Day and Night Closed-loop in Young People with Type 1 Diabetes (DAN05)

Statistical Analysis Plan
Version 6.0

Corresponds to Version UK 7.2/ US 7.1 of the Protocol

1. Objectives

The purpose of this study is to determine whether 24/7 automated closed-loop glucose control (CL) will improve glucose control as measured by glycated hemoglobin (HbA1c) and reduce the burden of hypoglycemia compared with insulin pump therapy alone. The detailed study objectives include:

1. Efficacy: to assess efficacy of day and night automated closed-loop glucose control in improving glucose control as measured by HbA1c, as compared with usual care (conventional or sensor-augmented insulin pump therapy).
2. Safety: to evaluate the safety of day and night automated closed-loop glucose control, in terms of episode of severe hypoglycemia and other adverse events.
3. Utility: to determine the frequency and duration of the use of the automated closed-loop system.
4. Human factor: to assess cognitive, emotional, and behavioral characteristics of participating subjects and family members and their response to the closed-loop system and clinical trial using validated surveys and focus groups.
5. Health economics: to perform a cost utility analysis to inform reimbursement decision-making.

2. General

Data from the DAN05 study will be analyzed by three groups. Jaeb Center will be responsible to analyzing the efficacy, safety, and utility outcomes, Korey Hood will be responsible to analyzing human factor data, and Steven Fox will be responsible for conducting health economics analysis.

A brief review of study is provided below.

2.1. Study Overview

The study will include 130 subjects with equal proportion of those aged 6 to 12 years and 13 to 18 years, a minimum quota of 25% participants with baseline HbA1c >8.5%. These 130 subjects will be randomized into CL and usual care two study arms in 1:1 ratio.

The automated closed loop system (FlorenceM) to be used in this study will consist of next generation sensor augmented Medtronic insulin pump 640G with pump suspend feature, and Medtronic continuous glucose monitoring (CGM) Transmitter with Enlite 3 sensor.

The maximum duration of study for a subject is 8 months. Eligible participants will first undergo a screening evaluation where blood samples will be taken. At screening/baseline assessment phase, surveys investigating participants' quality of life, psychosocial and cognitive functioning,

and response to their current treatment will be distributed. Participants will also be fitted with a blinded CGM device to assess baseline glycemic control.

After screening, subjects will undergo a 1-2 week of run-in period, and they will continue using their own insulin pump. Data obtained from blinded CGM and pump downloads may be used for therapy adjustment.

After run-in, eligible subjects will be randomized to one of the two intervention arms: 6 months use of day and night closed-loop system or 6 months use of their usual care. Those subjects randomized to the CL arm will receive additional training sessions covering the use of the study insulin pump and read-time CGM, prior to starting closed-loop insulin delivery. After that, subjects will receive training required for safe and effective use of the closed-loop system for 2-4 weeks. Competency on the use of closed-loop system will be evaluated. Thereafter, subjects are expected to use closed-loop system for 6 months. For those subjects randomized to the usual care arm, refresher training on key aspects of insulin pump therapy will be provided, then subjects will continue using their own insulin pump and CGM if a regular CGM user for 6 months.

For both treatment arms, all the subjects will undergo 2 assessments during the intervention period at 3 and 6 months, respectively. At each assessment time, blood sample will be taken and validated surveys will be completed by all the subjects. Participants in both study arms will be fitted with blinded CGM system at the end of each follow up visit. The sensors will be worn at home for up to 14 days. If the sensor fails or sensor life gets interrupted prematurely, another sensor may be inserted. At the end of the study, subjects/guardians will be invited to join follow-up focus groups to gather feedback and reactions to their current treatment, the clinical trial, and quality of life changes.

3. Study Cohort

3.1. Included Subjects

The primary analysis will follow intention-to-treatment approach, which means subjects will be analyzed in the treatment arm assigned by randomization regardless of compliance. All randomized subjects will be included in the primary analysis and safety analyses. For secondary efficacy analyses, only subjects with available data at 6 months will be included in statistical comparisons.

4. Metrics and Analyses

4.1. Primary Efficacy Endpoint

The main endpoint in this study is HbA1c level at 6 months.

Mean (SD) will be tabulated by treatment group at baseline, 3 and 6 months. The analysis time window is specified in below table. A linear regression model will be used to compare the difference in central lab HbA1c between the two intervention arms at end of treatment period (6 months), while adjusting for baseline HbA1c, age and random site effect. A 95% confidence interval will be reported for the difference between the treatment groups based on the linear regression model. Normality of the residuals will be assessed. If the residuals have highly

skewed distribution, then ranked normal score transformation of outcome data will be applied in the regression model. However, previous experience suggests that HbA1c values will follow an approximately normal distribution.

Imbalances between groups in important covariates are not expected to be of sufficient magnitude to produce confounding. However, the presence of confounding will be evaluated in the sensitivity analyses by including factors potentially associated with the outcome for which there is an imbalance between groups (section 4.6).

All variables obtained on a continuous scale will be entered into the models as continuous variables (e.g., baseline HbA1c and age), unless it is determined that a variable does not have a linear relationship with the outcome. In such a case, categorization and/or transformation will be explored.

Table 1. Analysis time window for HbA1c

	Time from Randomization
Baseline[a]	NA
3 months	2.5-<5.5 months[b]
6 months	5.5-<8.5 months[b]

a- Baseline HbA1c is measured at the visit 1 (before randomization), thus no analysis window applicable. If the randomization visit (visit 3) is more than 28 days after visit 1, then HbA1c will be measured again, and in such case, HbA1c measured at visit 3 will be reported as baseline HbA1c.

b- A ±1.5 months analysis window is given at each target time, since there will be 1 month of training after randomization date, the target time for each visit is shifted for 1 month.

Missing Data

In the primary analysis, missing data will be handled using multiple imputation with pattern mixture models assuming the dropout trajectory of the treatment subjects was that of the control arm.

It is worth noting that all statistical methods for handling missing data rely on untestable assumptions and there is no one correct way to handle missing data. Our goal is to minimize the amount of missing data so that the results will not be sensitive to which statistical method is used. This will be checked with the sensitivity analyses in section 4.6.

4.2. Other Key Endpoints

For following key endpoints at 6 months, 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 $\alpha = 0.05$ (Huque and Alosh, 2008).

- HbA1c (primary endpoint)
- Time spent in the target glucose range from 3.9 to 10.0 mmol/l (70 to 180 mg/dl)
- Mean glucose
- Time spent above target glucose 10.0 mmol/l (180 mg/dl)

- Time spent below target glucose 3.9 mmol/l (70 mg/dl)

The 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 become exploratory. For example, in the hypothetical scenario depicted in the table below, the first three outcome variables have significant results so testing continues to the fourth variable (%time CGM above 10.0 mmol/l). The result is not significant for that fourth variable ($p = 0.06$) so testing stops. No formal hypothesis test is conducted for the fifth variable on the list in this example scenario.

Table 2. Example for Fixed-Sequence Testing Results

Order	Outcome Variable	Treatment Arm P-value	Statistically Significant?	Action
1st	HbA1c (primary outcome)	0.001	Yes	Test next variable
2nd	% time CGM 3.9-10.0 mmol/l	0.02	Yes	Test next variable
3rd	mean glucose	0.007	Yes	Test next variable
4th	% time CGM above 10.0 mmol/l	0.06	No	Stop formal testing
5th	% time CGM below 3.9 mmol/l	Not tested	Unknown	N/A

Regardless of the results of the fixed-sequence testing, summary statistics appropriate to the distribution will be tabulated by treatment arm for each outcome. A 95% confidence interval for the treatment arm difference will also be calculated for all five 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.

For above key endpoints other than the primary outcome, linear regression models will be used to estimate treatment difference while adjusting for corresponding baseline values (i.e. adjust for baseline % glucose in range when comparing % glucose in range at 6 months between two treatment groups) for each outcome, baseline HbA1c, age and random site effect. The % time CGM below 3.9 mmol/l usually has skewed distribution in which case nonparametric methods will be used.

4.3. Secondary Efficacy Endpoints

The following endpoints will be considered exploratory. For these analyses, the false discovery rate (FDR) will be used to account for multiple comparisons (see below).

CGM-Measured:

- Glucose standard deviation
- Glucose coefficient of variation.
- Time with glucose levels <3.5 mmol/l (63 mg/dl)
- Time with glucose values <3.0 mmol/l (54 mg/dl)
- AUC of glucose below 3.5 mmol/l (63 mg/dl)

- Time with glucose levels > 16.7 mmol/l (300 mg/dl)
- AUC of glucose above 10.0 mmol/l (180 mg/dl)

Following CGM metrics will also be calculated separately for day-time period (06:00-23:59) and night-time period (00:00-05:59):

- Percentage of CGM glucose from 3.9 to 10.0 mmol/l (70-180 mg/dl)
- Mean glucose
- Glucose variability as measured by standard deviation
- Percentage of CGM glucose values <3.5 mmol/l (63 mg/dl)

HbA1c Binary Endpoints

- HbA1c <7.0%
- HbA1c <7.5%
- Relative reduction ≥10% from baseline
- Absolute reduction ≥0.5% from baseline
- Absolute reduction ≥1% from baseline
- Absolute reduction ≥1% from baseline or HbA1c <7.0%

Insulin and Other Endpoints

- Total daily insulin
- Daily basal insulin
- Daily bolus insulin
- Change of body weight (BMI z-score) from screening to 6 months visit
- Change of blood pressure from screening to 6 months visit

4.3.1. Metrics Calculation

CGM Metrics

All CGM metrics will be calculated separately using blinded CGM sensors wear data at 3 months visit and 6 months visit for up to 14 days. Up to 2 weeks of blinded sensor data pre-randomization will be used to calculate baseline CGM indices. Minimum of 5 days (120 hours) of CGM data are required to calculate CGM indices at each time point. For daytime and nighttime glycemic indices, at least 80 hours of daytime use and 40 hours of nighttime use will be required.

4.3.2. Analysis of the Secondary Endpoints

Cross-sectional Analyses

For CGM and insulin endpoints, mean (SD) or median (IQR), whichever is appropriate for the distribution, will be tabulated by treatment group at baseline, 3 months visit and 6 months visit. For binary HbA1c endpoints, number and percentages will be tabulated. For body weight and blood pressure, mean (SD) or median (IQR) will be tabulated by treatment group at baseline and 6 months visit.

For all the secondary efficacy endpoints, P-value will be only calculated for treatment group difference at 6 months visit. Linear regression models as described in section 4.1 will be used to

compare continuous outcomes. Residuals will be inspected for approximately normal distribution. If showing non-normal distribution, ranked normal score transformation of outcome data will be applied in the regression model. Logistic regression models will be used to compare the treatment difference at 6 months for binary HbA1c outcomes while adjusting for baseline HbA1c as a continuous variable, age and random site effect. For comparing change of body weight from baseline to end of study between treatment groups, age, gender, baseline weight and baseline HbA1c will be adjusted in the model.

Longitudinal Analyses

Longitudinal analyses will be conducted to compare each of following selected endpoints at baseline, 3 and 6 months between the two treatment groups. The analyses will be performed using a 3-level repeated measures least square model with random intercept and slope within subjects and sites, while adjusting for age and baseline HbA1c (for non-HbA1c outcomes) as fixed effect covariates.

- HbA1c
- Percentage of CGM glucose from 3.9 to 10.0 mmol/l (70-180 mg/dl)
- Mean glucose
- Glucose variability as measured by standard deviation and by coefficient of variation
- Percentage of CGM glucose values <3.5 mmol/l (63 mg/dl)
- Amount of delivered insulin

4.4. Safety Evaluation

The following events will be tabulated and statistically compared by treatment group:

- Number of SH events per subject and SH event rate per 100 person-years
- Number of DKA events per subject and DKA event rate per 100 person-years
- CGM-measured hypoglycemic events per week (>15 minutes with glucose concentration <54 mg/dL)
- CGM-measured hyperglycemic events per week (>15 minutes with glucose concentration >300 mg/dL
- Proportion of subjects with worsening of HbA1c from baseline to 6 months by >0.5%

In addition the following will be tabulated without formal statistical testing between groups

- Proportion of subjects with 1 or more SH events
- Proportion of subjects with 1 or more DKA events
- Any ketonemia event
- Any hyperglycemia event other than DKA and ketonemia
- Any adverse event other than hypoglycemia, hyperglycemia, DKA, and ketonemia which have been reported separately
- Any serious adverse events (SAE) other than hypoglycemia, hyperglycemia, DKA, and ketonemia which have been reported separately
- Adverse device effects (ADE)
- Unanticipated adverse device effects (UADE)

Comparison of safety outcomes between the two treatment groups will only include those events occurring on or after randomization until the completion of the 6-month visit or if the subject

drops from the study prior to the 6-month visit, then the last visit date or the last event date (whichever is later). The event rates will be calculated by dividing the number of events by the duration of (last visit/AE date – randomization date + 1).

If there are enough observed events to allow formal statistical modelling the following analyses will be conducted. For comparing event rates, Poisson regression models will be constructed to compare the treatment group difference for event rates by adjusting for age, baseline HbA1c and random site effect. If any outlier exist, a robust Poisson regression model will be used instead. For binary HbA1c outcome, logistic regression models will be used to compare the treatment group difference by adjusting for age, baseline HbA1c and random site effect. No adjustment of p-values will be made on above safety outcomes.

4.5. Closed Loop System Use/Function

The amount of CGM use in the CL arm will be calculated over the period starting from the day after intervention started until the earlier of the 6 month visit date or Day 183 from the day of intervention start (whichever comes first). If a subject drops out of the study without completing the 6 month visit, then they will be counted as zero use from that point forward until Day 183. The amount of system use in the CL arm will be calculated in a similar manner.

Boxplots will be created for % time closed loop system use and % time CGM use overall and by monthly period in the CL treatment arm.

Number of LGS events over the whole study period and in a monthly window from closed-loop initiation date will be tabulated.

4.6. Sensitivity Analyses

Missing Data

Sensitivity analyses will be conducted for the primary outcome only. It is not expected that a significant number of subjects will have missing HbA1c at 6 months that will meaningfully change the main conclusion. Any subjects who drop out the study will be also invited back to take HbA1c sample at 6 months. The following sensitivity analyses will be conducted:

- Only subjects who completed the 6-month HbA1c exam will be included (this will include any subjects who withdraw early but subsequently return to take HbA1c sample at 6 months). HbA1c values from previous outcome visits will be ignored.
- Direct likelihood method (MMRM model) with all available lab HbA1c values at baseline, 3 and 6 months included as dependent variable and local HbA1c values included as auxiliary variable
- Rubin's multiple imputation with treatment group in the imputation model
- Rubin's multiple imputation separately by treatment groups

Confounding

Repeat primary analysis by including other potential confounding factors beyond age, baseline HbA1c and site at enrollment. Following potential confounding factors will be assessed for imbalance between groups:

- gender

- race/ethnicity
- diabetes duration
- BMI
- baseline C-peptide
- total daily insulin
- blood glucose
- number of SH and DKA in last 12 months
- CGM use
- Pump use

The imbalance will be assessed based on clinical judgement reviewing the distributions in the two treatment arms, not on a p-value.

4.7. Subgroup Analysis

The treatment effect in subgroups based on baseline factors will be assessed in pre-planned subgroup analyses. These analyses will be conducted to determine whether a similar trend to the overall treatment effect is seen in these subgroups. The study is not expected to have sufficient statistical power for definitive conclusions in subgroups and statistical power will be low to formally assess for the presence of interaction. Interpretation of subgroup analyses will depend on whether the overall analysis demonstrates a significant treatment effect. In the absence of any significant treatment effects in the primary analysis, assessment of subgroups will be considered exploratory and used to suggest hypotheses for further investigation in future studies.

HbA1c Outcome

The planned subgroups analysis for HbA1c at 6 months are as follow:

- Baseline HbA1c: <8.0% and ≥8.0% (*cutpoint to be reassessed after enrollment completed*)
- Baseline CGM measured hypoglycemia: <5% vs. ≥5% time <3.9 mmol/l (70 mg/dl)
- Age: 6-12 years and 13-18 years
- Gender
- Race/Ethnicity
- Clinical site
- Baseline C-peptide level: below median vs. above median

CGM Endpoints

The planned subgroups analysis for CGM endpoints will include age subgroups (6-12 years, 13-18 years) analysis for following selected outcomes:

- Percentage of CGM glucose values 3.9-10.0 mmol/l (70-180 mg/dl)
- Percentage of CGM glucose values <3.5 mmol/l (63 mg/dl)

For each subgroup, mean change of HbA1c or other CGM endpoints from baseline to 6 months between CL and usual care groups will be tabulated by treatment group. Interactions between the subgroup factor and the treatment group will be tested in the linear regression models. The general approach for these exploratory analyses will be to add an interaction term for the subgroup factor by treatment into the linear regression model used for the primary/secondary

analyses. For continuous variables, the interaction between the continuous term and treatment group will be entered into the linear regression model. For % glucose <3.5 mmol/l (63 mg/dl) outcome, since it usually has skewed distribution, ranked normal score transformation will be used in regression models to calculate p-value. FDR-adjusted p-values for the interaction terms will be calculated among these subgroup analyses.

For the HbA1c outcome, a test for random treatment by center interaction effects will be performed. If significant then a forest plot of the estimated random treatment effect and its 99% CI for each site will be evaluated.

4.8. Per-Protocol Analyses

Per-protocol analyses will be performed to replicate the primary analysis on the subjects who meet the following criteria if at least 10% of the sample will be excluded:

- Did not withdraw from the study or miss the 6 months visit (withdrawals excluded even if they return for a 6 month HbA1c measurement)
- CL arm: Closed loop active for at least 70% of the time

Analyses will use the same models described above in section 4.1.

4.9. Multiple Comparisons

Key Endpoints

For comparison of the key endpoints listed in section 4.2, the familywise type I error rate (FWER) will be controlled at two-sided $\alpha = 0.05$. A gatekeeping strategy will be used as described above in Section 4.2, where the primary endpoint will be tested first, if passing the significance testing, other key endpoints will be tested in the order listed using the fixed-sequence method at $\alpha = 0.05$.

Other Efficacy Endpoints

For comparison of other efficacy endpoints considered exploratory, the false discovery rate (FDR) will be calculated using the Benjamini-Hochberg method (Benjamini, 2001) adapted using the two-stage test (TST; Benjamini, 2006). FDR adjusted p-values will be calculated separately for the following categories:

- CGM metrics over all 24 hours
- CGM metrics by daytime
- CGM metrics by nighttime
- HbA1c binary endpoints & longitudinal analyses
- Insulin/BMI/blood pressure
- Subgroup analyses

The longitudinal analyses described in section 4.3.2 will be placed into the categories above according to the outcome metric (e.g., longitudinal analysis of percentage of CGM glucose from 3.9 to 10.0 mmol/l will be included in the CGM metrics over 24 hours category).

P-values from safety analyses, sensitivity analyses and per-protocol analyses will not be adjusted for multiple comparisons.

4.10. Analysis of Human Factors/Questionnaire Data

Human Factors data collected during study assessments will be scored based on established procedures and descriptive statistics (e.g., mean, range) will be calculated for each questionnaire at each time point and across treatment groups. In addition, questionnaires that have established clinical thresholds (i.e., score above a certain number indicates clinically elevated scores), crosstabs will be used to display scores by treatment group. This includes the EQ-5D-Y, CESD, and Clarke Hypoglycemia Unawareness scale.

Next, these data will be analyzed in two ways. First, questionnaire scores (or dichotomized values based on clinical thresholds) will be used as baseline covariates in designated efficacy analyses described earlier to determine if outcomes (e.g., time in target range) differ based on levels at study entry. For example, these analyses will document whether someone with higher diabetes distress at baseline achieves poorer outcomes than those with less distress. The second way human factors data are analyzed is to determine if changes occur over time and between treatment groups. Using SAS v9.4, we will construct predictive models in the general linear framework (with repeated measures) to examine change in each set of human factors (e.g., diabetes distress, INSPIRE survey) over time. To test differences between groups, treatment group assignment will be the primary covariate and other variables potentially associated with scores (e.g., diabetes duration) will also be included as covariates. Common confounding variables (e.g., age) are limited in this sample and measures are restricted to narrow age ranges, thus interactions are not expected and interaction terms will not be included in these models.

4.11. Cost Utility/Economic Analysis

Data collection for the cost utility analysis will be broken into several components. One component will utilize two health related quality of life status surveys – the Child Health Utility-9D and EQ-5D-Y. These multidimensional status measures will be converted into univariate preference-based health status measures for each patient using previously established population-derived preference values. Descriptive statistics (e.g., mean, range, etc.) will be reported for these measures, stratified by treatment group, as well as their temporal variation over the course of the study. This health preference status information will also be incorporated into the utility calculation, denominated in quality adjusted life years (QALYs), comparing changes in health status between the treatment and control groups over the course of the study. Standard statistical significance tests will be applied to those differences, specifically including 2-sided t-tests, both paired (for pre- versus post- measurements) and unpaired (for treatment versus control measurements). Results will be reported by treatment group and time.

We will also conduct a discrete choice experiment (DCE) to assess participants' (patients' and parents') strength of preference trade-offs for various potential features of a future hybrid closed loop ('artificial pancreas') device. The four features of interest that will be tested are: out of pocket cost, daily time burden required for device management, effectiveness of diabetes control, and the sense of life 'normality' that using the device provides. Each participant will be

presented with a series of roughly ten 'A versus B' binary choices, with the two alternatives each describing a hypothetical device offering some relatively superior and some relatively inferior feature performance levels. Participants' response sets will be amalgamated into overall group preference levels for each feature using standard conjoint analysis regression techniques. Those preferences will be reported both relative to each other, and as a willingness to trade off out of pocket expense for additional performance on the other three features. Comparisons will be reported between the treatment and control groups, and over time within those groups including the results of standard significance tests.

Additional data will be collected on specific changes in each patient's health status, including all of the pre-specified outcome measures, such as HbA1c, as well as any adverse events. All such changes in health status will each be converted, using previously published data regarding the health utility value associated with the relevant health conditions, into a univariate preference-based health status, then multiplied by the duration of that change to calculate QALY increments or decrements. These QALY increments and decrements will then be aggregated for both the control and treatment groups to calculate the average incremental change in effectiveness (QALYs) achieved by the intervention over the course of the study. Future incremental effectiveness (beyond the study period) will be modeled using a multi-period Markoff model, using as its main initial input the average HbA1c levels of the treatment and the control groups, with other model parameters derived from prior published studies. Data on incremental costs will similarly be calculated by analyzing the categorical data on all health events experienced by each participant, including routine care provided; devices, medication, and consumable supplies used; and any adverse health events that occurred which required treatment. U.S. national standard costs for each of those items (derived from previously published data) will be applied to calculate the average incremental cost difference between the treatment and control groups. A similar approach will be applied to denominate incremental cost increases and savings beyond the study period using results from the numerical predictive model. These cost and effectiveness data will then be combined to report incremental cost effectiveness ratios (ICERs) for both the one year study period, and for a longer term (i.e., 10+ year, future discounted) interval.

4.12. Protocol Adherence and Additional Tabulations/Plots

The following tabulations and analyses will be performed according to treatment group:

- Baseline demographic and clinical characteristics
- A flow chart accounting for all subjects according to treatment group for all visits
- Visit completion rates for each follow-up visit according to treatment group
- Protocol deviations
- Device issues
- Number of and reasons for unscheduled visits and phone calls
- Scatter plots for close-loop use and selected CGM indices at 12 months

4.13. Exploratory Analysis of the Closed Loop System Version

The automated closed loop control algorithm was implemented on Medtronic 640G system called the FlorenceM at the beginning of the study; however, due to multiple device issues found which caused low usage of the CL system, it has been decided and approved that current and

future participants at UK sites will switch to the Dana-RS insulin pump and Dexcom G6 sensor with the same closed loop control algorithm installed also called the CamAPS FX (first date of switching CL system was 16Aug2009 and with earliest randomization date of 18Jun2019 among those participants). The main analyses will be based on pooled data (all participants in the CL group), and analyses of the two different systems separately will also be conducted as exploratory analyses. It is worth emphasizing that the sample size planned for this study is not powered for the exploratory analyses in this section and the conclusion of the study will not be drawn based on the findings from below exploratory analyses.

The exploratory analyses will be conducted on the primary endpoint and key secondary endpoints (listed below) by comparing these outcomes from two subgroups separately: 1) CL group participants who used the CamAPS FX at UK sites vs UK participants in standard care group who were randomized on or after 18Jun2019; 2) CL group participants who used the FlorenceM throughout the study vs the rest of the participants in standard care group. Similar regression models as described above with be used for these exploratory analyses and no formal adjustment for multiple comparisons will be conducted. Analysis will be available cases only.

- HbA1c
- Time spent in the target glucose range from 3.9 to 10.0 mmol/l (70 to 180 mg/dl)
- Mean glucose
- Time spent above target glucose 10.0 mmol/l (180 mg/dl)
- Time spent below target glucose 3.9 mmol/l (70 mg/dl)

Additionally, closed loop usage and glucose sensor usage will be tabulated.

5. Post-hoc Analysis Added After Version 5.0

Exploratory Analysis of Closed Loop System Version

In addition to the analysis described in section 4.13, the secondary outcomes listed below will be compared cross-sectionally between treatment groups at 6 months separately in the following subgroups: 1) CL group participants who used the CamAPS FX at UK sites vs UK participants in standard care group who were randomized on or after 18Jun2019; 2) CL group participants who used the FlorenceM throughout the study vs the rest of the participants in standard care group.

CGM-Measured:

- Glucose standard deviation
- Glucose coefficient of variation.
- Time with glucose levels <3.5 mmol/l (63 mg/dl)
- Time with glucose values <3.0 mmol/l (54 mg/dl)
- AUC of glucose below 3.5 mmol/l (63 mg/dl)
- Time with glucose levels > 16.7 mmol/l (300 mg/dl)
- AUC of glucose above 10.0 mmol/l (180 mg/dl)

Following CGM metrics will also be calculated separately for day-time period (06:00-23:59) and night-time period (00:00-05:59):

- Percentage of CGM glucose from 3.9 to 10.0 mmol/l (70-180 mg/dl)
- Mean glucose
- Glucose variability as measured by standard deviation
- Percentage of CGM glucose values <3.5 mmol/l (63 mg/dl)

HbA1c Binary Endpoints

- HbA1c <7.0%
- HbA1c <7.5%
- Relative reduction ≥10% from baseline
- Absolute reduction ≥0.5% from baseline
- Absolute reduction ≥1% from baseline
- Absolute reduction ≥1% from baseline or HbA1c <7.0%

Insulin and Other Endpoints

- Total daily insulin
- Daily basal insulin
- Daily bolus insulin
- Change of body weight (BMI z-score) from screening to 6 months visit
- Change of blood pressure from screening to 6 months visit

Similar regression models as described in section 4.3.2 with be used for these exploratory analyses. Adjustment for multiple comparisons will follow the same approach as the overall analysis with the following FDR categories separate for each subgroup:

- CGM metrics over all 24 hours
- CGM metrics by daytime
- CGM metrics by nighttime
- HbA1c binary endpoints
- Insulin/BMI/blood pressure

Additionally, the human factors/questionnaire analysis and cost utility/economic analysis will be conducted separately for each subgroup.

The following will be tabulated by treatment group separately for each subgroup with no formal testing:

- HbA1c by age and baseline HbA1c

Comparison of CGM Data Sources

For the participants in the closed-loop group who used the CamAPS FX system, CGM metrics will be recalculated at 3 and 6 months using data obtained from the Dexcom G6 CGM. Data

from the Dexcom G6 will be selected to match the dates of available data from the blinded Abbott Libre.

Summary statistics of the CGM metrics calculated based on the Dexcom G6 will be tabulated by period. Scatterplots will be created to compare CGM metrics calculated for the same individual from the Abbott Libre and the Dexcom G6.

Tabulation of Adverse Events

The number of adverse events and the number of participants experiencing at least one adverse event will be tabulated by treatment group. In addition, adverse events will be categorized as one of the following:

- Severe hypoglycemia
- Diabetic ketoacidosis
- Other reportable hypoglycemia
- Other reportable hyperglycemia
- Other serious adverse events
- Other reportable adverse events

The number of adverse events and the number of participants experiencing at least one adverse event will be tabulated by treatment group and above category. Finally, adverse events in the closed-loop group will be tabulated overall and separately by system in use at the time of the event.

Tabulation of Unscheduled Contacts and Device Issues

Unscheduled contacts and device issues will be tabulated in the CL group by system in use.

Reference

Benjamini YB, Yekutieli D: The control of the false discovery rate in multiple testing under dependency, 2001; *Ann Stat*, 29(4):1165-1188.

Benjamini YB, Krieger AM, Yekutieli D: Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 2006; 93(3):491-507.

Huque MF, Alosh M: A flexible fixed-sequence testing method for hierarchically ordered correlated multiple endpoints in clinical trials, 2008; *J Stat Plan Inference*, 138(2):321-335.