Fluid REStriction in Heart failure versus liberal fluid Uptake: the FRESH-UP study

Summary of changes; Appendix

Protocol ID	NL75112.091.20 / NCT04551729
Short title	FRESH-UP
Appendix A	Comments and rebuttal letter Journal of Cardiac Failure

Appendix A: Comments and rebuttal letter Journal of Cardiac Failure

Fluid REStriction in Heart failure versus liberal fluid UPtake: Rationale and design of the randomised FRESH-UP study.

We thank the statistical reviewer for the critical appraisal of our manuscript. In response to the comments, we adjusted the manuscript and feel that the quality has improved substantially. Please find below a point-by-point response to the reviewer' comments.

Statistical Reviewer Comments:

Summary Statement:
The authors submitted a paper that provided the rationale and design of the FRESH-UP study, a randomized, controlled, open-label trial that seeks to determine the effectiveness and safety of liberal fluid intake vs. fluid restriction on QoL in outpatient HF patients over a 3-month period.

Major Concerns:
1. Some correction for multiplicity of hypothesis testing on the two primary endpoints (KCCQ-OSS and TDS-HF) is needed in the statistical analysis plan. In the last paragraph of the Sample Size Calculation section, it initially appears that the authors are proposing hierarchical testing of the two endpoints. However, the authors write "In the case of non-significant findings for the first primary endpoint (KCCQ), the second primary endpoint will be tested at the 2-sided 5% significance level but will be regarded as an exploratory endpoint, and a nominal p-value will be produced." There are a few concerns. First, this deviation is not consistent with the hierarchical testing method used to protect inflation of the type I error rate. Second, TDS-HF is a primary endpoint and should not be switched to an exploratory endpoint after results for KCCQ-OSS become known. Third, how would such findings be communicated to the public? Please provide adequate control of type I error rate inflation. If adequate control cannot be provided, please consider TDS-HF as a key secondary endpoint. If KCCQ-OSS and TDS-HF are co-primary endpoints, then no multiplicity adjustment is needed since one would need to succeed on both endpoints in order to consider the intervention a success.
We thank the reviewer for the critical appraisal, which perfectly reflects the discussions we had during the design phase of the study. In line with the reviewer's suggestion, we have adjusted the primary and secondary endpoints. In the revised manuscript, the KCCQ-OS is the primary endpoint. The TDS-HF has been changed to a key secondary endpoint (abstract: page 3, line 14-15, manuscript: page 4, line 8 / page 5, line 22-24 / page 7, line 19 / page 8, line 15-17 and the Visual Take Home Graphic).

The KCCQ is a well validated and widely accepted questionnaire that can quantify the QoL (and its changes) in the heart failure population. Although, the TDS-HF is also validated and may quantify an important symptom as thirst. The questionnaire reflects only one aspect of the study and it may not be as established as the KCCQ.

We based our sample size calculation on the KCCQ-OSS, with adequate power and with just KCCQ-OSS as our primary endpoint there is no inflation of the type I error rate. Moreover, with simplification of the protocol with only KCCQ-OSS as the primary endpoint, the interpretation of the results of the trial will be much clearer for the public.
Accordingly, we feel that the quality of the study design and manuscript have improved due to your valuable suggestion.

2. The authors state that the intention-to-treat principle will be used for the primary analysis (re: Analysis populations section). However, the authors also state that "For the primary analysis, only those patients of whom both baseline and follow-up are available will be analysed." Written as such, the analysis population for the primary analysis is not the intention-

to-treat, since it does not include all randomized subjects. It is in fact a complete-case analysis. Please modify.
In accordance with the reviewer's rightful remark, we consider it important to include the ITT population in the primary analysis. We adapted the manuscript accordingly (page 7, line 21-22).

Therefore, we have proposed an analysis for which any missing baseline values are assumed to be missing completely at random and any missing follow-up values are assumed to be missing at random. For this, multiple imputation will be used to handle missing baseline, however unexpected, and outcome data on the KCCQ-OSS. We would like to refer to the answer on the next comment regarding handling missing QoL measurements.

3. The authors also state that sensitivity analyses will be performed to assess the impact of missing primary endpoints and suggested using multiple imputation (MI). There are a few concerns. First, MI assumes missing at random mechanism and missing QoL measurements are notoriously known to be missing not at random. Sensitivity analyses for non-ignorable missingness should be developed. Second, missingness due to death is not a missing data problem, since the outcome values do not exist and so, one should be cautious about using methods to impute these unobserved values. Third, the authors proposed single imputation for missing baseline scores. Why not multiple imputation for baseline scores? With MICE, one can develop a model for baseline scores and another model for 3-month scores that are dependent on the baseline scores (plus other auxiliary variables).
The reviewer has raised some important considerations related to the challenges related to the analysis of QoL data with missing values, and we agree with all of these concerns.
We do not expect any missing baseline QoL measurements because they must be completed before randomisation and are filled out in a digital manner in which completion is mandatory and entry errors are impossible. Besides, due to the method of scoring, the KCCQ will always provide an Overall Summary Score, based on a mean of the available other summary scores. For the unlikely case where a baseline value is missing, we have taken on the advice of the reviewer and proposed that multiple imputation is also used for these values (page 7, line 22-27).

Regarding the missing QoL measurements at follow up, it is agreed that the assumption that these data are missing at random is a strong assumption, and agree that sensitivity analyses for which these data are assumed to be missing not at random should be performed. We have now specified these in the protocol (page 8, line 3-6).
It is agreed that missing data due to death is not a missing data problem, however particularly for the KCCQ, which is predictive of clinical outcomes including death, excluding patients from the analysis who died is likely to bias the results. Therefore for our primary and sensitivity analyses we would like to retain the patients who died in the analysis. We will ensure that this issue is part of our discussion and keep an eye out for any emerging methodological research that may help to resolve this issue while allowing us to address the research question of interest.

4. The sample size calculation accounts for a 10% drop-out rate, increasing N from 454 subjects to 498 subjects. Can the authors clarify how 498 subjects were obtained? Applying the standard LTF adjustment, one gets around 506 patients (after rounding up).
We thank the reviewer for his/her suggestion and applied the LTF adjustment as suggested. We performed the calculation: $N = n / (1-(z/100)) \rightarrow N = 454 / (1-(10/100)) = 504.4$. Therefore, we adjusted the sample size to 506 subjects to maintain an equal distribution of subjects across both study arms (manuscript: page 9, line 12 / DSMB Charter: page 1, page 6, page 10 / Figure 1 and the Visual Take Home Graphic).

5. In the Data Safety Monitoring Board section, the authors write "The DSMB has two mandates. First, to perform two interim analyses when data is available of 33% and 66% of the patients for safety on the occurrence of the clinical endpoints....." This is not completely accurate since the DSMB charter lists "Early stopping due to harm of a treatment or clear

superiority" as one of the possible recommendations after an IA. Therefore, the IA consists of both efficacy and safety assessments especially since KCCQ scores (the primary endpoint) is listed as one the main parameters of interest for the DSMB. Since interim looks for efficacy on the primary endpoint will increase the type I error rate of a clinical trial, please clarify how the type I error rate will be controlled. Furthermore, it may no longer be the case that the significance level for the final analysis of the primary endpoint(s) will be at 0.05 (re: "Both primary hypotheses on QoL by the KCCQ-OSS and TDS-HF will be tested at a 2-sided 5% significance level.)

We agree with the reviewer that there is a discrepancy between the original manuscript and the DSMB charter. To be clear, the objective of the DSMB is to solely assess safety. Accordingly, we have adjusted the DSMB charter. There will be no interim analyses on efficacy. We have removed the KCCQ scores as parameters of interest from the DSMB charter to avoid any misunderstanding in this regard. In addition, we have adjusted the possible recommendation "Early stopping due to harm of a treatment or clear superiority" in the DSMB charter as well. The DSMB will assess safety based on clinical event rates. We thank you for addressing this inconsistency in our documents and feel that the revised version of the DSMB charter provides more clarity.

Since no interim analyses on efficacy will be performed, there is no increase in type I error rate and no additional measures are required to control type I error rate.

6. Based on prior experience, do the authors have a sense of cross-over rates (re: Discontinuation of study treatment section) between study arms? Why wasn't cross-over accounted for in the sample size calculation?

During the design phase of study we expected the rates of discontinuation of study treatment to be very low. In our experience, participants are highly motivated to adhere to the study protocol. As expected, most patients already adhere to the control therapy and are excited to particate in the interventional group. The drop-out rate of the control group is very low and there are, till further notice, no drop-outs in the interventional group. Our objective is to study the effect of a fluid advice in a real-world situation and a pragmatic design was chosen in which the cross-over rate was assumed neglectable. In addition, because the primary analysis is based on the intention-to-treat population we are essentially specifying a treatment policy estimand approach for the primary analysis, in which outcomes are measured regardless of whether patients divert from their advised fluid intake or not.

7. The authors acknowledged in the Limitations that patient-reported fluid intake at week 6 may not be reliable and yet this will form the basis of the per-protocol analysis. I worry about the utility of the per-protocol analysis if there will be misclassification on adherence. Is it feasible to record fluid intake daily? Furthermore, regardless of misclassification, per-protocol is a biased analysis and the authors may wish to consider causal inference methods (e.g., instrumental variable analysis) to address non-adherence.

We agree with the reviewer that the true fluid intake cannot be determined with full certainty. In that regard, we feel it is important to state that our primary analysis is based on an intention-to-treat principle. Our per-protocol analysis is only considered as an additional supporting analysis, but is indeed quite limited due to uncertainty on actual fluid intake.

It is essential to note that one of the main negative effects of a fluid restriction on QoL is the fact that patients need to monitor this on a day-to-day basis. The major upside of liberal fluid intake is the fact that there is no need to monitor this every day. Therefore, we do not want the patients in the intervention arm of liberal fluid intake to measure their intake for the whole study period. In addition, reporting daily fluid intake may even be more troublesome for patients which may lead to an increase in incorrect documentation of the actual fluid intake. Thus, we chose to ask patients for a dedicated week rather than the full study period.

We agree that a per-protocol analysis is biased given the potential for imbalance between the treatment arms and will explore alternative options for inclusion in the Statistical Analysis Plan. We thank the reviewer for referring us to causal inference methods and will explore this as an option.

<u>Minor Concerns:</u>

We thank the reviewer for the attention to details. We have changed the manuscript according to the remarks above.

1. Under Patients, the authors state that there are currently four participating centres. Please update the DSMB Charter as it notes 2 sites.
This has been corrected to 4 sites (DSMB Charter: page 1).

2. I assume that the composite clinical endpoint and AKI will be assessed at both 3 & 6 months and not only 6 months - is this correct? If so, please make clear in Outcome section.
The reviewer is correct and the manuscript has been adjusted accordingly (page 6, line 4).

3. Under Study Organisation, missing a word between "secondary" and "will": The review and adjudication of all secondary will be conducted by an independent event adjudication committee. This has been adjusted in the revised version of the manuscript (page 10, line 2).