Population Health Sciences Institute
Biostatistics Research Group

CONFORM-OH

Control, **F**ludrocortisone **or** **M**idodrine for the treatment of **O**rthostatic **H**ypotension: A Randomised Controlled Trial

Statistical Analysis Plan
SAP Version number: 1.0
SAP Date: 15/06/2023

This statistical analysis plan is based on protocol version 6.0 [21/12/2022]*

ISRCTN Number: 87213295
EudraCT Number: 2020-000794-25
REC Reference: 21/NE/0083

Sponsor: The Newcastle upon Tyne Hospitals NHS Foundation Trust
Sponsor protocol number: 09389

Funder: National Institute of Health Research (NIHR) Health Technology Assessment (HTA) Programme
Funder reference number: NIHR127385

Prepared by:

Name	Helen Mossop	Role	Trial Statistician
Signature	H. Mossop	Date	15/06/2023

Authorised by:

Name	Prof James Wason	Role	Lead Statistician
Signature		Date	20/06/2023
Name	Dr James Frith	Role	Chief Investigator
Signature		Date	21/06/2023

*MHRA approval for version 6.0 of the study protocol was still outstanding at the time of finalising this SAP. See File Note for more detail.

This statistical analysis plan (SAP) provides a framework and guidelines for the statistical analysis and reporting of the CONFORM-OH trial.

The SAP applies to a clean and validated dataset. Detailed information on data collection tools, data validation, consistency and accuracy checks, and data storage and archiving can be found in the current version of the Data Management Plan and Data Validation Plan.

Any deviation from the methods outlined in this SAP will be documented in the statistical end of trial report. Example Tables, Figures and Listings are for illustrative purposes only and are subject to change.

The current version of the SAP, any preceding agreed versions and all other documents relating to the analysis of this trial will be stored in the Statistical Section of the Trial Master File held by the PHSI Biostatistics Research Group.

Revision history

Version	Date	Changes made	Justification for change	Timing of change
1.0	15/06/2023	First version	Not Applicable	Not Applicable

Abbreviations

ADL	Activities of daily living
AE	Adverse Event
BP	Blood pressure
CDMS	Clinical Data Management System
CI	Confidence interval
DMC	Data Monitoring Committee
FWER	Family wise error rate
HR	Hazard Ratio
IMP	Investigational medicinal product
IQR	Interquartile range
IR	Incidence Rate
IRR	Incidence Rate Ratio
ITT	Intention to treat
LEDD	Levodopa equivalent daily dose
MAR	Missing at random
MCID	Minimally clinically important difference
MD	Mean difference
MedDRA	Medical dictionary for regulatory activities
MI	Multiple imputation
MNAR	Missing not at random
NEADL	Nottingham Extended Activities of Daily Living
OH	Orthostatic Hypotension
OHDAS	Orthostatic Hypotension Daily Activity Scale
OHQ	Orthostatic Hypotension Questionnaire
OHSAS	Orthostatic Hypotension Symptom Assessment Scale
OT	On-treatment
PPI	Patient Public Involvement
REML	Restricted maximum likelihood
SAE	Serious Adverse Event
SAP	Statistical analysis plan
SAS	Safety Analysis Set
SD	Standard deviation
TSC	Trial Steering Committee
UPDRS	Unified Parkinson's Disease Rating Scale

Contents

1. INTRODUCTION 5
1.1 Background and rationale 5
1.2 Objectives 5

2. STUDY METHODS 6
2.1 Trial design 6
2.2 Study setting and patient population 6
2.3 Randomisation and blinding 6
2.4 Definition of outcome measures 7
2.4.1 Primary endpoint 7
2.4.2 Secondary endpoints 7
2.4.3 Exploratory endpoints 9
2.5 Study assessments 9
2.6 Sample size and power 10

3. STATISTICAL CONSIDERATIONS 11
3.1 Timing of final analysis 11
3.2 Interim analyses, stopping guidelines and data monitoring 11
3.3 Confidence intervals and p-values 11
3.4 Analysis sets 11

4. STUDY POPULATION 12
4.1 Participant flow through trial 12
4.1.1 Screening, eligibility and recruitment 13
4.1.2 Protocol deviations 13
4.1.3 Follow-up 13
4.2 Baseline characteristics 14
4.3 Treatment compliance 16

5. ANALYSIS METHODS 19
5.1 Primary outcome measure 19
5.1.1 Main analysis methods 19
5.1.2 Supplementary analyses 20
5.2 Analysis of secondary outcomes 20
5.3 Additional / Exploratory analyses 22
5.4 Missing data 22

6. SAFETY 23
6.1 Adverse Events 23
6.2 Serious Adverse Events 26
6.3 Other safety measures 26

7. STATISTICAL SOFTWARE 27

APPENDIX 28

REFERENCES 40

1. INTRODUCTION

1.1 Background and rationale

Orthostatic hypotension (OH) is a common and disabling condition characterised by a significant reduction in blood pressure (BP) on standing upright, typically causing dizziness and falls.

There is little good quality evidence to support the management of OH. First line treatment is usually lifestyle advice and non-drug therapies. Where these are not effective there are two pharmacological options, fludrocortisone and midodrine; however, there is a lack of robust evidence on the clinical and cost-effectiveness of these treatment strategies and long-term efficacy and safety is unclear.

The CONFORM-OH trial, commissioned by NIHR HTA, was designed to evaluate the clinical and cost-effectiveness of fludrocortisone and midodrine for the management of symptomatic OH in comparison with conservative management (lifestyle advice and non-drug therapies).

The trial included a 10-month internal pilot phase to assess recruitment and retention with defined progression criteria. At a meeting of the Trial Steering Committee (TSC) on 30th August 2022, nine months into the internal pilot phase, it was agreed the that the trial should close to recruitment on the basis of very low recruitment rates.

This statistical analysis plan describes the trial design, outcome measures and plans to descriptively summarise available data. Details of analyses which would have been performed, had the trial continued to full enrolment, are described in an Appendix.

1.2 Objectives

This trial was designed to evaluate the clinical and cost-effectiveness of three different treatment strategies for the management of symptomatic OH:

I. Control: Conservative management (lifestyle advice and non-drug therapies)
II. Conservative management plus fludrocortisone
III. Conservative management plus midodrine

1.2.1 Primary objective

To determine whether the treatment strategies of conservative management plus fludrocortisone, and conservative management plus midodrine, improve symptoms of OH compared to conservative management alone, as measured by change in the Orthostatic Hypotension Questionnaire (OHQ) score at six months.

1.2.2 Secondary objectives

To determine how the treatment strategies of conservative management plus fludrocortisone, and conservative management plus midodrine, affect the following outcomes compared to conservative management alone over a 12-month period:

1. Activities of daily living (ADLs) measured by the Nottingham Extended ADL (NEADL) scale
2. Falls and syncope (number of falls, number of fallers/non-fallers, fall rate per person year, time to first fall, fall-related injuries, number of syncopal events)
3. Standing blood pressure and postural blood pressure drop
4. Side effects and the safety data associated with each treatment strategy
5. *Health-related quality of life measured by the EQ-5D-5L*
6. *Quality adjusted life years (QALYs) estimated from responses to the EQ-5D-5L and data derived from the literature*
7. *Costs to the NHS, personal social services and patients*
8. *Cost-effectiveness of each treatment strategy modelled from a patient and NHS and personal social services perspective measured in terms of the incremental costs per QALY gained*

Objectives 5-8 are outside the scope of this analysis plan and will be covered in the Health Economics Analysis Plan.

1.2.3 Exploratory objectives

An exploratory objective was to compare the clinical effectiveness (measured by change in the OHQ score at six months) of fludrocortisone with midodrine.

2. STUDY METHODS

2.1 Trial design

CONFORM-OH is a pragmatic, multi-arm, multi-stage, parallel group, prospective, randomised, open label, superiority trial. The trial was designed to assess the effectiveness of two drug therapies to improve the symptoms of OH compared to conservative management.

Adult patients presenting with symptomatic OH refractory to lifestyle modification were allocated (1:1:1) to receive either conservative management alone, conservative management plus fludrocortisone or conservative management plus midodrine for a period of 12 months.

The primary outcome is assessed after a period of 6 months. This timing was chosen based on a number of factors, including patient consultation, allowing adequate time for dose titration, and balancing adequate exposure with likely adherence and retention rates. Longer term follow-up will continue to 12 months.

An interim analysis, based on available three- and six-month primary outcome data, was planned to take place after the 200th patient had been recruited. If an intervention arm showed no benefit compared to the control arm (see Appendix for further details) it would be recommended to be dropped from the study, with control (conservative management) and the alternative intervention arm continuing to the planned recruitment target of 366 participants in a 1:1 ratio. If both intervention arms showed lack of benefit, the study would be recommended to stop. Recruitment would continue to all three arms while the interim analysis was being conducted.

The trial also included a 10-month internal pilot phase to assess recruitment and retention rates.

2.2 Study setting and patient population

Adult patients with symptomatic OH refractory to a minimum of 4 weeks of lifestyle modification were recruited. Patients were to be recruited from approximately 20 NHS trusts across the UK, typically from secondary care settings such as falls clinics, day hospitals, geriatric medicine clinics and movement disorder clinics.

For a full list of inclusion and exclusion criteria refer to section 4 of the study protocol.

2.3 Randomisation and blinding

Participants were allocated in a 1:1:1 ratio to receive conservative management alone, conservative management plus fludrocortisone or conservative management plus midodrine. A minimisation algorithm with a random element was used to assign treatment allocation. Minimisation factors were age (≥80 vs <80 years), aetiology (neurogenic vs non-neurogenic OH) and recruiting site/centre (to account for possible difference in usual care practice).

The minimisation system was provided by SealedEnvelope™ as a 24-hour, central, secure, web-based system accessed by delegated members of the research team at each site to perform randomisation.

The minimisation algorithm incorporates a random element such that there is an 80% chance the participant is allocated to the arm which minimises imbalance, with the remaining arms chosen with 10% probability each. In the case of ties, one of the tied arms is chosen at random to be the 'preferred' arm and assigned an 80% probability of being chosen with the remaining arms assigned a 10% probability each.

This is an open-label trial and there will be no blinding of participants, clinicians or research staff. It was planned that the trial statistician would not have access to outcome data by treatment group until the end of the trial. The interim analysis and closed reports to the Data Monitoring Committee (DMC), i.e. containing data presented by randomised treatment group, would have been performed/prepared by a statistician (a member of the Biostatistics Research Group) not otherwise involved in the study and reviewed by the Lead Statistician. However, following the decision to close the trial early it was agreed the trial statistician no longer needed to remain blinded. The perceived risk of having an unblinded statistician in an otherwise open-label trial was deemed to be minimal given that data would only be summarised descriptively and no interim or final inferential analyses would be performed.

2.4 Definition of outcome measures

2.4.1 Primary endpoint

The Orthostatic Hypotension Questionnaire (OHQ) [1] is completed at baseline, three, six and 12 month follow-up. The primary endpoint is the overall composite OHQ score at six months. The OHQ is validated for use in both clinical and research settings [2] and will be scored according to the validated scoring method [1].

The OHQ is composed of two sections: the first section consists of six questions, which rate the severity of six different symptoms on a scale of 0 to 10 (Orthostatic Hypotension Symptom Assessment Scale – OHSAS). The second section is composed of four questions, which rate the impact of symptoms on standing and walking (Orthostatic Hypotension Daily Activity Scale – OHDAS). The questions on the OHDAS are scored on a scale of 0 to 10 but also include an option of 'cannot be done for other reasons'.

The OHSAS score is calculated by averaging the responses to the six questions on the OHSAS. Similarly, the OHDAS score is calculated by averaging responses to the four questions on the OHDAS. Items which are scored as zero or answered 'cannot be done for other reasons' at baseline are not included in the scoring. Post-baseline scores are calculated using only those items which were included in the baseline score. If any item on the OHDAS which was included in the baseline score is missing or answered 'cannot do for other reasons' at a post-baseline assessment a value will be assigned using last observation carried forward.

An overall composite OHQ score will be calculated by averaging the OHSAS score and the OHDAS score. This score takes a value between 0-10, with higher scores indicating worse symptoms / interference.

At each follow-up time point the change from baseline will also be calculated as the follow-up score minus the baseline score.

2.4.2 Secondary endpoints

<u>Activities of daily living (ADLs) measured by the Nottingham Extended ADL scale</u>

The Nottingham Extended ADL (NEADL) scale is a 22 item questionnaire designed to assess the level of independence in carrying out social and domestic activities [3]. There are four sections/subscales which are mobility (six items), kitchen (five items), domestic (five items) and leisure (six items). Each item is scored as "on my own"; "own my own with difficulty"; "with help"; and "not at all/no".

When no more than 2 items are missing within subscales we will use simple imputation methods by replacing the missing item with the median response from the respondent specific completed questions within the subscale. This is can provide a valid approach for psychometrically validated questionnaires where responses to items within subscales are correlated [4]. A similar approach has also been used in other randomised controlled trials involving the NEADL scale [5].

One point is awarded per item if the participant selects either "on my own" or "on my own with difficulty." A score of zero is awarded if the participants selects "with help" or "not at all/no". All items are then summed to give an overall NEADL score which ranges from 0 – 22. Higher scores indicate greater independence. The score will not be calculated if any items are missing, after using the imputation method described above. The NEADL

questionnaire is measured at baseline, three, six and 12 month follow-up. At each follow-up time point the change from baseline will also be calculated as the follow-up score minus the baseline score.

Falls and syncope (number of falls, number of fallers/non-fallers, fall rate per person year, time to first fall, fall-related injuries, number of syncopal events)

Falls and faints (syncopal events) are self-reported by the participant and collected in monthly falls diaries which should be returned at three, six and 12 month follow up time points.

The number of falls for each participant will be summed over the 12 month period. The fall rate per person year will be calculated in each arm as the total number of falls from all participants in that arm divided by the total observation time for all participants in that arm. For each participant their observation time will be measured as the time (in years) from randomisation to the date of the last completed fall diary.

Participants will be classed as fallers if they report at least one fall over the 12 month period and a non-faller if they did not report a fall and returned at least one fall diary; i.e. participants who do not return at least one fall diary will be excluded from the analysis. We will also categorise participants as a single faller if they fall once, a recurrent faller if they fall twice or more, and a non-faller if they do not report any falls.

The number of faints (syncopal events) for each participant will be summed over the 12 month period. The syncopal event rate will be calculated as described above. In addition, we will report the total number, and rate of, a combined outcome of fall and faints.

Time to first fall will be measured as the time from randomisation to the first reported fall.

Fall related injuries are reported as free-text fields. There will be a medical review of any coding of free-text fields required.

Standing blood pressure and postural blood pressure drop

The lowest (nadir) standing blood pressure (systolic and diastolic) is measured at baseline, three, six and 12 month follow-up.

Blood pressure will also be measured in the supine position (lying down) at baseline, three, six and 12 month follow-up.

Postural blood pressure drop will be defined as (supine – lowest standing blood pressure) for systolic and diastolic measurements at each time point.

For each measurement the change from baseline will also be calculated as the follow-up value minus the baseline value.

Hospital admissions

Hospital admissions during the 12 month trial period will be collected from medical records. The number of admissions per participant will be calculated.

Reasons for admission will be tabulated. ‘Other’ reasons will be coded/grouped where possible with a medical review of any coding.

The rate of hospital admissions per person year will be calculated in each arm as the total number of admissions from all participants in that arm divided by the total observation time for all participants in that arm. For each participant their observation time will be measured as the time (in years) from randomisation to their last completed follow-up visit.

Side effects and the safety data associated with each treatment strategy

Adverse Events (AEs) will be collected over the 12-month trial period. At each follow-up visit participants will be asked about any side effects or adverse events they have experienced. This will be by open-ended questioning rather than a review of specific side effects or symptoms. Adverse events will be coded using the MedDRA dictionary (version 24) and summarised at the preferred term level. Severity (mild / moderate / severe),

seriousness (Yes / No) and relationship to study treatment (Unrelated / Unlikely to be related / Possibly related /Probably related / Definitely related) will also be collected.

An adverse event of the same type (i.e. same preferred term) will be counted as a separate occurrence if the start and end date of sequential events are separated by > 1 day.

Further detail on how adverse events will be summarised and reported is provided in section 6.

2.4.3 Exploratory endpoints

There are no planned exploratory endpoints, however, had the trial continued to full enrolment, an exploratory analysis would have compared the primary endpoint between the two intervention arms (conservative management plus fludrocortisone and conservative management plus midodrine).

2.5 Study assessments

Participants are assessed at three, six and 12 months from trial entry. As this is a pragmatic trial, assessments can take place within +/- 4, 6 and 8 weeks for the three, six and 12 month visits respectively. A simplified schedule of assessment is given in **Table 1**.

Table 1: Simplified schedule of assessments

Form	Visit			
	Baseline	Month 3 (+/- 4 weeks)	Month 6 (+/- 6 weeks)	Month 12 (+/- 8 weeks)
Demographics	X			
Medical History	X			
UPDRS*	X			
Blood pressure	X	X	X	X
Culprit medication review	X	X	X	X
OHQ	X	X	X	X
NEADL	X	X	X	X
EQ-5D-5L	X	X	X	X
Falls diary return		X	X	X
Hospital admission review		X	X	X

**Only for participants with Parkinson's disease, Dementia with Lewy bodies or Multi System Atrophy*

2.6 Sample size and power

The trial was designed to test two null hypotheses:

I. The mean difference in six-month OHQ score between conservative management plus fludrocortisone and conservative management alone is = 0
II. The mean difference in six-month OHQ score between conservative management plus midodrine and conservative management alone is = 0

Using standard sample size formula for a two sample t-test, a three-arm trial without an interim analysis would require 103 participants per arm to detect a difference of 1.0 point on the OHQ with 90% power, a two-sided type I error of 5% (equivalently a one-sided type I error of 2.5%), and assuming a standard deviation of 2.2.

A difference of 1.0 point on the OHQ is used as this represents an established minimally clinically important difference (MCID) [1] and was confirmed following discussion with patient and public involvement (PPI) representatives. A standard deviation of 2.2 was assumed based on local audit data from a cohort of 100 neurogenic and non-neurogenic patients.

Assuming attrition of 15% (based on comparable clinical trials in older people of similar duration [6,7] the recruitment target was 366 (122 per arm).

If at the interim analysis an intervention arm showed no benefit compared to the control arm (i.e. if the estimated mean OHQ score at 6 months was worse than control) it would be recommended to be dropped from the study. The multi-arm design would control the one-sided family-wise error rate (FWER), the total chance of falsely recommending an ineffective treatment, at 4.5%. With the possibility of early lack-of-benefit stopping, the FWER would be lower than this. The pair-wise error rate (PWER) would be controlled at 2.5% for each comparison.

Taking into account the interim analysis, which would be conducted once 200 participants had been recruited, the power of the design to recommend each treatment is displayed in the table below (as determined from one million simulation replicates per scenario).

We assumed participants would be recruited at a rate of 0.8 per site per month over a 30-month period with a staggered opening of sites (six sites open for the first six months with two new sites opening per month from month 7 onwards until a total of 20 sites are open). Under this assumed pattern of recruitment we would expect around 150 and 100 participants to have reached the three and six month follow-up time points respectively. The below table assumes only participants with six month data collected are included in the interim analysis. In practice including participants with three-month data but not yet six-month data would allow modest additional efficiency.

Scenario	Probability recommend fludrocortisone	Probability recommend midodrine	Probability recommend both	Probability recommend at least one
Mean effect of both = 0 (null scenario)	2.4%	2.4%	0.4%	4.4%
Fludrocortisone has MCID, midodrine effect = 0	91.8%	2.5%	2.5%	91.8%
Midodrine has MCID, fludrocortisone effect = 0	2.4%	91.7%	2.4%	91.7%
Both treatments have MCID effect	89.9%	89.7%	82.8%	96.8%

R code used to perform the power calculations can be found in S:\School Statistics\NCTU\CONFORM-OH\2. Study design\2.1 Statistical considerations\Sample size calculation

3. STATISTICAL CONSIDERATIONS

3.1 Timing of final analysis

The final analysis will take place once the last 12-month follow-up visit is complete. Once all data queries have been resolved (as far as possible) the database will be locked and the final analysis will commence.

3.2 Interim analyses, stopping guidelines and data monitoring

Had the trial continued past the internal pilot phase an interim analysis would have been performed based on available three and six month primary outcome data following recruitment of the 200th participant.

Analysis methods which would have been followed for the interim analysis can be found in the Appendix.

The DMC will meet periodically to review progress of the trial and the safety of trial participants. Further detail on the roles and responsibilities of the DMC can be found in the current version of the DMC charter. The DMC will have access to unblind outcome data. With the exception of the interim analysis time point (which will no longer be reached), outcome data will be summarised descriptively and no formal analyses will be conducted.

3.3 Confidence intervals and p-values

Had the trial continued to full enrolment all statistical tests for the final analysis would have been 2-sided and performed using a 5% significance level. For the interim analysis, an experimental arm would have been recommended to be dropped if the estimated mean between-group difference for an experimental arm is ≥0 (corresponding to the arm having worse outcome than control). This is equivalent to a one-sided p-value of >0.5.

Given that the trial has ended early no statistical testing will be performed. Any confidence intervals will be reported at the 95% level.

3.4 Analysis sets

The following analysis sets will be defined:

Analysis Set 1	• Participants will be analysed according to the treatment group they were randomised to receive, i.e. following the intention-to-treat (ITT) principle • All available outcome data will be included in the analysis
Analysis Set 2	• Participants will be analysed according to the treatment group they were randomised to receive, i.e. following the intention-to-treat (ITT) principle • Outcome data following cross-over to an alternative treatment group, or discontinuation of allocated trial treatment, will be set to missing
Safety Analysis Set (SAS)	• For each treatment group, the Safety Analysis Set will comprise all participants exposed to that treatment strategy • Safety data (adverse events) will be summarised according to the treatment strategy received at the time of onset of each adverse event. • Participants who cross-over between treatment groups will be included in the Safety Analysis Set for each treatment group

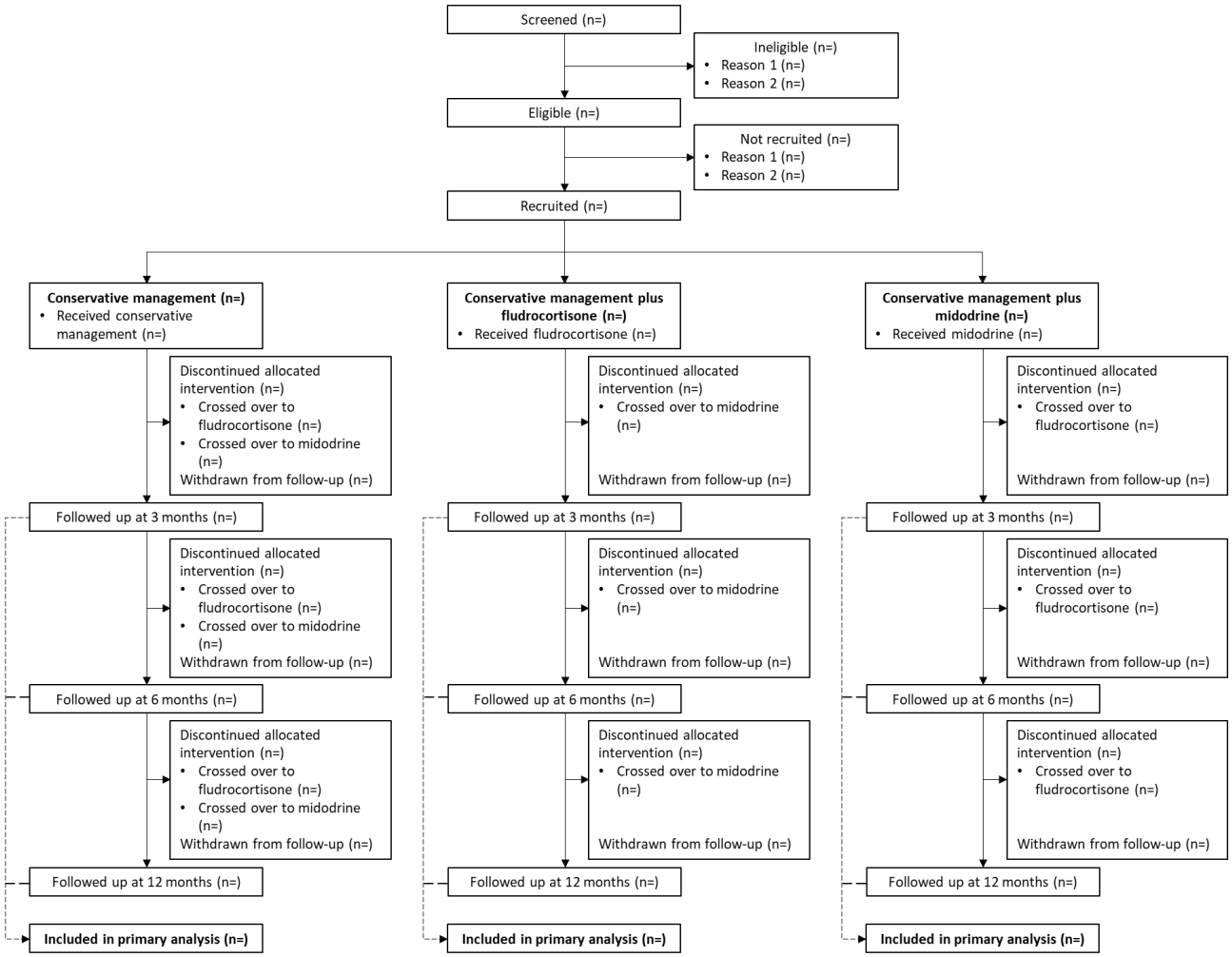
Primary and secondary clinical outcome measures will be reported according to randomised treatment group; safety data will be reported according to the treatment strategies received. Analysis Set 1 will be the main analysis set used to report the primary and secondary clinical outcome measures. Analysis Set 2 may also be used to provide supplementary information on the primary outcome measure. All safety data will be reported using the Safety Analysis Set.

4. STUDY POPULATION

4.1 Participant flow through trial

Participant flow through the trial will be presented using a CONSORT diagram, see Example Figure 1. Information will be provided on numbers and reasons for: screened patients not being eligible; eligible patients not being randomised; participants found to be ineligible after randomisation; participants deviating from allocated treatment; participants not evaluable for the primary endpoints and participant withdrawal from follow-up.

Example Figure 1: CONSORT flow diagram



4.1.1 Screening, eligibility and recruitment

The representativeness of the study sample will be assessed using the following data (shown in the CONSORT flow diagram):

- The number of patients identified at screening
- The number of patients screened and not meeting eligibility criteria (with reasons)
- The number of eligible patients identified at screening
- The number of eligible patients not taking part in the study (with reasons where available)
- The number of eligible patients randomised into the study

Data will be reported overall and by site. Observed and projected recruitment rates over time will be plotted.

4.1.2 Protocol deviations

Protocol deviations will be captured on a Deviation Tracking Log which will be held centrally by the Newcastle Clinical Trials Unit (NCTU).

Cross-over between treatment groups may occur during the course of this study and is permitted within the protocol; however, clinical teams will be asked to allow sufficient time, and ideally until the primary outcome is assessed at six months, before considering changing allocated treatment. Supplementary reporting of the primary outcome measure may be conducted using only data collected prior to cross-over or treatment discontinuation (see section 3.4). The number of participants changing treatment allocation or prematurely discontinuing their allocated trial treatment will be reported (see section 4.3).

Given the pragmatic nature of this trial we have not pre-specified any major protocol deviations which would affect the validity of the study or have a direct bearing on the primary outcome analysis or analysis populations.

A listing of all reported protocol deviations will be provided to the trial statistician at the end of the study. The number and type of protocol deviations will be summarised by treatment group.

4.1.3 Follow-up

Availability of outcome data at three, six and 12 month visits will be tabulated as frequency and percentage in each randomised group. The reasons for outcome assessments not being completed will be tabulated where available (e.g. due to withdrawal, death, loss to follow-up, participant too unwell etc.). The timing of outcome assessments (measured as days from randomisation) will be summarised by the median and range (minimum and maximum). The frequency and percentage of outcome assessments performed within protocol specified visit windows will be tabulated. Missing items within the OHQ and NEADL questionnaires will also be summarised by time point and treatment group.

Participants may withdraw consent to provide any further follow-up data. Unless the participant requests otherwise, routinely collected outcome data such as blood pressure measurements will be obtained where available from medical records. The number of participants withdrawing from follow-up (and whether they will allow continued collection of routine data) will be tabulated as frequency and percentage in each randomised group, with reasons where available. The timing of withdrawal (i.e. prior to month three, month six or month 12) will also be summarised, see example Table 1.

The number, and timing of, withdrawals will also be presented in a CONSORT diagram (Example Figure 1), with numbers of participants withdrawing between each stage (trial entry, month three, month six, month 12) shown.

Example Table 1: Withdrawals from trial by treatment group

	Conservative management N=	Conservative management plus fludrocortisone N=	Conservative management plus midodrine N=	Overall N=
By 3 months				
Reason 1				
Reason 2				
Reason 3				
Allowing routine data collection				
By 6 months				
Reason 1				
Reason 2				
Reason 3				
Allowing routine data collection				
By 12 months				
Reason 1				
Reason 2				
Reason 3				
Allowing routine data collection				

Data are n (%)

4.2 Baseline characteristics

Baseline characteristics will be summarised descriptively, both overall and by randomised treatment group. Categorical variables will be summarised by frequency and percentage. Continuous data will be summarised by the mean and standard deviation and/or median and range, as appropriate. No significance testing will be carried out due to the randomised nature of the study.

Parts I, II and IV of the Unified Parkinson's Disease Rating Scale (UPDRS) will be completed at baseline for participants with Parkinson's disease, dementia with Lewy bodies or Multi system atrophy. Part I and Part II each comprise 13 questions rated on a 5-point scale from 0 (normal) to 4 (severe). The Part I and Part II subscale scores are calculated by summing the responses to each of the 13 questions to obtain a score ranging from 0 to 52, with higher scores corresponding to worse outcomes. When no more than 20% of items are missing within subscales we will use simple imputation methods by replacing the missing item with the median response from the respondent specific completed questions within the subscale. If more than 20% of items are missing the subscale score will be set as missing. Part IV comprises six questions, rated on a 5-point scale from 0 (normal) to 4 (severe). The Part IV subscale score is calculated by summing the responses to each of the 6 questions to obtain a score ranging from 0 to 24, with higher scores corresponding to worse outcomes. When no more than 20% of items are missing we will use simple imputation methods by replacing the missing item with the median response from the respondent specific completed questions within Part IV. If more than 20% of items are missing the subscale score will be set as missing.

Details of characteristics to be reported are given in Example Table 2 below.

Example Table 2: Baseline characteristics

	Conservative management N=	Conservative management plus fludrocortisone N=	Conservative management plus midodrine N=	Overall N=
Demographics				
Age (years)[1]				
< 80 years[2]				
≥80 years[2]				
Sex				
Male[2]				
Female[2]				
Ethnicity				
Any white background[2]				
Mixed[2]				
Asian[2]				
African[2]				
Chinese[2]				
Other[2]				
Orthostatic Hypotension Questionnaire (OHQ)				
Overall OHQ score[1]				
Mean; SD				
Median (IQR); Range				
Blood pressure (mmHg)				
Supine blood pressure				
Systolic[1]				
Diastolic[1]				
Lowest standing blood pressure				
Systolic[1]				
Diastolic[1]				
Postural blood pressure drop				
Systolic[1]				
Diastolic[1]				
Medical history				
Disease aetiology				
Neurogenic[2]				
Non-neurogenic[2]				
Diabetes[2]				
Type 1[2]				
Type 2[2]				
Pure autonomic failure[2]				
Parkinson's disease[2]				
Disease duration (months)[1]				
Multi system atrophy[2]				
Disease duration (months)[1]				
Dementia with Lewy bodies[2]				
Disease duration (months)[1]				
UPDRS[3]				
Part I				
Part II				
Part IV				

Unless otherwise stated data are; [1] mean (SD) and/or median (range); [2] n (%)

[3]Only completed for participants with Parkinson's disease, dementia with Lewy bodies or Multi system atrophy

4.3 Treatment compliance

Participants will be randomised to receive conservative management, conservative management plus fludrocortisone or conservative management plus midodrine.

Conservative management is standard first-line care and forms the control arm of this study. Conservative management consists of non-pharmacologic therapy and will be implemented according to each site's usual clinical practice. The conservative, non-pharmacological measures advised at baseline will be tabulated by randomised treatment group.

Fludrocortisone and midodrine will be prescribed according to local clinical practice.

Given the pragmatic nature of this study adherence to prescribed medication and non-drug therapies will not be monitored. Participants may cross-over between treatment arms and this will be recorded. Participants who cross-over between treatment groups, or discontinue trial treatment should remain in the trial and be offered all follow-up visits.

The frequency and percentage of participants prematurely discontinuing their allocated trial treatment will be reported along with a reason, where available, and the timing of treatment discontinuation. Whether the participant crossed over to an alternative treatment group will be tabulated, see Example Table 3. Multiple cross-overs between treatment groups will be reported if this occurs.

For each pharmacological intervention (fludrocortisone and midodrine) the number of participants prescribed each treatment (regardless of allocated treatment group) will be reported along with the duration they received the treatment within the 12 month trial period and whether the treatment was prematurely discontinued, with reasons where available. See Example Table 4.

We will also report the frequency and percentage of participants taking 'culprit medications' at baseline and each follow-up visit and any changes to these since the previous visit (started, stopped, dose increased, dose decreased, or no change). This will be reported by randomised treatment group for each type of medication collected in the study database. In addition, for participants with Parkinson's disease, we will calculate the Levodopa equivalent daily dose (LEDD) at baseline and each follow-up visit according to the method proposed by Tomlinson et al [8]. This will be reported in each randomised treatment group as the median and range. See Example Table 5.

Example Table 3: Treatment adherence

	Conservative management N=	Conservative management plus fludrocortisone N=	Conservative management plus midodrine N=
Discontinued allocated treatment			
Yes			
No			
If yes, due to			
Lack of efficacy			
Side effects			
Other reason			
Time from randomisation to discontinuation of allocated treatment			
<3 months			
3-6 months			
>6 months			
Crossed over to another treatment group			
Yes			
No			
If yes, switched to			
Conservative management	NA		
Conservative management plus fludrocortisone		NA	
Conservative management plus midodrine			NA

Data are n(%)

Example Table 4: Trial pharmacological treatments received

	Fludrocortisone	Midodrine
Prescribed treatment		
Duration of treatment (months)		
Median; Range		
Prematurely discontinued treatment		
Yes		
No		
If yes, due to		
Lack of efficacy		
Side effects		
Other reason		

Data are n(%) unless otherwise stated

Example Table 5: Culprit medications

	Conservative management N=	Conservative management plus fludrocortisone N=	Conservative management plus midodrine N=
Culprit medication 1*			
Baseline[1]			
Month 3[1] *Started since baseline* *Stopped since baseline* *Dose increased* *Dose decreased* *No change*			
Month 6[1] *Started since Month 3* *Stopped since Month 3* *Dose increased* *Dose decreased* *No change*			
Month 12[1] *Started since Month 6* *Stopped since Month 6* *Dose increased* *Dose decreased* *No change*			
LEDD			
Baseline Number with data Median; Range			
Month 3 Number with data Median; Range			
Month 6 Number with data Median; Range			
Month 12 Number with data Median; Range			

**Will be repeated for each culprit medication.*

[1]Will be reported as Number taking at visit / number assessed at visit (%)

5. ANALYSIS METHODS

5.1 Primary outcome measure

5.1.1 Main analysis methods

Analysis Set 1 will be used for analysis (see Section 3.4). All data will be reported according to randomised treatment allocation and all available outcome data will be included.

The OHQ score and change from baseline at each follow-up visit will be summarised descriptively by randomised treatment group. The number with data, mean, standard deviation, median and range will be reported, see Example Table 6a. The OHSAS and OHDAS may also be summarised descriptively.

Example Table 6a: Summary of OHQ scores and change from baseline

OHQ score	Value at follow-up visit			Change from baseline		
	Control	Fludrocortisone	Midodrine	Control	Fludrocortisone	Midodrine
Month 3 N Mean (SD) Median (Range)						
Month 6 N Mean (SD) Median (Range)						
Month 12 N Mean (SD) Median (Range)						

Subject to sufficient numbers with data (at least three participants per group with outcome data), we will also estimate the mean difference in the change in OHQ score from baseline to each follow-up time point between each intervention and control group, see Example Table 6b. This will be reported with a standard deviation and 95% CI. The purpose of this is to inform potential future meta-analyses. No inferences will be drawn from this data.

Example Table 6b: Change in OHQ score - mean difference between groups

OHQ score	Mean difference (MD)	
	Fludrocortisone - control MD (SD; 95% CI)	Midodrine - control MD (SD; 95% CI)
Month 3 Month 6 Month 12		

To inform future sample size calculations will we report the correlation (Pearson's correlation coefficient) between baseline and follow-up OHQ scores. This will be reported within each randomised treatment group and overall.

Example Table 6c: Correlation between baseline and follow-up OHQ scores

OHQ score	Pearson's correlation coefficient			
	Control	Fludrocortisone	Midodrine	Overall
Month 3 Month 6 Month 12				

5.1.2 Supplementary analyses

Primary outcome data may also be reported using Analysis Set 2 (see Section 3.4). Data will be reported according to allocated treatment group but data collected after discontinuation of allocated treatment or initiation of non-allocated fludrocortisone or midodrine will be set as missing. This dataset will be reported descriptively as described in section 5.1.1. Data will only be reported in this way if there is at least three participants per group with outcome data at each follow-up visit.

5.2 Analysis of secondary outcomes

Activities of daily living (ADLs) measured by the Nottingham Extended ADL scale

Analysis Set 1 will be used for analysis. All available data will be reported according to randomised treatment allocation.

The NEADL score and change from baseline at each follow-up visit will be summarised descriptively by randomised treatment group. The number with data, mean, standard deviation, median and range will be reported, see Example Table 7.

Example Table 7: NEADL

NEADL score	Value at follow-up visit			Change from baseline		
	Control	Fludrocortisone	Midodrine	Control	Fludrocortisone	Midodrine
Month 3 N Mean (SD) Median (Range)						
Month 6 N Mean (SD) Median (Range)						
Month 12 N Mean (SD) Median (Range)						

Falls and syncope (number of falls, number of fallers/non-fallers, fall rate per person year, time to first fall, fall-related injuries, number of syncopal events)

Analysis Set 1 will be used for analysis. Data from all participants with at least one falls diary available will be reported according to their randomised treatment allocation.

The number of falls reported per participant will be tabulated as frequency and percentage and summarised as mean, SD, median and range. The total number of falls and total follow-up time from participants allocated to each randomised group will be reported. The simple, unadjusted incidence rate (IR) of falls per person year (i.e. total number of falls / total follow-up time) will be calculated in each randomised treatment group and presented with 95% CIs.

The same approach as above will be used for reporting of syncopal events, and a combined outcome of falls and syncopal events.

The number of fallers (i.e. the number of participants reporting at least one fall over the 12 month period) will be reported as frequency and percentage out of those returning at least one fall diary.

Time to first fall will be calculated and tabulated by whether this was within the first 3 months from randomisation, between 3 and 6 month follow-up or after 6 month follow-up.

Fall-related injuries will be tabulated by type, with the number of participants affected and the total number of occurrences reported by randomised treatment group. The number and proportion of participants reporting at least one fall-related injury will be tabulated by treatment group. The total number of fall-related injuries reported per participant will be tabulated as frequency and percentage and as mean, SD, median and range. The total number of fall-related injuries reported in participants allocated to each treatment group will be presented. For the purpose of calculating percentages the number of participants returning at least one fall diary will be used as the denominator.

Example Table 8: Falls and syncopal events

	Control N =	Fludrocortisone N =	Midodrine N =
Number of falls			
0			
1			
>2			
Mean (SD)			
Median (Range)			
Fall rate			
Total number of falls			
Total follow-up time (years)			
Incidence rate (95% CI)			
Fallers / non-fallers			
≥1 fall / N (%)			
Time to first fall			
< 3months			
3-6 months			
>6 months			
Number of syncopal events			
0			
1			
>2			
Mean (SD)			
Median (Range)			
Syncopal event rate			
Total number of syncopal events			
Incidence rate (95% CI)			
Number of falls and syncopal events			
0			
1			
>2			
Mean (SD)			
Median (Range)			
Falls and syncope event rate			
Total number of events			
Incidence rate (95% CI)			

Standing blood pressure and postural blood pressure drop

The methods described below will be applied to the following variables; nadir standing systolic blood pressure (BP), nadir standing diastolic BP, systolic postural BP drop, diastolic postural BP drop.

Analysis Set 1 will be used for analysis. All data available will be reported according to randomised treatment allocation.

Each BP measurement and change from baseline will be summarised descriptively by randomised treatment group. The number with data, mean, standard deviation, median and range will be summarised at each visit.

Example Table 9: Blood pressure

Blood pressure	Value at follow-up visit			Change from baseline		
	Control	Fludrocortisone	Midodrine	Control	Fludrocortisone	Midodrine
Nadir standing Month 3* N Systolic Mean (SD) Median (Range) Diastolic Mean (SD) Median (Range)						
Postural drop Month 3* N Systolic Mean (SD) Median (Range) Diastolic Mean (SD) Median (Range)						

**Data will be presented in the same way for Month 6 and Month 12*

Hospital admissions

Hospital admissions will be reported similarly to falls.

Analysis Set 1 will be used for analysis. Data from all participants with at least one review of hospital admissions available will be reported according to randomised treatment allocation.

The number of admissions reported per participant will be tabulated as frequency and percentage and summarised as mean, SD, median and range. Reasons for admission will also be tabulated as frequency and percentage. The total number of admissions and total follow-up time from participants allocated to each randomised group will be reported. The simple, unadjusted incidence rate of admissions per person year (i.e. total number of admissions / total follow-up time) will be calculated in each randomised treatment group and presented with 95% CIs.

5.3 Additional / Exploratory analyses

No additional or exploratory analyses are planned. Details of exploratory analyses which would have been performed had the trial continued to full enrolment are described in the Appendix.

5.4 Missing data

The availability of outcome data will be summarised as described in section 4.1.3. Given the small sample size we do not plan any imputation of missing data (except where part of questionnaire scoring). Methods which would have been employed had the trial continued are described in the Appendix.

6. SAFETY

6.1 Adverse Events

Safety data will be reported in the Safety Analysis Set, with adverse events reported according to treatment received at the time of AE onset.

The following metrics will be used to provide an overall summary of adverse events reported while exposed to each treatment group, see Example Table 10:

- The number of participants exposed to each treatment group.
- The total exposure time (measured in years) to each treatment group, i.e. a sum across all participants of the amount of time they were exposed to each treatment group.
- Number and proportion of participants affected by at least one adverse event while exposed to each treatment group. Percentages will be calculated out of those exposed.
- Worst grade (mild, moderate, severe) adverse event reported while exposed to each treatment group. This will be tabulated with percentages calculated out of those exposed.
- Total number of adverse event occurrences reported by each participant while exposed to each treatment group. This will be tabulated (with percentages calculated out of those exposed) and/or reported as median and range, as appropriate
- Total number of unique adverse event occurrences (i.e. different preferred terms) reported by each participant while exposed to each treatment group. This will be tabulated (with percentages calculated out of those exposed) and/or reported as median and range, as appropriate
- Total number of adverse event occurrences reported across all participants while exposed to each treatment group. This will also be broken down by severity (mild, moderate, severe).
- The incidence rate (IR) of adverse event occurrence while exposed to each treatment group (i.e. total number of AE occurrences / total exposure time).

For each adverse event, at the preferred term level, we will also report the following metrics by treatment group, see Example Table 11:

- The number and proportion of participants experiencing the adverse event at any point while exposed to each treatment group. Percentages will be calculated out of those exposed.
- The worst reported severity (mild, moderate, severe) for the adverse event at any point while exposed to each treatment group. Percentages will be calculated out of those exposed.
- The total number of occurrences of the adverse event while exposed to each treatment group.
- The total number of occurrences of the adverse event, by severity (mild, moderate, severe), while exposed to each treatment group.

This data may also be presented graphically as dot plots of the proportion of participants affected by each adverse event. The number of participants affected and the total number of occurrences will be shown on the graph. See Example Figure 5. Had the trial continued to full enrolment the graph would have also shown a measure of relative risk with 95% CIs, with separate graphs produced for the comparison of fludrocortisone and control, midodrine and control, and fludrocortisone and midodrine. If there are many different adverse events reported we will focus on the graphical presentation of only those occurring in a specified proportion of participants (e.g. >3% participants in either group).

The same summaries will also be reported for adverse reactions (adverse events reported as possibly, probably or definitely related to trial treatment) and for non-serious events (as required for EudraCT reporting).

Adverse events which resulted in treatment discontinuation will also be tabulated. Percentages will be calculated out of those exposed to each treatment group.

Example Table 10: Summary of adverse events

	Control	Fludrocortisone	Midodrine
Number of participants exposed; N			
Total exposure time (years); N			
Number of AEs per participant; N (%) 0 1-3 4-6 >6 Median (IQR); Range			
Number of unique AEs per participant; N (%) 0 1-3 4-6 >6 Median (IQR); Range			
Worst grade reported across all AEs; N (%) *Mild* *Moderate* *Severe*			
Total number of AEs reported; N *Mild* *Moderate* *Severe*			
Incidence of AEs; IR (95% CI)			

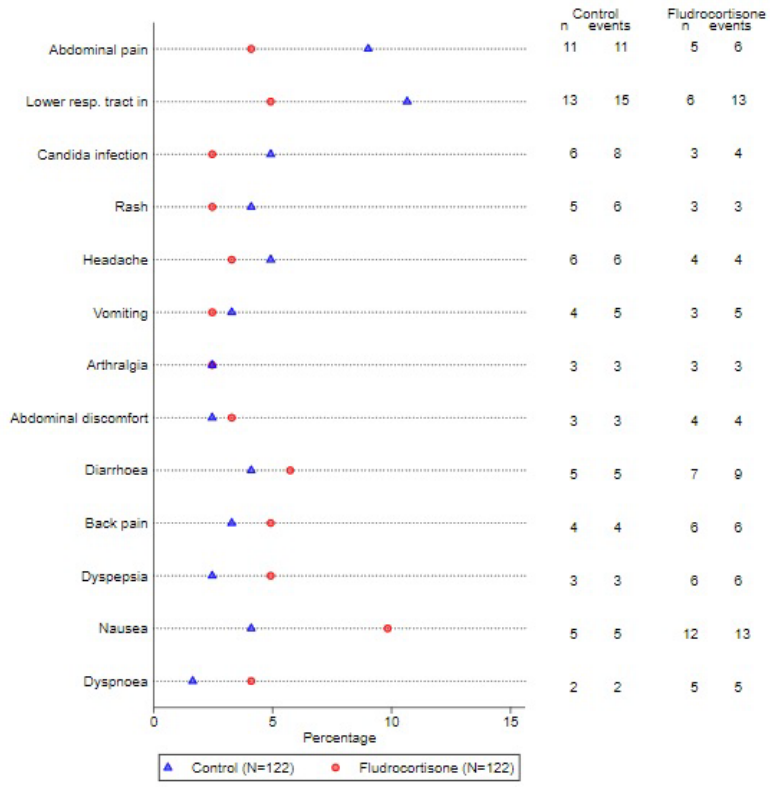
Example Table 11: Summary of adverse events by type

Adverse event term		Number (%) of participants affected[1,2]			Number of occurrences[1]		
		Control	Fludrocortisone	Midodrine	Control	Fludrocortisone	Midodrine
		N =	N =	N =	N =	N =	N =
Nausea	Overall						
	Mild						
	Moderate						
	Severe						
Headache	Overall						
	Mild						
	Moderate						
	Severe						
AE 1	Overall						
	Mild						
	Moderate						
	Severe						
AE 2	Overall						
	Mild						
	Moderate						
	Severe						

[1]*While exposed to each treatment group*

[2]*% calculated out of total number exposed. If >1 occurrence per participant the worst reported severity is tabulated*

Example Figure 5: Proportion of participants affected by each AE



6.2 Serious Adverse Events

A chronological listing of reportable serious adverse events (SAEs) will be presented, see Example Table 12.

The following metrics will also be reported:

- The number and proportion of participants reporting at least one SAE while exposed to each treatment group. Percentages will be calculated out of those exposed.
- The total number of occurrences of SAEs, across all participants, while exposed to each treatment group.
- For each SAE (using the preferred term), the total number of participants affected and the total number of occurrences while exposed to each treatment group.

Example Table 12: Line listing of all SAEs

ID	SAE no.	Rand. group	IMP at onset	IMP start + end dates	Description	SAE onset date	Severity[1]	Causality[2]	Serious criteria[3]	Outcome[4]	Outcome date

[1]*Mild; moderate; severe*

[2]*Related; unrelated; unable to determine*

[3]*Resulted in death; Life threatening; Inpatient hospitalisation / prolonged hospitalisation; Persistent or significant disability/incapacity; Congenital anomaly / birth defect; Important medical event*

[4]*Recovered; Condition improved; Condition deteriorated; Condition unchanged; Participant died; Recovered with sequelae; Condition stable and no change anticipated*

Non-reportable serious adverse events will also be summarised.

6.3 Other safety measures

We do not plan analyses of any other safety measures (e.g. laboratory values, vital signs etc.).

7. STATISTICAL SOFTWARE

Data will be exported from the electronic Clinical Data Management System (CDMS) into a STATA format by the NCTU Data(base) Manager at time points agreed by the Trial Management Group. Statistical analyses will be carried out by the Trial Statistician at the Biostatistics Research Group predominately using Stata but R software may also be used. All programs and output will be stored in the School Statistics folder on the IHS server.

APPENDIX

The planned analysis methods which would have been used had the trial continued to full enrolment are summarised below for both the interim and final analysis.

A. Interim analysis

A linear mixed-effects model would have been fitted to three and six month OHQ data, with fixed effects for baseline OHQ, age (<80 years or ≥80), and aetiology (neurogenic or non-neurogenic OH), and a treatment-by-time interaction term. We planned to include random effects for site and individual (nested within site), however if there were sites with few participants (<5) then site would have been excluded from the model.

Where Y_{ij} denotes the OHQ score for participant i at time j, the planned interim analysis model would be:

$$Y_{ij} = \beta_0 + \beta_1 Trt1_i + \beta_2 Trt2_i + \beta_3 OHQ_i^0 + \beta_4 Age_i + \beta_5 Aetiology_i + \beta_6 M6 + \beta_7 M6 * Trt1_i + \beta_8 M6 * Trt2_i + b_{1,i} + b_{2,k} + e_{ij}$$

Where $Trt1_i$ and $Trt2_i$ are dummy variables for treatment allocation ($Trt1_i$: 1 if allocated to conservative management plus fludrocortisone, 0 otherwise; $Trt2_i$: 1 if allocated to conservative management plus midodrine, 0 otherwise), OHQ_i^0 is the baseline OHQ score for participant i, $M6$ is a dummy variable for time (0 or 1) at month six. Note month three is represented by $M6$ = 0. $b_{1,i}$ and $b_{2,k}$ are random intercepts at the participant and site level respectively and are assumed to follow a normal distribution.

This model would have been fitted using restricted maximum likelihood (REML) and an unstructured covariance matrix. The following code in Stata would have been used:

```
mixed ohq i.arm##i.visit ohq_bl i.age i.aetiology || site: || id:,
residuals(unstructured, t(visit)) reml
```

This model accounts for missing outcome data under a missing at random (MAR) assumption. Participants would be included in the model if they had at least one OHQ score available from the three- or six-month follow-up visit.

From this model we would obtain the estimated mean difference in six-month OHQ between fludrocortisone and control (β_1+β_7) and between midodrine and control (β_2+β_8). This would be reported with a corresponding 95% confidence interval and p-value. The Wald test would be used to test the null hypothesis of no treatment difference at six months for each experimental treatment against control (reference). The following Stata code would be used:

```
lincom 1.arm + 2.visit#1.arm
lincom 2.arm + 2.visit#2.arm
```

If an experimental treatment arm showed no benefit compared to the control arm, i.e. if the estimated mean difference for an experimental arm was 0 or more (corresponding to the arm having worse outcome than control), it would be recommended to be dropped from the trial. This is equivalent to a one-side p-value > 0.5. Otherwise the trial would be recommended to continue as planned.

The table below shows the chance of each arm being stopped for lack of benefit at the interim analysis.

Scenario	Probability fludrocortisone stops at interim	Probability midodrine stops at interim
Mean effect of both = 0 (null scenario)	50.0%	50.0%
Fludrocortisone has MCID, midodrine effect = 0	2.1%	50.0%

Midodrine has MCID, fludrocortisone effect = 0	50.0%	2.1%
Both treatments have MCID effect	2.1%	2.1%

Any recommendation of the trial design to drop an experimental arm would be ratified by the DMC and the trial steering committee (TSC). A meeting of the DMC would have been held at the time of the interim analysis. The DMC would review the results of the interim analysis in context of other accumulating data such as adverse events, treatment crossovers and withdrawals. The DMC would make a recommendation, considering all aspects of the trial, on whether any experimental arm should be dropped at the interim analysis. The recommendation of the DMC would be passed onto the TSC who would then make a decision. Any decision to overrule the recommendation of the design would need to be clearly justified and documented. Unless the DMC specifically requested the release of some unblinded information, the TSC would not have access to any unblind outcome data.

If one experimental arm was recommended to be dropped at the interim analysis, and this was ratified by the oversight committees, the trial would continue to its planned enrolment of 366, but subsequent participants would be randomised in a 1:1 ratio between control and the remaining experimental arm. If both experimental arms were recommended for dropping at the interim analysis and this was ratified by the oversight committees, the trial would be terminated early. Recruitment would continue to all three arms while the interim analysis was being conducted.

It was planned that the interim analysis would have been conducted by a statistician (a member of the Biostatistics Research Group) not otherwise involved in the study and reviewed by the Lead Statistician.

B. Final analysis

a. Primary outcome

i. Main analysis methods

Analysis Set 1 would have been used for analysis (see Section 3.4). All participants would have been analysed according to their randomised treatment allocation and all available outcome data included in the analysis.

Using the estimand framework [9], we were primarily interested in the effect of being allocated to fludrocortisone or midodrine in addition to conservative management, regardless of treatment discontinuation or any other treatments received, i.e. a treatment policy estimand. The intended main primary outcome estimand is described in the table below:

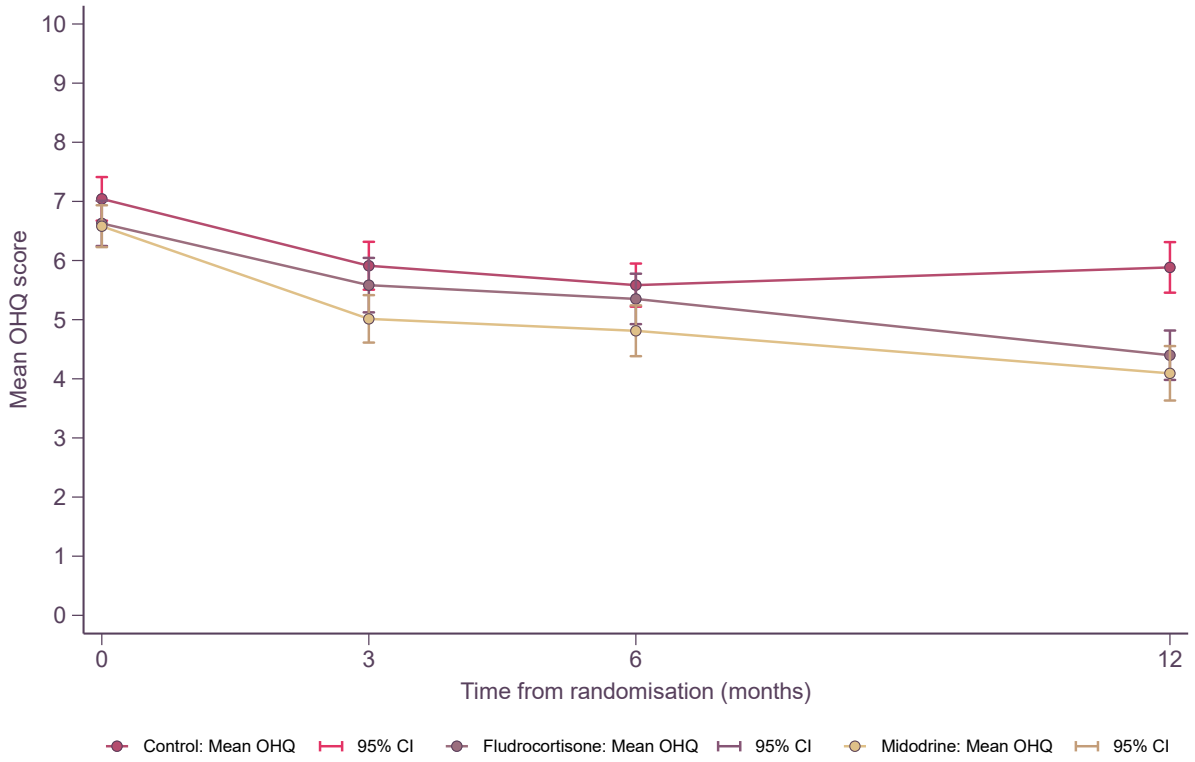
Estimand attribute	Description
Population	Patients with symptomatic OH refractory to lifestyle modification and meeting the CONFORM-OH eligibility criteria
Treatment	Allocated to six months of treatment with conservative management plus fludrocortisone or conservative management plus midodrine compared to conservative management alone
Outcome variable	Overall composite OHQ score at month six
Strategies used to handle intercurrent events	• Discontinuation of allocated treatment – treatment policy[1] • Use of non-allocated treatment with fludrocortisone or midodrine – treatment policy[1]
Population-level summary measure	Mean difference in OHQ score at six months (adjusted for baseline) between the intervention and control groups

[1] *A **treatment policy** strategy considers the occurrence of the intercurrent event as irrelevant, and participant data are analysed regardless*

The OHQ score would be summarised descriptively by randomised treatment group. The number with data, mean, standard deviation and the median, IQR and range would be summarised at each visit. Data would also

be presented graphically by randomised treatment group as the mean value with 95% CIs at each visit, see Appendix Figure 1. The OHSAS and OHDAS would have also been summarised descriptively.

Appendix Figure 1: OHQ score over the 12 month follow-up period (note this is not genuine data)



For the primary analysis of the overall OHQ score at six-months, a linear mixed-effects model would have been fitted to the three-and six-month OHQ data. Fixed effects would include baseline OHQ, baseline age (<80 years or >=80), aetiology (neurogenic or non-neurogenic) and a treatment-by-time interaction term. Random effects would include site and individual (nested within site).

Where Y_{ijk} denotes the OHQ score for participant i at time j from site k, the primary analysis model would be:

$$Y_{ijk} = \beta_0 + \beta_1 Trt1_i + \beta_2 Trt2_i + \beta_3 OHQ_i^0 + \beta_4 Age_i + \beta_5 Aetiology_i + \beta_6 M6 + \beta_7 M6 * Trt1_i + \beta_8 M6 * Trt2_i + b_{1,i} + b_{2,k} + e_{ijk}$$

Where $Trt1_i$ and $Trt2_i$ are dummy variables for treatment allocation ($Trt1_i$: 1 if allocated to conservative management plus fludrocortisone, 0 otherwise; $Trt2_i$: 1 if allocated to conservative management plus midodrine, 0 otherwise), OHQ_i^0 is the baseline OHQ score for participant i, $M6$ is a dummy variable for time (0 or 1) at month six. Note month three is represented by $M6$ = 0. $b_{1,i}$ and $b_{1,k}$ are random intercepts at the participant and site level respectively and are assumed to follow normal distributions.

This model would be fitted using restricted maximum likelihood (REML) and an unstructured covariance matrix. The following code in Stata would have been used:

```
mixed ohq i.arm##i.visit ohq_bl i.age i.aetiology || site: || id:,
residuals(unstructured, t(visit)) reml
```

Missing responses would be assumed to be missing at random (MAR). Participants would be included in the model if they had at least one OHQ score available from the three or six month follow-up visit.

From this model we would obtain the estimated mean difference in six-month OHQ between fludrocortisone and control (β_1+β_7) and between midodrine and control (β_2+β_8). This will be reported with a corresponding

95% confidence interval and p-value, see Appendix Table 1. The Wald test would be used to test the null hypothesis of no treatment difference at six months for each experimental treatment against control (reference). The following Stata code would be used:

```
lincom 1.arm + 2.visit#1.arm
lincom 2.arm + 2.visit#2.arm
```

We would have investigated the assumptions of the primary analysis model using the following methods:

- Plot of residuals versus fitted values
- Normal quantile plot of standardised residuals
- Normal quantile plot of the standardised predicted random effects

If the model failed to converge we would instead treat site as a fixed effect. Or, if there were sites with few participants (<5) then site would be excluded from the model. This may have been an issue particularly if recruitment to one (or all) arms is terminated at the interim analysis time point, resulting in a relatively small sample size (minimum 67 per arm).

Appendix Table 1: Primary outcome, OHQ

OHQ score	Control	Fludrocortisone	Midodrine	Mean difference (MD)	
				Fludrocortisone - control	Midodrine - control
	N; Mean (SD)	N; Mean (SD)	N; Mean (SD)	MD (95% CI); p-value	MD (95% CI); p-value
Month 3					
Month 6*					
Month 12					

**Primary outcome*

ii. Subgroup analyses

We would have explored whether the treatment effect for the 6-month OHQ score was consistent across subgroups. To do this a linear mixed-effects regression model would be fitted to the 6-month OHQ data. Fixed effects would include treatment group, baseline OHQ, baseline age (<80 years or >=80) and aetiology (neurogenic or non-neurogenic), and an interaction between treatment and the subgroup variable of interest. Site would be included as a random effect.

For each subgroup, the estimated mean difference (and 95% CI) in six-month OHQ between fludrocortisone and control and between midodrine and control would be presented using Forest plots which would also show the p-value from a test of interaction. Subgroups we would have considered were:

- Neurogenic OH versus non-neurogenic OH
- Age ≥80 years versus age <80 years
- Male versus female
- Presence of diabetes versus no diabetes*

**Subject to sufficient numbers within each group*

As the trial had not been powered to detect subgroup effects these results would have been considered exploratory and hypothesis generating. Conclusions would not have been drawn on the basis of these results.

Should any model have failed to converge we would instead have included site as a fixed effect, or excluded it from the model, as described above (*i. Main analysis methods)*.

iii. Sensitivity analyses

To assess the robustness of the results (i.e. that inferences do not change) we would conduct the following sensitivity analyses, which would target the same treatment policy estimand:

1) The primary analysis is valid under the assumption that data are missing at random. If the OHQ score is missing at six months in more than >15% of participants (the assumed attrition rate), either overall or in any arm, we would perform sensitivity analyses using multiple imputation methods to explore the robustness of the results to the MAR assumption. This is discussed in more detail below (*d. Missing data*).
2) The main analysis would use all OHQ data reported at the three, six and 12-month visits. If >10% of questionnaires, either overall or within any arm, were completed outside of the protocol specified visit windows we would repeat the analysis excluding these data.

iv. Supplementary analyses

A supplementary analysis would be performed targeting an 'on-treatment' or hypothetical estimand; the treatment effect if all participants had continued on their allocated treatment. The key aspects of this estimand are described in the table below:

Estimand attribute	Description
Population	Patients with symptomatic OH refractory to lifestyle modification and meeting the CONFORM-OH eligibility criteria
Treatment	Six months of treatment with conservative management plus fludrocortisone or conservative management plus midodrine compared to conservative management alone
Outcome variable	Overall composite OHQ score at month six
Strategies used to handle intercurrent events	• Discontinuation of allocated treatment – hypothetical[1] • Use of non-allocated treatment with fludrocortisone or midodrine – hypothetical[1]
Population-level summary measure	Mean difference in OHQ score at six months (adjusted for baseline) between the intervention and control groups

[1] *Under a **hypothetical** strategy we are interested in the value the outcome variable would have taken had the intercurrent event not happened*

This analysis would be performed using Analysis Set 2 (see Section 3.4). Participants would be analysed according to their allocated treatment group but data collected after discontinuation of allocated treatment or initiation of non-allocated fludrocortisone or midodrine would be set as missing. This dataset would be analysed using the same linear mixed-model as described above (*i. Main analysis methods)*. The model assumes data are missing at random, i.e. that the missing data would be similar to that of other participants in the same treatment group with similar baseline characteristics.

While the missing at random assumption is plausible under hypothetical, on-treatment conditions, it is possible participants who discontinued allocated treatment or initiated non-allocated fludrocortisone or midodrine would have had worse outcomes had they continued on allocated treatment than similar participants who did not experience these intercurrent events. We would therefore also perform sensitivity analyses implementing controlled multiple imputation (MI) using a δ-based pattern-mixture approach. This is discussed in more detail below (*d. Missing data*).

b. Secondary outcomes

All secondary outcomes would have been analysed using treatment policy estimands.

OHQ score at Month 12

Analysis Set 1 would be used for analysis. All participants with OHQ data available at three, six or 12-month time points would be analysed according to their randomised treatment allocation.

The estimated mean difference in 12-month OHQ would be estimated by fitting a liner mixed-effects model, as described above (*i. Main analysis methods*) but to three-, six- and 12-month outcome data. Where Y_{ijk} denotes the OHQ score for participant i at time j from site k, the analysis model would be:

$$Y_{ijk} = \beta_0 + \beta_1 Trt1_i + \beta_2 Trt2_i + \beta_3 OHQ_i^0 + \beta_4 Age + \beta_5 Aetiology + \beta_6 M6 + \beta_7 M6 * Trt1_i + \beta_8 M6 * Trt2_i + \beta_9 M12 + \beta_{10} M12 * Trt1_i + \beta_{11} M12 * Trt2_i + b_{1,i} + b_{2,k} + e_{ijk}$$

Where $M6$ is a dummy variable for time (0 or 1) at month six, $M12$ is a dummy variable for time (0 or 1) at month 12. Note month three is represented by $M6$ = 0 and $M12$ = 0. All other parameters are as previously described.

Activities of daily living (ADLs) measured by the Nottingham Extended ADL scale

Analysis Set 1 would be used for analysis. All participants with data available would be analysed according to their randomised treatment allocation.

The NEADL score would be summarised descriptively by randomised treatment group. The number with data, mean, standard deviation and the median, IQR and range would be summarised at each visit. Data would also be presented graphically by randomised treatment group as the mean value with 95% CIs at each visit.

As for the OHQ data, a linear mixed-effects model would be fitted to the three, six and 12 month data. Fixed effects would include baseline NEADL score, baseline age (<80 years or >=80), aetiology (neurogenic or non-neurogenic) and a treatment assignment x time interaction term. Random effects would include site and individual (nested within site). This model would be fitted using REML and an unstructured covariance matrix.

Missing responses would be assumed to be missing at random (MAR). Participants would be included in the model if they had at least one NEADL score available from the three, six or 12 month follow-up visit.

From this model we would obtain the estimated mean difference in the three, six and 12-month follow-up NEADL score between fludrocortisone and control and between midodrine and control. This will be reported with a corresponding 95% confidence interval, see Appendix Table 2.

Assumptions of the model would be investigated as described for the analysis of the primary outcome.

Appendix Table 2: NEADL

NEADL score	Control	Fludrocortisone	Midodrine	Mean difference (MD)	
				Fludrocortisone - control MD (95% CI); p-value	Midodrine - control MD (95% CI) ; p-value
	N; Mean (SD)	N; Mean (SD)	N; Mean (SD)		
Month 3 Month 6 Month 12					

Falls and syncope (number of falls, number of fallers/non-fallers, fall rate per person year, time to first fall, fall-related injuries, number of syncopal events)

Analysis Set 1 would be used for analysis. All participants with at least one falls diary available would be analysed according to their randomised treatment allocation.

The number of falls reported per participant would be tabulated as frequency and percentage and summarised as mean, SD, median, IQR and range. The number, or percentage, of falls may have also been presented graphically, see Appendix Figure 2. The total number of falls and total follow-up time from participants allocated to each randomised group would be reported. The simple, unadjusted incidence rate (IR) of falls per person year (i.e. total number of falls / total follow-up time) would be calculated in each randomised treatment group and presented with 95% CIs.

A mixed-effects negative binomial regression model would be fitted to the falls data. Site would be included as a random effect and treatment assignment, baseline age (<80 years or >=80) and aetiology (neurogenic or non-neurogenic) as fixed effects. The log of the exposure variable (follow-up time) would be included in the model with the coefficient constrained to be one. From this model we would obtain the estimated incidence rate ratio (IRR) and risk difference in the fall rate per person-year between fludrocortisone and control and midodrine and control. These estimates would be reported with corresponding 95% confidence intervals.

We may instead have used a Poisson or zero-inflated model if this was more suitable for the distribution of the data. We would explore this graphically by plotting the observed and predicted number of events using each model, using the Pearson chi-square goodness of fit test and using Akaike's and Bayesian information criterion (AIC and BIC).

The same approach as above would be used for the analysis and reporting of syncopal events, and a combined outcome of falls and syncopal events.

The number of fallers (i.e. the number of participants reporting at least one fall over the 12 month period) would be reported as frequency and percentage out of those returning at least one fall diary. A mixed-effects Poisson regression model would be fitted to this binary outcome measure. Site would be included as a random effect and treatment assignment, baseline age (<80 years or >=80) and aetiology (neurogenic or non-neurogenic) as fixed effects. The log of the exposure variable (follow-up time) would be included in the model with the coefficient constrained to be one. From this model we would obtain the incidence rate ratio for the treatment effect comparing fludrocortisone and control and midodrine and control. These estimates would be reported with corresponding 95% confidence intervals.

Time to first fall would be summarised graphically using Kaplan-Meier curves, see Appendix Figure 3. A Cox proportional hazards regression model with mixed effects (i.e. a Cox regression model with shared frailty) would be fitted to the time to first fall data. Treatment assignment, baseline age (<80 years or >=80) and aetiology (neurogenic or non-neurogenic) would be included as fixed effects and site as a shared frailty term. Shared frailty terms will be assumed to follow a gamma distribution. From this model we would obtain the hazard ratio (HR) for the treatment effect comparing fludrocortisone and control and midodrine and control. These estimates would be reported with corresponding 95% confidence intervals. The proportional hazards assumption would be investigated using Schoenfeld residuals and by testing the interaction of treatment assignment with log(survival time). If the proportional hazards assumption was violated an alternative approach, such as the restricted mean survival time, would be used.

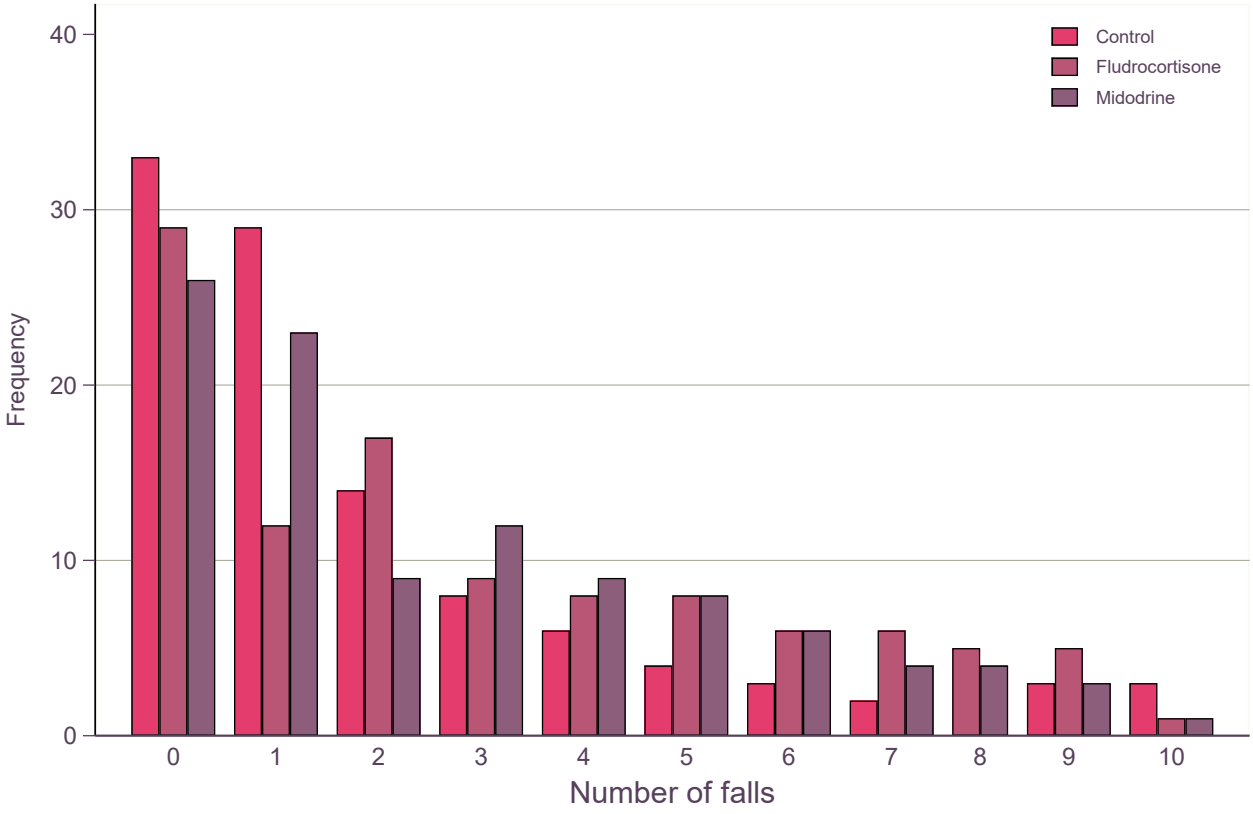
Fall-related injuries would be tabulated by type, with the number of participants affected and the total number of occurrences reported by randomised treatment group. The number and proportion of participants reporting at least one fall-related injury would be tabulated by treatment group. The total number of fall-related injuries reported per participant would be tabulated as frequency and percentage and as mean, SD, median, IQR and range. The total number of fall-related injuries reported in participants allocated to each treatment group would be presented. For the purpose of calculating percentages the number of participants returning at least one fall diary would be used as the denominator.

Appendix Table 3: Falls and syncopal events

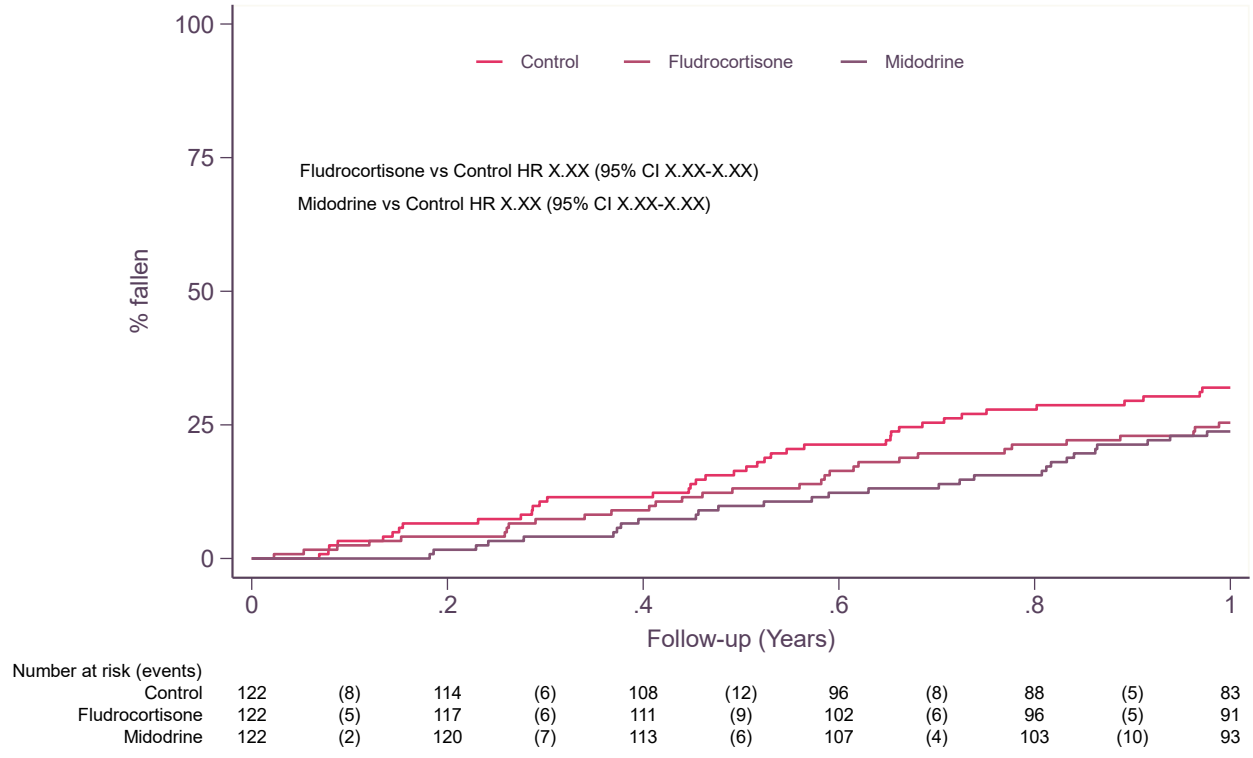
	Control N =	Fludrocortisone N =	Midodrine N =
Number of falls			
0			
1			
>2			
Mean (SD)			
Median (IQR)			
Range			
Fall rate			
Total number of falls			
Total follow-up time (years)			
Incidence rate (95% CI)			
Absolute difference* (95% CI)	NA		
Incidence Rate Ratio* (95% CI)	NA		
Fallers / non-fallers			
≥1 fall / N (%)			
Incidence Rate Ratio* (95% CI)	NA		
Number of syncopal events			
0			
1			
>2			
Mean (SD)			
Median (IQR)			
Range			
Syncopal event rate			
Total number of syncopal events			
Incidence rate (95% CI)			
Absolute difference* (95% CI)	NA		
Incidence Rate Ratio* (95% CI)	NA		
Number of falls and syncopal events			
0			
1			
>2			
Mean (SD)			
Median (IQR)			
Range			
Falls and syncope event rate			
Total number of events			
Incidence rate (95% CI)			
Absolute difference* (95% CI)			
Incidence Rate Ratio* (95% CI)			

**compared to control*

Appendix Figure 2: Number of falls (note this is not genuine data)



Appendix Figure 3: Time to first fall (note this is not genuine data)



Number at risk (events)

Control	122	(8)	114	(6)	108	(12)	96	(8)	88	(5)	83
Fludrocortisone	122	(5)	117	(6)	111	(9)	102	(6)	96	(5)	91
Midodrine	122	(2)	120	(7)	113	(6)	107	(4)	103	(10)	93

Standing blood pressure and postural blood pressure drop

The analysis methods described below would be applied to the following variables; nadir standing systolic blood pressure (BP), nadir standing diastolic BP, systolic postural BP drop, diastolic postural BP drop.

Analysis Set 1 would be used for analysis. All participants with data available would be analysed according to their randomised treatment allocation.

Each BP measurement would be summarised descriptively by randomised group. The number with data, mean, standard deviation and the median, IQR and range would be summarised at each visit. Data may have also been presented graphically by randomised treatment group as the mean value with 95% CIs at each visit.

As for the OHQ data, a linear mixed-effects model would have been fitted to the three, six and 12 month data. Fixed effects would include baseline BP measurement, baseline age (<80 years or >=80), aetiology (neurogenic or non-neurogenic) and a treatment assignment x time interaction term. Random effects would include site and individual (nested within site). This model would be fitted using REML and an unstructured covariance matrix.

Missing responses would be assumed to be missing at random (MAR). Participants would be included in the model if they had at least one BP measurement available from the three, six or 12 month follow-up visit.

From this model we would obtain the estimated mean difference in the three, six and 12-month follow-up BP measurements between fludrocortisone and control and between midodrine and control. This would be reported with a corresponding 95% confidence interval, see Appendix Table 4.

Assumptions of the model would be investigated as described for the analysis of the primary outcome.

Appendix Table 4: Blood pressure

Blood pressure	Control	Fludrocortisone	Midodrine	Mean difference (MD)	
				Fludrocortisone - control MD (95% CI) ; p-value	Midodrine - control MD (95% CI) ; p-value
	N; Mean (SD)	N; Mean (SD)	N; Mean (SD)		
Nadir standing					
Month 3					
Systolic					
Diastolic					
Month 6					
Systolic					
Diastolic					
Month 12					
Systolic					
Diastolic					
Postural drop					
Month 3					
Systolic					
Diastolic					
Month 6					
Systolic					
Diastolic					
Month 12					
Systolic					
Diastolic					

Hospital admissions

Hospital admissions would be analysed similarly to falls.

Analysis Set 1 would be used for analysis. All participants with at least one review of hospital admissions available would be analysed according to their randomised treatment allocation.

The number of admissions reported per participant would be tabulated as frequency and percentage and summarised as mean, SD, median, IQR and range. Reasons for admission would also be tabulated as frequency and percentage. The number, or percentage, of admissions may also have been presented graphically. The total number of admissions and total follow-up time from participants allocated to each randomised group would be reported. The simple, unadjusted incidence rate of admissions per person year (i.e. total number of admissions / total follow-up time) would be calculated in each randomised treatment group and presented with 95% CIs.

A mixed-effects negative binomial regression model would be fitted to the admissions data. Site would be included as a random effect and treatment assignment, baseline age (<80 years or >=80) and aetiology (neurogenic or non-neurogenic) as fixed effects. The log of the exposure variable (follow-up time) would be included in the model with the coefficient constrained to be one.

From this model we would obtain the estimated incidence rate ratio and risk difference in the admission rate per person-year between fludrocortisone and control and midodrine and control. These estimates would be reported with corresponding 95% confidence intervals.

As for falls data, a Poisson or zero-inflated model may have been used if this was more suitable for the distribution of the data.

c. Additional / Exploratory analyses

A pre-specified exploratory analysis would compare the clinical effectiveness of fludrocortisone with midodrine for the primary outcome. This would use the same model as for the primary analysis but with fludrocortisone as the reference group.

d. Missing data

i. Intention-to-treat / treatment policy estimand

The availability of outcome data would be summarised as described in section 4.1.3.

The characteristics of participants with and without a six month OHQ score would be explored. Characteristics to be summarised would have included baseline OHQ score, age, aetiology, gender, three month OHQ score (where available) and postural blood pressure drop (where available).

If more than 15% of participants had missing primary outcome data (OHQ score at six months), either overall or in any arm, then a sensitivity analysis of the primary outcome would be undertaken to explore the impact of departures from the MAR assumption, i.e. that the data are missing not at random (MNAR).

To do this we would implement reference-based multiple imputation (MI) using a jump-to-reference approach, following the guide proposed by Cro *et al* [10]. Missing data would be imputed assuming participants jump to behave like the usual care arm following their last observed timepoint. This would provide a conservative approach to missing data under a MNAR assumption. Note that any interim missing values would be imputed under a MAR assumption.

The variables to be included in the imputation model would be the same as those included in the primary analysis model. We would initially plan to impute 50 datasets, which would be very likely to be higher than the simple rule of thumb of one imputation per percent of missing data [11]. However, the number of imputations may have been increased if 50 imputations was not felt to provide adequate precision. Each imputed dataset would be analysed using the primary analysis model and Rubin's rules used to combine treatment estimates across datasets to give a single value [12].

ii. 'On-treatment' / hypothetical estimand

For the 'on-treatment' estimand we would implement controlled multiple imputation (MI) using a δ-based pattern-mixture approach, again following the guide proposed by Cro *et al* [8]. Briefly, missing primary outcome data would be imputed under the assumption that it is MAR, conditional on the other variables in the imputation model. The variables in the imputation model would be the same as those included in the primary analysis model. We would initially plan to impute 50 datasets, which would be very likely to be higher than the simple rule of thumb of one imputation per percent of missing data [11]. However, the number of imputations may have been increased if 50 imputations was not felt to provide adequate precision. A fixed value, δ, would then be added to the imputed values to increase the mean response beyond that predicted under MAR. Each imputed dataset would be analysed using the primary analysis model and Rubin's rules used to combine treatment estimates across datasets to give a single value [12]. We planned to use the following values of δ: +0.5, +1.0, +1.5, +2.0 and +2.5, but bounded to lie in the range 0 to 10. For example, for δ = 2.0 implies a missing OHQ score would be 2.0 points higher than expected if the data were missing at random. We would also explore how extreme a value a value of δ (positive or negative) would be needed to change the interpretation of the results (i.e. from $p < 0.05$ to $p \geq 0.05$ or vice-versa); this 'tipping point' value for δ would be reported and consideration given to how realistic or plausible this value would be in practice.

For each value of δ we would report the estimated mean difference in six-month OHQ between each experimental arm and control, along with 95% CIs and corresponding p-value, e.g. see Appendix Table 4 below.

Appendix Table 4: Exploring the impact of data being MNAR for the 'on-treatment' analysis of OHQ data

	Mean difference (MD)	
	Fludrocortisone - control	Midodrine - control
	MD (95% CI); p-value	MD (95% CI); p-value
δ = 0 (i.e. MAR)		
δ = 0.5		
δ = 1.0		
δ = 1.5		
δ = 2.0		
δ = 2.5		
δ = ? ('tipping point*')		

**may be different for each comparison*

We would also explore the possibility that data were only MNAR in the experimental arms by modifying the imputed values only for those allocated to conservative management plus fludrocortisone or midodrine. We would also determine the 'tipping point' value for δ in this scenario.

We did not anticipate missing baseline data for the variables included in the primary analysis model as these are either stratification factors and hence required for randomisation or required to confirm eligibility (i.e. the OHQ score). If however, the baseline OHQ score was missing we would use simple imputation and replace the missing value with the mean of all non-missing baseline values [13].

REFERENCES

1. Kaufmann H, Malamut R, Norcliffe-Kaufmann L, Rosa K, Freeman R: **The Orthostatic Hypotension Questionnaire (OHQ): validation of a novel symptom assessment scale**. *Clin Auton Res* 2012, **22**(2):79-90.
2. Frith J, Newton JL: **Validation of a questionnaire for orthostatic hypotension for routine clinical use**. *Geriatr Gerontol Int* 2016, **16**(7):785-790.
3. Nouri FM, Lincoln NB. **An extended activities of daily living scale for stroke patients.** Clinical Rehabilitation. 1987;**1**:301-5.
4. Bell ML, Fairclough DL. **Practical and statistical issues in missing data for longitudinal patient-reported outcomes.** *Statistical Methods in Medical Research* 2014, **23**(5):440-459.
5. Rodgers H, Howel D, Bhattarai N, Cant R, Drummond A, Ford GA, et al. **Evaluation of an Extended Stroke Rehabilitation Service (EXTRAS): A Randomized Controlled Trial and Economic Analysis.** *Stroke* 2019, **50**(12):3561-3568.
6. Spice CL, Morotti W, George S, Dent THS, Rose J, Harris S, Gordon CJ: **The Winchester falls project: a randomised controlled trial of secondary prevention of falls in older people**. *Age & Ageing* 2009, **38**(1):33-40.
7. Ciaschini PM, Straus SE, Dolovich LR, Goeree RA, Leung KM, Woods CR, Zimmerman GM, Majumdar SR, Spadafora S, Fera LA *et al*: **Community-based intervention to optimise falls risk management: a randomised controlled trial**. *Age Ageing* 2009, **38**(6):724-730.
8. Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE: **Systematic Review of Levodopa Dose Equivalency Reporting in Parkinson's Disease.** *Movement Disorders* 2010, **25**(15): 2649-2684.
9. **ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials.** EMA/CHMP/ICH/436221/2017.
10. Cro S, Morris TP, Kenward MG, Carpenter JR: **Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputaion: A practical guide.** *Statistics in Medicine* 2020; **39**:2815-2842.
11. White IR, Royston P, Wood AM: **Multiple imputation using chained equations: issues and guidance for practice.** *Statistics in Medicine* 2011;**30**(4):377-399.
12. Rubin, DB: **Multiple Imputation for Nonreponse in Surveys.** *Wiley Series in Probability and Mathematical Statistics.* New York, NY: John Wiley & Sons; 1987.
13. White IR, Thompson SG: **Adjusting for partially missing baseline measurements in randomized trials.** *Statistics in Medicine* 2005;**24**(7):993-1007.