

SUPPLEMENTARY TEXT S3**SUMMARY AND EXPLANATION OF DEVIATIONS FROM PROTOCOL AND STATISTICAL ANALYSIS PLAN**

- (i) Final follow-up assessments originally scheduled at week 24 could not be conducted by home visit due to change in administrative regulations regarding Covid-19, but had to be performed at the participating hospitals. Consequently, follow-up assessments could not be completed at 24 weeks as planned but were conducted at 28 weeks after baseline on average. *Explanation:* Due to a new policy of the National Health Commission of the PR China announced in the time period between post-treatment and planned follow-up assessments, home visits of patients were no longer permitted for health professionals. Instead, assessments could only be performed after the patients provided negative results of serum antibody and nucleic acid tests for SARS CoV-2. For the final assessment, participants were thus invited to return to the hospitals where they had originally received treatment. Free of charge serum antibody and nucleic acid tests for SARS CoV-2 were provided there. As new appointments for hospital visits had to be scheduled with the patients, this adjustment led to a delay in the final assessment point for about four weeks on average.
- (ii) The primary outcome 6MWD at post-treatment assessment was analysed simultaneously with the secondary outcome 6MWD at follow-up, i.e. a joint model was fit as opposed to two separate models. *Explanation:* This was done because two participants who had missed the post-treatment assessment returned for the follow-up assessment and using this information for model estimation was considered important in light of the principles of ITT analysis.
- (iii) Perceived dyspnoea (favourable outcome) as assessed with the modified Medical Research Council (mMRC) scale was analysed with log-linear Poisson regression with cluster-robust standard errors instead of longitudinal logistic regression. *Explanation:* The Poisson model was chosen over a longitudinal logistic model originally specified in the SAP due to ease of interpretation of rate ratios and derivation of population-averaged probabilities.
- (iv) For sensitivity analysis employing multiple imputation with chained equations based on an extended MAR assumption, models were estimated on 70 imputed data sets instead of the originally-specified 50 data sets. *Explanation:* This was done because multiple imputation for 50 data sets did not yield a satisfactory upper limit for the fraction of missing information for at least one of the estimated models¹.

Deviations from the SAP (ii-iv) had no effect on the statistical significance of estimates when compared with analyses as originally planned.

References

1. von Hippel PT. How Many Imputations Do You Need? A Two-stage Calculation Using a Quadratic Rule *Sociological Methods and Research* 2020;49(3):699-718.