Endpoints and design of clinical trials in patients with decompensated cirrhosis: Position paper of the LiverHope Consortium

Elsa Solà, Elisa Pose, Daniela Campion, Salvatore Piano, Olivier Roux, Macarena Simon-Talero, Frank Uschner, Koos de Wit, Giacomo Zaccherini, Carlo Alessandria, Ulrich Beuers, Paolo Caraceni, Claire Francoz, Rajeshwar Prosad Mookerjee, Jonel Trebicka, Victor Vargas, Miquel Serra, Ferran Torres, Sara Montagnese, Aleksander Krag, Ruben Hernáez, Marko Korenjak, Hugh Watson, Juan G.Abraldes, Patrick S. Kamath, Pere Ginès, for the LiverHope Consortium Investigators

Table of contents

Supplementary methods

1. METHODOLOGICAL ASPECTS FOR THE DESIGN OF CLINICAL TRIALS IN DECOMPENSATED CIRRHOSIS

1.1. Design and reporting results of clinical trials

Careful planning and reporting of the design of clinical trials is essential to achieve robust and comparable results. Important issues that need to be considered when designing clinical trials include aspects such as randomization and stratification, blinding and placebo methods, inclusion and selection of appropriate control groups, characteristics of the target population, selection of adequate endpoints and sample size calculation. Adherence to standard guidelines for reporting results of clinical trials is essential to provide comparable results between studies (https://www.equator-network.org) (Table 1).

1.2. Sample size calculation

Overall, sample size should be based on the primary outcome and the design of the study[1]. For the calculations of sample size it is necessary to predefine an expected effect on the primary endpoint in the intervention arm with respect to control group. The difference between the two groups should be based on a clinically important difference. In addition, it is important to consider the type of trial (i.e., superiority, noninferiority) as all have different sample size calculation requirements (details can be found at: https://www.equator-network.org/). Sample size calculation could be based both on a relative effect or absolute effect of the tested intervention; however, this effect needs to be clinically meaningful in that specific trial design. It is difficult to define what can be considered as a "clinically meaningful" effect. We consider that a specific reduction in relative mortality risk (or other primary endpoint) cannot be recommended here and the "meaningful effect" should be defined considering the details of study design and previous literature.

1.3. Statistical considerations

From a statistical perspective, it is noted that most phase 3 trials in patients with decompensated cirrhosis are designed and analyzed using time to event techniques. In some cases, another event may change the likelihood of the occurrence for the primary event of interest, and then that variable is considered to be a competing event (i.e., liver transplantation, TIPS). Competing-risk handling methods are then recommended so as to avoid upward-biased incidence estimates[2]. When recurrence of endpoints is considered important from a clinical endpoint, then techniques that collect recurrence of the endpoint of interest should be then used instead of the time to first event strategies. Recurrent-event methods consistently provide greater statistical power than time-to-first event analyses, particularly in the presence of highly heterogeneous study populations[3]. When the aims of the RCT include assessing the potential predictive role of post-baseline predictors including intermediate events, then methods assessing time-dependent covariates may be useful. It is of utmost importance to guarantee the complete follow-up of patients until death or until the end of the study period regardless of treatment and protocol adherence, for minimizing bias[4] and thus permitting the assessment of several types of treatment estimates as well as sensitivity and complementary analyses[5]. In addition, it is important to provide both intention to treat and per protocol analysis.

From this year on, there is a formal regulatory requirement to formulate the scientific questions in terms of an estimand framework at the stage design[4]. An estimand should include a clear definition on the target population, the endpoint, the population-level treatment effect measure and the post-randomisation (intercurrent) events during the trial (and the strategies for handling them)[6]. The handling of missing data should therefore be aligned with the study estimands and some good papers with examples have been recently been published[7,8]. A detailed discussion on the methods for handling missing data has been extensively handled in the literature and it is out of the scope of this manuscript clearly focused on clinical issues. For further details, specific recommendations from medical journals[9–11], as well as from expert's panel and regulatory guidances[8,12], are available for consultation.

The analysis of subgroups has a differentiated role when they are pre-planned in the context of exploratory or confirmatory trials, and a number of publications in scientific journals[13–18] and regulatory guidances[12] make clear recommendations on this topic. Even when the subgroups are pre-planned, there is a consensus in the sense that it should never serve as the basis to rescue a trial that has formally failed, and that any claim for a given subgroup should be substantiated by a clear biologic rational and supported with a mandatory strategy aimed to control the type I error. Any other situation should be considered only under an exploratory perspective and the conclusions should be therefore aligned with this principle. Particular warning is given to post-hoc data driven analyses, and this is applicable beyond the subgroup analysis issues. The type I error is not controlled and in fact, inferential analyses are biased when hypothesis were not pre-specified and tested data-driven[4].

Finally, clinicians and statisticians should predefine when an interim analysis will be performed and specify the stopping rules for that specific trial. This information needs to be included in the protocol design[19,20].

2. COST-EFFECTIVENESS ANALYSIS

Importantly, HRQOL assessment allows to carry out cost-effectiveness and cost-utility analysis, estimating the incremental cost of a treatment or intervention as compared with the treatment's incremental effects on health, which are usually measured by adjusting a clinical outcome such as survival by the HRQOL (Quality Adjusted Life Years, QALY)[21,22]. Such health-technology assessment strategies are used to decide the allocation of healthcare resources within limited resource settings by estimating the cost per QALY or incremental cost-effectiveness ratio (ICER) associated with medical intervention[23].

There are specific cost-effectiveness modeling guidelines for the design of RCTs[24]. In short, models ought to include both clinical outcome measures and health state utilities or QALYs alongside patient-level health resource usage. ICERs should be modeled with intention-to-treat estimates (rather than per-protocol) alongside a comprehensive uncertainty characterization. Both univariate and probabilistic sensitivity analyses are the preferred methods for that purpose. Conclusions for longer

time horizons than the trial follow-up period should be interpreted with caution. Finally, multinational trials require special care to tackle differences in both study populations and care patterns.

It is anticipated that several health technology assessment tools including tools that use artificial intelligence will be available in the near future. These tools need validation before being introduced for assessment of endpoints in clinical trials.

Supplementary references

1. Cook, J.A.; Julious, S.A.; Sones, W.; Hampson, L. V; Hewitt, C.; Berlin, J.A.; et al. DELTA 2 Guidance on Choosing the Target Difference and Undertaking and Reporting the Sample Size Calculation for a Randomised Controlled Trial. *BMJ* 2018, 5: 363-3750.

2. Austin, P.C.; Lee, D.S.; Fine, J.P. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* 2016, *133*: 601–609.

3. Claggett, B.; Pocock, S.; Wei, L.J.; Pfeffer, M.A.; McMurray, J.J.V.; Solomon, S.D. Comparison of time-to-first event and recurrent-event methods in randomized clinical trials. *Circulation* 2018. *138*: 570–577.

4. ICH E9 statistical principles for clinical trials | European Medicines Agency Available online: https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials (accessed on Jul 7, 2020).

5. Medicines Agency, E. *ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials - Step 2b*; 2017.

6. Pétavy, F.; Guizzaro, L.; Antunes dos Reis, I.; Teerenstra, S.; Roes, K.C.B. Beyond "Intent-to-treat" and "Per protocol": Improving assessment of treatment effects in clinical trials through the specification of an estimand. *Br. J. Clin. Pharmacol.* 2020.

7. Mallinckrodt, C.H.; Bell, J.; Liu, G.; Ratitch, B.; O'Kelly, M.; Lipkovich, I.; et al. Aligning Estimators With Estimands in Clinical Trials: Putting the ICH E9(R1) Guidelines Into Practice. *Ther. Innov. Regul. Sci.* 2020, *54*: 353–364.

8. Ratitch, B.; Bell, J.; Mallinckrodt, C.; Bartlett, J.W.; Goel, N.; Molenberghs, G.; et al. Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) Into Practice. *Ther. Innov. Regul. Sci.* 2020, *54*: 324–341.

9. Fleming, T.R. Addressing missing data in clinical trials. *Ann. Intern. Med.* 2011, *154*, 113–117.

10. Little, R.J.; D'Agostino, R.; Cohen, M.L.; Dickersin, K.; Emerson, S.S.; Farrar, J.T.; et al. The prevention and treatment of missing data in clinical trials. *N. Engl. J. Med.* 2012, *367*, 1355–1360.

11. Newgard, C.D.; Lewis, R.J. Missing data: How to best account for what is not known. *JAMA - J. Am. Med. Assoc.* 2015, *314*, 940–941.

12. Medicines Agency, E. *Committee for Medicinal Products for Human Use (CHMP) Guideline on the investigation of subgroups in confirmatory clinical trials*; 2019.

13. Assmann, S.F.; Pocock, S.J.; Enos, L.E.; Kasten, L.E. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000, *355*, 1064–1069.

14. Wang, R.; Lagakos, S.W.; Ware, J.H.; Hunter, D.J.; Drazen, J.M. Statistics in medicine - Reporting of subgroup analyses in clinical trials. *N. Engl. J. Med.* 2007, 22;357(21):2189-94

15. Brookes, S.T.; Whitley, E.; Peters, T.J.; Mulheran, P.A.; Egger, M.; Davey Smith, G. Subgroup analyses in randomised controlled trials: Quantifying the risks of false-positives and false-negatives. *Health Technol. Assess. (Rockv).* 2001, 5(33):1-56

16. Tanniou, J.; Van Der Tweel, I.; Teerenstra, S.; Roes, K.C.B. Subgroup analyses in confirmatory clinical trials: Time to be specific about their purposes. *BMC Med. Res. Methodol.* 2016, 18: 16-20

17. Rothwell, P.M. Treating individuals 2: Subgroup analysis in randomised controlled trials: Importance, indications, and interpretation. *Lancet* 2005, *365*: 176–186.

18. Schulz, K.F.; Grimes, D.A. Multiplicity in randomised trials II: Subgroup and interim analyses. *Lancet* 2005, *365*, 1657–1661.

19. Tyson, J.E.; Pedroza, C.; Wallace, D.; D'Angio, C.; Bell, E.F.; Das, A. Stopping guidelines for an effectiveness trial: What should the protocol specify? *Trials* 2016, *17*.

20. Consort-Statement > CONSORT 2010 > Interim analyses and stopping guidelines Available online: http://www.consort-statement.org/checklists/view/32--consort-2010/84-interim-analyses-and-stopping-guidelines (accessed on Apr 1, 2020).

21. Testa, M.A.; Simonson, D.C. Assessment of quality-of-life outcomes. *N. Engl. J. Med.* 1996, *334*: 835–840.

22. Williams, R.; Alexander, G.; Armstrong, I.; Baker, A.; Bhala, N.; Camps-Walsh, G.; et al. Disease burden and costs from excess alcohol consumption, obesity, and viral hepatitis: fourth report of the Lancet Standing Commission on Liver Disease in the UK. *Lancet* 2018, *391*: 1097–1107.

23. Weinstein, M.C.; Stason, W.B. Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. *N. Engl. J. Med.* 1977, *296*, 716–721.

24. Ramsey, S.; Willke, R.; Briggs, A.; Brown, R.; Buxton, M.; Chawla, A.; et al. Good research practices for cost-effectiveness analysis alongside clinical trials: The ISPOR RCT-CEA Task Force report. *Value Heal.* 2005, *8*: 521–533.

Table S1. Definition and diagnostic criteria of Refractory ascites.

Definition	
Diuretic-resistant ascites	Ascites that cannot be mobilized or the early recurrence of which cannot be prevented because of a lack of response to dietary sodium restriction and diuretic treatment.
Diuretic-intractable ascites	Ascites that cannot be mobilized or the early recurrence of which cannot be prevented because of the development of diuretic-related complications that preclude the use of an effective diuretic dosage.
Diagnostic criteria	
Treatment duration	Patients must be on intensive diuretic therapy (spironolactone 400 mg/day and furosemide 160 mg/day) for at least one week and on a salt-restricted diet of less than 90 mmol/day
Lack of response	Mean weight loss of <0.8 kg over four days and urinary sodium output less than the sodium intake
Early recurrence	Reappearance of grade 2 or 3 ascites within four weeks of initial mobilization
Diuretic-related complications	- Diuretic-related hepatic encephalopathy: development of encephalopathy in the absence of any other precipitating factor - Diuretic-related renal impairment: increase of SCr by >100% to a value >2 mg/dl (177 lmol/L) in patients with ascites responding to treatment - Diuretic-related hyponatremia: decrease of serum sodium by >10 mmol/L to a serum sodium of <125 mmol/L - Diuretic-related hypo- or hyperkalemia: change in serum potassium to <3 mmol/L or >6 mmol/L despite appropriate measures - Invalidating muscle cramps

Adapted from EASL clinical practice guidelines for the management of decompensated cirrhosis. J Hepatol. 2018 Aug;69(2):406-460.

SCr, serum creatinine.

Table S2. Definition and diagnostic criteria of Acute Kidney Injury (AKI) in cirrhosis.

Subject		Definition
Definition of AKI		Increase in SCr ≥0.3 mg/dL (≥26.5 µmol/L) within 48 hours; or, a percentage increase in SCr ≥50% from baseline which is known, or presumed, to have occurred within the prior 7 days
Stages of AKI	**Stage 1**	Increase in sCr ≥0.3 mg/dL (26.5 µmol/L) or an increase in sCr ≥1.5-fold to 2-fold from baseline. **-Stage 1A:** SCr at diagnosis <1.5 mg/dL **-Stage 1B:** SCr at diagnosis > 1.5 mg/dL
	Stage 2	Increase in SCr >2-fold to 3-fold from baseline.
	Stage 3	Increase of SCr >3-fold from baseline or sCr ≥4.0 mg/dL (353.6 µmol/L) with an acute increase ≥0.3 mg/dl (26.5 µmol/L) or initiation of renal replacement therapy.
Regression of AKI		Regression of AKI to a lower stage.
Progression of AKI		Progression of AKI to a higher stage and/or need for RRT.
Response to treatment		**No response:** no regression of AKI
		Partial response: Regression of AKI stage with a reduction of sCr to ≥0.3 mg/dL (26.5 µmol/L) above the baseline value.
		Complete response: Return of sCr to a value within 0.3 mg/dL (26.5 µmol/L) of the baseline value.

Adapted from EASL clinical practice guidelines for the management of decompensated cirrhosis. J Hepatol. 2018 Aug;69(2):406-460.

RRT, renal replacement therapy; SCr, serum creatinine.

Table S3. Diagnostic criteria of Hepatorenal syndrome - AKI (HRS-AKI).

Diagnostic criteria HRS-AKI
- Diagnosis of cirrhosis and ascites. - Diagnosis of AKI according to ICA-AKI criteria. - No response after 2 consecutive days of diuretic withdrawal and plasma volume expansion with albumin (1 g per kg of body weight). - Absence of shock. - No current or recent use of nephrotoxic drugs (NSAIDs, aminoglycosides, iodinated contrast media, etc.). - No macroscopic signs of structural kidney injury, defined as: o absence of proteinuria (>500 mg/day) o absence of microhaematuria (>50 RBCs per high power field), o normal findings on renal ultrasonography

Adapted from EASL clinical practice guidelines for the management of decompensated cirrhosis. J Hepatol. 2018 Aug;69(2):406-460.

NSAIDs, non-steroidal anti inflammatory drugs; RBCs, red blood cells.

Table S4. Definitions of bacterial infections and clinical/microbiological criteria of resolution of bacterial infections in cirrhosis.

Type of infection	Definition	Resolution
Spontaneous bacterial peritonitis	PMN cells count in ascitic fluid≥250/µL in absence of an evident intra-abdominal surgically treatable source	Reduction of PMN count in ascitic fluid > 25% of baseline after 48 hours AND Normalization of PMN count (<250 cells/µL) after 5-7 days
Urinary tract infections	At least one among fever>38°C; urgency; frequency; dysuria; suprapubic tenderness; worsening of consciousness AND positive urine culture (>10^5 CFU per cc of urine) OR At least 2 among fever>38°C; urgency; frequency; dysuria; suprapubic tenderness; worsening of consciousness AND at least one among: a) pyuria (urine specimen with ≥ 10 white blood cells/mm3); b) positive dipstick for leukocytes and/or nitrates	Resolution of the symptoms of urinary tract infection AND Demonstration that the bacterial pathogen found at diagnosis of infection is reduced to fewer than 10^3 CFU/mL on urine culture Time of assessment of resolution 5-7 days
Pneumonia	Chest X-ray findings (new or progressive pulmonary infiltrate, consolidation or cavitation) AND At least one sign/symptom (new onset of purulent sputum or change in character of sputum; new onset of cough, dyspnea or tachypnea; rales or bronchial breath sounds; worsening of gas exchange) AND At least one among: a) Fever ≥ 38°C b) Leukocytosis(>12,000/mm3) or leukopenia (<4,000/mm3) c) Altered mental status	Clinical resolution of signs and symptoms of infection Time of assessment of resolution: 7-10 days

Bloodstream infections*	One positive blood cultures showing a noncommon skin contaminant OR At least 2 positive blood cultures showing a common skincontaminant (e.g. coagulase negative staphylococcus, micrococcus, diphtheroid, Bacillus species, etc.) drawn from separate sites and/or on separate occasions.	Combination of survival, resolution of fever and symptoms related to BSI source, stable or improved CLIF-SOFA score and negative blood cultures Time of assessment of resolution: 7-10days

Fig. S1. Algorithm for the use of Quick Sequential Organ Failure Assessment score (qSOFA) and Sepsis-3 criteria for the evaluation of severity and prognosis of bacterial infections in cirrhosis. *(Adapted from EASL clinical practice guidelines for the management of decompensated cirrhosis. J Hepatol. 2018 Aug;69(2):406-460).*