The International Diabetes Closed Loop Protocol 5 (DCLP5) Trial: Clinical Acceptance of the Artificial Pancreas in Pediatrics

Primary Study Phase

Statistical Analysis Plan

Version 2.1

December 20, 2019

Based on Protocol Version 2.5

Note: The table shells will be included in a separate document

Version History

Version	Author	Approvers	Effective Date	Study Stage	Protocol Version
1.0	Peiyao Cheng	Craig Kollman	3/26/2019	Planning, enrollment not started yet	2.3
1.1	Peiyao Cheng	Craig Kollman	3/28/2019	Planning, enrollment not started yet	2.4
2.0	Peiyao Cheng	Craig Kollman	6/7/2019	Planning, enrollment not started yet	2.5
2.1	Lauren Kanapka	Craig Kollman	12/20/2019	Enrollment complete, follow-up in progress	2.5

Version	Revision Description
1.0	Original Version
1.1	Made correction on Section 4.2.2 for HFS subscales
2.0	Moved analyses for extension phase into a stand-alone document; revised per-protocol analyses criteria, safety analyses, and subgroup analyses; clarification made on various sections
2.1	Made clarifications to the calculation of CGM metrics, insulin metrics, and analysis windows for insulin metrics; increased the minimum hours needed to calculate CGM metrics at certain time points; added the calculation of CGM metrics by 1-week periods; removed treatment group comparison for INSPIRE survey which is not given to both groups at follow-up; added some exploratory analysis for a binary CGM outcome and tabulation of additional insulin metrics

Approvals

Role	Digital Signature or Handwritten Signature/Date
Author: Lauren Kanapka, JCHR	
Senior Statistician: Craig Kollman, JCHR	
Coordinating Center Director: Katrina Ruedy, JCHR	
Protocol Chair: R. Paul Wadwa, Barbara Davis Center	
Sponsor: Marc Breton, University of Virginia	

1. Study Overview

This document outlines the statistical analyses to be performed for the DCLP5 Trial and covers the analyses for the primary study phase. The analysis plan for the extension phase is detailed in a separate document.

The following table gives an overview of the study.

Table 1. Study Overview

PARTICIPANT AREA	DESCRIPTION
Title	The International Diabetes Closed Loop (iDCL) trial: Clinical Acceptance of the Artificial Pancreas in Pediatrics- A study of t:slim X2 with Control-IQ Technology
Précis	A randomized controlled trial of at-home closed loop system vs. standard of care (defined as either sensor-augmented pump or any kind of low predictive low blood glucose suspend [PLGS; LGS] if participant is currently using) in youth age 6 to 13 years old.
Investigational Device	t:slim X2 with Control-IQ and Dexcom G6 system
Objectives	The objective of the study is to assess efficacy and safety of a closed loop control (CLC) system (t:slim X2 with Control-IQ Technology) in a randomized controlled trial.
Study Design	A 16-week parallel group randomized clinical trial with 3:1 randomization to intervention with the closed loop system vs. standard of care (SC).
Number of Clinical Centers	Up to 4 US clinical centers
Endpoint	The primary outcome for the first phase is time in target range 70-180 mg/dL measured by CGM in CLC group vs. SC group over 16 weeks
Population	**Key Inclusion Criteria** • Type 1 Diabetes • Ages ≥ 6 and ≤ 13 years old **Key Exclusion Criteria** • Use of any non-insulin glucose-lowering agents except metformin • Actively using any other closed-loop system
Sample Size	Up to 150 screened participants with the goal of randomizing 100 participants in this 16-week randomized trial.
Treatment Groups	• Intervention Group: t:slim X2 with Control-IQ Technology and Study CGM. • Control Group: Standard of care (SC) (defined as either sensor-augmented pump or any kind of low or predictive low blood glucose suspend [PLGS; LGS] if participant is currently using), and study CGM
Participant Duration	16-20 weeks (depending on duration of run-in phase)
Protocol Overview/Synopsis	After consent is signed, eligibility will be assessed. Eligible participants not currently using an insulin pump and Dexcom G4, G5 or Dexcom G6 CGM with minimum data requirements will initiate a run-in phase of 2 to 4 weeks that will be customized based on whether the participant is already a pump or CGM user. Participants who skip or successfully complete the run-in will be randomly assigned 3:1 to the use of closed-loop control (CLC group) system using Tandem t:slim X2 with Control-IQ Technology vs SC for 16 weeks.

The following table provides overview of the schedule of study visits, phone contacts, and key procedures.

Table 2. Schedule of Study Visits and Procedures

	Pre	Pre	0	1w	2w	4w	6w	8w	10w	12w	14w	16w
Visit (V) or Phone (P)	V	V	V	P	V	P	P	V	P	P	P	V
Comment	Screen/ Enroll	Run-in	Rand									
Eligibility Assessment	X	X	X									
HbA1c (DCA Vantage or similar point of care device, or local lab)	X		X									X
HbA1c (Central lab)			X									X
C-peptide (Central lab) and blood glucose assessment			X									
Pregnancy test (females of child-bearing potential)	X		X					X				X
Device Data download(s)	X	X	X	X	X	X	X	X	X	X	X	X
Review diabetes management and AEs		X	X	X	X	X	X	X	X			X
Questionnaires			X									X

2. Statistical Hypotheses

The primary outcome for this study is CGM-measured % in range 70-180 mg/dL over a 16-week period. The intervention will be considered effective if the Closed-Loop Control [CLC] is superior to the Standard Care [SC] using a statistical significance of $\alpha=0.05$ and the model specified below in Section 6 (i.e., $p < 0.05$).

The null/alternative hypotheses are:

1. *Null Hypothesis*: There is no difference in mean CGM-measured % in range 70-180 mg/dL over 16 weeks between SC and CLC
2. *Alternative Hypothesis*: The mean CGM-measured % in range 70-180 mg/dL over 16 weeks is different for SC and CLC.

3. Sample Size

Sample size has been computed for the primary outcome (CGM-measured % in range 70-180 mg/dL). Data from IDE G170267; Device Name: t:slim X2 with Control-IQ Technology; "Real-Time Monitoring and Glucose Control During Winter-Sport Exercise in Youth with Type 1 Diabetes: The AP Ski Camp Continued" were used to calculate sample size specific to this age group. In this study, which was completed in the winter of 2018, 24 school-aged children (6-12 years) with type 1 diabetes participated in a 3-day ski camp (~5 h skiing/day), followed by an additional 72 hour at-home phase under parental supervision. Study participants were randomized 1:1 to SAP and t:slim X2 with Control-IQ Technology. The data from the 72-hour home phase was used for this sample size calculation – *note that the closed-loop control system and the age range of the participants are identical to those proposed in this application:*

Results from home phase of G170267	Control IQ	SAP	F	p value
Percent between 70 and 180mg/dl	71 ± 6.6	52.8 ± 13.5	16.4	0.001

From the DCLP1 study using the same algorithm in an older cohort, the effective standard deviation (after adjusting for the correlation between baseline and follow up) for time in range 70-180 mg/dL over the course of 6 months was 6% (95% CI 5% to 7%) for the CLC group and 7% (95% CI 6% to 8%) for the SAP group.

A total sample size was computed to be N=60 for the following assumptions: (1) 3:1 [CLC:SC] randomization, (2) 90% power, (3) a 10% absolute increase in % time in range 70-180 mg/dL, (4) an effective SD of 10%, and (5) 2-sided type 1 error of 0.05.

The total sample size has been increased to N=100 to account for dropouts and to increase the number of participants who will be exposed to the CLC system for an enhanced safety and feasibility assessment.

4. Efficacy Outcome Measures

4.1. Primary Endpoint:

- CGM-measured % in range 70-180 mg/dL

4.2. Secondary Endpoints

4.2.1. Hierarchical Endpoints

The following secondary endpoints will be tested in a hierarchical fashion between the two intervention arms as described in Appendix.

- CGM-measured % above 180 mg/dL
- CGM-measured mean glucose
- HbA1c at 16 weeks
- CGM-measured % below 70 mg/dL
- CGM-measured % below 54 mg/dL
- CGM-measured % above 250 mg/dL
- Glucose variability measured with the coefficient of variation (CV)

4.2.2. Other Secondary Endpoints

The following endpoints are considered exploratory.

- CGM-Measured
  - % in range 70-140 mg/dL
  - glucose variability measured with the standard deviation (SD)
  - % <60 mg/dL
  - low blood glucose index (LBGI)*
  - hypoglycemia events (defined as at least 15 consecutive minutes <70 mg/dL)
  - % >300 mg/dL
  - high blood glucose index (HBGI)*
  - % in range 70-180 mg/dL improvement from baseline to 16 weeks ≥5%
  - % in range 70-180 mg/dL improvement from baseline to 16 weeks ≥10%
- HbA1c
  - HbA1c <7.0% at 16 weeks
  - HbA1c <7.5% at 16 weeks
  - HbA1c improvement from baseline to 16 weeks >0.5%
  - HbA1c improvement from baseline to 16 weeks >1.0%
  - HbA1c relative improvement from baseline to 16 weeks >10%
  - HbA1c absolute improvement from baseline to 16 weeks >1.0% or HbA1c <7.0% at 16 weeks
- Insulin
  - Total daily insulin (units/kg)
  - Basal: bolus insulin ratio
- Weight and body mass index (BMI)
- Questionnaires
  - Fear of Hypoglycemia Survey (HFS-II) – total score, 2 subscales and 4 factor scores:
    - Behavior
      - Avoidance
      - Maintain high BG
    - Worry
      - Helplessness
      - Social consequences
  - Clarke Hypoglycemia Awareness Scores

- Problem Areas In Diabetes Survey (PAID)
- INSPIRE survey scores
- PedsQL Diabetes Module – total score and 5 subscales:
  - Diabetes
  - Treatment I
  - Treatment II
  - Worry
  - Communication
- Pittsburgh Sleep Quality Index (Parents only)
- System Usability Scale (SUS) (Closed-Loop participants only)

*Note that LBGI and HBGI will be calculated using all available CGM readings as described below. Therefore, they may not be comparable to the same metrics calculated with SMBG data.

4.3 Calculation of CGM Metrics:

- Baseline: CGM data to calculate baseline metrics will either come from the run-in period, or from the participant's personal CGM device if the run-in is not necessary:
  - If an enrolled participant is currently using an insulin pump and a Dexcom G4, G5 or G6 with CGM data captured on at least 11 out of the 14 days prior to enrollment, then he/she can proceed directly to randomization.
  - Otherwise, the participant will need go through 2-6 weeks of run-in that includes CGM use, prior to randomization.
  - In either case, the last 2 weeks of CGM data prior to randomization will be used in the calculation of baseline CGM metrics. If <168hr of CGM data are available for any reason (e.g., lost data or device failure), then the baseline metrics will not be calculated and will be set to missing.
- Follow-up:
  - CGM metrics will be calculated by pooling all CGM readings in the 16-week period starting from the randomization visit up through the 16 week visit. If a participant drops out before completing the 16-week visit, all available data through the last visit date will be included for calculating CGM metrics. Minimum 168 hours of CGM data will be required to calculate CGM metrics.
  - A separate set of CGM metrics will be calculated by excluding the first two weeks of CGM data following the randomization visit for sensitivity analyses only with ≥168 hours of CGM data required.
- All CGM metrics at baseline and follow-up will be calculated giving equal weight to each sensor reading for each participant.
- HIGH and LOW values will be imputed as 39 and 401, respectively.
- CGM metrics above will also be calculated for daytime period (06:00AM to 11:59PM) and overnight period (00:00AM to 05:59AM) separately. Minimum 126 hours of CGM data will be required to calculate daytime metrics and minimum 42 hours of CGM data will be required to calculate overnight metrics. If <168 hours of CGM data available for combined day and night, then CGM metrics will not be calculated separately for daytime and overnight periods.
- CGM metrics by 4-week periods: Starting from the randomization date, CGM metrics will be calculated within each 4-week period (Week 1-4, Week 5-8, Week 9-12 and Week 13-16). The

16 weeks visit date will be used as end date of Week 13-16. Minimum 168 hours of CGM data in each 4-week period will be required for the calculation.

- CGM metrics by 1-week periods: Starting from the randomization date, CGM metrics will be calculated within each 1-week period. The 16 weeks visit date will be used as end date of Week 13-16. Minimum 132 hours of CGM data in each 1-week period will be required for the calculation.

4.4 Calculation of Insulin Metrics

- Insulin metrics will be calculated at randomization and 16 weeks using Tandem pump data where available, otherwise using data reported on the CRF.
- Insulin metrics will be calculated from the pump data using data in the 7 days prior to the visit.

4.5. Questionnaires

All questionnaires will be administered online and participants/parents can skip specific questionnaires or items within a questionnaire. All questionnaires will be scored according to the instructions given in the manual. In case no manual exists for a given questionnaire or the manual does not provide guidance on how to handle missing data, then the following criteria will be applied:

- At least 75% of the questions must be completed to be included in the analysis.
- This 75% rule will be applied separately for the total score and each subscale so it is possible the sample size will be different for some subscales.
- The score used for analysis will be based on the average among the questions that were answered and then scaled accordingly.

4.6 Analysis Windows

Analysis windows apply to the following outcomes measured at 16 week visit:

- HbA1c
- Insulin metrics if calculated from CRF data
- Weight/BMI
- Questionnaires

This does not apply to the CGM metrics which are calculated as described above.

Data from 16 weeks visit occurring in the following windows will be included in analysis:

Visit (Target Date)	Metrics[a]	From Day[b]	Thru Day[b]
16 weeks (112 days)	H,I,B,Q	98	126

a – H = HbA1c, I = Insulin metrics, B=BMI (and weight), Q = Questionnaires.
b – Days from randomization, inclusive.

5. Analysis Datasets and Sensitivity Analyses

5.1 Analysis Datasets

All analyses comparing the CLC arm with SC arm will follow intention-to-treatment approach, which means participants will be analyzed in the treatment arm assigned by randomization regardless of compliance. All randomized participants will be included in the primary analysis and secondary

hierarchical analyses of CGM metrics. For other secondary outcomes, only participants with non-missing outcome data will be included.

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

5.2 Sensitivity Analyses

The following sensitivity analyses (except for the per-protocol analyses) will be conducted for primary endpoint only.

Per Protocol (PP) Analyses:

The following per-protocol analyses will be performed for the primary outcome and secondary hierarchical outcomes only if >5% of participants will be excluded:

- CLC group: Closed loop mode active for at least 80% of the time
- SC group: CGM use for at least 80% of the time

Confounding:

A sensitivity analysis will also be conducted if potential confounding factors collected at baseline are detected. The baseline factors listed in Section 11 will be assessed for imbalance between treatment groups.

The imbalance will be assessed based on clinical judgement reviewing the distributions in the two treatment arms, not on a p-value. The person making this judgement will be unaware of whether there is an association between baseline variables and study outcome. All variables obtained on a continuous scale will be entered into the models as continuous variables, 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.

Exclude First 2 Weeks of CGM Data

The primary analysis will be repeated by excluding the first 2 weeks of post-randomization CGM data.

Missing Data:

Missing data will be handled using direct likelihood method for the primary analysis. 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.

To that end, sensitivity analyses will be performed to explore whether results are similar for primary analysis when using different methods. The following methods will be applied:

- Rubin's multiple imputation with treatment group in the imputation model
- Available cases only
- Multiple imputation with pattern mixture model assuming the dropout trajectory of the CLC group was that of the SC group (Mallinckrodt and Clark, 2003)

6. Efficacy Analysis

6.1 Primary Analysis

This study primary outcome is CGM measured % time in range 70-180 mg/dL over 16 weeks.

Summary statistics (mean ± SD or median (quartiles)) will be reported by treatment group for the CGM-measured % in range 70-180 mg/dL at baseline, 16 weeks intervention and change from baseline to 16-week intervention.

Changes from run-in pre-randomization CGM wear to the 16-week post-randomization period in CGM-measured % in range 70-180 mg/dL between two treatment arms will be compared using a linear mixed effects regression model while adjusting for baseline CGM-measured % in range 70-180 mg/dL, age, prior CGM and pump use, and clinical center (random effect). A point estimate, 95% confidence interval and two-sided p-value will be reported for the treatment effect based on the linear regression model and a 5% level will be used to declare statistical significance. Residual values will be examined for an approximate normal distribution. If values are highly skewed then a transformation or robust statistical methods will be used instead. However, previous experience suggests that the residual values for % time glucose in target range 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 5).

Missing Data

In the primary analysis, any missing data at baseline or follow-up will be handled using direct likelihood. A longitudinal linear regression model will be fit with the percent of time in range at baseline and follow-up as the dependent variable. This model will adjust for age, prior CGM use and pump use as fixed effects and clinical center as a random effect. This model adjusts for baseline time in range by forcing the treatment groups to have the same mean value at baseline.

6.2 Analysis of the Secondary Endpoints

6.2.1 Hierarchical Analyses

To preserve the overall type 1 error for selected key secondary endpoints, a hierarchical testing procedure will be used. If the primary analysis for time in range described above results in a statistically significant result ($p < 0.05$), then testing (similar to the model described above for the primary outcome) will proceed to the next outcome metric in the following order:

- CGM-measured % in range 70-180 mg/dL (primary outcome)
- CGM-measured % above 180 mg/dL
- CGM-measured mean glucose
- HbA1c at 16 weeks
- CGM-measured % below 70 mg/dL
- CGM-measured % below 54 mg/dL
- CGM-measured % above 250 mg/dL
- Glucose variability measured with the coefficient of variation (CV)

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 eight 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 (see example in the Appendix).

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 seven hierarchical secondary 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.

Analysis for each of the CGM metrics listed above for the hierarchical analysis will parallel the analysis described for the primary outcome in Section 6.1.

For the HbA1c analysis, change in HbA1c from baseline to 16 weeks will be compared between the two treatment arms using a linear model while adjusting for baseline HbA1c, age, prior CGM and pump use, and clinical center (random factor). Missing data will be handled using direct likelihood in a regression model including all available central laboratory HbA1c measurements at baseline and 16-week visits. When available, the local HbA1c measurement will be included in the regression model as an auxiliary variable.

For all above analyses, regression diagnostics will be employed analogous to as described in Section 6.1 for the primary outcome.

6.2.2 Other Secondary Analyses

For all other secondary endpoints, only participants with non-missing data will be included in analyses (available cases method). Summary statistics (mean ± SD, median (IQR) or n (%)) appropriate to the distribution will be tabulated for them at baseline, 16 weeks and for the changes from baseline to 16 weeks. For continuous outcomes, linear regression models will be used to compare the treatment effects while adjusting for corresponding baseline values (e.g., baseline % in range 70-140 mg/dL for comparing change in % in range 70-140 mg/dL from pre-randomization CGM wear to 16 weeks post-randomization period), age, prior CGM and pump use, and clinical center as a random effect. Comparisons of body weight and BMI will also be adjusted for gender.

Binary CGM and HbA1c Outcomes

For the binary HbA1c outcomes listed in Section 4, risk-adjusted percentages by treatment group will be computed at 16 weeks from a logistic regression model. The logistic regression will adjust for baseline HbA1c (as a continuous factor), age, prior CGM and pump use as fixed effects, and clinical site as a random effect. Similar analyses will be done for the binary CGM outcomes in Section 4.

Questionnaires

For each questionnaire, mean ± SD or percentiles appropriate to the distribution will be given by treatment group at baseline and 16 weeks. Group comparisons will be conducted for the total score (mean score) and subscales from participant version and parent version separately using similar linear models as described above. The SUS survey and the INSPIRE post-treatment survey will only be administered to the CLC group at the 16-week visit, and thus the scores will only be tabulated.

Boxplots

Boxplots stratified by treatment group will be given for the primary outcome and each of the key secondary endpoints listed in Section 6.2.1 in 4week and 1-week periods over the course of 16 weeks intervention.

7. Safety Analyses

All randomized participants will be included in these analyses and all their post-randomization safety events will be tabulated. Events occurring from the randomization date until and including the 16-week visit date or randomization date + 126 days whichever is earlier will be included.

Any pre-randomization adverse events will be tabulated separately and will include participants who were never randomized.

For the following outcomes, mean ± SD or summary statistics appropriate to the distribution will be tabulated by treatment group and formal statistical comparisons will be performed if there are enough events (at least 5 events combined over both treatment groups):

- Number of SH events and SH event rate per 100 person-years
- Number of DKA events and DKA event rate per 100 person-years
- Any adverse event rate per 100 person-years
- Number of calendar days with any ketone level ≥1.0 mmol/L (if ≥5 total calendar days combined)
- CGM-measured hypoglycemic events (≥15 minutes with glucose concentration <54 mg/dL)
- CGM-measured hyperglycemic events (≥15 minutes with glucose concentration >300 mg/dL)

If there are at least 5 SH events combined over both treatment groups, the rates will be compared using a robust Poisson regression. The regression will adjust for the participant-reported number of SH events 12 months prior to the start of the study and clinical center as random effect. The amount of follow up will be included as an offset covariate to compare the rates. If one of the treatment groups has zero events then a Poisson model will not converge and Fisher's exact test will be used instead. A similar analysis will be done for DKA if there are at least 5 total DKA events among both treatment groups combined. For any adverse event and number of calendar days with ketone event, similar analyses will be done except that clinical center will be the only covariate to be adjusted in the model.

The calculation for the two continuous CGM-measured outcomes will be based on the same inclusion criteria mentioned above for the primary outcome. The event rates per week will be compared using similar linear mixed effects regression models as described above for the primary outcome.

The following will be summarized and tabulated by treatment group:

- Other serious adverse events (SAE)
- BG-measured hypoglycemic events (days with at least one BG record <54 mg/dL)
- BG-measured hyperglycemic events (days with at least one BG record >350 mg/dL)
- Worsening of HbA1c from baseline to 16 weeks by >0.5%
- Related to Investigational Device Usage (intervention group only):
    - Adverse device effects (ADE)
    - Serious adverse device events (SADE)

- Unanticipated adverse device effects (UADE)

8. Intervention Adherence

Closed Loop System

The amount of closed loop system use for the CLC group will be calculated from downloaded data over the period from the day after randomization until the day before the 16 week visit. Any dropouts will be counted as zero use for the remainder of the study.

Percent time in different operation modes overall and by 4-week periods will be tabulated and boxplots will be created by 4-week periods.

CGM Use

The amount of CGM sensor use for both treatment groups will be calculated from downloaded data in the same way as closed-loop system use described above.

Percent time sensor use overall and by 4-week periods will be tabulated by treatment group and boxplots will be created by 4-week periods.

Blood Glucose Monitoring

Average daily frequency of downloaded BGM use for both treatment groups will be calculated similarly in the same time window. If a participant drops out of the study or no meter download is available for a specific period, average daily BGM use will be set to missing for that period.

The average daily frequency of downloaded BGM use overall and by 4-week periods will be tabulated by treatment group.

9. Protocol Adherence and Retention

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

- Number of protocol and procedural deviations
- Flow chart accounting for all enrolled participants up to randomization
- Flow chart of all randomized participants at all scheduled visits and phone contacts post treatment initiation
- Number of and reasons for unscheduled visits and phone calls
- Number of participants who stopped treatment and reasons

10. Baseline Descriptive Statistics

Baseline demographic and clinical characteristics of the cohort of all randomized participants will be summarized in a table using summary statistics appropriate to the distribution of each variable. Descriptive statistics will be displayed by treatment group for the following:

- Age

- Gender
- Race/ethnicity
- Income, education, and/or insurance status
- Diabetes duration
- Insulin method before enrollment (pump vs. MDI)
- CGM use before enrollment
- HbA1c
- BMI
- C-peptide
- Participant-reported number of SH and DKA 12 months prior to the start of the study
- Baseline CGM metrics including:
  - % in range 70-180 mg/dL
  - % time >180 mg/dL
  - mean glucose
  - % time <70 mg/dL
  - % time <54 mg/dL

11. Device Issue

The following tabulations and analyses will be performed by treatment group to assess device issues:

- Device malfunctions requiring study team contact and other reported device issues.
- Sensor performance metrics (difference, absolute relative difference, and International Organization for Standardization criteria) – if applicable, by sensor version. All BGM values except those for QC tests will be used as reference and CGM data within ±5 minutes of each BGM value will be paired together. Each BGM value will be paired only to one sensor value and vice versa.
- Rate of different failure events and alarms per 24 hours recorded by the Control-IQ system – overall and by month.

12. Planned Interim Analyses

No formal interim efficacy analyses are planned. The analysis of the RCT will be performed on completion of the RCT prior to the completion of the extension phase.

The DSMB will review safety data at intervals, with no formal stopping rules other than the guidelines provided in the participant-level and study-level stopping criteria (as defined in the protocol).

13. Subgroup Analyses

In exploratory analyses, the primary outcome (% time 70-180 mg/dL), % time <70 mg/dL and HbA1c at 16 weeks will be assessed separately in various subgroups. Subgroups will be defined according to the baseline value of the factors listed below. Statistical analysis will involve tests for interaction with treatment group.

Interpretation of subgroup analyses will be viewed with caution, particularly in the absence of an overall significant difference. For continuous variables, results will be displayed in subgroups based on cutpoints although the analysis will utilize the variable as continuous, except for age which will be

analyzed both as a continuous variable and in two age groups. If there is insufficient sample size in a given subgroup, the cutpoints for continuous measures may be adjusted per the observed distribution of values. Cutpoint selection for display purposes will be made masked to the outcome data.

- Baseline HbA1c
- Baseline CGM time spent <70 mg/dL
- Baseline CGM time spent >180 mg/dL
- Baseline CGM time 70-180 mg/dL
- Device use before the enrollment: pump/MDI, CGM/no CGM, and combinations of both
- Age (<10 vs. ≥10)
- Sex
- Race/Ethnicity
- Clinical center (test for random effects interaction with treatment group)
- Body mass index
- Income, education, and/or insurance status
- C-peptide level

14. Multiple Comparisons

Primary Analysis

Since there will be a single comparison for the primary outcome (CGM-measured % 70-180 mg/dL), no adjustment is needed.

Secondary Hierarchical Analyses

The hierarchical testing procedure described above will be used to control the overall type I error for the primary outcome plus seven key secondary outcomes identified above.

All Other Secondary Analyses

For comparison of all other efficacy endpoints, the false discovery rate (FDR) will be controlled using the adaptive Benjamini-Hochberg procedure (Benjamini and Hochberg, 2000). FDR adjusted p-values will be calculated separately for the following categories:

- All other CGM metrics
- HbA1c analyses
- Insulin, weight and BMI
- Questionnaires
- Subgroup analyses for % time 70-180 mg/dL
- Subgroup analyses for % time <70 mg/dL
- Subgroup analyses for HbA1c at 16 weeks

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

15. Exploratory analyses

CGM Metrics

The following composite binary outcome will be tabulated at baseline and follow-up and compared between treatment groups in a logistic model as described in section 6.2.2.

- % time in range 70-180 mg/dL > 70% and % time <70 mg/dL <4%

The following selected CGM metrics will be reported with the appropriate statistics for daytime (6:00am – 11:59pm) and nighttime (00:00am – 5:59am) separately. No p-values will be calculated for following analyses.

- % time in range 70-180 mg/dL
- mean glucose
- % above 180 mg/dL
- % below 70 mg/dL
- coefficient of variation

Above selected CGM metrics will also be reported by restricting the CGM data in the CLC arm based on following criteria. No p-values will be calculated for following analyses.

- using only the CGM data when the closed-loop is active

Additional Insulin Metrics

The following insulin metrics will be calculated using data described in section 4.4 and tabulated by treatment groups at baseline, 16 weeks and for the changes from baseline to 16 weeks. No p-values will be calculated for these metrics.

- Total daily basal insulin (units/kg)
- Total daily bolus insulin (units/kg)
    - Total daily manual bolus (units/kg)
    - Total daily automated bolus (units/kg)

The following will be calculated for the CLC group in the 1-week prior to randomization and by 1-week follow-up periods from pump data only:

- Total daily insulin (units/kg)
- Total daily basal insulin (units/kg)
- Total daily bolus insulin (units/kg)
    - Total daily manual bolus (units/kg)
    - Total daily automated bolus (units/kg)
- Number of manual insulin doses per day
- Number of manual insulin doses with carb announcement per day

References

Benjamini, Y. and Hochberg, Y.: On the adaptive control of the false discovery rate in multiple testing with independent statistics, *Journal of Educational and Behavioral Statistics*, 2000; 25: 60–83.

Mallinckrodt CH, Clark WS, et al: Assessing responses profiles from incomplete longitudinal clinical trial data under regulatory considerations, *J. Biopharm. Stat.*, 2003; 13(2): 179-190.

Appendix: Example for Hierarchical Testing

In the hypothetical scenario depicted in the table below, the first four outcome variables all have a significant result so testing continues to the fifth variable (CGM % below 70 mg/dL). The result is not significant for that fifth variable (p = 0.06) so testing stops. No formal hypothesis test is conducted for the last three variables on the list in this example scenario.

Table. Example Hierarchical Test Results

HIERARCHICAL ORDER	OUTCOME VARIABLE	TREATMENT ARM P-VALUE	SIGNIFICANT?	ACTION
1st	CGM % 70-180 mg/dL (primary outcome)	0.001	Yes	Test next variable
2nd	CGM % above 180 mg/dL	0.02	Yes	Test next variable
3rd	CGM mean glucose	0.007	Yes	Test next variable
4th	HbA1c at 16 weeks	0.03	Yes	Test next variable
5th	CGM % below 70 mg/dL	0.06	No	Stop formal testing
6th	CGM % below 54 mg/dL	Not tested	Unknown	N/A
7th	CGM % above 250 mg/dL	Not tested	Unknown	N/A
8th	Glucose CV	Not tested	Unknown	N/A