Statistical Analysis Plan (SAP)
for the
final 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: 2020-05-13

Version: final

EudraCT-No.: 2015-002212-32

ClinicalTrials.gov - no.: NCT02669680

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

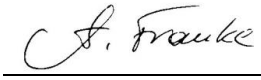
Coordinating Investigator		
	Date	Signature
Biometrician	13.05.2020	
	Date	Signature

Table of contents

1 Introduction

The purpose of this document is to provide a detailed elaboration of the final statistical analysis in the GRAFT trial which investigated the efficacy of serial G-CSF applications in patients with acute on chronic liver failure (ACLF), including detailed procedures for the confirmatory analysis of the primary and all secondary endpoints as well as other variables.

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 to the development of the SAP. Analysis programmes/routines developed for the ASRs, DMC reports and the interim analysis are re-used. The SAP is finalized before freezing the study data base for the final analysis, without more detailed knowledge on the complete data to as was available by DB snapshot of the interim analysis in March 2019, or presented within the last safety analyses (of January 2020). However, in preparation of the analyses statistical monitoring procedures and complex data checks were performed to ensure data completeness and consistency in front of the latest ASR.

IBM® SPSS® statistics, version 25 and 26 [1], PASS sample size®, version 14 [2], and R version 3.6.1 (or higher) 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 performed after approx. 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 conducted the analyses for the 3rd Annual Safety Report and the scheduled interim analysis and by using the same database (DB) freezing from January 2019.

Regarding the results of this interim analysis the coordination investigator decided on recommendation of the trials’ DMC to prematurely terminate the GRAFT trial without further recruitment of patients after April 2019 but to follow-up all patients until their regular or premature end of observation after 360 days. Given a regular end of trial of the last patient included LPO date will be the on March, 12th, 2020. The final analyses will be conducted after completion of data management procedures.

2.2 Purpose of the Statistical Analysis Plan

The purpose of this SAP is to outline the planned analyses of the final analysis.

Any post-hoc or unplanned analyses are clearly labelled as post hoc in the respective Clinical Study Report (CSR).

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.

The number of patients in various pre-defined analysis populations.

3.2 Impact on Statistical Analysis

Due to the possible heterogeneity of the trial centres, stratified randomisation by trial centres was done 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 Secondary and Safety Endpoints

4.2.1 Time-to-event endpoints

Beside the major secondary endpoint

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

which was already analysed in the scheduled interim analysis, all further secondary endpoints named within the trial protocol are part of the final analysis:

- Transplant-free survival at 360 days: Status and time see 4.1

4.2.2 Assessment of complications during the course of observation

Observed complications of ACLF are documented on the EP CRF-pages (yes/no; ongoing since last visit/ new ones) and refer to

- liver transplantation
- oesophageal and/or gastric varices
- gastrointestinal bleedings,
- ascites,
- hepatorenal syndrome (HRS),
- hepatic encephalopathy, and
- proven bacterial infections (yes/no) necessitating systemic use of antibiotics;
  for the latter their manifestation/localization will be provided for descriptive reasons:
  - urinary tract infection,
  - peritonitis
  - pneumonitis
  - cellulitis
  - sepsis
  - Clostridium d. infection, or
  - others.

For all complications the scheduled horizons of analysis refer to all observations until Day 90 (~_V6) and over the total period of observation (until V8= EoS; ~_iV). Therefore, the following detailed description of operationalization is valid for both quantifications per complication. During the FU period (after V5) however, it is possible to suffer from more than one episode of the respective complication. Therefore, these numbers reported on EP-CRF pages are included in any counting of re-occurrence.

In special analyses, a possibly deviating time horizon is explicitly stated if necessary for valid results.

4.2.2.1 Characteristics regarding HRS during the course of trial

Independently of whether or not a HRS exists at BL, any remission will be identified (i.e. grad=0/no HRS reported; HRSremis_~) together with the respective time from randomisation [days]. In case of repeated remissions during observation, the time/duration until first and the last remission is stored (Dur 1HRSremis_~; Dur_lastHRSremis).

An existing HRS at BL is counted as 1st episode.

In case a remission of an early episode occurred or no HRS during the early course was observed at BL, a subsequent re-occurrence is counted as next episode (over all EP-1 CRF pages per visit during the V6 resp. the end of observation (HRSepis~). To ensure correct quantification, patterns of ongoing & new occurrence (per visit) will be used to identify new re-occurrences. Ongoing HRS from preceding visits (which was already counted at that visit) is ignored to count only new episodes.

In case of repeated (new) episodes the time/ duration until first and the last episode is stored (Dur_1HRSreep_iV, Dur_lastHRSreep_iV).

Cases without any occurrence of HRS will have no episode. Cases with ongoing HRS from BL or from a later visit until end of observation (without remission) will have one episode without *remission during the course*.

Cases with initially reported HRS (and possibly continuing HRS until a few later visits) but remission thereafter may have one (or more) *further episode* of HRS, provided the episodes are separated by at least one visit without HRS (and experienced various remissions). The same is true for cases without HRS at BL but occurrence in an early visit followed be possible remission.

Occurrence resp. number of and time to next episode(s) of hepato-renal syndrome (HRS) will be reported.

4.2.2.2 Characteristics regarding ascites during the course of trial

Analogue sec. 4.2.2.1

4.2.2.3 Characteristics regarding HE during the course of trial

Analogue sec. 4.2.2.1
Documentation of “no HE” or “HE grade=0” are regarded a remission of HE.

4.2.2.4 Characteristics regarding portal-hypertensive bleeding during the course of trial

Analogue sec. 4.2.2.1; but just to quantify frequencies of occurrence and without times to event (AnzPhBleeding_~).
Occurrences of bleeding of oesophageal and gastric varices are separately quantified according to CRF EP-page and aggregated for every time horizon to analyse the portal-hypertensive bleeding which is expected to occur rather seldom.

4.2.2.5 Characteristics regarding gastrointestinal bleeding during the course of trial

Analogue sec. 4.2.2.1,; but just to quantify frequencies of occurrence and without times to event (AnzGIBleeding_~).

4.2.2.6 Characteristics regarding bacterial infections during the course of trial

Analogue sec. 4.2.2.1; but just to quantify frequencies of occurrence and without times to event (AnzInfect_~).

For a detailed data descriptions patterns of various (new and ongoing) infections per visits (KomInfNew_~/ KomInfOng_~) are derived (in string variables of 7 digits, with Yes=1, No=0, or "not reported"=9) to gain detailed information on bacterial infection per visit, patient, and arm.

4.2.3 Liver function during the course of treatment and follow-up endpoints

- MELD-Score and
- Child-Pugh-Score.

MELD score is derived, using 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]. Neverthelss, the score's range may exceed the usual maximum of 40 which ist used for emergency lists before OLT. If any compound is missing in spite of queries, the score per 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.

For the final analysis, both scores will be calculated for every study visit (if data are available). In case of death, early dropout or partially missing data the missing scores will be imputed by explicit codes which allow an analysis until the regular end of observation with an adjustment for possible imbalances between the arms.

4.2.4 Serious Adverse Events and deaths

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.

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

- Occurrence/number 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-coded as described for SAE.

Causes of early death occurred before transplantation will be extracted for between-group comparisons. MedDRA Preferred terms will be used and higher aggregation levels if suitable.

See sec. 6.6.7 for analysis.

4.2.5 Derivation of organ failures in non-liver organ systems

Organ failures are according to Jalan R et al. [6] (yes / no) are derived per patient and visit using the laboratory, and clinical data as presented and classified in *Table 1***Fehler! Verweisquelle konnte nicht gefunden werden.**. All necessary information is collected on baseline CRF and at every visit's CRFs until V8.

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

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).

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.

4.2.6 ACLF re-episode(s)

An ACLF re-episode requires that (at least) a single visit with ACLF remission (grade=0) was observed during the course but an increase to ACLF grad>0 occurred in a subsequent visit.
Occurrence/number of ACLF re-episode(s) will be identified based on all the organ failures observed and the respective ACLF grade per visit calculated. If the identification of an organ failure is impossible due to missing data in the underlying item(s), this fact will be reported to allow a validity assessment. Additionally, the centre-reported ACLF grade will be used instead of the calculated grade (see also sec. 4.3).

4.2.7 Duration of initial hospital stay

Duration of initial hospitalisation is derived from date of randomisation (Rando CRF) to the date of discharge from hospital as requested on the visit's regular CRF page in-between V2 and V6. This information will be aggregated over all visits available and the days from randomisation to the discharge after the initial stay will be calculated. Time to discharge will be defined as time from randomization to day of discharge, censoring patients deceased during the initial hospital stay. Furthermore, the time from the reported start of ACLF symptoms until discharge will be calculated. In case of a reported discharge but missing date the respective date of visit will be used to calculate the duration until discharge. As described in sec. 4.1.1 a one-day tolerance policy was applied if duration in [days] are calculated.

4.3 Covariates for confirmatory and sensitivity analyses

According to the trial protocol, Cox regression models within strata of Acute on Chronic Liver Failure (ACLF) grades are used to estimate the treatment effect. Hazard ratios for treatment

effect and their confidence intervals are presented within the full analysis set (FAS) of patient for confirmatory analysis.

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. Therefore, all clinical and/or laboratory sub-criteria essential to assess the ACLF grades had to be documented and ACLF grade was calculated according to published calculation schemes.

However, ACLF grade was not used as stratification criterium.

Analyses of agreement between centre-reported and calculated ACLF grades in front of previous DMC meetings revealed that for about 20 to 25% of patients deviating grades were reported. Therefore, we decided to use the program-based and validated ACLF grades (whenever possible) in the statistical analyses and report on the rate of agreement.

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*)

The same model will be applied within all predefined analyses populations, see sec. 5.3.

Further sensitivity analyses will focus on

- gender
- ACLF grade at BL documentation: ≤1 versus 2/3.
- Infection vs. no infection at randomisation
- Liver failure vs. no liver failure at baseline documentation

These binary characteristics serve as covariates in further Cox regression model/s.

Tests for interaction/s are performed adding an interaction term of the respective covariate with treatment to the primary Cox regression model (compare sec. 6.5).

Additional exploratory sub-group analyses will clearly be labelled as post hoc in the clinical study report.

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.

Results of pregnancy tests during the course of trial will be reported.

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 (FIO_2); in case of missing values (MV) and no supportive measures a rate of 0,21 is substituted in database.
- partial pressure of arterial oxygen (PAO_2) or peripheral oxygen saturation (measured via pulse oximeter; SPO_2)

The ratios of PAO_2/FIO_2 and SPO_2/FIO_2 are derived (per visit) and a respiratory organs failure identified according to *Table 1* from either or the other ratio per visit.

The precipitating event(s) of ACLF development is(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 namings is derived with items not seleceted 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] are calculated.

4.4.3 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** (GCSFappl=yes/no) as recorded on the STHER CRF per visit;
- **Cumulative G-CSF dose** (kumDose_IU, kumDose_ml): 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** (AnzInj_pPt)**:** see above
- **Total duration of intervention** (TreatDur)**:** interventional period [in days], defined as time differences from randomisation to the date of last injections documented. As described in sec. 4.1.1 a one-day tolerance policy was applied when duration in [days] are calculated.
- Whether or not a premature termination (GCSFAbbruchV) or an interruption of the G-CSF application occurred and the underlying reason (GCSFUnterbr_Gr), including an observed leucocytosis during the course of application or non-medical reasons.

Intervention as randomized (Interv_asRd=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 intented to treat.

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

- whether or not the series of injections was shortened resp. *relevantly* shortened *(<12 [GCSF_ISvk] resp. <9 injections [GCSF_relvk]);*
- whether or not the regimen of injections (regarding the 1d/ 3d distances between subsequent injections) was performed acc. to protocol or (one or more) deviations were identified (PV_GCSF_IR)
- whether or not at least 75% of the regularly planned GCSF dose per patients was applied (PPSapplGCSF_suff)

and if the letter is met

- whether medical (MedR_premaEoT) or non-medical reason/s have caused any premature End of Treatment (EoT) with a relevantly shortened G-CSF series(acc. to CRFs THER-END as well as THER per visit).

These items will contribute to classify a patient's belongingness to the per-protocol population (PPP). In contrast to medical reasons, non-medical reasons for relevantly shortened G-CSF series are regarded a „major" protocol violation (PV) and will lead to the patient's exclusion from per-protocol analysis (see sec. 5.4.2).

Deviations of the scheduled regimen of injections are classified as „minor" PV and will not lead to an exclusion from the PPP.

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; STh_alle_BL, STh_all_iV: with every option named once) to described the standard intervention/s of ACLF within the centres.
- The total **number of TAU** reported (AnzSTh_BL, AnzSth_iV) as well as mean number per visit (& AnzSth_MWpV) is derived from pre-defined options on STHER CRF page (yes/no) to describe comparable intensity of basic treatment.

4.4.4 Ruptures of the spleen

Beside the ACLF complications named above, **size of spleen, ruptures or other complication** if occurred (N/Y) are reported at every visit to assess any possible influence of the G-CSF treatment. While changes of spleen's size are descriptively reported during the course of observation, occurrence of ruptures or other complication are derived analogously to sec. 4.2.2.1 (SplCompl_~).

4.4.5 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. 6.6.7 for analysis.

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.

Further numbers of patient, e.g.
- included in the intent to treat analysis and evaluable for 28-day and 90-day all cause mortality, and
- included in the per protocol analysis and evaluable for 28-day and 90-day all cause mortality

are presented by a consort diagram along with reasons for exclusion.

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.

To explore possible benefits of G-CSF treatment in various sub-populations, additional subsets for exploratory analyses are pre-defined :

1. **Patients with bacterial infections as precipitating condition of ACLF (InfPrecEv_Pop)**: derived from the ACLF page at baseline (BL), section ACLF reasons, if the item RINF=”ja”; in case multiple reasons were chosen as precipitating event, all patients in whom a bacterial

infection was documented, alone or in combination with further reasons, comprise this sub-population.

2. **Patients with alcoholic hepatitis as precipitating condition of ACLF (AlcHep_Pop)**: derived from the ACLF page at BL), section ACLF reasons, if the item RALC=”ja”; in case multiple reasons were chosen as precipitating event, all patients in whom an alcoholic hepatitis was documented, alone or in combination with further reasons, comprise this sub-population.
3. **Patients without an existing bacterial infections as one out of various ACLF complication at baseline (NoInf_Pop)**: derived from the ACLF page at Baseline; section ACLF complications; all patients in whom the item INF=”nein” was documented, independently from the precipitating ACLF event.
4. **Patients** fulfilling the APASL criteria of ACLF according to Sarin et al. [7] (**APASL_Pop**):

„…serum bilirubin level >5 mg/dL and coagulopathy (international normalized ratio >1.5)”, verified by bilirubin and INR levels provided at BL. Furthermore, present ascites will be derived from the randomisation page of CRF (E01A=”ja”) and AND-combined baseline information from the VIT- and ACLF-complication documentation. Present encephalopathy come from the randomisation data (E01B=”ja”) and will be OR-combined with reported ACLF-complication documentation at BL.

5. **7d Survivor-Set (D7SV_Pop)**: all patients who are alive at least 7 days after randomisation: this condition will be derived from the patients’ documentation on the ES-CRF and Rando-CRF pages and refer to the major secondary endpoint, see also sec. 4.1.1.
6. **Best-Case Set (BestCase_Pop)**: all patients who are followed-up at least 14 days after visit V5 and without OLT until this time point (in both arms, BCS_minFUP) and received at least 75% of the intended total dose of G-CSF resp. 9 G-CSF injections (PPSapplGCSF_suff, as for inclusion into the PPS, see sec. 5.4) if allocated to the interventional arm.

   Prerequisite is that no relevant discrepancies in patient numbers per group at that time were observed and at least 20% of the number per arm reached an observation period that long.

While an analysis within the best-case sub-population primarily serves to describe any potential benefits of G-CSF treatment ***in the final analysis***, all former sub-sets of patients were already defined in front of the scheduled interim and futility analyses of TFS90 and OSV to avoid an unjustified premature termination of the trial, in case of potential benefits in the either or other subgroup of patients with better prognosis than that defined by the present inclusion/exclusion criteria for the trial population in total. Nevertheless, these sub-populations will also be used in the final analysis to

5.4 Major protocol violations

The following protocol violations are considered major according to protocol:

- Violation of inclusion or exclusion criteria (EK_OK, not AK_Viol)
- Intervention (G-CSF+TAU/TAU alone) not performed according to randomization result (Interv_asRd): see sec. 4.4.3
- Patients who did receive less than 75% of the intended total dose of G-CSF (PPSdose_GCSF; kumDose_suff) or who have received G-CSF on less than 9 treatment days/injections (AnzInj_suff, OR-combined in PPSapplGCSF_suff); for reasons others than early death, early OLT, or (serious) adverse events or relevant clinical reasons (MedR_premaEoT, NoMedR_premaEoT); see also sec. 4.4.3 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 randomised 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. However, according to protocol these patients were not excluded from further trial participation or from primary analysis.

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 contemporary changes observed between registration and baseline assessment or due to other reason.* Patients with reported ACLF-Grade>0 at registration and/or at baseline but with calculated ACLF grad=0 at BL will be excluded from the PPS.
- *The exclusion criterion whether a WBC>50 Gpt/L was present*, will be cross-checked with data recorded in the baseline CRF (not at the time of randomisation) and will lead to patients’ exclusion from the PPS.

5.4.2 Intervention not per protocol (without medical reason)

Further major protocol violations with respect to the trial intervention are considered. If no medical but non-medical reasons were provided for

- *An applied cumulative G-CSF dose lower than 75% of the planned dose (kumDose_suff) or*
- *An applied number of G-CSF injections lower than 9 of 12 planned (AnzInf_suff; OR-combined in PPSapplGCSF_suff), see also sec. 4.4.3*

This will lead to the patient’s exclusion from per-protocol analysis, see also sec. 4.4.3.

5.5 Data quality and protocol compliance

Deviations from the scheduled regimen will be presented. For protocol deviations regarding the G-CSF-treatment see sec. 4.4.3.

6 Planned analysis

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

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 issues on the planned confirmatory analyses

Each treatment comparison is reported as a point estimate of the intervention effect on a meaningful scale, together with its 95% confidence interval and a respective p-value.

For each treatment comparison in the major primary and secondary endpoints, both a simple and a model-based analysis are provided. The simple, easy to communicate analyses will use widely known standard methods like the chi²-test or t-test and the associated confidence intervals for the underlying measure of difference.

Advanced analyses will use Cox or logistic regression models to adjust the treatment comparison for relevant covariates.

If a relevant discordance between simple and advanced analysis arises, the model-based approach is given preference in general; but the conflict has to be explored and the statistical report and any publication will mention and discuss the discrepancy.

All statistical methods for analyses as predefined within the trial protocol are adhered to, without changes (see also sec. 6.5). However, additional analyses are planned (as described in laters sections) to

Due to the explorative nature of secondary analyses, missing data in secondary endpoints (beside the time- to event ones) and exploratory characteristics will not be imputed.

Significance level within the for final analysis

The scheduled interim analysis after outcomes of ~50 % of the patients would have allowed to detect marked superiority of the experimental treatment if present. Multiplicity adjustment followed O'Brien [8] with nominal significance levels α_1=0.005 at the interim analysis resulted in a remaining significance level of α_2=0.048 for the final analysis.

Premature termination of the trial due to futility

Based on the results of the scheduled interim analysis regarding transplant-free and overall survival (and further analyses within various sub-populations of assumed most beneficial prognosis) the Data Monitoring and Safety board (DMSB) recommended to prematurely terminate the trial due to futility in march 2019 given the rather high numbers of early deaths observed in both arms and the low conditional power regarding later successful results of the trial. The principle investigator followed this recommendation and decided on a premature end of recruitment in April 2019. Without any indication of safety problems the already included patients were followed-up until their regular end of observation or until a primary endpoint was documented. The last patient's last visit was in March 2019.

6.1 Demographic and other baseline parameter

Demographic and other baseline parameter are described for the FAS and by randomization arm by means of adequate descriptive statistics.

6.2 Concomitant Diseases and Medication

Frequencies of concomitant diseases and medication are described for the FAS and by randomization arm, based on MedDRA coding level of preferred terms to obtain statistically informative frequencies for arm comparisons. 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 Pattern of components of the primary endpoint

We will describe the frequencies of the observed patterns of components of the primary endpoint (death or OLT) within the FAS and by randomization arm.

6.5 Primary endpoint

6.5.1 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 within ACLF grades as strata to gain power… Confirmatory analysis follows the intention to treat principle and are based on the full analysis set (FAS). The hazard ratio for treatment effect will be presented with 95%

confidence intervals. Previous data on TFS suggest that the proportional hazard assumption is justified [Garg 2012, Duan 2013].

Furthermore, the cumulative time-to-event data per ACLF-grade and the Kaplan-Meier curves per arm will be presented.

6.5.2 Further specification of the analyses

Because of the frequent occurrences of early deaths in both arms and the premature end of recruitment, for exploratory reasons we will analyse TFS90 within predefined six additional sub-populations, see sec. 5.3.

Although the criteria defined there are not used in a stratified randomisation it might be valuable to report the primary endpoint for these sub-populations, too, for sensitivity reasons, given these populations deviate to >10% from the FAS sample or more than 20% of the FAS sample was observed beyond end of treatment *and* was censored at D90/visit 6 (V6).

6.6 Secondary endpoints

6.6.1 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 as in the analysis of the primary endpoint.

Furthermore, the cumulative survival curves per ACLF-grade will be presented.
Impact of further covariates can be explored in supplementary analyses.

Analysis populations: FAS, PPS, Inf, AlcHep, NoInf, APASL, 7dSV, BCS

6.6.2 Analysis of time to event and transplant-free survival at 360 days

Time to TFS 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, PPS, Inf, AlcHep, NoInf, APASL, 7dSV, BCS

6.6.3 Analysis of incidence of complications within 90 days/360 days

To analyse the incidence of complication (HRS, ascites, and HE) against the background of observed deaths, a first a descriptive analysis will quantify of the frequencies of complication per arm at BL. In a next step, all patients per arm will be identified a) without any occurrence of the respective complication as well as those b) without any remission during their entire course of observation, therefore a permanently persisting complication.

In a further step, a subset of patients with a remission during the course will be identified and the time of remission set to t_0 to analyse the incidence of/ time to the first re-occurrence per arm. Patients with death occurred before re-occurrence will be censored in this analysis. The incidences will be reported with 95% confidence interval.

> *CAVE: Due to the possibly small numbers and unbalanced arms this results can only be considered descriptively.*

In all survivors from V6 to V8, Venn diagrams per visit present pattern of complications including death, transplantation, HRS, ascites and HE.

Analysis populations: FAS, PPS, BCS

If the sample sizes per arm became known a decision on further sensitivity/exploratory analysis will be made for more detailed results.

6.6.4 Analysis of frequency of bleedings and infections within 90 days/360 days

A first descriptive analysis will quantify of the frequencies of complication per arm at BL. In a next step and given comparable survival curves per arm (as expected from the interim analysis), data will be cross tabulated and compared by Chi² test. Categorisations of 0 vs >0 bleeding episodes and 0 vs 1 vs >1 bacterial infection will be done before.

Analysis populations: FAS, PPS, BCS

If the sample sizes per arm became known a decision on further sensitivity/exploratory analysis will be made for more detailed results.

6.6.5 Analysis of liver function during the course of treatment

6.6.5.1 Analysis of MELD-score

Courses are analysed by repeated-measures ANCOVA until V6, with or without transformation of scale depending on a potentially skewed distribution. If a patient deceaded his/her MELD score for the last visit available will be imputed. Because of the relevant number of death before V6 the results should be regarded rather descriptively.

Analysis populations: FAS, PPS, BCS

6.6.5.2 Analysis of Child-Pugh-Score

Frequencies (per group) during the course of treatment and follow-up as well as comparisons between subsequent time points will be provided descriptively. Death and dropouts will be imputed by codes which allow.

Analysis populations: FAS, PPS, BCS

6.6.6 Further characteristics for evaluation

6.6.6.1 Analysis of organ failures

Frequencies (per group) of organ failures contributing to ACLF during the course of treatment and follow-up will be reported per arm, i.e. of

- liver function
- renal function,
- cerebral function,
- coagulation,
- respiratory function,
- cardiovascular function.

For further descriptive analyses see sec. 6.6.5.2.

Analysis populations: FAS, PPS, BCS

6.6.6.2 Analysis of ACLF-re-episode(s)

The analysis strategy ACLF-re-episode(s) is similar to that explained in sec. 6.6.3 (for observed complications).

Since all patients should have ACLF grades ≥1 at BL, in all patients a remission of ACLF must have occurred before a possible re-episode:

1. Calculation of the cumulative incidence function per arm for the first ACLF remission with an indicator variable containing: 0=censored (no remission observed/ 1= remission/ 2=death occurred before, using the time horizon from randomisation to last observation.
2. Calculation of the cumulative incidence function per arm for an ACLF re-episode with an indicator variable containing: 0=censored (no re-episode, no death) / 1= re- episode / 2=death occurred before, using the time horizon from the time point of remission to the last observation.
   CAVE: This analysis refers to that sub-population of patients per arm in which at least at a single visit with absence of ACLF (grade=0) was reported.

The between-groups difference of incidences will be reported with 95% confidence interval.

Analysis populations: FAS, PPS, BCS

6.6.6.3 Analysis of length of initial hospitalisation

The times to from randomization to day of discharge (DurInitHosp), and from start of ACLF symptoms (SyStrt_toDischarge) will be provided by cumulative incidence functions per arm for the initial discharge with an indicator variable containing: 0=censored (no discharge observed/ 1= discharge / 2=death during the initial hospital stay, using the time horizon from the respective starting points.

Analysis populations: FAS, PPS, BCS

6.6.7 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

6.6.7.1 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.

6.6.7.2 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

6.7 Sensitivity analyses

For sensitivity reasons, the consideration of additional covariates of possible impact within the model may be regarded as useful and may extend those which are outlined in sec. 4.3 depending on the results found. Additional exploratory analyses will clearly be labelled as post hoc in the clinical study report and or publication(s).

7 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. D'Amico G et al. J Hepatol 2018; 68:563-76
10. Jepsen P et al. Hepatol 2015; 62:292-302
11. Trinchet JC et al. Hepatol 2015; 62:737-750
12. Sully BGO, Julious SA, Nichols J. An investigation of the impact of futility analysis in publicly funded trials. Trials 2014;15:61-70
13. Jennison, C., and Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC. New York, (from PASS [1]).
14. Proschan, M., Lan, K.K.G., Wittes, J.T. 2006. Statistical Monitoring of Clinical Trials. Springer. New York, (from PASS [1]).
15. Chang, Mark. 2008. Classical and Adaptive Clinical Trial Designs. John Wiley & Sons. Hoboken, New Jersey, (from PASS [1]).

8 Appendix

8.1 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

8.2 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.)

8.3 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