Automated Insulin Delivery in Elderly (AIDE):

A Randomized Cross-over Trial Evaluating Automated Insulin Delivery Technologies on Hypoglycemia and Quality of Life in Elderly Adults with Type 1 Diabetes

Primary Study Phase

Statistical Analysis Plan

Version 1.3

September 6, 2023

Version History

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

Version Number	Protocol Version	Author	Approver	Effective Date	Study Stage
1.0	2.0	Lauren Kanapka	Craig Kollman	4/20/20	Enrollment has not started.
1.1	4.1	Lauren Kanapka	Craig Kollman	5/14/21	Enrolling
1.2	5.0	Lauren Kanapka	Craig Kollman	8/19/22	Enrolling
1.3	5.0	Lauren Kanapka	Craig Kollman	9/6/23	Follow-up. No data have been shared with the study group.

Version Number	Revision Description
1.0	Original Version
1.1	Safety analysis section was modified to specify how adverse events that occur between treatment periods will be handled.
1.2	The section describing the analysis windows for HbA1c was reformatted to be clearer and the analysis window for HbA1c at baseline was updated to be wider.
1.3	• The section describing the analysis of the questionnaires was modified to say that the technology acceptance questionnaire will not be formally compared between treatment groups. • The tabulations that will be performed for sleep and exercise mode use have been edited.

Approvals

Role	Digital Signature or Handwritten Signature/Date
Author: Lauren Kanapka	Lauren Kanapka I agree to the terms defined by the placement of my signature in this document 2023-09-06 12:08-04:00
Senior Statistician: Craig Kollman	Craig Kollman I am electronically signing this document 2023-09-07 10:54-04:00
JCHR Coordinating Center Director: Robert Henderson	Robert Henderson Digitally signed by Robert Henderson DN: cn=Robert Henderson ou=South Wing Reason: I have reviewed this document Location: Date: 2023-09-06 15:39-04:00
Principal Investigator: Yogish Kudva	Yogish C Kudva 06-SEP-2023

1. Study Overview

This document outlines the statistical analysis to be performed for the randomized trial phase of the AIDE T1D study. The approach to sample size and statistical analyses for this study are summarized below.

This is a multicenter, randomized, three period crossover study to assess the effectiveness and safety of both hybrid closed loop control (HCL) technology and predictive low-glucose insulin suspension (PLGS) compared to sensor-augmented pump (SAP) therapy in older adults >65 years with type 1 diabetes. Eligible participants will receive all three interventions and the order of receiving them will be randomized on a 1:1:1:1:1:1 ratio.

Randomization will be preceded by a run-in period where participants must demonstrate competency and compliance in using the study insulin pump and CGM device. After randomization, the subjects will enter the three 12 week study periods and will test one intervention per study period.

2. Comparison to Protocol

The author of this SAP has verified that the analyses described in this document are consistent with the version of the protocol listed in the version history table above except for the following:

- Removed tabulation of percent of time spent in sleep mode outside of nighttime hours
- Added tabulation of days per week of sleep mode use and exercise mode use

3. Statistical Hypotheses

A. *Null hypothesis*: There will be no difference in percentage of CGM-measured glucose values <70 mg/dL between PLGS vs SAP.

   Alternate hypothesis: There is a non-zero difference in the percentage of CGM-measured glucose values <70 mg/dL between PLGS vs SAP.

B. *Null hypothesis*: There will be no difference in percentage of CGM-measured glucose values <70 mg/dL between HCL vs. SAP.

   Alternate hypothesis: There is a non-zero difference in the percentage of CGM-measured glucose values <70 mg/dL between HCL vs. SAP.

4. Sample Size

Data from the CGM group in the WISDM randomized clinical trial of older adults ≥ 60 years of age were used to estimate the standard deviation and frequency of time <70 mg/dL in the SAP control period. Data from the two weeks prior to the 4 week visit were used to mimic the run-in period in this study where the minimum hypoglycemia eligibility criteria will be assessed. Only WISDM CGM group participants who had ≥2% of time <70 mg/dL at the 4 week time point were included. Data from the two weeks prior to the 26-week visit for these participants were then used to estimate standard deviation and frequency of time <70 mg/dL. N=50 WISDM participants were included of which 28 were injection users and 22 were pump users.

The point estimate for the simple standard deviation was 2.24%. Percent time <70 mg/dL was skewed, so a robust estimate of the mean, 3.14%, was used to calculate the size of a 33% and 50% relative reduction in hypoglycemia.

Since the primary outcome of the study involves two comparisons, the sample size calculations assume an alpha of 0.025 in order to control the overall type 1 error rate at 0.05. In addition, a two-tailed test and 90% power are assumed. We assume the standard deviation is 2.24% in all three periods.

Table 1: Sample Size (Alpha=0.025, Power=90%, SD=2.24%)

Correlation between Periods	Relative Reduction in Hypoglycemia	
	33%	50%
0	119	54
0.3	84	38
0.5	61	28

The Tandem pivotal PLGS study, which was a 3 week period crossover of PLGS and SAP, observed a correlation of about 0.8 between the treatment arms (unpublished data). However, since the periods in this study are longer, we assume a lower correlation of 0.3 to estimate sample size. A sample size of 84 will give us 90% power to detect a 33% reduction in % time <70 mg/dL. This is increased to 90 to account for 5% loss to follow-up. Loss to follow-up and missing data is expected to be minimal in this population as the retention rate in the WISDM study was 99% at 6 months.

5. Outcome Measures

Primary Efficacy Endpoint:

- CGM % time <70 mg/dL

Secondary Efficacy Endpoints:

Hypoglycemia

- CGM % time <54 mg/dL
- Frequency of CGM-measured hypoglycemic events (see definition below)

Glucose Control

- CGM mean glucose
- CGM % time in range 70 to 180 mg/dL
- Coefficient of variation (CV)

Hyperglycemia

- CGM % time >180 mg/dL
- CGM % time >250 mg/dL

HbA1c

- HbA1c

Hypoglycemia Unawareness

- Gold survey

Patient-reported Outcomes

- Hypoglycemia Fear Survey (HFS-II)
    - Total score
    - Worry subscale

- Hypoglycemia confidence
- Diabetes Distress Scale (DDS)
- Technology acceptance
- System usability

5.1. CGM Outcomes

Baseline

Participants will complete a CGM training period and/or SAP training period as necessary prior to randomization. All participants will wear the study pump and CGM for at least 14 days prior to randomization even if they do not require any device training. Therefore, the last 14 days of CGM data prior to randomization will be used in the calculation of baseline CGM metrics. If <24hr 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

The first 4 weeks of CGM data in each period will be excluded to reduce the chance of a carryover effect since there is no washout period. Therefore, CGM metrics will be calculated for each period by pooling all sensor readings between midnight on the date of the treatment initiation visit + 29 days and midnight on the date of the end of treatment visit. If a participant drops out before completing the period, all data between midnight on the date of the treatment initiation visit + 29 days and midnight on the dropout date will be included. A minimum of 168 hours of CGM data will be required to calculate CGM metrics.

Hypoglycemic Events

A CGM-measured hypoglycemic event will be defined as at least 2 sensor values <54 mg/dL that are 15 or more minutes apart plus no intervening values >54 mg/dL; at least 2 sensor values >70 mg/dL that are 15 or more minutes apart with no intervening values <70 mg/dL are required to define the end of an event, at which point the study participant becomes eligible for a new event.

Daytime vs. Nighttime

Each of the CGM metrics listed will be calculated over 24 hours and separately for daytime (6am-<12am) and nighttime (12am-<6am). Daytime and nighttime versions of CGM % time <70 mg/dL will be considered secondary. 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 nighttime metrics.

5.2. HbA1c

The analysis windows for HbA1c at each time point are given in the table below:

Visit	Target Date	Window
Baseline	Randomization date	Screening visit date to randomization date +7 days
End of period 1	Period start date + 84 days	±21 days from target date. Values collected >7 days after the start of period 2 will be excluded.

End of period 2	Period start date + 84 days	±21 days from target date. Values collected >7 days after the start of period 3 will be excluded.
End of period 3	Period start date + 84 days	±21 days from target date. Values collected >7 days after the start of the extension period will be excluded.

The measurement within the analysis window that is closest to the indicated target date will be used for analysis. If no measurement is available within the analysis window, the endpoint will be treated as missing.

5.3. Questionnaires

The hypoglycemia fear, hypoglycemia confidence, and diabetes distress surveys will be administered online or on paper at the screening visit and at the end of each treatment period. The technology acceptance survey and system usability scale will be administered at the end of each treatment period. Only responses obtained within ±21 days of the end of period target visit dates will be included in the analyses. Questionnaires collected >7 days after start of next period will not be included for the previous period. The baseline questionnaires must be completed prior to the initiation of period 1 visit. Participants 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 75% rule will not include any questions marked as "N/A"
- The score used for analysis will be based on the average among the questions that were answered.

6. Analysis Datasets and Sensitivity Analysis

All analysis will follow the intention-to-treat principle with each period analyzed according to the treatment assigned by randomization regardless of actual system utilization. The Intention-to-Treat (ITT) Analysis Dataset will include all randomized participants for any period in which they meet the minimum data requirement (≥168 hours of CGM data after excluding the first 4 weeks for CGM metrics, non-missing for all other outcomes).

The Safety Analysis Dataset will include all enrolled participants, irrespective of whether the participant was randomized or the study was completed.

6.1. Per-protocol Analysis

The Per-Protocol Analysis Dataset will include participants for any period in which they meet the following criteria after excluding the first 4 weeks:

- ≥168 hours of CGM data

- CGM use ≥80%
- Control-IQ active ≥80% for the HCL period
- Basal-IQ active ≥80% for the PLGS period
- SAP active (Control-IQ and Basal-IQ not active) ≥80% for the SAP period

A per-protocol analysis will be performed for the primary outcome to provide additional information regarding the magnitude of treatment effect. The per-protocol analysis will only be performed if at least 10% of participants in any period would be excluded by these criteria.

The intent-to-treat analysis is considered primary and if the results of the per-protocol analysis and intent-to-treat analysis differ, the per-protocol analysis will be interpreted with caution.

6.2. Other Sensitivity Analysis

Missing Data

It is worth emphasizing that any statistical method for handling missing data makes a number of untestable assumptions. The goal will be to minimize the amount of missing data in this study so that results and conclusions 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 the primary analysis when using different methods. The following methods will be applied:

- Direct likelihood (primary analysis described below)
- Rubin's multiple imputation
- Available cases only

Carryover

A period by treatment interaction will be added to the primary analysis model to assess for the presence of a carryover effect. We do not expect a carryover effect to be present because we expect the effect of the treatment administered in the prior period to wear off during the first 4 weeks, which are not included in the calculation of CGM metrics.

7. Analysis of the Primary Efficacy Endpoint

The primary analysis for CGM % time <70 mg/dL will involve two comparisons: PLGS vs. SAP and HCL vs. SAP. Control of the type 1 error for multiple treatment group comparisons will be handled as described in section 16.

Participants will be included for any period in which they have at least 168 hours of CGM data in the last 8 weeks.

Summary statistics appropriate to the distribution will be calculated for % time <70 mg/dL separately by treatment arm. A repeated measures regression model with an unstructured covariance will be fit including data from baseline and all three treatment periods. The model will adjust for period as a covariate. It is expected that % time <70 mg/dL will have a skewed distribution. Therefore, values will be winsorized at the 10^{th} and 90^{th} percentiles. If winsorization is not sufficient, other transformations will be explored.

There will be no imputation of missing data in the primary analysis. Missing data will be handled by using a direct likelihood approach which will allow participants to be included even if they only have baseline and no follow-up data.

8. Analysis of the Secondary Endpoints

8.1. Secondary Hypoglycemia CGM Endpoints

Summary statistics appropriate to the distribution will be given by treatment group for the secondary hypoglycemia CGM endpoints (time <54 mg/dL and rate of hypoglycemic events). The primary hypoglycemia endpoint and the secondary hypoglycemia outcomes will be evaluated in a hierarchical approach to control the type 1 error (see section 16 below for details). If the endpoint is to be formally compared between treatment groups, analysis will parallel that of the primary outcome. It is also expected that time <54 mg/dL and rate of hypoglycemic events will have a skewed distribution. Therefore, values will be winsorized at the 10^{th} and 90^{th} percentile. If winsorization is not sufficient, other transformations will be explored.

Separate day and night versions of all three hypoglycemia CGM endpoints will be summarized by treatment group but will only be formally compared between groups in a secondary analysis if the overall version of the endpoint was formally compared and was statistically significant.

8.2. Additional CGM Endpoints

Summary statistics appropriate to the distribution will be given by treatment group for time in range 70-180 mg/dL, mean glucose, time >180mg/dL, and time >250 mg/dL. These metrics will be evaluated overall and separately for daytime and nighttime. Analysis will parallel the analysis of the primary endpoint above. The outcomes will be winsorized at the 10^{th} and 90^{th} percentile if they appear to be skewed. If winsorization is not sufficient, other transformations will be explored.

8.3. HbA1c

HbA1c will be measured by central lab at baseline and following each 12 week period. Summary statistics appropriate to the distribution for HbA1c will be reported by treatment group. A repeated measures regression model with an unstructured covariance will be fit including the data from baseline and all three treatment periods. The model will adjust for period as a covariate. HbA1c will be winsorized at the 10^{th} and 90^{th} percentile if the residuals from the model fit on the un-transformed outcome appear to be skewed. If winsorization is not sufficient, other transformations will be explored.

8.4. Questionnaires

For each patient-reported outcome, summary statistics appropriate to the distribution will be given by treatment group and at baseline (if applicable). The technology acceptance questionnaire will not be formally compared between treatment groups. All other patient-reported outcomes will be compared between treatment groups using a model similar to the primary outcome model described above. The outcomes will be winsorized at the 10^{th} and 90^{th} percentile if the residuals from the model fit on the un-transformed outcome appear to be skewed. If winsorization is not sufficient, other transformations will be explored.

9. Safety Analyses

Details of all reportable adverse events will be provided in a listing by treatment group. Adverse events that occur pre-randomization or between treatment periods will be listed separately and will not be included in any treatment group comparisons. It is intended that a period will usually end and the next one will start on the same day. The situation where the start of the next period is delayed and an AE occurs between periods is expected to be rare. Each treatment period will inclusively consist of all days in between the treatment initiation visit and the end of treatment visit. If an adverse event occurs on a border day (including the randomization date, treatment initiation visit dates, and end of treatment visit dates), the site staff or medical monitor will review the event details and determine which treatment period the event occurred in. This determination will be documented in the study database. In the unlikely situation where it is not possible to tell which period the event occurred in, the event will be excluded from treatment arm comparisons but will be included in the listing with treatment arm specified as "undetermined". If the subject withdraws from the study in the middle of a period and the end of treatment visit for that particular period does not occur, then the later of the last visit date, last AE date or final status date (if known) will be used as the last day of the period for the purpose of calculating event rates.

For the following outcomes, summary statistics appropriate to the distribution will be tabulated by treatment arm.

- Number of adverse events
- Number of serious adverse events
- Number of unexpected device events
- Number of SH events and SH incidence rate per 100 person-years
- Number of hospitalizations related to a SH event
- Number of ER visits related to a SH event
- Number of fractures related to a SH event
- Number of falls related to a SH event
- Number of DKA events and DKA incidence rate per 100 person-years
- Number of hospitalizations related to a DKA event or severe hyperglycemia
- Number of ER visits related to a DKA event or severe hyperglycemia

If there are at least 5 SH events across treatment arms, the SH incidence rate will be compared pairwise between all treatment arms using a repeated measures Poisson regression model adjusting for period and whether the subject had an event in the 12 months prior to the study as a covariate. If there are zero events in one treatment arm, Poisson regression will not converge and so the number of events will be compared pairwise using Barnard's test instead. A similar analysis will be done for DKA events if there are at least 5 events across the treatment arms.

10. Protocol Adherence and Retention

The following will be performed according to treatment arm:

- A flow chart accounting for all participants for all visits
- Tabulation of visit completion rates for each follow-up visit

- Tabulation of protocol deviations
- Tabulation of number and reasons for unscheduled visits and phone calls
- Tabulation of device issues

11. Baseline Descriptive Statistics

The following baseline demographic and clinical characteristics will be summarized in a table:

- Age
- Gender
- Race/ethnicity
- Income, education, employment, and/or insurance status
- Diabetes duration
- Age at diagnosis
- Insulin method before enrollment (pump vs. MDI)
- CGM use before enrollment
- AID use before enrollment
- HbA1c
- BMI
- C-peptide
- Participant reported number of SH and DKA 12 months prior to the start of the study
- MoCA total score
- WAIS-IV Processing Speed Index
- FAQ score
- Frailty walk time
- Baseline CGM metrics including:
    - % time <70 mg/dL
    - % time <54 mg/dL
    - % in range 70-180 mg/dL
    - % time >180 mg/dL
    - Coefficient of variation

For continuous variables, summary statistics appropriate to the distribution will be given. For discrete variables, number and percentage will be reported for each category.

12. Planned Interim Analyses

No formal interim efficacy analysis is planned for this study. Safety data tabulations will be performed at least every 6 months for review by the Data and Safety Monitoring Board (DSMB).

13. Subgroup Analyses

Subgroup analyses/assessments of effect modification (interaction) will be conducted for the primary outcome. These analyses will be considered exploratory. Additionally, interpretation of the analyses will depend on whether the overall analysis demonstrates a significant treatment group difference; in the absence of such an overall difference, subgroup analyses will be interpreted with caution. The general approach for these exploratory analyses will be to add an

interaction term for the subgroup factor by treatment into the models used for the primary analyses.

The baseline factors listed below will be assessed:

- Race/Ethnicity
- Gender
- Baseline % time <70 mg/dL
- Age
- Education (≤ Bachelor's vs. >Bachelor's)
- Employment (Retired vs. not retired)
- Duration of Diabetes
- Hypoglycemia Unawareness
- MoCA total score
- WAIS-IV Processing Speed Index
- Functional activities questionnaire
- Frailty
- Prior CGM experience
- Prior Insulin Delivery Method
- C-peptide

14. Exploratory Analyses

All of the primary and secondary outcomes will be compared between PLGS and HCL in an exploratory analysis that parallels the analysis described above. A difference, if any, in the hypoglycemia outcomes between PLGS and HCL is expected to be small, so the power will be low.

15. Additional Tabulations and Analysis

15.1. Device use

The percent of time in each system control mode and the percent of time using CGM will be tabulated by treatment arm overall between treatment initiation and the end of treatment visit and by 4-week intervals. These tabulations will be repeated separately over 24-hours, daytime, and nighttime. Dropouts will be counted as zero use for the remainder of any period that was initiated but not completed and will be counted as missing use for any periods that were never started.

Overall 24-hour CGM use will be compared pairwise between all treatment arms using the same model described for the primary outcome.

In addition, we will tabulate the following for the HCL period overall between treatment initiation and the end of treatment visit and by 4-week intervals:

- Percent of time spent in sleep mode
- Average days per week with any sleep mode use
- Average days per week with any exercise mode use

15.2. Insulin

Average total daily insulin per kg, basal insulin per kg, and bolus insulin per kg will be tabulated at baseline and by treatment arm. Pump download data in the 2 weeks prior to the baseline/end of treatment visit will be used where available. If pump download data is not available for at least 7 out of 14 days, data reported on the CRF will be used instead. Insulin metrics will be compared pairwise between all treatment arms using the same model described above for the primary outcome.

15.3. Pump Alert and Alarms

The number and rate per 24 hours of different alerts and alarms from the Tandem pump will be tabulated by treatment arm between treatment initiation and the end of treatment visit.

15.4. BMI

Height and weight will be measured at baseline and the end of each period. BMI will be tabulated at baseline and by treatment arm.

15.5. BG Checks

Average blood glucose checks per day will be reported on the CRF at baseline and the end of each period. BG checks per day will be tabulated at baseline and by treatment arm. Download data will be used where available, otherwise we will use self-report.

16. Multiple Comparison/Multiplicity

16.1. Primary Analysis and Other Key Hypoglycemia Outcomes

For the primary comparisons of interest, PLGS vs. SAP and HCL vs. SAP, the three CGM-measured hypoglycemia metrics will be evaluated in a hierarchical approach to control the type 1 error at 0.05. Each comparison will be allocated alpha 0.025. The outcomes will be evaluated in the following order:

- % time <70 mg/dL
- % time <54 mg/dL
- Rate of hypoglycemic events

The process moves to the next variable down on the list until a non-significant result ($p \geq 0.025$) is observed, or all outcomes have been tested. If a non-significant result is encountered, then formal statistical hypothesis testing is terminated and any comparisons below on the list are not formally tested. If for either comparison of interest, PLGS vs. SAP or HCL vs. SAP, all three outcomes are rejected at the level of 0.025, the alpha can be recycled and the other comparison can be tested down the hierarchy at an alpha level of 0.05 rather than 0.025.

See the next page for examples of possible scenarios.

16.2. Other Secondary and Exploratory Analysis

For all other secondary and exploratory outcomes, the false discovery rate will be controlled using the Benjamini-Hochberg procedure method adapted using the two-stage test. For these analyses,

the adjusted p-value and 95% confidence interval will be reported. The categories for FDR correction will be:

- HCL vs. SAP and PLGS vs. SAP
    1. Glycemic control outcomes (overall, daytime, and nighttime CGM metrics and HbA1c), CGM use, and insulin metrics
    2. Patient-reported outcomes
    3. Subgroups
- HCL vs. PLGS
    4. Glycemic control outcomes (overall, daytime, and nighttime CGM metrics and HbA1c), CGM use, and insulin metrics
    5. Patient-reported outcomes

There will be no adjustment for safety outcomes.

Example 1

All tests for the PLGS vs. SAP comparison are significant, so we can recycle the alpha 0.025 and use it for the HCL vs. SAP comparisons.

HIERARCHICAL ORDER	OUTCOME VARIABLE	HCL VS. SAP				PLGS VS. SAP			
		ALPHA	P-VALUE	SIGNIFICANT?	ACTION	ALPHA	P-VALUE	SIGNIFICANT?	ACTION
1st	% time <70 mg/dL	0.05	0.04	Yes	Test next variable	0.025	0.001	Yes	Test next variable
2nd	% time <54 mg/dL	0.05	0.06	No	Stop formal testing	0.025	0.02	Yes	Test next variable
3rd	Rate of hypoglycemic events	0.05	Not tested	Unknown	N/A	0.025	0.007	Yes	Recycle alpha to HCL vs. SAP

Example 2

HIERARCHICAL ORDER	OUTCOME VARIABLE	HCL VS. SAP				PLGS VS. SAP			
		ALPHA	P-VALUE	SIGNIFICANT?	ACTION	ALPHA	P-VALUE	SIGNIFICANT?	ACTION
1st	% time <70 mg/dL	0.025	0.04	No	Stop formal testing	0.025	0.001	Yes	Test next variable
2nd	% time <54 mg/dL	0.025	Not tested	Unknown	N/A	0.025	0.02	Yes	Test next variable
3rd	Rate of hypoglycemic events	0.025	Not tested	Unknown	N/A	0.025	0.07	No	Stop formal testing