An open-label, multi-centre, randomised, two-period, crossover study to assess the efficacy, safety and utility of 16 week day and night automated closed-loop glucose control under free living conditions compared to sensor augmented insulin pump therapy in older adults with type 1 diabetes

DAN06 Study

Statistical Analysis Plan

Randomized Crossover Trial

Version: 2.0

Version Date: 30/07/2020

Protocol Version: 5.0

Revision History

The following table outlines changes made to the Statistical Analysis Plan.

Version Number	Author	Approver	Effective Date	Study Stage	Revision Description
1.0	Charlotte Boughton		18/05/2020	Protocol development	Original Version
2.0	Charlotte Boughton		30/07/2020	Active	Adjustment to statistics in line with protocol

19

20

1. Overview

This document outlines the statistical analyses to be performed for the DAN06 study. The approach to sample size and statistical analyses for this study are summarized below.

This is an open-label, multicenter, randomized, two period crossover study to assess the efficacy and safety and utility of closed loop (CL) insulin delivery in comparison with sensor augmented pump (SAP) therapy with continuous glucose monitoring (CGM) over 16 weeks in older adults with type 1 diabetes aged over 60 years. Approximately 36 subjects are expected to be randomized and enter the trial. All participants will receive both interventions, and the order of receiving them will be randomized based on a 1:1 ratio. Randomization will be preceded by a 3-4 week run-in period where subjects must demonstrate competency and compliance in using the study insulin pump and CGM device. After randomization, the subjects will enter the two 16 week study periods and will test one intervention per study period. The two periods will be separated by a 4 week washout period.

2. Statistical Hypotheses

- *Null Hypothesis:* There is no difference in the mean time spent in the target range (3.9 to 10.0 mmol/L) over the 16 week period between the two treatment groups.
- *Alternative Hypothesis:* There is a non-zero difference in the mean time spent in the target range over the 16 week period between the two treatment groups.

3. Sample Size

The study is projected to randomize 36 subjects. The sample size was calculated assuming 80% power, a treatment effect of 10% in the percentage of time in the target range, and a standard deviation of 18% for an individual measurement.

4. Outcome Measures

Primary Efficacy Endpoint:

- Time spent in the target range (3.9 to 10.0 mmol/L) over the 16 week period

Key Secondary Endpoints:

- Percent time spent with glucose levels above 10.0 mmol/L
- HbA1c at 16 weeks
- Mean of glucose levels
- Percent time spent with glucose levels below 3.9 mmol/L

Secondary Efficacy Endpoints:

CGM Metrics

Glucose variability

- Standard deviation of glucose levels
- Coefficient of variation of glucose levels

Hyperglycemia

- Percent Time spent with glucose levels above 16.7 mmol/L

Hypoglycemia

- Percent Time spent with glucose levels below 3.5 mmol/L
- Percent Time spent with glucose levels below 3.0 mmol/L

Insulin Delivery

- Total insulin dose (units/kg/day)
- Basal insulin dose (units/kg/day)
- Bolus insulin dose (units/kg/day)

Questionnaires

- WHO-5 quality of life measurement
- Diabetes Distress Scale
- Glucose Monitoring Satisfaction Survey
- Hypoglycemia Confidence
- INSPIRE Survey
- Pittsburgh Sleep Quality Index (PSQI)

Cognitive testing (CogState)

- Detection Task Score
- Identification Task Score
- One Card Learning Task Score
- One Back Task Score

Actiwatch data

Holter data

4.1 Calculation of CGM Metrics

For the primary outcome and all secondary CGM metrics, a single value will be calculated for each subject for each period by pooling all CGM readings between the treatment initiation visit and up to 112 days post-initiation visit or the end of treatment visit, whichever comes first. All glucose sensor readings will be weighted equally in the pooled percentages regardless of how they distribute across weeks. Data will not be truncated due to protocol deviations.

Baseline CGM metrics will be calculated by pooling all readings up to the last 14 available days of CGM readings prior to randomization.

5. Analysis Datasets and Sensitivity Analyses

5.1 Analysis Cohorts

The primary analysis and all secondary analyses will be performed on an intention-to-treat basis with each day included in the treatment group assigned by randomization.

A per-protocol analysis restricted to randomized participants with a minimum of 60% of available CGM readings during the control period and 60% CL system use during the CL period will be conducted for the primary outcome.

Safety outcomes will be reported for all enrolled participants, regardless of whether the study was completed.

6. Analysis of the Primary Endpoint

6.1 Included Subjects

Only subjects with at least 168 hours of CGM data in at least one period will be included. If a subject has more than 168 hours of data in period 1 and then drops out of the study without any data in period 2, then he or she will be included in the analysis.

6.2 Missing Data

Missing data will not be imputed for the primary analysis in this study.

6.3 Statistical Methods

Mean ± SD or summary statistics appropriate to the distribution will be reported for the primary outcome and each of the key secondary outcomes listed below over the 16 week period by treatment intervention. The treatment interventions will be compared using a linear mixed model adjusting for period as a fixed effect and site as a random effect. The analysis dataset will be three records per subject (one for baseline and one for each period). Inclusion of the pre-randomization baseline value as a third observation for each subject in the model gives a variance reduction analogous to adjusting for it as a covariate. Baseline is not modeled as a covariate in this analysis because there is no corresponding baseline for period 2, only pre-randomization. Note that adjusting for a post-randomization period 2 baseline can introduce a bias so that is not done here. The model will account for correlated data from the same subject.

A 95% confidence interval will be reported for the difference between the interventions based on the linear mixed model.

Residual values will be examined for an approximate normal distribution. If values are highly skewed, then a ranked normal score transformation will be used instead. However, previous experience suggests that the primary outcome will follow an approximately normal distribution. A two-sided p-value will be reported.

For the primary endpoint and other key endpoints listed in section 4, the familywise type I error rate (FWER) will be controlled at two-sided $\alpha = 0.05$. A gatekeeping strategy will be used, where the primary endpoint will be tested first, if passing the significance testing, other key endpoints will be tested in the order listed below using the fixed-sequence method at $\alpha = 0.05$:

- Time spent with sensor glucose levels between 3.9 to 10.0 mmol/l
- Time spent above target glucose (10.0 mmol/l)
- HbA1c
- Average of glucose levels
- Time spent below target glucose (3.9 mmol/l)

This process continues iteratively moving to the next variable down on the list until a non-significant result ($p \geq 0.05$) is observed, or all five variables have been tested. If a non-significant result is encountered, then formal statistical hypothesis testing is terminated and any variables below on the list are not formally tested and analysis of these variables becomes exploratory.

Regardless of the results of the hierarchical testing, summary statistics appropriate to the distribution will be tabulated by treatment arm for each hierarchical outcome. A 95% confidence interval for the treatment arm difference will also be calculated for all five hierarchical outcomes listed above. However, a confidence interval that excludes zero will not be considered a statistically significant result if an outcome variable higher on the hierarchical list failed to reach statistical significance.

7. Analysis of the Secondary Endpoints

7.1 Included Subjects

In the analyses involving HbA1c, all subjects with an available measurement within the analysis windows specified in section 7.3 will be included.

For secondary CGM metrics, inclusion criteria will be the same as the primary analysis.

For the secondary insulin outcomes, at least 168 hours of insulin data in at least one period will be required for inclusion. If a subject has more than 168 hours of insulin data in period 1 and then drops out of the study without any data in period 2, then he or she will be included in the analysis.

7.2 Missing Data

For the secondary CGM and insulin metrics, missing data will not be imputed in this study.

7.3 Analysis Windows

Only HbA1c obtained within ±14 days of the end of treatment visit dates during each period will be included in the analyses as the outcome. The baseline measurements must be within ±14 days of the recruitment visit.

7.4 Statistical Methods

7.4.1 Secondary CGM Outcomes

For all secondary CGM outcomes, summary statistics appropriate to the distribution will be tabulated by treatment group over the four month period. Analysis of all secondary CGM endpoints will parallel the primary analysis. A ranked normal score transformation will be applied to all highly skewed secondary outcomes.

7.4.2 Secondary Insulin Outcomes

For all secondary insulin outcomes, summary statistics appropriate to the distribution will be tabulated by treatment group over the four month period. Analysis of all secondary insulin endpoints will parallel the primary analysis. A ranked normal score transformation will be applied to all highly skewed secondary outcomes.

7.4.3 Secondary HbA1c Outcomes

For HbA1c, a longitudinal model adjusting for period as a fixed effect will be constructed to compare treatment arms. The model will include three time points: (1) baseline, (2) period 1 outcome, and (3) period 2 outcome.

7.5 Secondary Analyses by Time of Day

Summary statistics for the following outcome metrics will also be tabulated separately for daytime (defined as 6am to less than 12am) and nighttime (defined as 12am to less than 6am) over the four month period:

- Percent time with glucose levels spent in the target range (3.9 to 10.0 mmol/L)
- Mean of glucose levels

- Standard deviation of glucose levels
- Percent time with glucose levels below 3.9 mmol/L
- Total insulin dose

For each of these outcome metrics, the same model described above for the primary and secondary analyses will be fit with the inclusion of a treatment by time of day interaction. The p-value for the interaction term will be reported. These analyses will be conducted to determine whether a similar trend to the overall treatment effect is seen in the different times of day.

The study is not expected to have sufficient statistical power for definitive conclusions in the secondary analyses by time of day, and statistical power will be low to formally assess for the presence of a treatment by time of day interaction. Interpretation of the analyses by time of day will depend on whether the overall analysis demonstrates a significant treatment effect. In the absence of any significant treatment effects in the overall analyses, assessment of secondary analyses by time of day will be considered exploratory and used to suggest hypotheses for further investigation in future studies.

7.6 Questionnaire Analyses

For each questionnaire (and their corresponding subscales), total scores will be calculated and reported at each time point. They will also be compared between treatment arms using the same model described above for the primary outcome. The distribution of responses for each individual question at baseline and for each treatment arm will also be reported in separate tables.

For the INSPIRE Survey and the CL Experience Survey, a treatment arm comparison will not be done, because the surveys are only completed at the end of the CL arm. For these questionnaires, only summary statistics for the total scores and the distribution of responses for each question will be reported.

Analysis will be limited to subjects who submit a questionnaire (no imputation).

7.7 Cogstate Analyses

For the Cogstate, the distribution of responses for the individual questions will not be tabulated, because the electronic testing system does not provide them. Summary statistics for each of the four test scores and the total score for the entire survey will be reported at each time point. They will also be compared between treatment arms using the same model described above for the primary outcome. The distribution of responses for each individual question at baseline and for each treatment arm will also be reported in separate tables. The analysis will be conducted by Korey Hood.

7.8 Focus Groups Analyses

Focus groups will take place at the end of study. A script of open ended questions will be used to gather feedback and reactions to the clinical trial, use of the closed-loop system, and quality of life changes. There will also be time for discussion of content raised by participants. Qualitative data will be analysed using Atlas.ti (release 6.0; Scientific Software Development GmbH, Berlin, Germany) to organise and manage the entire corpus of focus group data. Analysis begins with an initial coding procedure to capture and describe the range of responses to the intervention. A second, more focused and detailed level of coding will be applied to major categories of findings in the initial review to determine themes in response to the clinical trial, use of the closed-loop system, and quality of life changes. The analysis will be conducted by Korey Hood.

7.9 Sleep Analyses

The Pittsburgh Sleep Quality Index (PSQI) and actigraphy data will be used to calculate mean total sleep quality score, sleep duration, time in bed, sleep disturbance (including wake after sleep onset and number of awakenings), latency, efficiency, quality, and daytime dysfunction.

Sleep will be automatically scored by Actiware software using previously described and validated algorithms. Sleep duration will be calculated as the sum of all epochs scored as sleep during the time in bed. Variability across nights in a participant's sleep duration will be summarised using the coefficient of variation. Sleep data will be averaged across nights in each participant for each study period. The analysis will be conducted by Eleanor Scott

7.10 Holter Data Analyses

Holter monitor data at the fourth month in the two treatment groups will be analysed at the end of the study. Arrhythmic events analysed will include atrial fibrillation, atrial ectopic beats, bradycardia, ventricular premature beats, complex ventricular premature beats, non-sustained ventricular tachycardia and measures of heart rate variability and QT interval. All identified arrhythmic events will be manually verified for accuracy.

Investigators will be masked to glucose values during arrhythmia analysis. The frequency of arrhythmic events during periods of hypoglycaemia (sensor glucose ≤3.9 mmol/L and 3.0 mmol/L), will be compared to the frequency of arrhythmic events during normoglycaemia (3.9-10 mmol/L) within each study arm. Analyses will be calculated for daytime and night-time periods to take into account diurnal variation. Any incidental finding will be referred by the Investigator to the clinical team and be managed as per local site policy.

8. Safety Analyses

All safety outcomes will be tabulated by participant for all events from enrollment to the final study visit.

8.1 Definitions

Reportable adverse events for this protocol include any untoward medical occurrence, unintended disease or injury, or untoward clinical signs (including abnormal laboratory findings) in a subject who has received an investigational device, whether or not related to the investigational medical device. These include severe hypoglycemia (SH) and diabetic ketoacidosis (DKA).

Hypoglycemic events will be considered severe if the event requires assistance of another person due to altered consciousness to actively administer carbohydrate, glucagon, or other resuscitative actions. If plasma glucose measurements are not available during such an event, neurological recovery attributable to the restoration of plasma glucose to normal is considered sufficient evidence that the event was induced by a low plasma glucose concentration.

Definite DKA is defined as having all of the following:

- Hyperglycemia (blood glucose >11 mmol/L)
- with either low pH (<7.3) or low serum bicarbonate (<15 mmol/L)
- and ketonemia or ketonuria

8.2 Adverse Events Summary

All episodes of SH and of DKA along with any other reportable adverse events will be listed by treatment group.

Separate listings will be provided for pre-randomization and post-randomization adverse events.

8.3 Comparison of Safety Outcomes between Treatment Groups

The following safety analyses will be performed if enough events occur for formal statistical analyses.

For each of the following safety outcomes, mean ± SD or summary statistics appropriate to the distribution will be tabulated by treatment group:

- Number of subjects with any DKA events
- Number of episodes of DKA events per subject and incidence rate per 100 person years
- Number of subjects with any SH events
- Number of episodes of SH events per subject and incidence rate per 100 person years
- Number of adverse events per subject

- Number of serious adverse events per subject

All of the above safety outcomes will be tabulated for all subjects (including dropouts and withdrawals), regardless of whether CGM data are available or whether the closed loop system was operational (if the event occurred during the CL period). Any adverse events that occurred before the treatment initiation visit in period 1 or during the washout period will not be included in the rate calculations or treatment group comparisons listed above. In all safety analyses, each period will inclusively consist of all days in between the treatment initiation visit and the end of treatment visit.

The number of person-years for the incidence rate calculations in each period will be inclusively defined as the number of person-years in between the treatment initiation visit date and the end of treatment visit date.

For each of DKA and SH (if enough events), the event rates will be compared using a repeated measures Poisson regression model adjusting for period and whether the subject has ever had a prior event. Binary variables will also be compared using a repeated measures logistic regression model adjusting for period and whether the subject has ever had a prior event.

9. Adherence and Retention Analyses

9.1 Utility Analysis

The amount of CGM use will be tabulated for each treatment arm, in addition to the amount of closed loop system use in the CL arm. Summary statistics appropriate to the distribution and range will be reported for the percentage of time using the CGM over the 16 week period (as defined above) for each treatment group. The same will be done for the percentage of time using the closed loop system in the CL arm. Tabulations of summary statistics will also be performed for the percentage of time spent using the closed loop system while using the CGM in the CL arm.

The percentage of time spent using the CGM will be calculated by dividing the total number of CGM readings by the expected number of readings during the 16 week period. The percentage of time using the closed loop system in the CL arm will be calculated by dividing the total amount of time that temporary basal infusion lasts no more than 30 minutes by the maximum possible amount of time that the system could have been used. The percentage of time using the closed loop system while using the CGM (in the CL arm) will then be computed by dividing the time that the closed loop system was operational by the amount of time that the CGM was available.

If a subject drops out of the study in the middle of a period, then the subject will be counted as not using the CGM or the closed loop system at all during the remainder of the study. Thus, these time points will be counted as zero use in the calculation of CGM use and closed loop system use.

9.2 Protocol Adherence and Retention

The following tabulations and analyses will be performed to assess protocol adherence for the study:

- Number of protocol and procedural deviations per subject along with the number and percentage of subjects with each number of deviations
- Number of protocol and procedural deviations by severity with brief descriptions listed
- Flow chart accounting for all subjects at all visits post randomization to assess visit completion rates
- A flow chart accounting for the number of subjects enrolled, the number of dropouts pre- and post-randomization, and the number of subjects eligible to be included in the primary analysis
- Number of and reasons for unscheduled visits

10. Baseline Descriptive Statistics

Baseline demographic characteristics of the cohort of all randomized subjects will be summarized in a table. Descriptive statistics will be tabulated overall and by randomization group. For continuous variables, summary statistics appropriate to the distribution will be given. For discrete variables, number and percentage will be reported for each category. The following baseline CGM metrics will be included in the table:

- % Time in Range (3.9-10.0 mmol/L)
- Mean of sensor glucose levels
- Standard deviation of glucose levels
- Coefficient of variation of glucose levels
- % Time >10.0 and >16.7 mmol/L
- % Time <3.9 and <3.0 mmol/L

11. Planned Interim Analyses

No formal interim analyses or stopping guidelines are planned for this study.
The DSMB will review data collected for the study every six months. The data to be reviewed will include information regarding all of the following:

- Status of randomized participants
- Recruitment rates by month and by site
- Baseline demographic characteristics
- Dropped participants and reasons for discontinuing

- Reportable adverse events

12. Subgroup Analyses

No subgroup analyses are planned for this study.

13. Multiple Comparisons/Multiplicity

13.1 Primary analysis and other key secondary outcomes

For the primary endpoint and other key endpoints listed in section 4, the familywise type I error rate (FWER) will be controlled at two-sided $\alpha = 0.05$. A gatekeeping strategy will be used, where the primary endpoint will be tested first, if passing the significance testing, other key endpoints will be tested in the order listed below using the fixed-sequence method at $\alpha = 0.05$:

- Time spent with sensor glucose levels between 3.9 to 10.0 mmol/l
- Time spent above target glucose (10.0 mmol/l)
- HbA1c
- Average of glucose levels
- Time spent below target glucose (3.9 mmol/l)

Additional details are provided in section 6.3.

13.2 Other Secondary Analyses

For the other secondary endpoints listed in section 4, Benjamini-Hochberg false discovery rate (FDR) adjusted p-values will be calculated within each subcategory below:

CGM derived indices:

- Standard deviation, and coefficient of variation of glucose levels
- Time with glucose levels <3.5 mmol/l and <3.0 mmol/l
- Time with glucose levels in significant hyperglycaemia (glucose levels > 16.7 mmol/l)

Insulin Endpoints:

- Total, basal, and bolus insulin dose

Questionnaires:

- WHO-5 quality of life measurement
- Diabetes Distress Scale
- Glucose Monitoring Satisfaction Survey

- Hypoglycemia Confidence
- INSPIRE Survey
- Pittsburgh Sleep Quality Index (PSQI)

14. Exploratory Analyses

No exploratory analyses will be performed for this study.