IOBP Pivotal Trial
Statistical Analysis Plan
BP aspart/lispro vs. Standard Care

Version 4.0
Protocol Version 10.0
Author: Zoey Li
Date: December 9, 2021

Revision History

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

Version Number	Author	Senior Statistician	Effective Date	Study Stage
1.0	Zoey Li	Peter Calhoun	December 18, 2020	Pre-randomization.
2.0	Zoey Li	Peter Calhoun	March 19, 2021	Enrollment.
3.0	Zoey Li	Peter Calhoun	May 25, 2021	Enrollment.
4.0	Zoey Li	Peter Calhoun	December 9, 2021	Enrollment completed. Changes made before unblinding of results to study investigators

Version Number	Revision Description
1.0	Original version
2.0	- Removed requirement of baseline data to be 60 days prior to randomization. - Added sub-analyses excluding participants using closed loop systems at baseline. - Updated per-protocol CGM wear requirements for Control group
3.0	- Added sensitivity analyses for participants not randomized due to protocol deviation - Added sub-analyses for participants with HbA1c >7.0% - Account for lab HbA1c measurements measured at alternate labs - Added analysis restricting to periods of iLet use only for participants who discontinued iLet use but continued in the study.
4.0	- Added statistical tests for site effects. - Clarified subgroup analyses by different age cohorts - Added additional clarification on which outcomes will have confidence intervals. - Revised non-CL cohort at baseline sub-analyses. - Moved transition phase analyses, BP aspart/lispro vs. SC comparison, and BP Fiasp vs. BP aspart/lispro comparisons to separate SAPs. - Removed condition to combine the BP aspart/lispro and BP Fiasp groups; the BP groups will be analyzed separately. - Added additional CGM-measured hyperglycemic event definition.

	- Added safety analyses for pediatric cohort - Added baseline SH/DKA covariate to SH/DKA safety analyses.

Approvals

Role	Digital Signature or Handwritten Signature/Date
Author: Zoey Li	
Senior Statistician: Peter Calhoun	
JCHR Sponsor Representative: Roy Beck	

1. Contents

2. Consistency of Statistical Analysis Plan with Protocol/Analyses Planned for Primary Manuscript

This SAP is consistent with the study protocol stats chapter (version indicated on the title page) with the exception of the following change:

- The BP Fiasp Group and the BP aspart/lispro Group will not be pooled for secondary analyses

Separate analysis plans will be written for the analyses involving the BP Fiasp Group, the Transition Phase, and the BG Run Test Ancillary Study.

All outcomes are intended to be included in the primary manuscript unless flagged with '*'.**

3. Study Overview

The IOBP Pivotal trial is a multi-center randomized control trial that tests the efficacy and safety of an insulin-only configuration of the iLet Bionic Pancreas (BP) system versus a Control group using CGM during a 13-week study period.

The sample size is expected to include ~220 in the BP Group (using aspart/lispro) and ~110 in the Standard Care Control Group with 50% adults and 50% pediatric participants.

Following the initial screening visit, there will be the following visits:

a) Randomization Visit
b) Day 1-2 Phone Call
c) Week 1 Phone Call
d) Week 2 Visit
e) Week 6 Visit
f) Week 10 Visit
g) Week 13 Visit

4. Statistical Hypotheses

The primary outcome is:

- Superiority in HbA1c at 13 weeks

Only the primary outcome needs to be met to declare device efficacy. The key secondary outcome is:

- Non-inferiority in CGM-measured time <54 mg/dL calculated over 13 weeks

Primary and key secondary outcome analyses will combine pediatric and adult participants into a single analysis. The study hypotheses can be stated as follows:

HbA1c Outcome:

- *Null Hypothesis*: There is no difference in the mean HbA1c at 13 weeks between BP aspart/lispro and Control Group
- *Alternative Hypothesis*: There is a nonzero difference in the mean HbA1c at 13 weeks between BP aspart/lispro and Control Group

Time <54 mg/dL Outcome:

- *Null Hypothesis*: There is a mean difference of at least 1% in the percentage of time spent with a sensor glucose level below 54 mg/dL between the BP aspart/lispro and Control Group over the 13 weeks
- *Alternative Hypothesis*: There is a mean difference of less than 1% in the percentage of time spent with a sensor glucose level below 54 mg/dL between the BP aspart/lispro and Control Group over the 13 weeks

5. Sample Size

The sample size of 440 for the RCT was selected to provide sufficient exposure to the iLet BP system for regulatory purposes, with respect to different age groups and for use of both insulin aspart/lispro and Fiasp. RCT completion is expected for 200 participants randomized to BP aspart/lispro Group and 100 to the Control Group.

The primary and key secondary analyses for BP aspart/lispro vs. Control will include both the pediatric and adult participants in a single analysis. Superiority for HbA1c at 13 weeks and non-inferiority for time <54 mg/dL measured with CGM at intervals over the 13 weeks are considered the primary and key secondary endpoints, respectively. Statistical power for each endpoint was computed assuming the following:

- HbA1c: Statistical power is >99%, assuming true mean HbA1c difference of 0.4% between BP aspart/lispro and Control Group, standard deviation of 13-week HbA1c of 0.8%, correlation between baseline and 13-week HbA1c of 0.40, two-sided type 1 error of 5%

- Time <54 mg/dL: Statistical power is 99%, assuming no true difference in mean time <54 mg/dL between BP aspart/lispro and Control Group, a non-inferiority margin of 1%, standard deviation of percent time <54 mg/dL of 2.0%, correlation between baseline and follow-up of 0.40, and one-sided type 1 error of 0.025%

A separate document details the justification of these power calculations.

6. Efficacy Outcome Measures

6.1 Primary and Key Secondary Efficacy Endpoints

1. HbA1c at 13 weeks (superiority, primary endpoint)
2. CGM time < 54 mg/dL (non-inferiority, key secondary endpoint)

Efficacy of the tested device will be declared if superiority in HbA1c is met.

To preserve the overall type 1 error, a hierarchical gatekeeping testing procedure will be used with HbA1c at 13 weeks tested first. If the HbA1c analysis results in a statistically significant result ($p < 0.05$), then testing will proceed to the CGM time <54 mg/dL analysis.

6.2 Additional Secondary Efficacy Endpoints

6.2.1 Secondary Efficacy Endpoints Included in Hierarchical Analysis

Assuming both the primary and key secondary endpoints meet statistical significance as described above, the following CGM-measured secondary endpoints will be tested for superiority in a hierarchical fashion as described in Section 9.4.

3. Mean glucose
4. Time 70-180 mg/dL
5. Time >180 mg/dL
6. Time >250 mg/dL
7. Standard deviation
8. Time <70 mg/dL
9. Time <54 mg/dL
10. Coefficient of variation

6.2.2 Other Secondary Efficacy Endpoints

The following endpoints are considered exploratory. Type 1 error for these endpoints will be controlled using the false discovery rate (FDR).

Percentage improvements/increases/decreases are absolute unless otherwise noted.

HbA1c:

- HbA1c <7.0% at 13 weeks
- HbA1c <7.0% at 13 weeks in participants with baseline HbA1c >7.5%
- HbA1c <7.5% at 13 weeks
- HbA1c <8.0% at 13 weeks
- HbA1c >9.0% 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 improvement from baseline to 13 weeks >1.0% or HbA1c <7.0% at 13 weeks

CGM-Measured:

Continuous Outcomes:

- Time in range 70-140 mg/dL
- Time in range 70-120 mg/dL
- Time <60 mg/dL
- Area over the curve (70 mg/dL)
- Low blood glucose index (LBGI)
- CGM-measured hypoglycemic events
- CGM-measured hyperglycemic events
- Time >300 mg/dL
- Area under the curve (180 mg/dL)
- High blood glucose index (HBGI)
- Mean of daily difference

Binary Outcomes:

- Time in range 70-180 mg/dL >70%
- Time in range 70-180 mg/dL improvement from baseline to 13 weeks ≥5%
- Time in range 70-180 mg/dL improvement from baseline to 13 weeks ≥10%
- Time <70 mg/dL <4%
- Time <54 mg/dL <1%

Combined Secondary Outcomes:

Continuous Outcomes:

- BGRI = LBGI + HBGI

Binary Outcomes:

- Improvement in HbA1c > 0.5% without an increase in time < 54 mg/dl by > 0.5% OR improvement in time < 54 mg/dl by > 0.5% without an increase in HbA1c by > 0.5%
- Improvement in time 70–180 mg/dl by >10% without an increase in time < 54 mg/dl by > 0.5% OR improvement in time < 54 mg/dl by > 0.5% without a decrease in time 70–180 mg/dl by > 10%
- Mean glucose <154 mg/dL and time <54 mg/dL <1%
- Time in range 70-180 mg/dL >70% and time <54 mg/dL <1%

Other Secondary Outcomes:

- Questionnaires scores on each questionnaire that is administered to both treatment groups (see Section 9.6)***
- Insulin
  - Total daily insulin (units/kg)
  - Percentage change in the TDD of insulin over the first two-week period relative to the TDD of insulin in the last two-week period (BP aspart/lispro group only)
- Weight and Body Mass Index (BMI)
- From the weekly questionnaires, number of hypoglycemic events requiring carbohydrate treatment per 24 hours
- From the weekly questionnaires, grams of carbohydrate taken specifically to prevent or treat hypoglycemic events per 24 hours

7. Outcome Calculations

7.1 Analysis Windows

The target dates for pertinent visits used in defining efficacy analysis, safety, HbA1c, height and weight, and questionnaire allowable analysis windows are as follows:

Visit	Target Date	Allowable Analysis Window	Central Lab HbA1c Measured	Height/Weight Measured	Questionnaires Given
Screening	Same as Screening Visit day	-	No	Yes*	Yes
Randomization	Same as Randomization Visit day	-	Yes	Yes, if randomization date different from screening date*	No

Week 13	Randomization Visit day + 91 days	-14 to +28 days	Yes	Yes	Yes

*For measurements that are taken at both screening and randomization, prioritize the measurement at randomization to be used as the baseline value.

For weekly questionnaires, target dates will be every 7 days after the Randomization Visit date. These target dates will be given a window of ±3 days.

The analysis windows for device data (i.e. CGM and insulin) inclusion for each period are as follows:

	Device Data Analysis Windows	
Period	**Start Date-Time**	**End Date-Time***
Baseline/Pre-Randomization†	End date-time minus 14 24-hour days	Final CGM reading prior to midnight of the Randomization Visit day.
RCT (Weeks 1-13)	Midnight of Randomization Visit day + 1 day	Earliest of: - Midnight of Week 13 Visit day - Start date-time + 91 full days
RCT First 2 Weeks	Midnight of Randomization Visit day + 1 day	Start date-time + 14 full days
RCT Last 2 Weeks	End date-time minus 14 days.	Earliest of: - Midnight of Week 13 Visit day - Start date-time + 91 full days
RCT First 4 Weeks	Midnight of Randomization Visit day + 1 day	Start date-time + 28 days
RCT 2nd and 3rd 4-Week Periods	End date-time of previous 4-week period	Start date-time + 28 days

*With the exception of the "RCT Last 2 Weeks" period, if midnight of latest contact date occurs earlier than the specified end date-time, then midnight of the latest contact date will be used.
†Assumes that CGM data are available for the 14 days immediately preceding randomization. If at least 10 days of data are not present within the 14 days, then the time period is extended to include 14 days of data.

The safety analysis windows for each period are as follows:

- Pre-randomization: Enrollment until randomization.
- RCT: Randomization Visit day to Week 13 visit day.

7.2 CGM Metrics

Baseline values for each CGM metric will be computed from either the participant's personal Dexcom G6 data or from the G6 wear prior to randomization. The most recent two weeks of CGM data prior to randomization will be included in the calculation of baseline CGM metrics.

If the subject has <168 hours of CGM data, then the baseline metrics will not be calculated and will be set to missing; this is not expected to occur as participants with <85% of CGM values during 14 days of CGM wear will need to complete a period of CGM wear to be eligible for randomization.

CGM metrics from the other periods will use all CGM data available. If a subject has <168 hours of CGM data, then post-randomization CGM metrics will not be calculated and will be set to missing. The primary and secondary analyses utilize a direct likelihood model to account for missing CGM metrics data, as described in Section 9.

Percentage of CGM values that fall within a specified range will be calculated by dividing the number of CGM values that fall within the range by the total number of CGM readings.

A CGM-measured hypoglycemic event is defined as ≥15 consecutive minutes with a CGM glucose value <54 mg/dL. The hypo event ends when there are ≥15 consecutive minutes with a CGM glucose value ≥70 mg/dL, at which point the participant becomes eligible for another hypo event.

Two separate definitions of CGM-measured hyperglycemic events will be calculated and assessed:

1) ≥15 consecutive minutes with a CGM glucose value >300 mg/dL. The hyper event ends when there are ≥15 consecutive minutes with a CGM glucose value ≤250 mg/dL, at which point the participant becomes eligible for another hyper event.
2) ≥90 cumulative minutes with a CGM glucose value >300 mg/dL within a 120-minute period. The hyperglycemia event ends when there are ≥15 consecutive minutes with a CGM glucose value ≤180 mg/dL, at which point the participant becomes eligible for another hyper event.

The amount of time with CGM data will be calculated as the number of CGM values multiplied by 5 minutes, with the exception that there can only be one value within a 5-minute period. Percentage of time CGM is used will be calculated as the amount of time of CGM readings divided by the total possible time the participant could have used the CGM (end date-time minus

start date-time from the analysis window tables in Section 7.1). Note that this denominator includes sensor warm up periods with no CGM values, so this metric will serve as a conservative estimate of CGM use.

Mean of daily differences is the mean of the absolute differences between daily mean glucose.

7.3 Closed Loop Use

For calculating closed loop use, the numerator will be number of hours in which the CL insulin delivery system was on and available, and the denominator is the period end date-time minus start date-time. Note the denominator includes time in which the pump is turned off or not in use, thus this metric would reflect a real-world percentage of time in which the participant actively used the device.

8. Description of Statistical Methods

8.1 General Approach

For all outcomes, mean and standard deviation or summary statistics appropriate to the distribution will be reported.

All p-values will be two-sided. Standard residual diagnostics will be performed for all analyses. If residuals are highly skewed, then an appropriate alternative transformation or nonparametric analysis based on ranks will be performed instead.

All analyses will compare the BP with aspart/lispro group with the Control Group.

8.2 Analysis Cohorts

Primary and Secondary Analyses:

- All randomized participants and participants assigned to an RCT treatment group will be analyzed for the intention-to-treat (ITT) efficacy analyses and all safety analyses.
- Safety outcomes will be reported for the pre-randomization period for all enrolled participants.

Per Protocol Analysis:

- A per-protocol analysis will be restricted to participants with:
  a) Baseline and 13-week HbA1c central lab measurements available,
  b) For participants in the Control Group, at least 80% of CGM data available in the 13-week RCT wear period defined in Section 7.1,

c) No glucose-lowering medications used other than those acceptable in the protocol,

d) No major protocol deviation that could impact outcome measures, and

e) For participants in the BP group, closed loop mode is active for at least 80% of the time during the 13 weeks and announce a meal bolus on average at least 2 times per day when the system is being used.

This analysis will only involve the primary and key secondary outcomes and will only be done if >5% of subjects are excluded based on this criterion.

While-on-Treatment Analysis:

- A while-on-treatment analysis will include all participants assigned/randomized to an RCT treatment group. For participants who discontinued the iLet device in the BP group, HbA1c and CGM metrics will only include data prior to iLet discontinuation.

This analysis will only involve the hierarchical outcomes. Any missing endpoints will be handled by the method of direct likelihood.

9. Analysis of the Primary and Secondary Efficacy Endpoints

The primary analysis will include both the pediatric and adult participants in a single analysis. All analyses comparing the BP aspart/lispro with the Control Group will follow the intention-to-treat (ITT) principle with the data from each participant analyzed according to the treatment assigned by randomization. Participants not randomized due to a protocol deviation and assigned to the BP Group will be analyzed according to the BP group assignment and a sensitivity analysis will be performed reanalyzing these participants as part of the Control group.

9.1 HbA1c Analyses (Superiority)

Only lab HbA1c measurements will be used in the analyses. If a hemoglobin variant is present such that the primary central lab is unable to measure HbA1c using its methodology, the sample may be analyzed at a separate lab using the boronate affinity method or another method which is not affected by the hemoglobin variant. Summary statistics appropriate to the distribution will be reported for HbA1c at baseline, at 6 weeks, and at Week 13 as well as for differences from baseline by treatment group. If both venous and capillary central lab measurements are available for the same timepoint, the venous value will be used in the analyses. Scatterplots of the baseline HbA1c vs. week 13 HbA1c and baseline HbA1c vs. change in HbA1c will be produced for the Control and BP aspart/lispro groups.

HbA1c at 13 weeks will be compared between the BP aspart/lispro and Control Groups using a linear mixed effects regression model adjusting for baseline HbA1c, age, and clinical center

(random factor). A 95% confidence interval will be reported. HbA1c is expected to be normally distributed, but regression diagnostics will be performed to check the residuals and an appropriate alternative transformation or nonparametric analysis based on ranks will be performed if the residuals have a skewed distribution. In the event that some HbA1c values are not available at 13 weeks, then the linear mixed effect regression model will use the method of direct likelihood to incorporate information from baseline and 6 weeks measurements to calculate the maximum likelihood at 13 weeks. Further details on direct likelihood are provided in Section 15.

9.2 Time <54 mg/dL (Noninferiority)

Summary statistics appropriate to the distribution will be reported for % time <54 mg/dL at baseline and over the 13-week follow-up period as well as for differences from baseline by treatment group. Scatterplots of the baseline time <54 mg/dL vs. 13-week time <54 mg/dL and baseline time <54 mg/dL vs. change in time <54 mg/dL will be produced for the Control and BP aspart/lispro groups. If % time <54 mg/dL is skewed and does not display well in a scatterplot, then boxplots will be produced instead.

Time below 54 mg/dL will be calculated over 13 weeks for each subject as described in Section 7. A two-sided 95% confidence interval on the mean difference in % time <54 mg/dL between BP aspart/lispro and Control Group will be performed based on a linear mixed effects regression model adjusting for baseline % time <54 mg/dL, age, and clinical center (random factor). Noninferiority will be assessed by comparing the upper bound of this confidence interval to a noninferiority limit of 1%. Missing data will be handled by the method of direct likelihood (see Section 15). Residuals values will be examined for an approximate normal distribution. If the values are highly skewed, then an appropriate alternative transformation or nonparametric analysis based on ranks will be performed instead. A two-sided p-value will be reported, and a 5% significance level will be used to declare significance.

Since noninferiority is typically framed in terms of a one-sided test, it is worth noting that the left half of a two-sided test at alpha = 0.05 gives the same rejection region as a one-sided test at alpha = 0.025. Therefore, reporting a two-sided 95% confidence interval will provide flexibility to test for inferiority if noninferiority cannot be declared while maintaining the overall type 1 error rate of 5%.

9.3 Secondary CGM and HbA1c Metrics (Superiority)

Summary statistics appropriate to the distribution will be reported for the CGM-measured metrics at baseline and during follow up as well as for differences from baseline by treatment

group. Scatterplots or boxplots of the secondary efficacy endpoints included in the hierarchy will be produced at baseline and over 13 weeks for the Control and BP aspart/lispro groups.

Secondary continuous CGM metrics will be calculated as described in Section 7. CGM metric differences between BP and Control Groups will be compared using a linear mixed effects regression model adjusting for the baseline value of the metric, age, and clinical center (random effect). A 95% confidence interval for the treatment group difference will be reported. Residual values will be examined for an approximate normal distribution. If residuals are highly skewed, then an appropriate alternative transformation or nonparametric analysis based on ranks will be performed instead. Missing data will be handled using direct likelihood (see Section 15).

For the binary secondary outcomes, *p*-values will be produced from logistic regression models adjusting for the same factors mentioned above for the primary analyses (i.e., baseline value of the outcome, age, and clinical center as a random effect). A 95% confidence interval for the treatment group adjusted risk difference will be produced using parametric bootstrapping.

Participant-level cumulative distribution plots of HbA1c and % time in range 70-180 mg/dL by treatment group will be produced.

9.4 Hierarchical Analyses

To preserve the overall type 1 error for the primary and key secondary efficacy endpoints as defined in Section 6.1 and additional secondary efficacy endpoints listed in Section 6.2.1, a hierarchical testing procedure will be used. If the primary analysis for HbA1c results in a statistically significant result ($p < 0.05$), then testing at the 0.05 level will proceed to the next outcome metric. This process continues iteratively moving to the next variable down on the list until a non-significant result is observed, or all 10 variables have been tested. If a non-significant result is encountered, then formal statistical hypothesis testing is terminated and any variables lower on the list will not be formally tested.

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

9.5 Sensitivity Analyses

- Covariate Adjustment: The primary analysis includes a pre-specified list of covariates. As an additional sensitivity analysis, any baseline demographic or clinical characteristics observed to be imbalanced between treatment groups will be added as covariates to the analyses of the primary endpoint. The determination of a meaningful baseline imbalance will be based on clinical judgment and not a p-value.
- Missing Data: All randomized participants will be included in the primary and key secondary analyses according to ITT and any missing data will be handled using direct likelihood. 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, the following methods will be performed:
    - Multiple imputation with a pattern mixture model assuming the dropout trajectory of the BP Group was that of the Control Group.
    - Available cases only.
- The primary analysis will be recalculated switching non-randomized participants assigned to the BP Group due to a protocol deviation to the Control group.

9.6 Questionnaires and Other Outcomes Analyses

Questionnaire scoring will be detailed in the Questionnaire Scoring Appendix.

Summary statistics of total and subscale scores appropriate to the distribution will be tabulated by visit and treatment group. For questionnaires administered to both randomization groups, comparisons will be made using similar linear models as described above for the primary and key secondary outcomes. If questionnaires include a total score, separate models will be run for the total score and any subscales, with the exception of the Clarke Hypoglycemia Awareness questionnaire. The models for comparing DTSQ scores between treatment groups will use a combined score calculated as the DTSQ-S score plus the DTSQ-C score. It will still adjust for the baseline DTSQ-S score.

Similarly, for carbohydrate, total daily insulin, weight and BMI metrics, comparisons will be made using similar linear models as described above for the primary HbA1c analysis (Section 9.1). Percentage change in the TDD of insulin over the first two-week period relative to the TDD of insulin in the last two-week period will not be tested because this metric is limited to BP group only. A 95% confidence interval will be reported for treatment group mean difference in total daily insulin, body weight, and body mass index.

10. Safety Analyses

All randomized participants will be included in the safety analyses and all of their post-randomization safety events will be reported. Separately, any adverse events occurring between screening and randomization will be reported.

All reportable adverse events will be tabulated by treatment group. Details will be provided in a listing of each event, including Medical Dictionary for Regulatory Activities (MedDRA) term and MedDRA System Organ Class. Safety analyses for the RCT will include events occurring on or after randomization until and including the 13-week visit, latest contact date, or Day 98 from randomization, whichever occurs first. Analysis windows for the RCT and other periods are further detailed in Section 7.1.

Formal statistical testing only will be performed for selected safety endpoints. For the following outcomes, mean and standard deviation (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 between the BP aspart/lispro and Control Group):

- 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
- Worsening of HbA1c from baseline to 13 weeks by >0.5%

If enough events occur for the severe hypoglycemia (SH), DKA, and other serious adverse event (SAE) outcomes, the numbers of events will be compared between the two treatment groups during the RCT using a Poisson regression model adjusting for baseline age and HbA1c, and site as a random effect. For SH, the model will additionally control for whether or not the participant experienced at least 1 SH event prior to randomization. For DKA, the model will additionally control for whether or not the participant experienced at least 1 DKA event prior to randomization. If distribution is zero-inflated or the variance is greater than its mean, then a zero-inflated Poisson or negative binomial regression model will be considered. For worsening of HbA1c from baseline to 13 weeks by >0.5%, a logistic regression model adjusting for baseline age and HbA1c and site as a random effect will be used to compare the two treatment groups.

Since the Control Group is not provided with a study blood glucose meter or blood ketone meter, no treatment group comparisons of meter data will be performed. Additionally, no formal statistical comparison will be made of all reported adverse events combined since there are

specific requirements in the protocol for reporting certain events in real-time for the BP Group but not the Control Group, and adverse device effects are only reported for the BP Group.

No adjustments for multiple comparisons will be made for any safety analyses.

10.1 Safety Tabulations Specific to the BP Group

For the BP group, the following will be tabulated :

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

11. Additional Tabulations and Analyses

The following tabulations will be performed according to treatment group:

- Baseline demographics and clinical characteristics
- A flow chart accounting for all participants for all visits
- Visit completion rates for each follow-up visit
- Protocol deviations
- Modifications in insulin delivery type in the Control Group (e.g. change between MDI/CSII) during the study
- Number and reasons for unscheduled visits and phone calls
- CGM usage

11.1 Tabulations Specific to the BP Group

The following will be tabulated in the BP group:

- Number of participants who stopped BP use and reasons
- % time in closed loop
- % CGM capture while iLet is on
- % time device was autonomously dosing while iLet is on
- Total daily insulin dose by week during the first month and then by month, and also according to baseline HbA1c
- Participant completed surveys specific to BP Group
- Device malfunctions requiring study team contact and other reported device issues

11.2 Baseline Descriptive Statistics

Baseline demographic and clinical characteristics of the study cohort will be summarized. The following descriptive statistics will be tabulated overall and by randomization group:

- Age in years at randomization
- Gender
- Race
- Ethnicity
- Type I diabetes duration in years at randomization
- BMI (if available and participant is ≥18 yr)
- BMI percentile (if available and participant is <18 yr)
- HbA1c (% mmol/L)
- *C*-peptide
- Current insulin modality
- Past amount of pump use for patients using pump at enrollment
- Daily insulin units
- CGM use status at enrollment
- Time range of most recent severe hypoglycemic event
- Number of severe hypoglycemic events in the last 12 months
- Number of severe hypoglycemic events involving seizure/coma over lifetime
- Number of severe hypoglycemic events involving seizure/coma in the last 12 months
- Time range of most recent DKA event
- Number of DKA events in the last 12 months
- Number of glucose tests per day
- Other non-insulin blood sugar control medications taken

For continuous variables, summary statistics appropriate to the distribution will be given. For discrete variables, number and percentage will be reported for each category. For participants who have multiple screening visits due to an extended period between initial screening and randomization, the baseline data closest to randomization will be used.

12. Planned Interim Analyses

No formal interim efficacy analyses are planned as study recruitment is expected to be rapid and the duration of follow up short. 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 detailed in the protocol.

Upon completion of the RCT, the efficacy and safety analyses will be performed in preparation for PMA submission.

13. Subgroup Analyses

In exploratory analyses, the primary and key secondary outcomes plus % time in range 70-180 mg/dL and mean glucose will be assessed separately for interaction with certain baseline variables. Subgroup analyses will assess the effectiveness of BP by the following patient characteristics:

- Age (6-<13, 13-<18, 18-<26, 26-<50, ≥50 years old)
- Baseline HbA1c (<7.0%, 7.0 to <8.0%, 8.0 to <9.0%, 9.0 to <10.0%, ≥10.0%)
- CGM use and insulin modality prior to enrollment. Subgroups will be determined based on CGM and insulin delivery mode (i.e. hardware and how it is being used) after all participants have been randomized and before data analyses are conducted.
- Baseline % time <70 mg/dL
- Baseline % time >180 mg/dL
- Baseline % time in range 70-180 mg/dL
- Sex
- Combined race/ethnicity
- T1D Duration
- BMI subgroups as follows:
    - Underweight – BMI index <18.5 for adults or BMI percentile <5th percentile for pediatric participants
    - Normal weight – BMI index 18.5 to <24.9 for adults or BMI percentile in the 5th to <85th percentile for pediatric participants
    - Overweight – BMI index ≥24.9 for adults or BMI percentile ≥85th percentile for pediatric participants.
- Education
- *C*-peptide
- Site (tests for random effects interaction with treatment group)

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 in the pre-defined age groups specified above. 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.

A *p*-value for the subgroup-by-treatment interaction effect will be produced for each outcome assessed using a linear mixed effects regression model similar to the one for the primary HbA1c

analysis (Section 9.1).

13.1 Sub-Analyses

All primary and hierarchical secondary variables will be evaluated within the following subgroups:

- Adults ≥18 years old
- Pediatric participants <18 years old
- Non-closed loop users at baseline, if this subgroup makes up <75% of the randomized cohort
- Participants with HbA1c >7.0% at baseline

p-values and 95% confidence intervals for the treatment group difference for these outcomes will be produced within each subgroup using a linear mixed effects regression model similar to the one for the primary HbA1c analysis (Section 9.1).

13.2 Missing Data in Subgroup Analyses

For all subgroup analyses, missing data will be handled using direct likelihood (Section 15). If residuals are skewed, then an appropriate alternative transformation or nonparametric analysis based on ranks will be performed instead.

Interpretation of the subgroup analyses will be made with caution, particularly if the primary analysis is not significant.

14. Multiple Comparisons/Multiplicity

Hierarchical Analyses

The hierarchical testing procedure described above in Section 9.4 will be used to control the overall Type 1 error for the primary and key secondary endpoints plus the eight additional hierarchical secondary efficacy outcomes identified in Section 6.

All Other Secondary Analyses

For the other secondary analyses, the false discovery rate will be controlled using the adaptive Benjamini-Hochberg procedure. Analyses will be grouped into the following categories:

- Other Secondary Exploratory Analyses – HbA1c, Insulin, Carbohydrate, and BMI/Weight

- Other Secondary Exploratory Analyses – CGM-Measured
- Sub-Analyses – Adults, Pediatric Participants, Non-CL Users at Baseline Participants, Participants with HbA1c >7.0%
- Subgroup Analyses – Analyses by Baseline Characteristics and Site Effects
- Questionnaires

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

15. Direct Likelihood

The direct likelihood method uses baseline data to estimate missing 13-week values via maximum likelihood. This involves adding a visit variable and a treatment by visit interaction effect to the model. The baseline version of the outcome will be included as part of the outcome (i.e. not as a fixed effect covariate), but the treatment effect will be constrained to be zero at baseline. Other baseline covariates added to the model will be treated as fixed effects.

16. Additional Exploratory Analyses

Additional analyses comparing treatment groups will include:

- ***Comparison of aspart/lispro BP Group versus Control Group in adults ≥18 years old and separately for participants <18 years old for the safety outcomes described for the main RCT analyses. *p*-values will be produced using similar models as detailed in Section 10.

- Separate CGM analyses for daytime and nighttime. The following CGM metrics will be tabulated separately for daytime and nighttime by period (i.e. baseline and follow-up), and also by 4-week follow-up periods:
  - Mean glucose
  - Glucose SD
  - Glucose coefficient of variation
  - % time in range 70-180 mg/dL
  - % time above 180 mg/dL
  - % time above 250 mg/dL
  - % time below 70 mg/dL
  - % time below 54 mg/dL

For the time-of-day analyses, at least 126 hours of daytime CGM data and at least 42 hours of nighttime CGM data will be required. Boxplots of % time 70-180 mg/dL and % time <54 mg/dL by time of day and by 4-week period will be produced.

- ***Intersubject variability of HbA1c

16.1 Other Analyses, Tabulations, and Figures

- ***Scatterplots of HbA1c and % time 70-180 mg/dL vs. CGM and closed loop use and change in HbA1c and time 70-180 mg/dL versus CGM and closed loop use.

IOBP Pivotal Trial
Statistical Analysis Plan
BP Fiasp

Version 1.0
Protocol Version 10.0
Author: Zoey Li
Date: October 28, 2021

Revision History

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

Version Number	Author	Senior Statistician	Effective Date	Study Stage
1.0	Zoey Li	Peter Calhoun	October 28, 2021	Enrollment completed. SAP created before unblinding of results to study investigators.

Version Number	Revision Description
1.0	Original version

Approvals

Role	Digital Signature or Handwritten Signature/Date
Author: Zoey Li	
Senior Statistician: Peter Calhoun	
JCHR Sponsor Representative: Roy Beck	

1. Contents

2. Consistency of Statistical Analysis Plan with Protocol/Analyses Planned for Manuscript

This SAP is consistent with the study protocol stats chapter (version indicated on the title page) with the exception of the following change(s):

- CGM-measured hyperglycemic events definition revised.
- The BP Fiasp Group and the BP aspart/lispro Group will not be pooled for secondary analyses.

3. Study Overview

The IOBP Pivotal trial is a multi-center randomized control trial that tests the efficacy and safety of an insulin-only configuration of the iLet Bionic Pancreas (BP) system versus a Control group using CGM during a 13-week study period.

The sample size is expected to include ~110 adults in the BP Fiasp Group, ~110 adults in the BP aspart/lispro group, and ~55 adults in the Standard Care Control Group.

Following the initial screening visit, there will be the following visits:

a) Randomization Visit
b) Day 1-2 Phone Call
c) Week 1 Phone Call
d) Week 2 Visit
e) Week 6 Visit
f) Week 10 Visit
g) Week 13 Visit

4. Statistical Hypotheses

The primary outcome is:

- Superiority in HbA1c at 13 weeks between the BP Fiasp and Control Group.

Only the primary outcome needs to be met to declare device efficacy. The key secondary outcome is:

- Non-inferiority in CGM-measured time <54 mg/dL calculated over 13 weeks between the BP Fiasp and Control Group.

The study hypotheses can be stated as follows:

HbA1c Outcome:

- *Null Hypothesis*: There is no difference in the mean HbA1c at 13 weeks between BP Fiasp and Control Group
- *Alternative Hypothesis*: There is a nonzero difference in the mean HbA1c at 13 weeks between BP Fiasp and Control Group

Time <54 mg/dL Outcome:

- *Null Hypothesis*: There is a mean difference of at least 1% in the percentage of time spent with a sensor glucose level below 54 mg/dL between the BP Fiasp and Control Group over the 13 weeks
- *Alternative Hypothesis*: There is a mean difference of less than 1% in the percentage of time spent with a sensor glucose level below 54 mg/dL between the BP Fiasp and Control Group over the 13 weeks

5. Sample Size

The sample size of 440 for the RCT was selected to provide sufficient exposure to the iLet BP system for regulatory purposes, with respect to different age groups and for use of both insulin aspart/lispro and Fiasp. RCT completion is expected for 100 adults randomized to the BP Fiasp group, 100 adults randomized to BP aspart/lispro Group, and 50 adults randomized to the Control Group.

Superiority for HbA1c at 13 weeks and non-inferiority for time <54 mg/dL measured with CGM at intervals over the 13 weeks are considered the primary and key secondary endpoints, respectively. Statistical power for each endpoint was computed assuming the following:

- HbA1c: Statistical power is 91%, assuming true mean HbA1c difference of 0.4% between BP Fiasp and Control Group, standard deviation of 13-week HbA1c of 0.8%, correlation between baseline and 13-week HbA1c of 0.40, two-sided type 1 error of 5%
- Time <54 mg/dL: Statistical power is 86%, assuming no true difference in mean time <54 mg/dL between BP Fiasp and Control Group, a non-inferiority margin of 1%, standard deviation of percent time <54 mg/dL of 2.0%, correlation between baseline and follow-up of 0.40, and one-sided type 1 error of 0.025%

A separate document details the justification of these power calculations.

6. Analysis of Efficacy and Safety Endpoints

All efficacy analyses (including exploratory analyses and tabulation-only analyses) and safety analyses comparing the BP Fiasp vs. Control group will be the same as described in the BP aspart/lispro vs Control Group SAP detailed separately in the Appendix, with the following difference:

- All analyses comparing BP Fiasp vs. Control are restricted to adults only

7. Comparison of the BP Fiasp vs. BP aspart/lispro Groups

All efficacy analyses (including exploratory analyses and tabulation-only analyses) and safety analyses comparing the BP Fiasp vs. BP aspart/lispro group will be the same as described in the BP Aspart/lispro vs Control Group SAP detailed separately in the Appendix, with the following differences:

- All analyses comparing BP Fiasp vs. BP aspart/lispro are restricted to adults only
- The study was not formally powered for these analyses, so primary and hierarchical secondary outcomes will be in its own FDR category when comparing BP Fiasp vs. BP aspart/lispro.
- Sensitivity, per-protocol, and while-on-treatment analyses will not be performed for BP Fiasp vs BP aspart/lispro comparisons.

All analyses comparing the BP Fiasp vs. BP aspart/lispro group will be considered exploratory.

8. Multiple Comparisons

The BP Fiasp vs. Control hierarchical analyses detailed in Section 6 will be controlled for multiple comparisons using the hierarchical method as described in the BP Aspart/lispro vs Control Group SAP in the Appendix. This BP Fiasp vs. Control hierarchy will be given an initial α=0.05.

The false discovery rate for all other BP Fiasp vs. Control group efficacy comparisons detailed in Section 6 and the BP Fiasp vs. BP aspart/lispro group comparisons detailed in Section 7 will be controlled using the adaptive Benjamini-Hochberg false discovery rate correction procedure. The BP Fiasp vs. Control analyses and BP Fiasp vs. BP aspart/lispro analyses will be considered separate, each with their own FDR categories that parallel the BP aspart/lispro vs. Control FDR categories described in the BP Aspart/lispro vs. Control Group SAP detailed separately in the Appendix.