,38=75,98≈76;

in total N=125,8≈126 presents the target number of events (2).

The hazard ratio to observe either death or OLT until day 90 (4) is the reciprocal of the assumed HR_{TFS90}=1.815 named in the trial protocol, and therefore HR_{event}=1/1.815=0.551 (as assumed in sample size calculation).

7.1.3 Preparation of interim data for futility analysis

Since both the Cox regression model and the SPSS-based Logrank test and associated Kaplan-Meier curves provide Chi^2-distributed test statistics, which are not easily to be transferred into the z score of the logrank test needed for the futility analysis the following method is chosen:

Within the time intervals i between to subsequent GRAFT visits (from Baseline to visit 8= End of observation: 8 intervals) the number of observed events (O_i) and censored cases are counted per arm and in total so that both the number of patients under observation, and events are available per interval. Furthermore, the ratio O_i/N_i, the expected values $E(O_i)$ and variances $Var(O_i)$ will be calculated and the expected standard deviation (as square root of the total variance over all intervals).

The (standard-normal distributed) z score of the logrank test statistic is derived when summing-up all differences between observed and expected numbers (given the null hypothesis is true) over all intervals and divide it by the expected standard deviation.

All these calculations necessary to derive the condition power for the major endpoints will be done by a programme and applied within all sub-populations (1) … (5), see sec. 5.3, in whom knowledge of the conditional power is essential to decide on continuation or premature termination of the GRAFT trial.

7.1.4 Conditional power as pre-requisites of potential premature termination

Since no safety concerns were detected during the complete observational period until the previous safety analysis, we further assume no safety problems associated with the study intervention and no ethical reasons for a premature termination.
Furthermore, no statistical problems regarding an inflation of the β (or α) error is to be expected from futility analyses as planned.
Therefore, it is not necessary to pre-define a cut-off value, which the observed conditional power should exceed to support a decision in favour of continuation or premature termination of the trial. Based on the conditional power values from the interim data within the FAS *and all exploratory sub-populations* and following the DMSB's recommendation, a decision will be made to either continue or prematurely abort the GRAFT trial after consideration of all ethical, scientific, procedural and financial aspects by the principal investigator, biostatistician, and the management of the ZKS.

8 Further steps until final analysis

If the primary analysis within the FAS population supports a significant superiority of the G-CSF treatment based on α level defined of the interim analysis ***or*** if the conditional power results as whole picture of all futility analyses lead to the decision to stop the GRAFT trial, fhe regulatory authorities have to be informed on the decision of a premature end of recuritment..

Independently from stopping further recruitment all patients included until this time point will be further treated and observed until their regular end of trial or death (with or without OLT). In doing so we will have all endpoints completely available for final analysis, which may be performed after LPLV, finalised DM procedures and DB closure.

In case neither superiority nor futility was demonstrated based on the results of the interim analysis, the GRAFT trial will be continued with or without changes (as recommended by the DMSB) and decided by the principal investigator.

An extended scope of final analysis will be described within a separate statistical analysis plan and comprise additional aspects of secondary and exploratory data analyses to gain as much information as possible from the data as aquised.

9 References

1. IBM: SPSS V25. https://www.ibm.com/de-en/marketplace/spss-statistics; September 2018
2. NCSS: PASS Sample Size V14. http://www.ncss.com/software/pass/; December 2018
3. Hofmann WP, Rädle J, Moench C, Beschstein W, Zeuzem S. Prädiktion der perioperativen Mortalität bei Patienten mit fortgeschrittener Lebererkrankung und abdominalchirurgischen Eingriffen anhand unterschiedliche Score- und Testsysteme. Z Gastoenterol 2008; 46:1283-9
4. UNOS (2009-01-28). "MELD/PELD calculator documentation" (PDF).
5. Kamath, Patrick S; Kim, W. Ray (2007). "The model for end-stage liver disease (MELD)". Hepatology. 45 (3): 797–805.
6. Jalan R, Saliba F, Pavesi M et al (for the CANONIC study investigators of the EASL-CLIF consortium). Development and validation of a prognostic score to predict mortality in patients with acute-on-chronic liver failure. J Hepatol 2014;61:1038-1047
7. Sarin SK et al. Acute-on-chronic liver failure: consensus recommendations of the Asian Pacific Association for the study of the liver (APASL). Hepatol Int 2009;3(1):269-282)
8. O'Brien PC et al. Biometrics. 1979 Sep;35(3):549-56.
9. Sully BGO, Julious SA, Nichols J. An investigation of the impact of futility analysis in publicly funded trials. Trials 2014;15:61-70
10. Jennison, C., and Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC. New York, (from PASS [1]).
11. Proschan, M., Lan, K.K.G., Wittes, J.T. 2006. Statistical Monitoring of Clinical Trials. Springer. New York, (from PASS [1]).
12. Chang, Mark. 2008. Classical and Adaptive Clinical Trial Designs. John Wiley & Sons. Hoboken, New Jersey, (from PASS [1]).

10 Appendix

10.1 Abbreviations

AE	Adverse event
AlcHep	patients' sub-population with alcoholic hepatitis
APASL	patients' sub-population which fulfil the APASL criteria
CRF	Case Report Form
EC	Ethics Committee
FAS	Full analysis set
GCP	Good Clinical Practice
ICH	International Conference on Harmonisation
NoInf	patients' sub-population without bacterial infection at randomisation
PPS	Per protocol set
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Safety analysis set
3dSV	3d Survivor-Set (3dSV) of patients
7dSV	7d Survivor-Set (3dSV) of patients

10.2 Conversion of Laboratory and Clinical Values

Laborwert	**Umrechnung**		**Umrechnungsfaktor**
	von der Einheit	**in die Einheit**	
bilirubin	µmol/l	mg/dL	0.0585
creatinine	µmol/l	mg/dL	0.01131
albumin	g/L	g/dL	0.1
albumin	mg/dL	g/dL	0.001
FIO_2	%	(without dimension)	0.01

Mean arterial pressure= (2*diastolic blood pressure [mm Hg] + 1*systolic blood pressure [mm Hg])/3.
(Patients with deviations between calculated and centre-reported value are listed.)

10.3 Umrechnung von Dosisangaben

Acc. to protocol G-CSF doses should be guided by the body weight using a cut off value of 70 kg (≤ 70kg 30 Mio IU G-CSF, > 70 kg 48 Mio IU G-CSF).

30 Mio IU means 0.5 ml solution (=300 µg filgrastim) while 48 Mio IU refer to 0.8 ml solution (400 µg filgrastim) 0.8 ml. Therefore the cumulative dose [in ml] instead of [IU] may be calculated by division through 60 → [mL]=[IU]/60 according to SmPC.