Statistical Analysis Plan

APRIL Trial

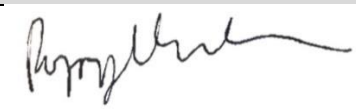
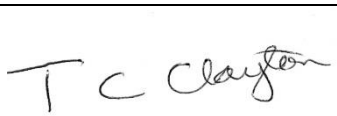
Study title: Clinical trial of Ashwagandha for Promoting Recovery in Long COVID

SAP Author: Poppy Mallinson

SAP Version number: 1.0

APRIL Protocol version: 4.1

FINAL VERSION APPROVED BY:

Name	Signature	Date
Poppy Mallinson (author)		18/07/2025
Sanjay Kinra (Chief investigator)		18/07/2025
Tim Clayton (Lead statistician)		21/07/2025

CURRENT VERSION APPROVED BY (IF CHANGED FROM FINAL):

Name	Signature	Date

Contents

1 Introduction

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for the APRIL Trial. APRIL is a multicentre randomised placebo-controlled trial of Ashwagandha, a popular herbal medicine, for improving functional status and symptoms in people diagnosed with Long COVID in the UK.[1]

The APRIL Trial initially planned to recruit up to 2500 patients via a remote recruitment model, however this protocol was not approved by the regulator (MHRA) so it was agreed that the trial would recruit through NHS GP surgeries until sufficient safety data was gathered (at least 100 patients) to support modifying the monitoring protocol to allow faster recruitment. However, it took longer than anticipated to recruit the first 105 patients by which time the Long COVID patient numbers, NHS support for Long COVID services and other factors had changed considerably. It was therefore decided by the Trial Steering Committee that the trial should be halted with the initial 105 patients recruited, and trial data analysed, to inform decisions around the design of a future trial and inclusion of the most appropriate individuals. The original set of analyses will still be conducted as laid out in the protocol although the primary outcome analysis is expected to be underpowered (based on the effect size assumptions in the original power calculation). The safety outcome analyses are anticipated to be particularly informative, given the absence of quality clinical data on the safety profile of Ashwagandha for this patient population. This may help to inform the general public regarding the safety of their usage of Ashwagandha (which is a popular over-the-counter supplement) as well as the design of future randomised trials of this product.

2 Study Objectives

Primary – to determine the effectiveness of Ashwagandha tablets (3-month course) for improving functional status at 3 months (measured by the Post-COVID-19 Functional Status Scale) among people experiencing ongoing symptoms of COVID-19.

Secondary – to determine the safety of Ashwagandha tablets (3-month course) and effectiveness for improving quality of life, reported fatigue, breathlessness, pain, sleep quality, mental health, cognitive function, work status and other symptoms at 3 months.

3 Study Design

3.1 Overall design

This is a randomised double-blind placebo-controlled trial. Interested individuals who have a diagnosis of Long COVID (and conform to all other eligibility criteria) were randomised to take Ashwagandha or placebo for 3 months (delivered at home by post). Participants were recruited through participating GP practices or NHS Long Covid Clinics (hereafter referred to as clinical trial sites). Follow-up for functional status, quality of life and other self-reported outcomes was by online or postal surveys (as per participant preference) conducted monthly. Follow-up for

safety was through clinical assessment by the study investigator's clinical team, conducted monthly.

3.2 Sample size

The original sample size planned for the trial was 2500. The justification for this given in the study protocol is as follows: "Published data from an online survey of Long COVID patients indicate that patients' grade distribution on the Post-Covid Functional Scale (scale 4 (severe) to 0 (none)) improved from of 3.4, 49.4, 35.1, 7.1 and 5.0% at 3 months to 2.1, 43.9, 35.1, 10.5 and 8.4% at 6 months[2]. If Ashwagandha can improve symptoms to a similar extent at 3 months, this translates to an overall odds ratio of 0.71 on the ordinal logistic regression analysis (with an odds ratio less than one indicating a shift to a lower grade of symptoms). In order to improve symptoms with a more modest odds ratio of 0.75, a sample size of 2500 gives in excess of 90% (or 80% power to detect an even more conservative odds ratio of 0.80). An odds ratio of 0.75 would result in the percentages in each grade (grades 4 to 0) being 2.6, 43.1, 38.8, 9.0 and 6.6% in the Ashwagandha arm at end-line. The primary analysis will include adjustment for the baseline value of the primary outcome scale which is expected to increase the power further."

This sample size was proposed initially alongside a more simplified trial protocol in which Long COVID patients registered their interest online, were randomised online, and then posted the medications directly from the trial coordinating centre. However this model (which would have allowed much more rapid recruitment) was not permitted by the regulator (MHRA) who required clinician oversight of each patient including regular liver function monitoring. The MHRA-directed approach was adopted, which involved modifying the protocol to recruit and monitor through NHS GPs. It was agreed with the MHRA that after safety and effectiveness data from at least ~100 patients' had been gathered, we would submit the safety information and apply to relax the monitoring procedures to speed up recruitment. This initial safety-focussed phase of the trial has been completed although we opted not to continue the trial because of barriers faced in identifying and recruiting Long COVID patients. The trial is therefore expected to be underpowered for the primary outcome, although they will still be presented in adherence with the pre-specified protocol, and emphasis will be placed on the safety-related trial outcomes.

3.3 Randomisation

In this trial we conducted blinded web-based randomisation in 1:1 ratio after initial consent. Block randomisation (randomly permutated blocks) with varying block size were used to ensure balance throughout the recruitment period.

3.4 Study Assessments

3.4.1. Trial outcomes

Primary outcome measure

Self-reported functional status measured using the Post-COVID-19 Functional Status Scale (PCFS), at 3 months[2]. Functional status is a key patient-centred indicator which measures the extent to which a condition limits a person's ability to conduct their usual activities. The PCFS has gained widespread popularity as a tool for measuring self-assessed functional status in Long COVID in trials and observational studies and has been translated into multiple languages. It was based on a similar, widely-used functional scale for stroke patients, and has been

validated in Long COVID patients[3]. The PCFS asks patients to consider a flow-chart of options that describe the extent of limitations to their daily activities over the past week. Based on the flow chart, they assign themselves to a grade between 0 (no functional limitations) to 4 (severe functional limitations).

Secondary outcome measures (all at 3 months)

- Quality of life by PROMIS 29+2 summary score,
- Self-reported fatigue, pain, sleep, anxiety, depression and social, physical and cognitive function by PROMIS 29+2 individual dimensions
- Self-reported breathlessness by modified MRC Dyspnoea Scale
- Health utility by EQ 5D-5L utility score
- Work status and productivity
- Other self-reported symptoms

Outcomes domains were selected to cover the most relevant key indicators for patients and policy makers as well as the most common symptoms of Long COVID. Outcome questionnaire instruments were chosen following thorough review of the literature and other ongoing studies, on the basis of wide use/validation, short length (to reduce participant burden), and relevance for this study population (ideally used successfully in Long COVID patients before). For outcome scores that required derivation, the approach used in the primary publication will be applied (PROMIS scores[4] and EQ 5D-5L utility score[5] using the most recent UK-based value set).

Safety outcomes

- SUSARs
- Severe adverse events (including pregnancies)
- All adverse events
- Adverse events related to liver function abnormalities
- Liver function test parameters

3.4.2 Follow-up for trial outcomes

Follow-up was conducted through a brief monthly online or postal survey of self-reported functional status, alongside secondary outcome measures such as quality of life, fatigue, breathlessness, pain, mental health, sleep quality, work status, and other symptoms, and brief questions on intervention adherence.

Participants were followed-up for 3 months (i.e., end of month 1, end of month 2, end of month 3). Automated email or text reminders were sent as per participant preference. In case participants were unable or unwilling to complete surveys online, paper questionnaires in pre-paid return envelopes were posted each month along with the supplement/placebo packages.

For the purposes of analysis, the target outcomes of interest are at 3 months, and this will be operationalised as the outcome data available nearest to completion of 90 days of trial medication, within a maximum of 30 days either side of this date. This means, for example, if both the second month and third month survey were delayed by 3 weeks, the second month survey will be used as this will be closer to the target outcome date of day 90.

3.4.3 Clinical monitoring

The investigator's clinical team assessed all participants at least monthly and conducted symptom-guided assessment as described in NICE Long Covid guidelines. Assessments took place at end of 30 days (±3 days, 1st interim assessment), 60 days (±3 days, 2nd interim assessment) and 91 days (final end of trial assessment) since treatment initiation. At each assessment, they actively elicited the participant's history related to a) any worsening of their Long covid, b) exacerbation of any other pre-existing conditions, and c) any side effects that might be related to the drug (including AE/SAE assessment if applicable). To guide assessment of c), we provided the clinical teams with full list of possible of side effects that have been observed for the trial product (regardless of whether causally attributed). Monthly assessments were conducted remotely or in-person, as per preference of the investigator or participant. Following the initial assessment, further in-person follow-up including any indicated tests and investigations were sought if needed based on participant's symptom assessment, and as guided by relevant local and national guidelines. In particular we highlighted the following areas of assessment: cardiac chest pain, symptoms suggestive of liver or kidney problems, and worsening of psychiatric symptoms.

In addition to these monthly clinical assessments, liver function tests (AST, ALT and bilirubin) were conducted at 30 days (i.e., along with first interim clinical monitoring assessment) and 91 days (i.e., at the end of trial assessment). The test at 30 days was intended to check for any early acute liver toxicity associated with the trial medication, while the test at 91 days was intended to assess for any longer-term effects associated with cumulative build-up of the product in the system. Additional liver function tests could be ordered by the study investigator at any point as deemed clinically indicated for the patient. All WOCBP had to complete an urine pregnancy test monthly during the trial. These tests were done after completing 1st and 2nd month of trial treatment, and after completing 3rd month's treatment at the final end of study assessment, to confirm that no pregnancies occurred during the trial treatment phase.

4 Statistical Analysis approach

4.1 General

Statistical analysis will be performed as pre-specified in this SAP. Any, post-hoc, exploratory analyses completed to support planned analyses, which were not identified in this SAP, will be documented and reported in the relevant trial reports. Any results from unplanned analyses will be clearly identified in the text of the trial reports.

4.2 Time points for analysis

All final, planned, analyses identified in the protocol and in this SAP will be performed only after the last patient has completed their 90-day follow-up. In addition, no database may be locked, randomisation code unblinded, or analyses completed until this SAP has been approved.

4.2.3 Interim analyses

There were no planned or conducted interim analyses for this trial. The Independent Data Monitoring Committee examined safety data stratified by trial arm (not unblinded) on a 3-monthly basis as part of their safety monitoring. No member or attendee of the closed part of the Data Monitoring Committee has any input into the content or writing of this SAP.

4.3 Pooling of sites

Data from each participating trial site (GP or Long COVID Clinic) will be pooled for all analyses.

4.4 Definition of study populations for analysis

4.4.1 Intention-to-treat

The primary analyses will be on an intention-to-treat basis, that is, participants will be considered to belong to the trial arm that they were assigned to (even if they did not end up completing the treatment).

4.4.2 Per protocol

Additional analyses on a per-protocol basis will also be conducted, in which participants will be stratified into groups in terms of whether they took the trial treatment or not. The following groups will be analysed:

- Adherers: People who took at least 50% of the intended dose of treatment for at least 60 days (2 months) out of the total 90 days of treatment.
- Non-adherers: All others.

4.5 Methods for handling withdrawals and missing data

There were two types of withdrawal in the trial – complete withdrawal and partial withdrawal.

- Complete withdrawal was when patients discontinued trial medication AND discontinued completing the monthly outcome questionnaires.
- Partial withdrawal was when patients discontinued trial medication but participants continued to complete the monthly outcome questionnaires.

Patients who withdrew from trial medication but continued to participate in follow-up will be analysed according to the intention-to-treat and per protocol definitions described above (4.4).

Complete withdrawals (people who did not complete an endline survey or clinical monitoring), or people who have endline data missing for any other reason, will be excluded from the primary analysis of outcomes at 3 months. Outcome data will not be imputed in the primary analysis models. Additional analyses will be undertaken using data from the 1st, 2nd and 3rd months' follow-up only to explore how timing of outcome assessment influenced the results (described in section 8.2 below).

4.6 Statistical analysis software

Statistical analysis will be conducted using R or Python Software and the version will be stated in the study report. Statistical analysis code will be published along with the study report.

5 Descriptive analyses

Continuous variables will be described by the means and standard deviation except for skewed variables which may be described by the median and inter quartile range (IQR). Categorical

variables will be described by frequency and percentages in each category by treatment group (see appendix table 1).

Description of the follow-up of participants by trial arm (i.e. number completing months 1, 2 and 3 medication and surveys) will be presented graphically in a flow chart.

Description of adherence to medication and timeliness of survey completion and clinical monitoring by trial month (for participants not completely withdrawn or lost) will be tabulated (see appendix table 2).

6 Evaluation of patient-reported outcomes

6.1 Primary outcome

The primary analysis of the primary outcome will be an intention-to-treat complete case analysis at 3-month follow-up, adjusted for baseline level of the same outcome. Post-Covid Functional Status Scale will be analysed using ordinal logistic regression. No other adjustments will be made for the primary analyses, although additional secondary analyses adjusting for age (above or below median) and gender will be conducted (see section 8.4 below).

6.2 Secondary outcomes

Due to the large number of secondary outcomes, the PROMIS-29+2 summary score and EQ-VAS score will be considered as 'key secondary outcomes' and only these (along with the primary outcome) will be taken forward into the additional analyses (section 8). For all secondary outcomes, the main analysis will be intention-to-treat complete case analysis at 3-month follow-up, and for continuous and ordinal outcomes these will be adjusted for baseline level of the same outcome. The statistical models that we plan to use for each outcome are listed below. Distributional checks will first be conducted to ensure that key assumptions for each planned model type are met, and if for any given outcome there are violations that threaten the validity of the proposed model type, that outcome will be binarized by the median value and analysed with binary logistic regression.

Ordinary least squares (linear) regression:

- Quality of life by PROMIS 29+2 summary score
- Self-reported fatigue by PROMIS 29+2 fatigue dimension
- Self-reported sleep disturbance by PROMIS 29+2 sleep dimension
- Self-reported anxiety by PROMIS 29+2 anxiety dimension
- Self-reported depression by PROMIS 29+2 depression dimension
- Self-reported social participation by PROMIS 29+2 social function dimension
- Self-reported physical function by PROMIS 29+2 physical function dimension
- Self-reported cognitive function by PROMIS 29+2 cognitive function dimension
- Self-reported pain interference by PROMIS 29+2 pain dimension
- Self-reported pain intensity by PROMIS 29+2 0-10 pain rating scale
- Health utility by EQ 5D-5L utility score
- EQ VAS score
- Self-reported work productivity by 0-10 scale on the work questionnaire (people out of work scored to 0)

Ordinal logistic regression:

- Self-reported breathlessness by modified MRC Dyspnoea Scale
- Other self-reported symptoms – number (count) of other symptoms experienced (grouped into 5 groups according to the distribution)

Binary logistic regression

- Whether any hours of work were missed in the past 7 days (yes/no) (people out of work set to no)
- Other self-reported symptoms – binary presence/absence of each of the 22 other symptoms asked (only including in the table those with at least 10 individuals reporting that symptom at baseline). If any other self-reported symptom (from the free-text field) was reported by at least 10 people at baseline, this can also be added.

Dummy table illustrating the presentation of the primary and secondary analysis results is given in appendix table 3.

7 Evaluation of safety outcomes

A table will be presented with counts of the following adverse safety outcomes, by trial arm (appendix table 4):

- SUSAR
- Serious adverse event (SAE)
- Pregnancy
- Adverse reactions involving abnormal liver test results (adverse event designated as probably or definitely related to trial medication)
- All other adverse reactions
- Adverse events (designated possible or unlikely related to trial medication) involving abnormal liver test results
- All other adverse events (designated possible or unlikely related to trial medication)

Further description and discussion of any SUSAR, SAE, pregnancy, adverse reactions and abnormal liver test results will be provided in the text (as the numbers of these are anticipated to be few). For each abnormal liver test result, the status of liver function at baseline will be mentioned, as abnormal liver function at baseline is expected to be linked to abnormal results during follow-up (although numbers are expected to be too small to formally test this). All adverse events will be broken down into specific categories (e.g. headache, nausea, etc) by trial arm in an appendix table. We will also present adverse events by trial arm stratified by gender in the same appendix table.

Changes over time in liver function test (LFT) results by arm will be described through data visualisation. Specifically, we will plot the mean and standard error values of AST, ALT and bilirubin at baseline, month 1 and month 3 by trial arm. T-test for the differences in each LFT will be calculated at each timepoint, crudely and with adjustment for baseline level. The corresponding numerical results will be given in an appendix table.

8 Additional analyses

The following additional analyses will only be applied to primary and key secondary to avoid multiplicity of analysis results. Apart from the specific aspects mentioned, each analysis will be the same as described in section 6.

8.1 Per protocol analyses

A per protocol analysis will be conducted using the per-protocol definition provided above (4.4.2) to define the treatment groups (rather than randomisation arm).

8.2 Different time-points

The following additional analyses will be conducted to explore the impact of Ashwagandha at different time-points:

- Analysis of the study outcomes at 1-month follow-up
- Analysis of the study outcomes at 2-month follow-up
- Analysis of the study outcomes at 3-month follow-up

We will also conduct a mixed-effect analysis treating each month of medication as a separate observation (i.e. dataset with one row per month), using the baseline value at the start of that month and endline value at the end, with observations clustered within participants using a multilevel model with participant-level random intercepts. This will increase the power of the analysis assuming that the trial medication will have measurable effect after 30 days of treatment, which some previous trials support[6].

In addition to these analyses, description of the outcome data by arm and time point (baseline, month 1, month 2, month 3) will be presented in graphical format.

8.3 Sub-group analyses

The following subgroups analyses were specified in the original study protocol, but will only be conducted if there are at least 10 participants falling into the minority category at baseline. They will be severely underpowered so are exploratory and will mostly be informative for describing safety outcomes across subgroups. Statistical models will be run as in the primary analyses but with interaction terms fitted for these factors and p-values for interaction will be calculated.

- Gender (male/female)
- Ethnicity (white/non-white)
- Symptom severity at enrolment (PCFSS 3+ or <3).
- Symptom duration at enrolment (<1.5 years or 1.5+ years).

8.4 Co-variate adjustment

We will present additional analyses adjusted for age (above/below median) and gender, in addition to baseline level of the outcomes. This will only be done if the models are sufficiently stable to accommodate additional covariates and model assumptions continue to be met.

References

1 Mallinson PAC, Joshi M, Mathpathi M, *et al*. Ashwagandha (Withania somnifera (L.) Dunal) for promoting recovery in long covid: protocol for a randomised placebo-controlled clinical trial (APRIL Trial). *BMJ Open* 2025; **15**: e094526.

2 Klok FA, Boon GJAM, Barco S, *et al*. The Post-COVID-19 Functional Status scale: a tool to measure functional status over time after COVID-19. *Eur Respir J* 2020; **56**: 2001494.

3 Machado FVC, Meys R, Delbressine JM, *et al*. Construct validity of the Post-COVID-19 Functional Status Scale in adult subjects with COVID-19. *Health Qual Life Outcomes* 2021; **19**: 40.

4 PROMIS. https://www.healthmeasures.net/explore-measurement-systems/promis (accessed July 9, 2025).

5 Herdman M, Gudex C, Lloyd A, *et al*. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res Int J Qual Life Asp Treat Care Rehabil* 2011; **20**: 1727–36.

6 Tandon N, Yadav SS. Safety and clinical effectiveness of Withania Somnifera (Linn.) Dunal root in human ailments. *J Ethnopharmacol* 2020; **255**: 112768.

APPENDIX

This appendix contains dummy tables corresponding to the main analyses described in the SAP. The exact format and presentation of the tables may vary in the final reports.

Appendix Table 1: Dummy table for baseline description of study participants, APRIL Trial.

Baseline characteristic		Intervention arm (N=)	Control arm (N=)
Age – mean (SD)			
Gender – N (%) female			
Ethnicity – N (%) White British			
Education – N (%) completed university			
Work status – N (%) in work			
Current smoker – N (%)			
Co-morbidities	Diabetes – N (%)		
	Hypertension – N (%)		
	Lung condition – N (%)		
	Other – N (%)		
Duration of Long COVID	<3 months - N (%)		
	3-12 months - N (%)		
	12+ months - N (%)		
PCFSS at baseline	1 – negligible limitations		
	2 – occasional limitations		
	3 – some limitations		
	4 – severe limitations		
Quality of life by PROMIS 29+1 at baseline – mean (SD)			
Quality of life by EQ-VAS at baseline – mean (SD)			
Symptom burden (count) at baseline – mean (SD)			

SD is standard deviation. PCFSS is post-COVID functional status scale.

Appendix Table 2: Dummy table for medication and protocol adherence among those initiating a given month medication (i.e. if someone withdrew half-way through month 2, they would be counted in the month 2 column but not month 3)

Medication and protocol adherence	Intervention arm (N=)			Control arm (N=)		
	M1	M2	M3	M1	M2	M3
N participants on medication at start						
Mean % monthly tablets taken (SD)						
Mean days after end of month that patient survey was completed (SD)						
Mean days after end of month that clinical monitoring was completed (SD)						

Appendix Table 3: Dummy table for primary and secondary patient-reported outcome analyses

	Outcome at 3-months		Effect size (95% CI) for intervention vs control*	P-value
	Intervention arm (N=)	Control arm (N=)		
Primary outcome: PCFSS (N and %) ♦				
0 – no limitations				
1 – negligible limitations				
2 – occasional limitations				
3 – some limitations				
4 – severe limitations				
Key secondary outcomes (mean and SD)				
Quality of life by PROMIS 29+2 summary score				
Quality of life by EuroQol Visual Analogue Scale (0-100)				
Secondary outcomes (mean and SD or N and %)				
Fatigue (PROMIS score)				
Sleep disturbance (PROMIS score)				
Anxiety (PROMIS score)				
Depression (PROMIS score)				
Social participation (PROMIS score)				
Physical function (PROMIS score)				
Cognitive function (PROMIS score)				
Pain interference (PROMIS score)				
Pain intensity				
Breathlessness♦				
Health utility (EQ-5D-5L)				
Missed work●				
Work productivity				
Symptom burden♦				
Symptom 1 (TBC) ●				
Symptom 2 (TBC) ●				
Symptom 3 (TBC) ●				

*Adjusted for baseline level of outcome. Effect sizes represent mean difference in outcome score between intervention and control arms unless otherwise specified: ♦ = Odds ratio from ordinal logistic regression, ● = Odds ratio from binary logistic regression.

Appendix Table 4: Dummy table for tabulation of safety outcomes

Safety outcome	Number of occurrences while on treatment		Comment
	Intervention arm (N=)	Control arm (N=)	
Total person-months in study			
Suspected unexpected serious adverse reaction (SUSAR)			
Serious adverse event			
Pregnancy (participant or participant's partner)			
Adverse reaction involving LFT			
Other adverse reaction			
Adverse event involving LFT			
Other adverse event			

LFT is liver function test. Study definitions of SUSAR, serious adverse event, adverse reaction and adverse event are provided in the APRIL Trial protocol (V4.1).