Statistical analysis plan – before data was unblinded

The University Of Sheffield.

Sheffield Teaching Hospitals NHS
NHS Foundation Trust

A randomised controlled trial in patients with Respiratory Muscle Weakness due to Motor Neurone Disease of the NeuRx/4 Diaphragm Pacing Trial

Statistical analysis plan
version 1.0, 30 October 2012

Authored by

_____ __/__/____
Mike Bradburn *Date*
Trial statistician
CTRU, University of Sheffield

Approved by

_____ __/__/____
Chris McDermott *Date*
Chief investigator
SITraN, University of Sheffield

_____ __/__/____
Lyn Taylor *Date*
TSC statistician
ScHARR, University of Sheffield

_____ __/__/____
Carolyn Young *Date*
TSC Chair
The Walton Centre, Liverpool

List of abbreviations used

AE	Adverse event
ALS	amyotrophic lateral sclerosis
ANCOVA	Analysis of Covariance
CI	confidence interval
CTRU	University of Sheffield Clinical Trial Research Unit
DMEC	Data Monitoring and Ethics Committee
DP	Diaphragm pacing
EQ-5D	EuroQol health utility questionnaire
FVC	forced vital capacity
GCP	Good Clinical Practice
ITT	intent-to-treat
MND	Motor neurone disease
NIV	non invasive ventilation
PH	Proportional hazards
PP	per-protocol
QoL	Quality of life
SAE	Serious adverse event
SAQLI	sleep apnoea quality of life index
SF-36	Short form-36 questionnaire
TMG	Trial Management Group
TSC	Trial Steering Committee
VAS	visual analogue scale

Introduction, study design and key trial objectives

Study outline

The DiPALS study is a two-arm, parallel group, open-label randomised controlled clinical trial in patients with motor neurone disease (MND) or amyotrophic lateral sclerosis (ALS) with respiratory muscle weakness. Patients will be allocated to either non invasive ventilation (NIV) with diaphragm pacing (DP), or to NIV alone. NIV alone is the current standard care, and is the control group.

This statistical analysis plan is written in conjunction with the International Conference on Harmonisation topic E9 (Statistical principles for clinical trials, 1998), applicable standard operating procedures from the University of Sheffield Clinical Trials Research Unit (CTRU) and trial documents referenced in section 6. The trial will be conducted in accordance with Good Clinical Practice (GCP) in Clinical Trials.[1]

Outcome measures

The primary objective of this trial will be to evaluate the effect of DP on overall survival.

Secondary objectives will be to evaluate the effect of DP on the following:

- Quality adjusted life years (QALYs, not covered in this document)
- Quality of life: sleep apnoea quality of life index (SAQLI), and the Short Form-36 (SF-36) version 1.
- Quality of life of the main carer, measured by Caregiver Burden Inventory (CBI).
- Safety and tolerability.
- Health economic objectives and resource use (not covered in this document).
- Qualitative user perspectives (not covered in this document).

Randomisation

Patients will be allocated to their treatment by minimisation which is carried out via a web-based interface hosted by the CTRU. The minimisation factors are baseline bulbar function (mild, moderate, severe),forced vital capacity (FVC) at baseline (50-59%, 60-69%, 70+%), age (<=39, 40-79, 80+) and sex.

Interim analyses, data monitoring committees etc.

Three committees will be established to govern the conduct of this study:

- Trial Steering Committee (TSC)
- Independent Data Monitoring and Ethics Committee (DMEC)
- Trial Management Group (TMG)

Periodic efficacy and safety analyses will be provided to the DMEC. These summaries will have the treatment groups labelled as “A” and “B” , but may be fully unblinded at the discretion of the DMEC. The TSC and TMG will see overall results only.

Sample size calculation

The sample size calculation is based on log-rank test, using Simpson's rule [2] as implemented in Stata version 11.1 [3] to allow for the unequal length of follow-up. The study duration comprises an 18-month recruitment period and a 12-month follow-up period, giving a maximum follow-up of 30 months and a minimum of 12 months. Assuming control group survival proportions of 45%, 20% and 10% at the minimum, average and maximum follow-up times respectively, a hazard ratio of 0.45 and an additional 10% loss-to-follow-up, a total of 108 patients (54 per group) are needed to ensure a power of 85% using a two-sided type I error of 5%. The control group figures are conservative estimates based on the sole randomised controlled trial of NIV, which is now considered standard care in the UK. A study carried out in the United States has estimated a one year survival of 86% after study entry for patients using DP and NIV. We have estimated the sample size on a conservative (but clinically important) 1-year difference in survival of 45% versus 70%, which produces the estimated hazard ratio of 0.45. It is anticipated that we will have complete survival data on all subjects recruited, based on previous experience in MND trials.

With regard to quality of life data we anticipate a low level of missing data due to loss to follow up. We have reviewed the patients who were initiated on NIV in the year up to Jun 2009 and we have maintained contact with 100% of those patients surviving at 12 months. The appointment of a research nurse at each study site will enable home visits if necessary to collect the quality of life data. We have however allowed for a 10% loss to follow up in the sample size/power calculation.

Additional patients may be recruited if less than 54 patients receive their allocated treatment.

Data sources, data and analysis populations

Data sources

The data used in this study will come from data entered onto the following sources:
Case Report Form (CRF) version 1, 11 November 2011
Screening log version 1, 26 October 2011
NIV group patient diary version 2, 3 August 2011
DP group patient diary version 2, 3 August 2011

The database will be stored on the CTRU database (PROSPECT) with the exception of the randomisation list which is held on the CTRU's randomisation system. Electronic data will be extracted from the system during the trial for the purpose of checking (validating) and trial progress reports; however, access to any data which would unblind the study (randomised group, resource usage) will be limited to members independent of the trial (e.g. DMEC statistician). Personal records will not be made available to CTRU staff.

Protocol Deviations

The following will be classified as protocol deviations in the statistical analysis:

1. Patients who are randomised in error (any of the inclusion/exclusion criteria are breached)
2. Patients who do not tolerate NIV
3. (DP group only) Patients who do not have a successful DP implantation
4. (DP group only) Patients who do not use DP

The above criteria will be used in explanatory analyses. All are objective criteria which will be determined from the clinical database and verified by the CI at database lock and before analysis is undertaken. Criteria 2 will also be used as a basis for subgroup analyses.

Analysis populations

Three analysis sets will be used:

Name	Patients included	Treatment group
Intent-to-Treat (ITT)	All patients for whom consent is obtained and who are randomised into the trial	As randomised
Per Protocol (PP)	The subset of the ITT patients who do not violate the protocol, as defined in section 2.2.	As randomised
Safety	All patients who proceed to NIV, with or without DP	As treated

All summaries will be based on the ITT set, with the following exceptions:

- Effectiveness analyses will also be undertaken on the PP set. Additional analyses to assess the impact of compliance, for example complier-average causal effect (CACE) methodology, may also be undertaken if there are a substantial number of withdrawals from treatment or crossing between groups.
- Analysis of adverse events (AEs) will be undertaken on both the ITT and safety sets. In the latter, patients will be analysed by the treatment they were using at the point of the adverse event.

Outline of analyses
General considerations

Data will be reported and presented according to the revised CONSORT statement.[4,5]

Complete details of data derivations and methods of handling multiplicity, multi-centre data and missing data are covered in section 4. No interim analyses or early stopping are planned.

All summaries will be provided on the full analysis set using an intention to treat (ITT) approach. If loss-to-follow-up and/or treatment crossover is greater than 5%, sensitivity analyses will be performed to assess the robustness of the results. If important discordances are observed between these and the ITT approach, both analyses will be reported.

All summary tables will present summary statistics within each treatment group and overall unless stated otherwise.

Summaries of continuous variables will comprise the number of observations used and either
i) mean, standard deviation, median, minimum and maximum, or
ii) median, inter-quartile range, minimum and maximum
as appropriate for the distributional form of the data.

Summaries of categorical variables will comprise the number of observations used, and the number and percentage of observations in each category.

Summaries of time-to-event outcome data will comprise the median and inter-quartile range of survival within each group, the hazard ratio and its 95% confidence interval.

Summaries of QoL outcomes will comprise the following:

- A summary (mean, standard deviation, median, minimum and maximum) of the QoL, by treatment group and time point.
- The least squared mean QoL change from baseline at each time point, together with its standard error, by treatment group and time point.
- The least squared mean difference between treatment groups ,together with its 95% confidence interval and p-value, by time point
- The overall comparison between the treatment groups, together with its 95% confidence interval and p-value, and a test for interaction between treatment group and time. Both are based on longitudinal modelling.

All treatment comparisons will use the NIV only group as the reference (comparator), all statistical exploratory tests of main effects will be two-tailed with alpha = 0.05; and all confidence intervals (CIs) will be two-sided, 95% intervals. As there is controversy with regards to the operating characteristics of minimisation, a permutation test will be used to confirm the p-value from the primary endpoint.[6] Since interaction tests have low statistical power, additional consideration will be given to p-values below 0.1 when testing interactions (treatment x centre and treatment x subgroups).

Disposition and data completeness

The following summary will be presented for all patients screened for entry to the study, by centre and overall:

Enrolment	The "CONSORT" diagram comprising the number of people screened, eligible, consented, randomised, receiving their allocated treatment, withdrawing/lost to follow-up, and included in ITT and PP analysis sets.
Characteristics of non-enrolled patients	The age, gender and reason for non-inclusion.

In addition the following by-patient line-listings will be provided

Reasons for non-randomisation	For people screened but not randomised: - A list of reasons for non-recruitment (any/all that apply) - A list of reasons for non-consent
Reasons for non-implantation	For people allocated pacing device: - A list of reason(s) for not proceeding with surgery (FVC not acceptable, failure of pre-operative checks, withdrawal, unable to stimulate diaphragm, others)

The inclusion of patients in each analysis population is outlined in section 2.3. Where the randomised intervention (as recorded on the randomisation list) differs from the intervention group as recorded on the case record form, the randomisation list will be assumed to be the correct data source. The following summary will be provided, by treatment group and overall:

Attrition, compliance and analysis sets	The number and percentage of patients who a. complete each visit b. are lost to follow-up c. do not receive their allocated treatment or switch treatment d. are in each analysis population as defined in 2.3

During the recruitment and follow-up periods, the following summaries will be made available to members of the TMG, TSC and DMEC.

Data completeness	The number of patients with complete data for each key parameter, by centre.

The inclusion of key parameters may be allowed to vary at the request of the TMG, TSC or DMEC during the trial. In order to allow time for data to be entered onto the system, data items will be considered complete if they have been entered within 30 days of scheduled visit date, and otherwise incomplete.

Demographics and baseline characteristics

The following summaries will be presented:

Demographics and vital signs	Centre, Age, gender, ethnic category, height, weight.
MND characteristics	Duration of MND symptoms, type of onset, site of onset, El Escorial diagnosis, bulbar function, ALSFRSr
Baseline severity	FVC, Supine VC, SNIP, PaCO2, and the number and percentage of patients with respiratory insufficiency, unacceptable phrenic nerve function, ECG abnormality, and blood test abnormality
Physical examination	The number of patients with abnormalities in each body system
Medical history	The number of patients with each medical history, past or present
Pre-operative checks and surgery details	The number and percentage of patients who meet checks for surgery at initial assessment, at repeat assessment, and who fail and are withdrawn by reason (parameter not met); the number and percentage of patients where DP was not fitted because diaphragm could not be stimulated; the number and percentage of patients with complications (by reason); and the number and percentage of patients with a surgical AE (DP only)

In addition, the following by-patient line listing will be provided:

Surgical complications	A list of all surgical complications recorded

Efficacy

Primary endpoint

The primary efficacy endpoint will be overall survival, defined as the duration from randomisation to death of any cause.

The following summaries will be presented:

Overall survival	A comparison of the overall survival by treatment group.
Overall survival by NIV tolerance	A comparison of the overall survival by treatment group and by NIV tolerance subgroup (tolerant versus intolerant)
Overall survival by bulbar function	A comparison of the overall survival by treatment group and by bulbar function subgroup (mild/moderate versus severe)

NIV tolerance is defined in section 3.5.2 and bulbar function is defined in section 4.3.

Analysis will be undertaken by Cox proportional hazards (PH) regression, with covariates including treatment group and the minimisation factors. The model fit will be assessed as described in section 4.4. Kaplan-Meier survival curves will be presented overall and for each subgroup. Differences in survivorship between subgroups will be tested by inclusion of the covariate in the model along with treatment group. Differences in the effect of treatment between subgroups will be tested using an interaction term between the two.

Secondary outcomes
Patient health utility and QoL

The patient will complete EQ-5D and EQ-5D VAS at six time points (screening, 2, 3, 6, 9 and 12 months), and the SAQLI and SF-36 questionnaires at five time points (screening, 2, 3, 6 and 12 months).

EQ-5D and EQ-5D VAS

Both will all be analysed in four ways as described below:

1. All patients, at end of follow-up: A change from baseline to 12-month analysis (all patients; deaths imputed with worst-case).
2. All patients, longitudinal analysis: A repeated measures analysis, in which the trajectory across time is modelled. This will include all patients, including those who die; deaths will have values imputed with death state case (zero) for both EQ-5D and EQ-5D VAS.
3. Survivors, at end of follow-up: A change from baseline to 12-month analysis (surviving patients only)
4. Survivors, longitudinal analysis: A repeated measures analysis, in which the trajectory across time is modelled (surviving patients only).

The 12-month change in EQ-5D and EQ-5D VAS will be compared between groups using analysis of covariance (ANCOVA) in which the treatment group and the baseline score are included as covariates along with minimisation factors. Repeated measures analyses will be undertaken using a generalised least squares model repeated measures ANCOVA with the same covariates.

EQ-5D and EQ-5D VAS will be summarised as described in 3.1 (QoL outcomes), with the following summaries presented:

Health utility, all patients	EQ-5D and EQ-5D VAS
Health utility, survivors	EQ-5D and EQ-5D VAS

The 12-month change in EQ-5D and EQ-5D VAS will be compared between groups using analysis of covariance (ANCOVA) in which the treatment group and the baseline score are included as covariates along with minimisation factors. Repeated measures analyses will be undertaken using a generalised least squares model repeated measures ANCOVA with the same covariates. The model will use an exchangeable correlation matrix structure, and fit of this model will be tested against an unstructured model (and potentially other alternatives) using Hausman's test.

EQ-5D will be further reported by subgroups (NIV tolerance and bulbar function). Testing for differential treatment effect between subgroups would necessitate a 3-way interaction (treatment group x subgroup x time), which given the sample size would produce potentially unstable coefficients. Therefore, the focus here will be on within-group summary statistics and graphical displays, separately by treatment group.

SF-36

The SF-36 will be used to calculate 8 scales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, social functioning and role-emotional) and 2 component summary measures (physical health and mental health). Each of the domains will be rescaled to be between 0-100. All 10 summary measures will be summarised by time, but the longitudinal modelling will only be applied to physical and mental health. Unlike EQ-5D, summaries will comprise only patients who are alive at each time point.

The following will be presented:

SF-36	SF-36 domains by time point

The analysis strategy will match that of the EQ-5D.

SAQLI

This one-domain questionnaire will be analysed in the same manner as SF-36 physical and mental health domains.

Carer QoL

The patient's main carer will complete EQ-5D and EQ-5D VAS at six time points (screening, 2, 3, 6, 9 and 12 months), and the CBI at five time points (screening, 2, 3, 6 and 12 months).

EQ-5D

Where the patient has a nominated carer, they will complete the EQ-5D and EQ-5D VAS at the same time points as the patient. Analysis will be as for patient EQ-5D survivors only: no attempt will be made to impute carer EQ-5D for whom the patient has died.

CBI

This one-domain questionnaire will be analysed in the same manner as carer EQ-5D.

Admissions / resource

The following will be presented by timepoint visit among patients for whom complete diary data are available:

Resource use	The number and percentage of patients having each of the following, and the number and percentage of episodes of each: 1) Services - hospital admission (with duration) - emergency department attendance - minor injury clinic or walk in centre - general practitioner referral 2) Devices - use of CoughAssist - use of Breath-Stacking - use of suction 3) Health and social care - physiotherapist (home visit, outpatient visit) - occupational therapist (home visit, outpatient visit) - other (home visit, outpatient visit) 4) Additional care/support - formal (e.g. home help) - informal (e.g. family/friends)

Safety outcomes
Compliance and machine settings (diaphragm pacing)

Compliance and machine settings for both NIV and DP will be collected in and summarised for in each assessment period separately. Assessments are carried out at baseline, 1 week post-surgery (DP group only), and at months 2,3,6,9 and 12. Compliance will be reported by time period between these visits, i.e. 0-2 months, 2-3 months, 3-6 months, 6-9 months and 9-12 months.

The following summaries will be provided, for the DP group only:

DP usage	The number and percentage of DP users and average DP usage, by time period
DP parameter settings	The number and percentage of patients with one or more channel X within each time period; and a summary of each parameter (amplitude, pulse, respiratory rate, inspiratory interval, pulse frequency and pulse ramp) for each channel, and for each parameter the percentage of patients whose parameters changed, within each time period

Patients are defined as DP compliant if their DP usage has reached at least 4 hours per day one month after implantation. The percentage of users in each time period will be defined as

Percentage of DP users =
100 x number of DP compliant patients in time period / number of patients alive at end of time period.

Compliance and machine settings (NIV)

The following summaries will be presented:

NIV usage	The number and percentage of patients who are NIV compliant, average NIV usage and percentage of target usage, by time period
NIV parameter settings	A summary of each parameter (minimum volume assured pressure support, maximum volume assured pressure support, inspiratory positive airways pressure, expiratory positive airways pressure, target tidal volume, respiratory rate, type of machine, type of mask, and humidification), and for each parameter the percentage of patients whose parameters changed, within each time period

NIV compliance is reported by time period defined identically to DP (above) Patients are defined as NIV compliant within a time period if their NIV usage is at least 4 hours per night. The percentage of users will be defined as

Percentage of NIV users =
100 x Number of NIV compliant patients in time period / number of patients alive at end of time period

Percentage of target time used will be defined *among users only* as

Percentage of target usage achieved =
100 x average use in time period / target usage within period

In addition, the following by-patient line-listing will be presented:

Device technical issues	Reported technical problems or other observations on DP and/or NIV, by time period

Adverse events

The safety and tolerability profiles will be reported by analysing the proportion of patients experiencing adverse outcomes. The following summaries will be presented:

AEs	The number and percentage of patients reporting an AE, by type
Serious AEs (SAEs)	The number and percentage of patients reporting an SAE, by type
Treatment-related AEs	The number and percentage of patients reporting a DP-related AE, by type The number and percentage of patients reporting a NIV-related AE, by type

"Related" will be defined as those AEs recorded as definite, probable or possible.

The following by-patient line listings will be presented:

All AEs	A listing of all AEs including - Treatment group (if the patient switches treatment groups, details will be included)

	- Description - Days from treatment commencement to AE onset (if patient switches treatment group, this will the be most recent treatment - Severity - Relationship - Outcome - Seriousness
All SAEs	A listing of all SAEs (as "all AEs" with the omission of "serious")
All treatment-related AEs	A listing of all treatment-related AEs (as "all AEs" with the omission of "relationship")

Concomitant medications

The following summary will be presented

Concomitant medication	The number and percentage of patients taking each medication

Detailed statistical methods and calculations
General considerations

4.1.1 Number and timing of analyses – adjustment for multiplicity

The study may stop prematurely on grounds of safety or futility. However, no formal interim analyses will be performed for efficacy, and consequently no adjustment for multiplicity will be made to the significance levels.

4.1.2 Missing, spurious and unused data

Missing data is not expected for the primary outcome (overall survival). If the patient is lost to follow-up, their survival time will be censored at the date last known alive. For the questionnaire-based QoL outcomes, missing data may arise at any timepoint in one of four ways:

1) Questionnaire not completed, patient has died
2) Questionnaire completed but not in the correct time frame (i.e. outside visit window)
3) Questionnaire incomplete (some questions missing)
4) Questionnaire not returned (all questions missing)

1) is addressed in section 3.4.2.

2) will be handled by imposing the following visit windows:

Time point	**Lower limit**	**Upper limit**
Screening	No limit	Date of randomisation
2 months	Date of randomisation	75 days post randomisation [2 months + 15 days]
3 months	76 days post randomisation	135 days post randomisation [3 + 1.5 months]
6 months	136 days post	225 days post randomisation [6 + 1.5 months]

9 months	226 days post randomisation	post	315 days post randomisation [9 + 1.5 months]
12 months	316 days post randomisation	post	405 days post randomisation [12 + 1.5 months]

Scenario 3) is covered in the individual sections below.

For scenario 4), the same approach will apply to all questionnaires:

i) If the questionnaire from an adjacent visit falls within the visit window, use this value.

ii) If i) has not imputed a value, but values are available both before and afterwards, this will be imputed using the trapezoid method:

$Q_t = [Q_1 x (t_2 - t) + Q_2 x (t - t_1)] / (t_2 - t_1)$

Where
Q_t is the imputed quality of life at time t
(t_1, t_2) are the time points immediately prior to and following time t ($t_1<t<t_2$) at which valid responses exist
Q_1 and Q_2 are the responses at times t_1 and t_2.

Illustration

Suppose a patient has data as follows, for which approaches 1), 2), 3) and 4i) have not dealt with.

Time point	EQ-5D
2 months	0.7
3 months	missing
6 months	0.5

The missing value is then calculated by

EQ-5D at t=3 = [0.7 x(6-3) + 0.5 x(3-2)] / (6-2) = [0.7x3 + 0.5x1]/4 = 0.65

iii) If missing data still persist, there is no consensus of how to proceed. Since MND is a deteriorating condition the use of last observation carried forward will lead to an overestimation of QoL and, depending on the drop-out patterns between the two groups, a potentially biased comparison (see, for example Saha and Jones 2009 [7]). Other approaches, such as multiple imputation or group mean imputation among survivors, have their own advantages and disadvantages. At the time of writing, research is ongoing within ScHARR to assess approaches to precisely this problem. The preliminary report is expected in late 2011, and an amendment to the analysis plan will be made in 2012 to cover this in more detail.

4.1.3 Analysis sets

The ITT will be the primary analysis population for effectiveness outcomes, with the results for the PP being supportive of it. If for any endpoint the populations confer inconsistent results, further analyses will attempt to investigate the reason for this.

4.1.4 Methods for dealing with multi-centre data

The consistency of outcomes among the treatment centres will be assessed by fitting a model which includes an interaction between treatment group and centre. If the test for interaction is not statistically significant the interaction term will be removed. If significant differences are found, further analyses will be undertaken to assess whether this may be due to differences in case mix (i.e. an artefactual centre effect) or not (i.e. real centre effect).

Disposition and data completeness

Recruitment data, data completeness and patient demographic & characteristics will be reported to the TSC, DMEC and TMG in an ongoing fashion.
Demographics and baseline patient characteristics

The baseline date is the date of randomisation. The centre will be defined as the centre at which the patient first attended. Age is defined as (date of baseline – date of birth).

ALSFR is a one-domain questionnaire comprising 24 questions, each of which is scored from 0-4. The overall score is the sum of these. If the questionnaire is incomplete but at least half of the questions answered (i.e.at least 12 of the 24), the overall score will be multiplied up by the formula

Overall score = (24 / Number of questions answered) x (total score among answered questions)

If fewer than 12 questions are answered, the questionnaire will be treated as missing and will not be used in summaries.

The bulbar function score is calculated from the ALSFRS-r. The answers to the first three questions (speech, swallowing and salivation) are summed, and the bulbar function is categorised from this sum into the following: mild (0-4), moderate (5-8) and severe (9-12).

Efficacy
Primary endpoint

The primary endpoint is overall survival, defined as the time between randomisation and death. If no notification of death has been received, the patient will be censored at the date last known alive.

Pre-trial modelling, undertaken at the proposal development stage, found PH to be the best fit to previous data, and therefore a Cox PH regression model will be fitted with ties handled by the Efron method. The PH assumption will be checked by adding time-dependent covariates and graphing scaled Schoenfeld residuals against time.[8] If PH is found not to fit the data adequately, an AFT alternative will be fitted and the adequacy of its fit assessed using Q-Q plots.[9] If this too does not fit, a residual life analysis [10] will be used as the basis for summarising the treatment effect

Secondary endpoints

SAQLI (Sleep Apnoea Quality of Life Index)
SAQLI is a one-domain questionnaire comprising 14 questions, each of which is scored from 1-7. The overall score is the average of these. If the questionnaire is incomplete (i.e. less than 14 questions are answered), the overall score will be defined as the average provided at least half of the questions (7 of the 14) have been answered. If fewer than 7 questions are answered, the questionnaire will be treated as missing; no further imputation will be undertaken.

SF-36
Version 1 of the SF-36 will be used. The domain scores will be calculated using the standard RAND organisation algorithm.

CBI
CBI is a one-domain questionnaire comprising 24 questions, each of which is scored from 0-4. The overall score is the total of these. Incomplete questionnaires will be handled in the same manner as SAQLI

Safety

The DP and NIV usage will be based on the values recorded in the CRF, which are in turn estimated from the patient diary. Additional exploratory analyses using the diary usage data may also be undertaken but are not described here.

For each AE, the report will comprise the number and percentage of patients affected, and the number of events (a patient may have more than one occurrence of the same AE)

Modifications to the original protocol analysis statement

Not applicable.

References

Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis Part III: multivariate data analysis -- choosing a model and assessing its adequacy and fit. *British Journal of Cancer* 2003; 89(4):605-11.

Committee for proprietary medicinal products (CPMP). Points to consider on adjustment for baseline covariates. *Statistics in Medicine (2004);* 23:701–709

ICH Harmonised Tripartite Guideline Topic E6(R1): Guideline for Good Clinical Practice (1996). Available at: http://www.ich.org/LOB/media/MEDIA482.pdf (last accessed 13 January 2010)

ICH Harmonised Tripartite Guideline Topic E9: Statistical Principles For Clinical Trials (1998). Available at: http://www.ich.org/LOB/media/MEDIA485.pdf (last accessed 13 January 2010)

Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, for the CONSORT Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. BMJ 2010;340:c869

RAND health Medical Outcomes Study: 36-Item Short Form Survey Scoring Instructions. http://www.rand.org/health/surveys_tools/mos/mos_core_36item_scoring.html (last accessed 15 November 2011)

Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine* 2011;30(19):2409-21.

Saha C, Jones MP Bias in the last observation carried forward method under informative dropout. *Journal of Statistical Planning and Inference* 2009; 139: 246-255.

Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39:499-503

Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal* 2010;340:c332.

StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.

Scott NW, McPherson GC, Craig R. Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. *Controlled Clinical Trials* 2002; 23:662–674.

Therneau TM, Grambsch PM. *Modeling Survival Data Extending the Cox Model*. Springer-Verlag, New York 2000.

Statistical analysis plan – after data was unblinded

A randomised controlled trial in patients with Respiratory Muscle Weakness due to Motor Neurone Disease of the NeuRx/4 Diaphragm Pacing Trial

Statistical analysis plan
version 2.0 Incorporating amendment 1, 18 November 2014

Authored by

_____ __/__/____
Mike Bradburn *Date*
Trial statistician
CTRU, University of Sheffield

Approved by

_____ __/__/____
Chris McDermott *Date*
Chief investigator
SITraN, University of Sheffield

_____ __/__/____
Lyn Taylor *Date*
TSC statistician
PAREXEL International, Sheffield[1]

[1] Lyn Taylor was a PhD student at the University of Sheffield between 2011 and 2014. PAREXEL International have had no involvement in the design or conduct of the study.

_____ __/__/____

Carolyn Young *Date*
TSC Chair
The Walton Centre, Liverpool

List of abbreviations used

AE	Adverse event
ALS	amyotrophic lateral sclerosis
ANCOVA	Analysis of Covariance
CI	confidence interval
CTRU	University of Sheffield Clinical Trial Research Unit
DMEC	Data Monitoring and Ethics Committee
DP	Diaphragm pacing
EQ-5D	EuroQol health utility questionnaire
FVC	forced vital capacity
GCP	Good Clinical Practice
ITT	intent-to-treat
MND	Motor neurone disease
NIV	non invasive ventilation
PEG	Percutaneous endoscopic gastrostomy
PH	Proportional hazards
PIG	Per-oral image guided gastrostomy
PP	per-protocol
QoL	Quality of life
SAE	Serious adverse event
SAQLI	sleep apnoea quality of life index
SF-36	Short form-36 questionnaire
TMG	Trial Management Group
TSC	Trial Steering Committee
VAS	visual analogue scale

Introduction, study design and key trial objectives

Study outline

The DiPALS study is a two-arm, parallel group, open-label randomised controlled clinical trial in patients with motor neurone disease (ALS) or amyotrophic lateral sclerosis (ALS) with respiratory muscle weakness. Patients will be allocated to either non invasive ventilation (NIV) with diaphragm pacing (DP), or to NIV alone. NIV alone is the current standard care, and is the control group.

This statistical analysis plan is written in conjunction with the International Conference on Harmonisation topic E9 (Statistical principles for clinical trials, 1998),[33] applicable standard operating procedures from the University of Sheffield Clinical Trials Research Unit (CTRU) and trial documents referenced in section 6. The trial will be conducted in accordance with Good Clinical Practice (GCP) in Clinical Trials (International Conference on Harmonisation, 1996).[34]

Outcome measures

The primary objective of this trial will be to evaluate the effect of DP on overall survival.

Secondary objectives will be to evaluate the effect of DP on the following:

Quality adjusted life years (QALYs, not covered in this document)
Quality of life: sleep apnoea quality of life index (SAQLI), and the Short Form-36 (SF-36) version 1.
Quality of life of the main carer, measured by Caregiver Burden Inventory (CBI).
Safety and tolerability.
Health economic objectives and resource use (not covered in this document).
Qualitative user perspectives (not covered in this document).

Randomisation

Patients will be allocated to their treatment by minimisation which is carried out via a web-based interface hosted by the CTRU. The minimisation factors are baseline bulbar function (mild, moderate, severe),forced vital capacity (FVC) at baseline (50-59%, 60-69%, 70+%), age (<=39, 40-79, 80+) and sex.
Interim analyses, data monitoring committee, and early termination of DiPALS

Trial governance committees

Three committees will be established to govern the conduct of this study:

Trial Steering Committee (TSC)
Independent Data Monitoring and Ethics Committee (DMEC)
Trial Management Group (TMG)

Periodic efficacy and safety analyses will be provided to the DMEC. These summaries were originally envisaged as semi-blinded summaries with the treatment groups labelled as “A” and “B” , but the DMEC subsequently exercised their discretion to see fully unblinded summaries. The TSC and TMG have seen overall results only and remain blinded.

Early stopping

In December 2013 the DMEC recommended that recruitment to DiPALS should cease on safety grounds, citing a discrepancy in survival between the two arms. In summary, they recommended that:
Recruitment should be suspended with immediate effect
Implantation of new pacing devices be suspended
Other aspects of the trial remain unaltered; in particular, that patients in the pacing arm should be encouraged to continue using their device.

In doing so, they acknowledged the sample size was relatively small (74 randomised, 24 deaths at the point of this recommendation), and that their decision would be reviewed as additional data became available. Their recommendations were upheld by the TSC, who requested the DMEC step up the frequency of their meetings to every three months.

The recommendation to formally stop recruitment was made by the DMEC in June 2014. This time, their recommendations went further:
That participants in the pacing arm be informed of the concern and advised to cease use of their device forthwith (unless the patient and their clinician believed there were just grounds to do otherwise).
That the trial follow-up continue until all participants had either died or completed 12 month follow-up
That the TSC and the trial team remain blind to the outcome data until this time.

The recommendation was accepted by the TSC, with the exception of the last point. The independent TSC members voted unanimously that some preliminary results be presented before the end of follow-up (expected to be December 2014). Although unusual, the reasons for this were that: the safety concern should not be withheld from a wider audience; the pacing device was becoming more widely used; and the information given to trial patients as above had revealed the concern about pacing to the ALS community. They believed there was good ethical reason to present preliminary data to the ALS community at the major annual international conference in December 2014. The DMEC subsequently agreed to this.

The scope of the interim analysis on preliminary data is outlined in section 3.6. At present, the DMEC and the responsible CTRU statistician remain the only persons to have seen unblinded outcome data.

Wider context: other studies of DP and ALS

Although the specific details (and unblinded data) remain confidential at this stage, the DMEC have indicated survival in the pacing arm is inferior to the control group. A natural follow-on from this is to attempt to reconcile the disparity between our findings and those which were submitted to (and approved by) the FDA. Unusually, the license for DP was granted on the basis of a relatively small, non-randomised cohort of patients with ALS with respiratory failure ("the FDA cohort"), recognising the rarity of the condition and the lack of curative therapies. A part of the analysis will be to compare – informally – the DiPALS trial data against the FDA cohort with a view to understanding the reasons behind the apparently different outcomes.

At present there are no randomised trials of DP in this therapeutic area, although one is planned; a second (albeit in ALS without respiratory failure) is currently ongoing. There is a possibility of future analyses combining data from these studies: these are not covered here.

Sample size calculation

The sample size calculation was based on log-rank test, using Simpson's rule [18] as implemented in Stata version 11.1 [19] to allow for the unequal length of follow-up. The study duration comprises an 18-month recruitment period and a 12-month follow-up period, giving a maximum follow-up of 30 months and a minimum of 12 months. Assuming control group survival proportions of 45%, 20% and 10% at the minimum, average and maximum follow-up times respectively, a hazard ratio of 0.45 and an additional 10% loss-to-follow-up, a total of 108 patients (54 per group) are needed to ensure a power of 85% using a two-sided type I error of 5%. The control group figures are conservative estimates based on the sole randomised controlled trial of NIV, which is now considered standard care in the UK. A study carried out in the United States and France (the aforementioned "FDA cohort") has estimated a one year survival of 86% after study entry for patients using DP and NIV. We have estimated the sample size on a conservative (but clinically important) 1-year difference in survival of 45% versus 70%, which produces the estimated hazard ratio of 0.45. It is anticipated that we will have complete survival data on all subjects recruited, based on previous experience in ALS trials.

With regard to quality of life data we anticipate a low level of missing data due to loss to follow up. We have reviewed the patients who were initiated on NIV in the year up to Jun 2009 and we have maintained contact with 100% of those patients surviving at 12 months. The appointment of a research nurse at each study site will enable home visits if necessary to collect the quality of life data. We have however allowed for a 10% loss to follow up in the sample size/power calculation.

Data sources, data and analysis populations

Data sources

The data used in this study will come from data entered onto the following sources:

Case Report Form (CRF) version 3, 4 December 2013
Screening log version 1, 26 October 2011
NIV group patient diary version 2, 3 August 2011
DP group patient diary version 2, 3 August 2011

The database will be stored on the CTRU database (PROSPECT) with the exception of the randomisation list which is held on the CTRU's randomisation system. Electronic data will be extracted from the system during the trial for the purpose of checking (validating) and trial progress reports; however, access to any data which would unblind the study (randomised group, resource usage) will be limited to the CTRU data management team, CTRU trial statistician and members independent of the trial (DMEC). Personal records will not be made available to CTRU staff.

Protocol Deviations

The following will be classified as protocol deviations in the statistical analysis:
Patients who are randomised in error (any of the inclusion/exclusion criteria are breached)
Patients who do not tolerate NIV
(DP group only) Patients who do not have a successful DP implantation
(DP group only) Patients who do not use DP

The above criteria will be used in explanatory analyses. All are objective criteria which will be determined from the clinical database and verified by the CI at database lock and before analysis is undertaken. Criteria 2 will also be used as a basis for subgroup analyses.

Analysis populations

Three analysis sets will be used:

Name	Patients included	Treatment group
Intent-to-Treat (ITT)	All patients for whom consent is obtained and who are randomised into the trial	As randomised
Per Protocol (PP)	The subset of the ITT patients who do not violate the protocol, as defined in section 2.2.	As randomised
Safety	All patients who proceed to NIV, with or without DP	As treated

All summaries will be based on the ITT set, with the following exceptions: Effectiveness analyses will also be undertaken on the PP set. Additional analyses to assess the impact of adherence, for example complier-average causal effect (CACE) methodology, may also be undertaken if there are a substantial number of withdrawals from treatment or crossing between groups.

Analysis of adverse events (AEs) will be undertaken on both the ITT and safety sets. In the latter, patients will be analysed by the treatment they were using at the point of the adverse event.

Outline of analyses
General considerations

Data will be reported and presented according to the revised CONSORT statement .[16,35]

Complete details of data derivations and methods of handling multiplicity, multi-centre data and missing data are covered in section 4. No interim analyses or early stopping are planned.

All summaries will be provided on the full analysis set using an intention to treat (ITT) approach. If loss-to-follow-up and/or treatment crossover is greater than 5%, sensitivity analyses will be performed to assess the robustness of the results. If important discordances are observed between these and the ITT approach, both analyses will be reported.

All summary tables will present summary statistics within each treatment group and overall unless stated otherwise.

Summaries of continuous variables will comprise the number of observations used and either
i) mean, standard deviation, median, minimum and maximum, or
ii) median, inter-quartile range, minimum and maximum
as appropriate for the distributional form of the data.

Summaries of categorical variables will comprise the number of observations used, and the number and percentage of observations in each category.

Summaries of time-to-event outcome data will comprise the median and inter-quartile range of survival within each group, the hazard ratio and its 95% confidence interval.

Summaries of QoL outcomes will comprise the following:
A summary (mean, standard deviation, median, minimum and maximum) of the QoL, by treatment group and time point.
The least squared mean QoL change from baseline at each time point, together with its standard error, by treatment group and time point.
The least squared mean difference between treatment groups ,together with its 95% confidence interval and p-value, by time point
The overall comparison between the treatment groups, together with its 95% confidence interval and p-value, and a test for interaction between treatment group and time. Both are based on longitudinal modelling.

All treatment comparisons will use the NIV only group as the reference (comparator), all statistical exploratory tests of main effects will be two-tailed with alpha = 0.05; and all confidence intervals (CIs) will be two-sided, 95% intervals. As there is controversy with regards to the operating characteristics of minimisation, a permutation test will be used to confirm the p-value from the primary endpoint .[25] Since interaction tests have low statistical power, additional consideration will be given to p-values below 0.1 when testing interactions (treatment x centre and treatment x subgroups).

Disposition and data completeness

The following summary will be presented for all patients screened for entry to the study, by centre and overall:

Enrolment	The "CONSORT" diagram comprising the number of people screened, eligible, consented, randomised, receiving their allocated treatment, withdrawing/lost to follow-up, and included in ITT and PP analysis sets.
Characteristics of non-enrolled patients	The age, gender and reason for non-inclusion.

In addition the following by-patient line-listings will be provided

Reasons for non-randomisation	For people screened but not randomised: A list of reasons for non-recruitment (any/all that apply) A list of reasons for non-consent
Reasons for non-implantation	For people allocated pacing device: A list of reason(s) for not proceeding with surgery (FVC not acceptable, failure of pre-operative checks, withdrawal, unable to stimulate diaphragm, others)

The inclusion of patients in each analysis population is outlined in section 2.3. Where the randomised intervention (as recorded on the randomisation list) differs from the intervention group as recorded on the case record form, the randomisation list will be assumed to be the correct data source. The following summary will be provided, by treatment group and overall:

Attrition, adherence and analysis sets	The number and percentage of patients who complete each visit are lost to follow-up do not receive their allocated treatment or switch treatment are in each analysis population as defined in 2.3

During the recruitment and follow-up periods, the following summaries will be made available to members of the TMG, TSC and DMEC.

Data completeness	The number of patients with complete data for each key parameter, by centre.

The inclusion of key parameters may be allowed to vary at the request of the TMG, TSC or DMEC during the trial. In order to allow time for data to be entered onto the system, data items will be considered complete if they have been entered within 30 days of scheduled visit date, and otherwise incomplete.
Demographics and baseline characteristics

The following summaries will be presented:

Demographics and vital signs	Centre, Age, gender, ethnic category, height, weight.
ALS characteristics	Duration of ALS symptoms, type of onset, site of onset, El Escorial diagnosis, bulbar function, ALSFRSr
Baseline severity	FVC, Supine VC, SNIP, PaCO2, and the number and percentage of patients with respiratory insufficiency, unacceptable phrenic nerve function, ECG abnormality, and blood test abnormality
Physical examination	The number of patients with abnormalities in each body system
Medical history	The number of patients with each medical history, past or present
Pre-operative checks and surgery details	The number and percentage of patients who meet checks for surgery at initial assessment, at repeat assessment, and who fail and are withdrawn by reason (parameter not met); the number and percentage of patients where DP was not fitted because diaphragm could not be stimulated; the number and percentage of patients with complications (by reason); and the number and percentage of patients with a surgical AE (DP only)

In addition, the following by-patient line listing will be provided:

Surgical complications	A list of all surgical complications recorded

Efficacy
Primary endpoint

The primary efficacy endpoint will be overall survival, defined as the duration from randomisation to death of any cause.

Primary overall survival analyses

The following summaries will be presented:

Overall survival	A comparison of the overall survival by treatment group.
Overall	A comparison of the overall survival by treatment group and by

survival by NIV tolerance	NIV tolerance subgroup (tolerant versus intolerant).
Overall survival by bulbar function	A comparison of the overall survival by treatment group and by bulbar function subgroup (mild/moderate versus severe)
Overall survival – per protocol population	A comparison of the overall survival by treatment group, with protocol deviations removed.

Note that NIV tolerance is a key subgroup as written in the protocol. In this context, "NIV tolerance" relates to NIV usage; "tolerant" is taken to mean "adherent" or "compliant", whilst "intolerant", "non-adherent" and "non-compliant" all suggest "non-usage".

For the remainder of this SAP the terms "use" or "usage" will be used when describing the extent to which participants use NIV (i.e. hours), whilst the term "adherence" will be termed to categorise NIV as binary (a subjective "yes or no" judgement). The same will be applied to DP usage in the intervention group. Specific details are provided in sections 3.5.2 and 4.5.1.1.

Bulbar function is defined in section 4.3.

Analysis will be undertaken by Cox proportional hazards (PH) regression, with covariates including treatment group and the minimisation factors. The model fit will be assessed as described in section 4.4. Kaplan-Meier survival curves will be presented overall and for each subgroup. Differences in survivorship between subgroups will be tested by inclusion of the covariate in the model along with treatment group. Differences in the effect of treatment between subgroups will be tested using an interaction term between the two

Supportive overall survival analyses – relationship to adherence

Supportive overall survival analyses – relationship to adherence

In light of the early stopping, additional exploratory (observational) analyses will be performed to better understand the mechanism of pacing.

Kaplan-Meier survival curves will be presented overall and for each subgroup listed below. Differences in survivorship between subgroups will be tested by inclusion of the covariate in a Cox proportional hazards model. These analyses are in additional to the pre-planned analyses according NIV adherence (tolerance) and overall protocol compliance (per-protocol) set out above.

Overall survival by NIV adherence and DP adherence	Analysis 1: Comparison of the overall survival in three subgroups: Adherent to pacing (participants randomised to DP)

	Not adherent to pacing (participants randomised to DP group only) Control group (participants randomised to NIV group only) Rationale: to help assess whether non-users of DP are have the same survival as those who use DP Analysis 2: Comparison of survival in three subgroups NIV-adherent patients (DP group only) Non NIV-adherent patients (DP group only) Control group Rationale: this will assess the association between NIV compliance and outcomes in DP participants. One plausible mechanism is that DP participants were using DP in place of NIV.

Differences in survivorship will be investigated by inclusion of the usage as a covariate in a Cox proportional hazards model.

Overall survival by typical NIV use	A comparison of the overall survival by treatment group, including a covariate of the number of hours of typical NIV usage in the first i) 3 months and ii) 6 months. This will entail including NIV use in hours as a) a continuous measure and b) by categories.
Overall survival by typical DP use	A comparison of the overall survival including a covariate of the number of hours of typical DP usage in the first i) 3 months and ii) 6 months. This will entail including DP use in hours as a) a continuous measure and b) by categories

The relationship between NIV and DP usage by time point will also be assessed and reported graphically. Since the relationship between typical adherence (in hours) and survival is not expected to be linear, fractional polynomials will be used to assess the fits of quadratic and other non-linear relationships.[24] Finally, usage will be defined in categories. For NIV we will follow the approach of Kleopa et al (1999),[13] who characterised participants as non-adherent (typical usage below 1 hour per day), low-adherent (typical usage 1 to less than 4 hours per day) and good-adherence (typically 4 or more hours per day. Adherence to pacing will also be categorised, but here there is no equivalent published data defining low and good adherence for pacing and so the cutpoints will need adjudication. However, since target usage for pacing on this study is generally lower than NIV target usage in this study, the cutpoints used for low- and good-adherence will be in the region of 1 and 3 hours respectively.

A key question is whether the assessment of adherence should start from randomisation or from first use. Both may be viewed as being of interest, and so exploratory analyses will take adherence as i) starting at randomisation

(adherence being zero in this period) and ii) starting at initiation (for NIV) or surgery (for pacing).

Participants whose NIV or DP adherence cannot be determined will be excluded from these exploratory analyses.

In doing this, the small number of participants is recognised. Conclusions will be tempered by this limitation, and also the post-hoc nature of the analyses. Nevertheless, given the circumstances of the trial termination there is sufficient interest in this for it to be justified.

Supportive overall survival analyses – analyses at the point of DMEC intervention

The overall survival will also be reported as of the point at which the DMEC made the decision to i) suspend the trial; ii) terminate the trial with advice to stop pacing. In both of these analyses, participants who were randomised to pacing but did not receive it as a result of the DMEC decision (two patients) will be excluded.

Supportive overall survival analyses – tracheostomy free survival

Finally, since "Tracheostomy-free survival" is considered an important outcome, we will look at the incidence of tracheostomy. Tracheostomy-free survival is defined as the time from randomisation to tracheostomy or death, whichever occurs first. However, at the time of writing, no tracheostomies have been performed within the DiPALS trial.

Secondary outcomes
Patient health utility and QoL

The patient will complete EQ-5D and EQ-5D VAS at six time points (screening, 2, 3, 6, 9 and 12 months), and the SAQLI and SF-36 questionnaires at five time points (screening, 2, 3, 6 and 12 months).

EQ-5D and EQ-5D VAS

Both will all be analysed in four ways as described below:

All patients, at end of follow-up: A change from baseline to 12-month analysis (all patients; deaths imputed with worst-case).
All patients, longitudinal analysis: A repeated measures analysis, in which the trajectory across time is modelled. This will include all patients, including those who die; deaths will have values imputed with death state case (zero) for both EQ-5D and EQ-5D VAS.
Survivors, at end of follow-up: A change from baseline to 12-month analysis (surviving patients only)
Survivors, longitudinal analysis: A repeated measures analysis, in which the trajectory across time is modelled (surviving patients only).

The 12-month change in EQ-5D and EQ-5D VAS will be compared between groups using analysis of covariance (ANCOVA) in which the treatment group and the baseline score are included as covariates along with minimisation factors. Repeated measures analyses will be undertaken using a generalised least squares model repeated measures ANCOVA with the same covariates.

EQ-5D and EQ-5D VAS will be summarised as described in 3.1 (QoL outcomes), with the following summaries presented:

Health utility, all patients	EQ-5D and EQ-5D VAS
Health utility, survivors	EQ-5D and EQ-5D VAS

The 12-month change in EQ-5D and EQ-5D VAS will be compared between groups using analysis of covariance (ANCOVA) in which the treatment group and the baseline score are included as covariates along with minimisation factors. Repeated measures analyses will be undertaken using a generalised least squares model repeated measures ANCOVA with the same covariates. The model will use an exchangeable correlation matrix structure, and fit of this model will be tested against an unstructured model (and potentially other alternatives) using Hausman's test.

EQ-5D will be further reported by subgroups (NIV tolerance and bulbar function). Testing for differential treatment effect between subgroups would necessitate a 3-way interaction (treatment group x subgroup x time), which given the sample size would produce potentially unstable coefficients. Therefore, the focus here will be on within-group summary statistics and graphical displays, separately by treatment group.

SF-36

The SF-36 will be used to calculate 8 scales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, social functioning and role-emotional) and 2 component summary measures (physical health and mental health). Each of the domains will be rescaled to be between 0-100. All 10 summary measures will be summarised by time, but the longitudinal modelling will only be applied to physical and mental health. Unlike EQ-5D, summaries will comprise only patients who are alive at each time point.

The following will be presented:

SF-36	SF-36 domains by time point

The analysis strategy will match that of the EQ-5D.

SAQLI

This one-domain questionnaire will be analysed in the same manner as SF-36 physical and mental health domains.

Carer QoL

The patient's main carer will complete EQ-5D and EQ-5D VAS at six time points (screening, 2, 3, 6, 9 and 12 months), and the CBI at five time points (screening, 2, 3, 6 and 12 months).

EQ-5D

Where the patient has a nominated carer, they will complete the EQ-5D and EQ-5D VAS at the same time points as the patient. Analysis will be as for patient EQ-5D survivors only: no attempt will be made to impute carer EQ-5D for whom the patient has died.

CBI

This one-domain questionnaire will be analysed in the same manner as carer EQ-5D.

Admissions / resource

The following will be presented by timepoint visit among patients for whom complete diary data are available:

Resource use	The number and percentage of patients having each of the following, and the number and percentage of episodes of each: 1) Services hospital admission (with duration) emergency department attendance minor injury clinic or walk in centre general practitioner referral 2) Devices use of CoughAssist use of Breath-Stacking use of suction 3) Health and social care physiotherapist (home visit, outpatient visit) occupational therapist (home visit, outpatient visit) other (home visit, outpatient visit) 4) Additional care/support formal (e.g. home help) informal (e.g. family/friends)

3.4.2.4 Additional outcomes: ALSFRS-r and respiratory function

In the light of the early stopping, the TSC requested additional respiratory function data to be collected to augment that which is already collected at baseline (and for DP, immediately pre-surgery). The following data is now collected: FVC, oxygen saturation, arterial carbon dioxide, SNIP, vital capacity and ALSFRS-r.
The ALSFRS-r score is derived from a questionnaire. In health, a score of 48 (the maximum) is expected, and consequently the overall rate of decline pre-baseline is given by
(48 – baseline ALSFRS-r score) /. (Date of randomisation – Date of symptoms onset)

The remaining five measurements are clinical measurements of lung function.

For each measure, the data will be used to address the following:
Provide summaries of the baseline function by group
Calculate the proportion of patients in the DP arm who decline between baseline and surgery
Respiratory function over time, particularly in relation to surgery for in the DP group.

The reasoning for these is are to assess the following aspects of the study:
To assess whether the groups are comparable. These differences should be small due the minimisation, but may imbalance may occur given the relatively small sample size.
To assess whether the extent to which DP patients decline over the period between randomisation and surgery, and an exploration of the association between this decline and survival. These findings will be compared with the aforementioned FDA cohort study.
n.b. An FVC between 50-75% (or equivalent on other tests) is an entry criteria for DiPALS, additionally, patients with FVC below 45% cannot undergo surgery. The FDA cohort used a similar criteria (<=85% at baseline and >45% at surgery), but unlike DiPALS excluded those who were never implanted.
The respiratory function over time may shed light on the mechanistic impact of DP in relation to control.

An important note is that problems will occur by virtue of missing data. Some centres do not standardly collect this, and the analyses will be restricted to those which do. More pertinent is the possibility that patients may be too ill to undergo tests; most obviously, the patient may have died. These limitations will not affect i) and will have minimal impact on ii), but iii) in particular will be compromised. Therefore, iii) will comprise mainly of graphical displays depicting change against time, with the date of NIV initiation, DP insertion, withdrawal and death superimposed.

Safety outcomes
Adherence and machine settings (diaphragm pacing)

Adherence and machine settings for both NIV and DP will be collected in and summarised for in each assessment period separately. Assessments are carried out at baseline, 1 week post-surgery (DP group only), and at months 2,3,6,9 and 12. Adherence will be reported by time period between these visits, i.e. 0-2 months, 2-3 months, 3-6 months, 6-9 months and 9-12 months.

The following summaries will be provided, for the DP group only:

DP usage	The number and percentage of DP users and average DP usage, by time period
DP parameter settings	The number and percentage of patients with one or more channel X within each time period; and a summary of each parameter (amplitude, pulse, respiratory rate, inspiratory interval, pulse frequency and pulse ramp) for each channel, and for each parameter the percentage of patients whose parameters changed, within each time period

The assessment of pacing adherence is detailed in section 4.5. Unlike NIV, there is no normative data or robust evidence to support a minimal "therapeutic dose" of pacing, and it is likely that several analyses of adherence will be proposed in addition to those outlined here.

The percentage of users in each time period will be defined as

Percentage of DP users =
100 x number of DP adherent patients in time period / number of patients alive at end of time period.
Adherence and machine settings (NIV)

The following summaries will be presented:

NIV usage	The number and percentage of patients who are NIV adherent, average NIV usage by time period, and percentage of target usage, by time period
NIV parameter settings	A summary of each parameter (minimum volume assured pressure support, maximum volume assured pressure support, inspiratory positive airways pressure, expiratory positive airways pressure, target tidal volume, respiratory rate, type of machine, type of mask, and humidification), and for each parameter the percentage of patients whose parameters changed, within each time period

NIV adherence is reported by time period defined identically to DP (above). As a guideline, patients are defined as NIV adherent within a time period if their NIV usage is at least 4 hours per night; this will be adjudicated as set out in section 4.5. The percentage of users will be defined as

Percentage of NIV users =

100 x Number of NIV adherent patients in time period / number of patients alive at end of time period

Percentage of target time used will be defined *among users only* as

Percentage of target usage achieved =
100 x average use in time period / target usage within period

In addition, the following by-patient line-listing will be presented:

Device technical issues	Reported technical problems or other observations on DP and/or NIV, by time period

NIV usage data will only be included from the point of NIV initiation onwards, but withdrawal or non-usage of NIV thereafter will be included (as zero).
Adverse events

The safety and tolerability profiles will be reported by analysing the proportion of patients experiencing adverse outcomes. The following summaries will be presented:

AEs	The number and percentage of patients reporting an AE, by type
Serious AEs (SAEs)	The number and percentage of patients reporting an SAE, by type
Treatment-related AEs	The number and percentage of patients reporting a DP-related AE, by type The number and percentage of patients reporting a NIV-related AE, by type

"Related" will be defined as those AEs recorded as definite, probable or possible.

The following by-patient line listings will be presented:

All AEs	A listing of all AEs including Treatment group (if the patient switches treatment groups, details will be included) Description Days from treatment commencement to AE onset (if patient switches treatment group, this will be the most recent treatment Severity Relationship Outcome Seriousness
All SAEs	A listing of all SAEs (as "all AEs" with the omission of "serious")
All treatment-related AEs	A listing of all treatment-related AEs (as "all AEs" with the omission of "relationship")

Concomitant medications

The following summary will be presented

Concomitant medication	The number and percentage of patients taking each medication

Interim analysis (December 2014)

Following the request of the TSC (see section 1.4.2), the following analyses will be produced for presentation at the 25th International Symposium on MND/ALS in December 2014:

Overall survival
DP and NIV adherence

The analysis will be undertaken using the database as of October 24th 2014. Site staff will be asked to enter all data pertaining to adherence and known deaths by October 17th 2014. Thereafter the database will be exported and the interim analysis commence. A copy of the database as used at this time will be retained, but data entry will continue until the end of study close date (anticipated December 2014).

These analyses will be descriptive rather than definitive, and will entail the following:

Overall survival: Kaplan-Meier survival curves, with hazard ratio as estimated based on a Cox regression model adjusted for minimisation covariates. Two additional, supportive analyses will be performed: i) logrank test (unadjusted), and ii) a Cox regression, stratified by treatment centre.

NIV adherence: the number and percentage of participants deemed adherent with NIV by time period and overall; and a summary of usage hours.

DP adherence: the number and percentage of participants deemed adherent with DP by time period and overall; and a summary of usage hours.

Detailed statistical methods and calculations
General considerations
Number and timing of analyses – adjustment for multiplicity

The study may stop prematurely on grounds of safety or futility. However, no formal interim analyses will be performed for efficacy, and consequently no adjustment for multiplicity will be made to the significance levels.

Missing, spurious and unused data

Missing data is not expected for the primary outcome (overall survival). If the patient is lost to follow-up, their survival time will be censored at the date last

known alive. For the questionnaire-based QoL outcomes, missing data may arise at any timepoint in one of four ways:

Questionnaire not completed, patient has died
Questionnaire completed but not in the correct time frame (i.e. outside visit window)
Questionnaire incomplete (some questions missing)
Questionnaire not returned (all questions missing)

1) is addressed in section 3.4.2.

2) will be handled by imposing the following visit windows:

Time point	**Lower limit**	**Upper limit**
Screening	No limit	Date of randomisation
2 months	Date of randomisation	75 days post randomisation [2 months + 15 days]
3 months	76 days post randomisation	135 days post randomisation [3 + 1.5 months]
6 months	136 days post randomisation	225 days post randomisation [6 + 1.5 months]
9 months	226 days post randomisation	315 days post randomisation [9 + 1.5 months]
12 months	316 days post randomisation	405 days post randomisation [12 + 1.5 months]

Scenario 3) is covered in the individual sections below.

For scenario 4), the same approach will apply to all questionnaires:
If the questionnaire from an adjacent visit falls within the visit window, use this value.
If i) has not imputed a value, but values are available both before and afterwards, this will be imputed using the trapezoid method:

$$Q_t = [\ Q_1 \times (t_2 - t) + Q_2 \times (t - t_1)\ ] / (t_2 - t_1)$$

Where
Q_t is the imputed quality of life at time t
(t_1, t_2) are the time points immediately prior to and following time t ($t_1<t<t_2$) at which valid responses exist
Q_1 and Q_2 are the responses at times t_1 and t_2.

Illustration

Suppose a patient has data as follows, for which approaches 1), 2), 3) and 4i) have not dealt with.

Time point	EQ-5D

2 months	0.7
3 months	missing
6 months	0.5

The missing value is then calculated by

EQ-5D at t=3 = [0.7 x(6-3) + 0.5 x(3-2)] / (6-2) = [0.7x3 + 0.5x1]/4 = 0.65

If missing data still persist, they will be imputed using multiple imputation using baseline covariates (age, gender, ALSFRS-r and SNIP/FVC group), previous questionnaire values, time point and treatment group.

Analysis sets

The ITT will be the primary analysis population for effectiveness outcomes, with the results for the PP being supportive of it. If for any endpoint the populations confer inconsistent results, further analyses will attempt to investigate the reason for this.

Methods for dealing with multi-centre data

The consistency of outcomes among the treatment centres will be assessed by fitting a model which includes an interaction between treatment group and centre. If the test for interaction is not statistically significant the interaction term will be removed. If significant differences are found, further analyses will be undertaken to assess whether this may be due to differences in case mix (i.e. an artefactual centre effect) or not (i.e. real centre effect).

Disposition and data completeness

Recruitment data, data completeness and patient demographic & characteristics will be reported to the TSC, DMEC and TMG in an ongoing fashion.
Demographics and baseline patient characteristics

The baseline date is the date of randomisation. The centre will be defined as the centre at which the patient first attended. Age is defined as (date of baseline – date of birth).

ALSFR is a one-domain questionnaire comprising 24 questions, each of which is scored from 0-4. The overall score is the sum of these. If the questionnaire is incomplete but at least half of the questions answered (i.e.at least 12 of the 24), the overall score will be multiplied up by the formula

Overall score = (24 / Number of questions answered) x (total score among answered questions)

If fewer than 12 questions are answered, the questionnaire will be treated as missing and will not be used in summaries.

The bulbar function score is calculated from the ALSFRS-r. The answers to the first three questions (speech, swallowing and salivation) are summed, and the bulbar function is categorised from this sum into the following: mild (0-4), moderate (5-8) and severe (9-12).

Efficacy
Primary endpoint

The primary endpoint is overall survival, defined as the time between randomisation and death. If no notification of death has been received, the patient will be censored at the date last known alive.

Pre-trial modelling, undertaken at the proposal development stage, found PH to be the best fit to previous data, and therefore a Cox PH regression model will be fitted with ties handled by the Efron method. The PH assumption will be checked by adding time-dependent covariates and graphing scaled Schoenfeld residuals against time .[20] If PH is found not to fit the data adequately, an AFT alternative will be fitted and the adequacy of its fit assessed using Q-Q plots [21] If this too does not fit, a residual life analysis.[22] will be used as the basis for summarising the treatment effect

Secondary endpoints

SAQLI (Sleep Apnoea Quality of Life Index)
SAQLI is a one-domain questionnaire comprising 14 questions, each of which is scored from 1-7. The overall score is the average of these. If the questionnaire is incomplete (i.e. less than 14 questions are answered), the overall score will be defined as the average provided at least half of the questions (7 of the 14) have been answered. If fewer than 7 questions are answered, the questionnaire will be treated as missing; no further imputation will be undertaken.

SF-36
Version 1 of the SF-36 will be used. The domain scores will be calculated using the standard RAND organisation algorithm.

CBI
CBI is a one-domain questionnaire comprising 24 questions, each of which is scored from 0-4. The overall score is the total of these. Incomplete questionnaires will be handled in the same manner as SAQLI

Safety
NIV and DP usage

The original analyses of DP and NIV usage were to be based primarily on diary data. In the course of the trial however, it became clear that fewer diaries were being completed than expected. As a consequence, alternative approaches to defining usage were required.
NIV usage

In addition to the patient diary and CRF, NIV usage data can be estimated from the NIV machine itself. The CRF was amended to collect this additional information in 2014.

NIV usage will be collected for each time period, using the following hierarchy:

NIV machine
if i) not available, NIV usage will be taken from the participant diary entries
if i) and ii) not available, NIV usage will be taken from the participants typical usage recorded in the CRF

For the purpose of the analyses, "time period" is defined as the duration between successive visits (0-2, 2-3, 3-6, 6-9 and 9-12 months). Where NIV initiation starts part way through an interval (for example, if NIV started 20 days after randomisation), usage will be defined over the period starting from NIV initiation.

For i), the CRF records the cumulative "blow hour count", or the number of hours use recorded at each visit date. The average NIV use (in hours) since last visit is then calculated as

(NIV count at visit – previous NIV count at previous visit)/number of days between visits

This will be used to calculate NIV usage for time periods for which NIV use is collected both before and after

For ii), the NIV use will be collected from the diary only if the participant has completed at least 80% of the scheduled diary entries. For example, if 95 days have elapsed between the 3&6 month visits, the participant would need to have completed a diary on at least 76 of these. The value taken will be the median value recorded.

For iii), the NIV use will be taken as the value "typical NIV usage" recorded in the CRF.

Based on this, participants will be defined as "adherent" or "not adherent", blinded to treatment group. The assessment will be made separately by the chief investigator and Dr Stephen Bourke. Cases for which the assessors cannot reach agreement will be referred to the TMG/TSC.
DP usage

DP adherence will be quantified in an analogous manner as NIV, but with two key differences. Firstly, the DP machine does not collect usage data (unlike NIV); DP usage will therefore be taken only from the diary (preferably) or the CRF. Secondly, the qualitative adjudication of adherence by definition cannot be blinded to treatment group. Assessment of adherence to pacing will be made in the same manner as NIV.

Adverse events

Adverse events will be coded as one of the following categories. Coding will be done by the chief investigator and will be done blinded to the participant's treatment group.

Adverse event categories

Cardiovascular system
Central nervous system
Dermatological
Gastro-intestinal
Genito-urinary
Infection of percutaneous endoscopic gastrostomy (PIG) or per-oral image guided gastrostomy (PEG)
Insertion/removal of PIG/PEG
Respiratory, further subdivided into
Breathless – unclassified
Chest infection
Cough
Infection
Pneumothorax/Capnothorax
Pulmonary Embolism
Respiratory failure
ALS symptoms
NIV specific
Pain
Psychiatric
Wire infection
Wire problems
Other

Additional categories or subcategories may be added at the request of the chief investigator, TSC or DMEC.

For each AE, the report will comprise the number and percentage of patients affected, and the number of events (a patient may have more than one occurrence of the same AE).

Modifications to the original protocol analysis statement

The original analysis plan was written by the trial statistician with access to unblinded, but very limited data: at the time of sign off, one patient death was recorded on the database. The updated analysis plan was written by the same statistician, again with access to unblinded data. The DMEC were asked to approve the plan, and did so (with minimal suggestions) having seen unblinded data summaries throughout the trial progress; all other reviewers/approvers were blinded to the data.

The following have been updated:
Section 1.4: Gives background to the early trial stopping, early data release and other studies of DP in ALS.

Section 2.1: Updated with new CRF version and details of individuals with data access.
Section 3.4.1: Additional exploratory analyses around the primary endpoint, particularly in relation to compliance/adherence/tolerance
Section 3.4.2: New outcome measures, using routinely collected data from clinic visits, have been added at the request of the TSC
Sections 3.5: Additional details of adherence are provided.
Section 3.6: Details of the interim analyses for presentation at the 25th International Symposium on MND/ALS have been added.
Section 4.1.2 Missing data imputation has been updated. Scenario 4(iii) has been amended.
Sections 4.5: Additional details of adherence and adverse event categorisation are provided.