Pediatric Artificial Pancreas Study (PEDAP) Trial

Primary Study Phase (RCT)

Statistical Analysis Plan

Version 3.1

Corresponds to Version 11.0 of the Protocol

27 Version History

SAP Version	Author	Approver	Effective Date	Revision Description	Study Stage	Protocol Version
1.0	Colleen Bauza	Craig Kollman	March 11, 2021	Original Version	Enrollment has not started yet	4.0
2.0	Zack Reed	Craig Kollman	August 4, 2021	Removed winsorization analysis. Added more sensitivity analyses. Added analysis plan for extension phase.	RCT phase recruitment and follow-up in progress. No data analyses performed yet.	8.0
3.0	Zack Reed	Craig Kollman	February 10, 2022	Moved Extension phase section to a separate SAP. Added analysis window for Baseline. Corrected RCT safety analysis endpoint so that it agrees with the Protocol. Added omitted questionnaires (Hypoglycemia Confidence Scale and System Usability Scale) to Section 5.2.2. Corrected typos and inconsistencies. Clarified DCLP5 reference.	RCT and Extension phase recruitment and follow-up in progress.	11.0
3.1	Zack Reed	Craig Kollman	May 11, 2022	Removed analysis window for questionnaires, insulin metrics, and weight/BMI at Baseline. Reverts to analysis plan for SAP v2.0 and earlier because these variables are not measured at randomization. Clarified that Baseline insulin metrics must come from the Diabetes Screening CRF. End date for CGM metrics by 1-day periods will be Day 18 rather than the 2-week visit date. Removed exploratory insulin analyses that require pre-randomization pump data. Corrected grammar and typos.	All randomized participants have completed the RCT phase. Extension and Extended Use phases in progress.	11.0

Author: Zachariah Reed
I am digitally signing this document
2022-05-18 09:48-04:00

Senior Statistician: Craig Kollman
I am digitally signing this document
2022-05-18 11:37-04:00

JCHR Project Director: John Lum
I am digitally signing this document
2022-05-18 10:41-04:00

Sponsor: 

1. Study Overview

The following table provides an overview of the PEDAP Trial.

Table 1. Study Overview

PARTICIPANT AREA	DESCRIPTION
Title	The Pediatric Artificial Pancreas (PEDAP) trial: A Randomized Controlled Comparison of the Control-IQ technology Versus Standard Care in Young Children in Type 1 Diabetes
Précis	A randomized controlled trial of at-home closed loop system vs. standard care (defined as either multiple daily injections of insulin [MDI] or use of an insulin pump without hybrid closed-loop control capabilities [low-glucose suspend or predictive low-glucose suspend functionality is permitted]) in youth age 2 to <6 years old.
Investigational Device	t:slim X2 with Control-IQ and Dexcom G6 system
Objectives	The objective of the study is to assess efficacy, quality of life, and safety of a closed loop control (CLC) system (t:slim X2 with Control-IQ Technology) in a randomized controlled trial with partial crossover.
Study Design[1]	First phase: a 13-week parallel group randomized clinical trial with 2:1 randomization to intervention with the closed loop system vs. standard care (SC); Second Phase: following the RCT, a 13-week period where the Standard Care (SC) group will transition to use CLC and the experimental arm will extend the use of CLC for the same period. After 26 weeks, participants may continue using CLC for an additional Extended Use period. A subset of participants will be invited to join an optional exercise/meal challenge ancillary study.
Number of Clinical Centers	~3 US clinical centers
Endpoints	Efficacy The primary outcome for the RCT is time in target range 70-180 mg/dL (TIR) measured by CGM in CLC group vs. SC group over 13 weeks The primary outcome for the extension phase assessed separately for each treatment group is change in TIR comparing extension phase with RCT phase. Quality of Life Patient-reported outcome questionnaires will be completed. Safety The key safety outcomes are severe hypoglycemia and ketoacidosis.
Population	**Inclusion Criteria** 1. Clinical diagnosis, based on investigator assessment, of type 1 diabetes for at least 6 months and using insulin for at least 6 months 2. Familiarity and use of a carbohydrate ratio for meal boluses. 3. Age ≥2 and <6 years old 4. Living with one or more parent/legal guardian knowledgeable about emergency procedures for severe hypoglycemia and able to contact emergency services and study staff. 5. Investigator has confidence that the parent can successfully operate all study devices and is capable of adhering to the protocol 6. Willingness to switch to lispro (Humalog) or aspart (Novolog) if not using already, and to use no other insulin besides lispro (Humalog)

PARTICIPANT AREA	DESCRIPTION
	or aspart (Novolog) during the study for participants using a study-provided Tandem pump during the study. 7. Total daily insulin dose (TDD) at least 5 U/day 8. Body weight at least 20 lbs 9. Willingness not to start any new non-insulin glucose-lowering agent during the course of the trial 10. Participant and parent(s)/guardian(s) willingness to participate in all training sessions as directed by study staff. 11. Parent/guardian proficient in reading and writing English 12. Live in the United States, with no plans to move outside the United States during the study period **Exclusion Criteria** 1. Concurrent use of any non-insulin glucose-lowering agent (including GLP-1 agonists, Symlin, DPP-4 inhibitors, SGLT-2 inhibitors, sulfonylureas). 2. Hemophilia or any other bleeding disorder 3. History of >1 severe hypoglycemic event with seizure or loss of consciousness in the last 3 months 4. History of >1 DKA event in the last 6 months not related to illness, infusion set failure, or initial diagnosis 5. History of chronic renal disease or currently on hemodialysis 6. History of adrenal insufficiency 7. Hypothyroidism that is not adequately treated 8. Use of oral or injectable steroids within the last 8 weeks 9. Known, ongoing adhesive intolerance 10. Plans to receive blood transfusions or erythropoietin injections during the course of the study 11. A condition, which in the opinion of the investigator or designee, would put the participant or study at risk 12. Currently using any closed-loop system, or using an insulin pump that is incompatible with use of the study CGM 13. Participation in another pharmaceutical or device trial at the time of enrollment or during the study 14. Employed by, or having immediate family members employed by Tandem Diabetes Care, Inc., or having a direct supervisor at place of employment who is also directly involved in conducting the clinical trial (as a study investigator, coordinator, etc.); or having a first-degree relative who is directly involved in conducting the clinical trial
Sample Size	Up to 150 screened participants with the goal of randomizing 102 participants.
Treatment Groups	• Intervention Group: t:slim X2 with Control-IQ Technology and Study CGM. • Control Group: Standard care (SC) - defined as either multiple daily injections of insulin (MDI) or use of an insulin pump without hybrid closed-loop control capabilities (low-glucose suspend or predictive low-glucose suspend functionality is permitted) in conjunction with study CGM All participants will be offered to extend the study for an additional 13 weeks, with the SC group switching to the t:slim X2 with Control-IQ System after the 13-week RCT period. After 26 weeks, participants may continue using CLC for an additional Extended Use period.

PARTICIPANT AREA	DESCRIPTION
Participant Duration	~26-32 weeks for RCT and Extension Phase, depending on duration of run-in phase; up to an additional ~7 months in Extended Use period
Protocol Overview/Synopsis	After consent is signed, eligibility will be assessed. Eligible participants not currently using an insulin pump and Dexcom G5 or Dexcom G6 CGM with minimum data requirements will initiate a run-in phase of 2 to 6 weeks that will be customized based on whether the participant is already a CGM user. Participants who skip or successfully complete the run-in will be randomly assigned 2:1 to an intervention using Tandem t:slim X2 with Control-IQ Technology or the standard care control group using existing insulin therapy in conjunction with study CGM. All participants will continue their participation by using the t:slim X2 with Control-IQ system in an Extension Phase. After 26 weeks, participants may continue using CLC for an additional Extended Use period through no later than July 31, 2022. A subset of participants will complete an optional exercise/meal challenge ancillary study.

[1] *This SAP is restricted to the RCT Phase. Analyses of the Extension Phase, Extended Use Period, and Ancillary Study are specified in a separate SAP.*

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

Table 2. Schedule of Study Visits and Procedures During the RCT and Extension Phases[1]

	Pre	Pre	0	3d[2]	1w	2w	6w	10w	13w	13w + 3d	14w	15w	19w	23w	26w
Visit (V), Videoconference (VC), or Phone (P)	VC/V	VC/V	VC/V	VC/ P	P	VC/ V	VC/ V	P	VC/ V	VC/ P	P	VC/ V	VC/ V	P	VC/ V
Comment	Screen/ Enroll	Run-in	Rand												
Eligibility Assessment	X	X	X												
HbA1c (Central lab)			X						X						X
Device Data download(s)	X	X	X	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	X	X	X	X	X
Questionnaires as defined in section 9.2 of the Protocol	X								X						X

[1] *This SAP is restricted to the RCT phase. Analyses of the Extension phase are specified in a separate SAP.*

[2] *Only for participants on MDI at enrollment assigned to the CLC group*

2. Comparison with the Protocol

This SAP is consistent with version 11.0 of the protocol except for the following:

- The Hypoglycemia Confidence Scale and System Usability Scale have been added to the list of questionnaires in Section 5.2.2.
- Changed sensitivity analysis in Section 6.3 to include all hierarchical outcomes instead of just % in range 70-180 mg/dL.
- A sensitivity analysis including CGM data after the 13-week visit has been added (Section 6.3).
- Exploratory analyses that require pre-randomization pump data have been removed.

3. Primary Statistical Hypotheses

The primary outcome for this study is CGM-measured % in range 70-180 mg/dL over a 13-week period. The intervention will be considered effective if the Closed-Loop Control [CLC] is superior to Standard Care [SC] using a statistical significance of $\alpha=0.05$ and the model specified below in Section 7 (i.e., $p < 0.05$).

The null/alternative hypotheses are:

a. *Null Hypothesis*: There is no difference in mean CGM-measured % in range 70-180 mg/dL over 13 weeks between SC and CLC

b. *Alternative Hypothesis*: The mean CGM-measured % in range 70-180 mg/dL over 13 weeks is different for SC and CLC.

4. Sample size

Based on data from the DCLP5 study (Breton et al.), we conservatively estimate the standard deviation (SD) for the primary outcome, time in range (TIR), to be 10%. A sample size of N=90 subjects (60 in the CLC arm and 30 in the SC arm) therefore will give 90% power to detect a 7.5% improvement in TIR with a two-sided test and type 1 error of 5%. Accounting for potential attrition rate of up to 10%, the total sample size for the PEDAP Study is estimated at N=102.

The following table shows the minimum detectable difference with a total sample size N=90 for key secondary outcomes. The observed standard deviations are taken from the DCLP5 study.

Outcome	Standard Deviation			Correlation with Baseline	Effective SD [a]	Detectable Difference with N=90 [b]
	CLC (n=78)	SC (N=22)	Pooled (N=100)			

% >250 mg/dL [c]	5.5%	9.9%	6.5%	0.69	4.7%	3.4%
Mean Glucose [d]	18	26	20	0.68	14	10
HbA1c	0.8%	0.9%	0.8%	0.70	0.6%	0.4%
% <70 mg/dL [c]	1.07%	1.13%	1.08%	0.61	0.86%	0.63%
% <54 mg/dL [c]	0.23%	0.23%	0.23%	0.54	0.19%	0.14%

a – After accounting for the baseline value as a covariate in the regression model.
b – With 90% power and two-sided type 1 error rate = 5%.
c – Outcomes with a skewed distribution winsorized at the 10th and 90th percentiles.
d – Units for mean glucose are mg/dL

5. Outcome Measures

5.1. Primary Efficacy Outcome

- The primary outcome is CGM-measured % in range 70-180 mg/dL

5.2. Secondary Efficacy Outcomes

5.2.1. Hierarchical Outcomes

The following secondary outcomes will be tested in a hierarchical fashion between the intervention arms as described in Section 7.2.1.

- CGM-measured % above 250 mg/dL
- CGM-measured mean glucose
- HbA1c at 13 weeks
- CGM-measured % below 70 mg/dL
- CGM-measured % below 54 mg/dL

5.2.2. Other Secondary Efficacy Outcomes

The following outcomes are considered exploratory.

- CGM-Measured:
  - % above 180 mg/dL
  - % in range 70-140 mg/dL
  - glucose variability measured with the coefficient of variation (CV)
  - glucose variability measured with the standard deviation (SD)
  - % <60 mg/dL
  - low blood glucose index (LBGI)*
  - hypoglycemic events (defined as at least 15 consecutive minutes <54 mg/dL)
  - hyperglycemic events (defined as at least 90 consecutive minutes >300 mg/dL)
  - % >300 mg/dL
  - high blood glucose index (HBGI)*
  - % in range 70-180 mg/dL improvement from baseline to 13 weeks ≥5%
  - % in range 70-180 mg/dL improvement from baseline to 13 weeks ≥10%

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

- HbA1c:
  - HbA1c <7.0% at 13 weeks
  - HbA1c <7.5% at 13 weeks
  - HbA1c improvement from baseline to 13 weeks >0.5%
  - HbA1c improvement from baseline to 13 weeks >1.0%
  - HbA1c relative improvement from baseline to 13 weeks >10%
  - HbA1c absolute improvement from baseline to 13 weeks >1.0% or HbA1c <7.0% at 13 weeks
- Questionnaires:
  - PedsQL Diabetes Module – total score and 5 subscales:
    - Diabetes
    - Treatment I
    - Treatment II
    - Worry
    - Communication
  - Pediatric Inventory for Parents (PIP) 2 domains each with a total score and 4 subscales for (5x2=10 difference scores)
    - Frequency
      - Total Score
      - Communication
      - Medical Care
      - Role Function
      - Emotional Functioning
    - Difficulty
      - Same total + 4 subscales as above for Frequency
  - INSPIRE (CLC arm only)
  - Pittsburgh Sleep Quality Index (PSQI) global score
  - Fear of Hypoglycemia Survey for Parents (HFS-P) – total score, 2 subscales and 4 factor scores:
    - Behavior
      - Avoidance
      - Maintain high BG
    - Worry
      - Helplessness
      - Social consequences
  - Hypoglycemia Confidence Scale (HCS)
  - System Usability Scale (SUS; CLC arm only)

*Note that LBGI and HBGI will be calculated using all available CGM readings as described below.

- Other:
  - Insulin
    - Total daily insulin (units/kg)
    - Percentage of total insulin delivered via basal

- Weight and Body Mass Index (BMI)

5.3. 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 a Dexcom 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 to 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 <168 hours 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 13-week period starting from the fourth day after the randomization visit up through the 13-week visit. If a participant drops out before completing the 13-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 a sensitivity analysis 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 from the sensor will be imputed as 401 and 39, respectively.

5.4. Calculation of Insulin Metrics

Insulin metrics describe the seven days prior to the visit.

- Baseline insulin metrics will use data reported from the Diabetes Screening CRF.
- Insulin metrics will be calculated at 13 weeks using Tandem pump data where available, otherwise using data reported on the CRF.
  - If pump data are available, insulin metrics will be calculated from the pump data using data in the 7 days prior to the visit. Insulin data on 5 of 7 days will be required to calculate insulin metrics.

5.5. Questionnaires

All questionnaires will be scored according to the instructions given in the reference manual/documentation for the questionnaire. 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 (where applicable) 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.

5.6. Analysis Windows

Analysis windows apply to HbA1c measured at randomization and the following outcomes measured at the 13-week visit:

- HbA1c
- Insulin metrics
- Weight/BMI
- Questionnaires

This does not apply to the CGM metrics using data from the Tandem pump which are calculated as described above.

Data from visits occurring in the following windows will be included in the analysis:

Visit (Target Date)	From Day[a]	Thru Day[a]
Randomization[b] (0 days)	−3	21
13 weeks (91 days)	76	112

a – Days from randomization, inclusive.
b – Window for randomization applies to HbA1c only.

6. Analysis Datasets and Sensitivity Analyses

6.1. 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 were randomized or the study was completed.

6.2. Per-Protocol 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

6.3. Other Sensitivity Analyses

The following sensitivity analyses will be conducted for the primary outcome and all outcomes listed in the hierarchy in Section 5.2.1.

Confounding:

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

Include CGM Data after the 13-week visit

It is anticipated that some subjects will spend up to two additional weeks under the RCT treatment while they await the equipment necessary to begin the extension phase. CGM data collected during this period will not be included in the calculations for the preceding analyses. To determine what impact these data may have had on the models, the data will be pooled together with the data from the RCT phase, and the models will be refit, with any major discrepancies noted.

Missing Data:

Missing data will be handled using direct likelihood method for the hierarchical analyses. 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)

7. Efficacy Analysis

7.1. Primary Analysis

This study's primary outcome is CGM measured % time in range 70-180 mg/dL over 13 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, 13 weeks intervention and change from baseline to follow-up.

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 age, prior CGM and pump use, and clinical center (random effect). Baseline CGM-measured % in range 70-180 mg/dL will be accounted for in the direct likelihood model as described below in the Missing Data paragraph. 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 robust regression using M-estimation 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 6.3).

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.

7.2. Analysis of Secondary Endpoints

Point estimates and confidence intervals for the treatment arm differences will be presented for all secondary metrics. Analyses will parallel those described above for the primary outcome.

7.2.1. 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 250 mg/dL
- CGM-measured mean glucose
- HbA1c at 13 weeks
- CGM-measured % below 70 mg/dL
- CGM-measured % below 54 mg/dL

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 six 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.

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 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 7.1.

Analyses for HbA1c at 13 weeks will parallel those of the primary outcome. Missing data will be handled using direct likelihood in a regression model including all available central laboratory HbA1c measurements at baseline and 13-week visits.

For all above analyses, regression diagnostics will be employed analogous to as described in Section 7.1 for the primary outcome.

7.3. 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, 13 weeks and for the changes from baseline to 13 weeks. For continuous outcomes noted in Section 5.2., analyses will parallel those described for the primary outcome. Comparisons of body weight and BMI will also be adjusted for gender.

For CGM-measured hypoglycemia/hyperglycemia events, event rates per week will be compared using similar linear mixed effects regression models as described above for the primary outcome.

Binary CGM and HbA1c Outcomes

For the binary HbA1c outcomes, risk-adjusted percentages by treatment group will be computed at 13 weeks from a logistic regression model (Kleinman and Norton, 2009). 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.

Questionnaires

For each questionnaire, mean ± SD or percentiles appropriate to the distribution will be given by treatment group at baseline and 13 weeks. Group comparisons will be conducted for the total score (mean score) and subscales from the parent version using similar linear models as described above. The INSPIRE and SUS post-treatment surveys will only be administered to the CLC group at the 13-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 7.2.1 (i.e., hierarchical CGM metrics) in 4 week and 1-

week periods over the 13 week course of follow-up. Additionally, boxplots will be given for the primary outcome and each of the key secondary CGM-measured endpoints in 1-day periods over the first 2 weeks of follow-up.

As noted above, the following periods will be given for hierarchical CGM metrics:

- CGM metrics by 4-week periods: Starting from the fourth day after the randomization visit, CGM metrics will be calculated within each 4-week period (Week 1-4, Week 5-8, and Week 9-13). The 13 week visit date will be used as end date of the Week 9-13 period. A minimum of 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 fourth day after the randomization visit, CGM metrics will be calculated within each 1-week period. The 13 week visit date will be used as end date of the Week 12-13 period. Minimum 72 hours of CGM data in each 1-week period will be required for the calculation.
- CGM metrics by 1-day periods: Starting from the fourth day after the randomization visit, CGM metrics will be calculated within each day. Day 18 of follow-up will be used as end date of the analysis period. Minimum 17 hours of CGM data in each 1-day period will be required for the calculation.

7.4. Formal Statistical Comparisons of Treatment Groups

With the exceptions of the INSPIRE and System Usability questionnaires, all secondary outcomes noted in Section 5.2 will have formal statistical comparisons of treatment groups.

8. Safety Analysis

All randomized participants will be included in these analyses and all their post-randomization safety events will be tabulated. Any pre-randomization adverse events will be tabulated separately and will include any participants who were never randomized. Safety analyses of the main study (randomized trial phase) will include events occurring on or after randomization until and including the 13-week visit date or randomization date + 105 days whichever is earlier.

For the following outcomes, the number of events will be tabulated by treatment group. Formal statistical comparisons (main study phase only) will be performed if there are enough events (at least 5 events combined between the two 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
- Other serious adverse events
- Any adverse event rate
- Number of calendar days with any ketone level ≥1.0 mmol/L (if ≥5 total calendar days combined)
- Worsening of HbA1c from baseline to 13 weeks by >0.5%
- Investigational device related (intervention group only):

- Adverse device effects (ADE)
- Serious adverse device events (SADE)
- Unanticipated adverse device effects (UADE)

For DKA and SH events, if enough events (i.e., at least 5 events combined over both treatment groups) for each outcome, the rates will be compared between the two treatment arms during the main study phase 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. If one of the treatment groups has zero events then a Poisson model will not converge and Barnard's exact test will be used instead. The amount of follow-up will be included as an offset covariate to compare the rates. A similar analysis will be done for DKA, if at least 5 total DKA events among both treatment groups.

9. 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 fourth day after randomization until the day before the 13-week visit. Any dropouts will be included prior to the time of dropout and their percentage will be prorated accordingly.

- Time in closed loop overall.
- Time in closed loop by 4-week periods: Starting from the fourth day after the randomization visit, percent time in closed loop will be calculated within each 4-week period (Week 1-4, Week 5-8, and Week 9-13). The 13-week visit date will be used as end date of the Week 9-13 period.
- Percent time in different operational modes overall and by 4-week periods (the last interval will contain Weeks 9-13) will be tabulated and boxplots will be created by these 4-week periods.

Sensor 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 will be tabulated by treatment group and boxplots will be created.

10. 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 contacts
- Number of participants who stopped treatment and reasons

11. 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
- Parent's income, education, and/or insurance status
- Diabetes duration
- Insulin method before enrollment (pump vs. MDI)
- CGM use before enrollment
- HbA1c
- BMI %
- 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 >250 mg/dL
    - Mean glucose
    - % time <70 mg/dL
    - % time <54 mg/dL

12. Device Issues

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
- Rate of different failure events and alarms per 24 hours recorded by the Control-IQ system – overall and by month

13. 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.

14. Subgroup Analyses

In exploratory analyses, the primary outcome (% time 70-180 mg/dL), % time <70 mg/dL and HbA1c at 13 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.

- Age (<4 vs. ≥4)
- Gender
- Race/Ethnicity
- Parent's income, education, and/or insurance status
- Diabetes duration
- Insulin method before enrollment (pump vs MDI)
- CGM use before enrollment
- Baseline HbA1c
- Body mass index (%)
- Participant-reported number of SH and DKA 12 months prior to the start of the study
- Clinical center (test for random effects interaction with treatment group)
- Baseline CGM metrics including:
    - % time in range 70-180 mg/dL
    - % time >250 mg/dL
    - Mean glucose
    - % time <70 mg/dL
    - % time <54 mg/dL
- The p-value for any categorical variable will only be calculated if there are a minimum of 10 participants per treatment group.

15. Multiple Comparisons

Hierarchical Analyses

The hierarchical testing procedure described above will be used to control the overall type 1 error for the primary outcome plus five 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 method adapted using the two-stage test. 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 and HbA1c at 13 weeks
- Subgroup analyses for % time < 70 mg/dL

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

16. Exploratory Analyses

CGM Metrics

In addition to the analysis for the hierarchical CGM-measured endpoints described earlier, separate analyses will be conducted with the appropriate statistics for daytime (6:00am – 9:59pm) and nighttime (10:00 PM – 5:59 AM) separately. No p-values will be calculated for following analyses.

- % time in range 70-180 mg/dL
- Mean glucose
- % above 250 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 criterion. No p-values will be calculated for following analyses.

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

Hierarchical CGM metrics noted above will also be calculated for daytime period (06:00AM to 9:59PM) and overnight period (10:00PM 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.

Additional Insulin Metrics

The following insulin metrics will be tabulated by treatment group at baseline, 13 weeks and for the changes from baseline to 13 weeks. No p-values will be calculated for these metrics.

- Total Daily Insulin for Pump Users (units/kg)
  - Total daily basal insulin (units/kg)
  - Total daily bolus insulin (units/kg)
- Total Daily Insulin for Injection Users (units/kg)
  - Total daily basal insulin (units/kg)
  - Total daily bolus insulin (units/kg)
  - Total daily short-acting injections

17. Additional Analyses after SAP Version 1.0

17.1. FDA Review

Following FDA review of this SAP, the following analyses were added.

- The plausibility of the missing at random assumption will be evaluated by comparing the baseline characteristics listed in Section 11 for completers versus dropouts.

- A two-way tipping point sensitivity analysis will be done to assess how strong a deviation from the missing at random assumption would need to be to affect study conclusions:
  - Start with Rubin's imputation as listed above initially assuming missing at random.
  - Shift imputed values by a fixed amount in each treatment arm to represent a bias from selective dropout.
  - Calculate estimated treatment delta and p-value on the resulting dataset using the model described in Section 7.
  - Repeat this process using various pairs of shift values for the treatment and control arms.
  - Characterize which pairs of shift values result in a statistically significant difference in the treatment arm comparison.

17.2. Extension Phase, Extended Use Period, Ancillary Study

Details for the analyses of the Extension Phase, the Extended Use Period, and the Ancillary Study are specified in a separate SAP.

18. References

Breton MD, Kanapka LG, Beck RW, Ekhlaspour L, Forlenza GP, Cengiz E, Schoelwer M, Ruedy KJ, Jost E, Carria L, Emory E, Hsu LJ, Oliveri M, Kollman CK, Dokken BB, Weinzimer SA, DeBoer MD, Buckingham BA, Cherñavvsky D, and Wadwa RP, for the iDCL Trial Research Group: A Randomized Trial of Closed-Loop Control in Children with Type 1 Diabetes. N Engl J Med. 2020 Aug 27;383(9):836-845. doi: 10.1056/NEJMoa2004736.

Kleinman LC, Norton EC. What's the risk? A simple approach for estimating adjusted risk measures from nonlinear models with logistic regression. *Health Services Research* 2009; 44(1): 2009 Feb.

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.