Statistical Analysis Plan

for

Efficacy and safety of the Omnipod 5 System compared to pump therapy in the treatment of type 1 diabetes: a randomized, parallel-group clinical trial

SAP: version 2.0 DRAFT, dated 17 October 2023

Protocol: version 1.1, dated 14 November 2022
version 3.0-FR, dated 25 July 2023

Document Approval

Approval	
Approved By	Kaisa Kivilaid, MS DocuSigned by Kaisa Kivilaid
Signature	I approve this document 17 Oct 2023 | 11:44:32 AM EDT
Date	44D9C626FC8A4993B8167137B423B198
Role	Biostatistician, Fortrea

Approval	
Approved By	Trang Ly, MBBS, FRACP, PhD DocuSigned by Trang Ly
Signature	Trang Ly I approve this document 17 Oct 2023 | 11:48:38 AM EDT
Date	5574B167ED01438CB4D9D3EBC2DDCFE5
Role	Senior Vice President, Clinical and Medical Director, Insulet Corporation

Table of Contents

1 Purpose 6
2 Study Design 6
2.1 Study Design Overview 6
2.2 Investigational Device 6
2.3 Study Schedule 6
2.4 Study Success 7
2.5 Randomization 7
2.6 Point of Enrollment 7
3 Study Objectives and Endpoints 7
3.1 Primary Objective and Endpoint 7
3.2 Secondary Effectiveness Objective and Endpoints 7
3.3 Exploratory Endpoints 8
3.4 Patient-Reported Outcomes 9
3.5 Period 2 Outcomes 9
4 Analysis of Primary Endpoint 10
4.1 Sample Size 10
4.2 Primary Analysis 10
4.3 Sensitivity analysis 11
4.3.1 Complete Cases Only 12
4.3.2 Multiple Imputation Assuming MAR 12
4.3.3 Two-way Tipping Point Analysis 12
4.3.4 Covariate Adjustment 13
4.4 Poolability 13
4.5 Subgroup and Stratified Analyses 13
5 Analysis of Other Endpoints 14
5.1 Analysis of Secondary Endpoints (PERIOD 1) 14
5.1.1 Percentage of Time <54 mg/dL (Non-Inferiority) 15
5.1.2 Percentage of Time >180 mg/dL 16

5.1.3 Mean Glucose 16
5.1.4 Change from Baseline in A1C 16
5.1.5 Percentage of Time <70 mg/dL 16
5.1.6 Change from Baseline in T1-DDS Total Score 17
5.1.7 Change from Baseline in HCS Total Score 17
5.1.8 Change from Baseline in DQOL-brief Total Score 17
5.1.9 Proportion of Participants with T1-DDS MCID (Total Score Improvement ≥0.19) after 13 Weeks 18
5.1.10 Proportion of Participants with DQOL-brief Clinically Meaningful Improvement after 13 Weeks 18
5.1.11 Proportion of Participants with HCS Total Score ≥3 after 13 Weeks 18
5.2 Analysis of Exploratory Endpoints (PERIOD 1) 19
5.2.1 A1C at 13 Weeks and Improvement from Baseline in A1C 19
5.2.2 Change from Baseline in Body Mass Index (BMI) 19
5.2.3 Change from Baseline in Total Daily Insulin (TDI) 20
5.2.4 Additional CGM-Derived Endpoints 20
5.2.5 Episodes of Severe Hypoglycemia and DKA 20
5.2.6 Analysis Methods for Exploratory Endpoints 21
5.3 Analysis of Period 2 Outcomes 21
6 Analysis of Safety 22
6.1 Medical Monitor 22
6.2 Relationship to Study Device and Procedure 22
6.3 Severity 23
6.4 Outcome 23
6.5 Analysis of Adverse Events 24
6.6 Evaluation of Device Deficiencies 24
7 General Statistical Considerations 24
7.1 Analysis Windows 24
7.1.1 Sensor-Derived Endpoints 24
7.1.2 Other Endpoints 25

7.2 Analysis Sets 25
7.2.1 ITT (Intention to Treat) Analysis Set 25
7.2.2 mITT (modified Intention to Treat) Analysis Set 25
7.2.3 PP (Per Protocol) Analysis Set 26
7.3 Control of Systematic Bias 26
7.4 Calculation of Percentage of Time 26
7.5 Length of Randomized Period 27
7.6 Missing Data Handling (Period 1 only) 27
7.6.1 Missing Data Handling for Sensor-Derived Endpoints 27
7.7 Patient-Reported Outcomes 28
7.8 Other Data Summaries 31
7.9 Statistical Software 31
7.10 Periodic Review of Controller Data 31
8 References 32

1 PURPOSE

The purpose of this Statistical Analysis Plan (SAP) is to describe the statistical methodology that is used to confirm the safety and effectiveness of the Omnipod 5 Automated Glucose Control System in patients with type 1 diabetes. Should there be a discrepancy between the OP5-003 study protocol(s) and this SAP, the content of the SAP shall prevail.

2 STUDY DESIGN

2.1 STUDY DESIGN OVERVIEW

This is a randomized, double-arm, multi-center prospective clinical study. The study will be conducted in up to 15 clinical sites across United States and France. Participants will be randomized in a 2:1 allocation to Intervention Arm (Omnipod 5 System with Dexcom G6 CGM) or Control Arm (participant's current insulin pump with Dexcom G6 CGM).

The study is expected to enroll 200 participants in order to obtain a minimum of 170 randomized participants with 131 participants completing the 13-week randomized period. Adults (aged ≥18 years) with type 1 diabetes currently on pump therapy with an A1C between 7-11% (inclusive) will be recruited for the study. At least 80% of participants must have an A1C ≥8%. At least 50% of participants must be using Omnipod for pump therapy at the time of enrollment.

2.2 INVESTIGATIONAL DEVICE

Devices used in the Intervention Arm include:

- Omnipod 5 System, comprised of the following components:
  - Omnipod 5 Pod
  - Omnipod 5 App (on the Insulet-provided Controller)
- Dexcom G6 Continuous Glucose Monitor

Devices used in the Control arm include:

- Participant's own insulin pump
- Dexcom G6 Continuous Glucose Monitor

2.3 STUDY SCHEDULE

The study will consist of:

1. Two-week standard therapy period, followed by
2. 13-week randomized period (Period 1)

Following Period 1, France participants of both groups will be offered the option to transition to use the CE-marked commercial Omnipod 5 System for 12 more months with French interface (Period 2) to collect additional data in real-world use.

2.4 STUDY SUCCESS

The study will be deemed successful if the primary endpoint is met for Period 1.

2.5 RANDOMIZATION

This is a randomized, double-arm, multi-center prospective clinical study. Participants will be randomized in a 2:1 allocation to Intervention Arm (Omnipod 5 System with Dexcom G6 CGM) or Control Arm (participant's current insulin pump with Dexcom G6 CGM). The participants who have met all the eligibility criteria, have completed all screening assessments, and have met the CGM device usage and data criteria will be randomized 2:1 using a computer-generated randomization scheme stratified by site, A1C (7% to <8% vs. ≥8% to 11%), and Omnipod use (user vs. non-user). A permuted block randomization scheme will be utilized to balance groups assignments.

2.6 POINT OF ENROLLMENT

A participant is considered enrolled into the study when they begin wearing the Dexcom G6 CGM dispensed for the study. Participants who do not meet the eligibility criteria will not continue in the study and will be considered screen failures.

3 STUDY OBJECTIVES AND ENDPOINTS

3.1 PRIMARY OBJECTIVE AND ENDPOINT

The primary objective is to demonstrate superior efficacy of the Omnipod 5 System compared to pump and CGM therapy in adults with type 1 diabetes The primary endpoint associated with this objective will be compared between the Intervention arm and the Control arm:

- Per-participant percentage time in range (TIR) 70-180 mg/dL as measured by study CGM during the 13 weeks of randomized period

3.2 SECONDARY EFFECTIVENESS OBJECTIVE AND ENDPOINTS

The secondary objective is to demonstrate additional measures of efficacy and safety of the Omnipod 5 System compared to pump and CGM therapy in adults with type 1 diabetes The secondary endpoints associated with this objective will be compared between the Intervention arm and the Control arm:

- Percentage of time <54 mg/dL (non-inferiority) during the 13 weeks of randomized period
- Percentage of time >180 mg/dL during the 13 weeks of randomized period
- Mean glucose during the 13 weeks of randomized period
- Change from baseline in A1C at 13 weeks
- Percentage of time < 70 mg/dL during the 13 weeks of randomized period
- Change from baseline in Type 1-Diabetes Distress Scale (T1-DDS) total score at 13 weeks
- Change from baseline in Hypoglycemia Confidence Scale (HCS) total score at 13 weeks
- Change from baseline in Diabetes Quality of Life-brief (DQOL-brief) total score at 13 weeks
- Proportion of participants achieving a minimally clinically important difference (MCID) of ≥0.19 points (improvement) on the T1-DDS questionnaire total score at 13 weeks
- Proportion of participants achieving a clinically meaningful improvement in the DQOL-brief total score after 13 weeks
- Proportion of participants achieving HCS total score of ≥3 after 13 weeks

3.3 EXPLORATORY ENDPOINTS

The following exploratory endpoints will be evaluated and compared in the Intervention group vs. Control group at the end of the 13 weeks:

- Proportion of participants with A1C <7%
- Proportion of participants with A1C <8%
- Proportion of participants with ≥1% improvement in A1C from baseline
- Proportion of participants with ≥10 percentage point relative improvement in A1C from baseline
- Proportion of participants with either ≥1% improvement in A1C from baseline OR ≥10 percentage point relative improvement in A1C from baseline
- Change from baseline in Body Mass Index (BMI)
- Change from baseline in total daily insulin (TDI) (units/kg)
- Additional CGM-derived endpoints, such as coefficient of variation, time above range >250 mg/dL, percentage of time < 54 mg/dL (superiority), TIR during daytime and nighttime hours
- Number of episodes of severe hypoglycemia
- Number of episodes of DKA

3.4 PATIENT-REPORTED OUTCOMES

The following questionnaires will be used to evaluate general and disease-specific quality of life, and device usability. Results will be assessed at 13 weeks compared to baseline (Intervention and Control group):

- Clarke Hypoglycemia Awareness
- EQ-5D-3L
- Type 1 Diabetes Distress Scale (T1-DDS)
- Diabetes Quality of Life-brief (DQOL-brief)
- DAWN Impact of Diabetes Profile (DIDP)
- Hypoglycemia Confidence Scale (HCS)
- Pittsburgh Sleep Quality Index (PSQI)
- System Usability Scale (SUS)
- Insulin Dosing Systems: Perceptions, Ideas, Reflections, and Expectations (INSPIRE)

3.5 PERIOD 2 OUTCOMES

The following outcomes will also be assessed at the end of Period 2 (France participants only):

- Ratio of participants joining Period 2 per group
- Maintenance of use
- Incidence rate of severe hypoglycemia
- Incidence rate of diabetic ketoacidosis (DKA)

The following outcomes will also be assessed at the end of Period 2 as compared to baseline values (2-week standard therapy period) (France participants only):

- HbA1c (at 3, 6, 9 and 12 months)
- Glycemic outcomes
    - Percentage of time <54 mg/dL
    - Percentage of time <70 mg/dL
    - Percentage of time >180 mg/dL
    - Percentage of time in range (70-180 mg/dL)
    - Percentage of patients having time in range ≥70%
    - Percentage of patients having time below range <4%
    - Mean glucose
    - Change from baseline in HbA1c
    - Proportion of participants with HbA1c <7%

  - Proportion of participants with HbA1c <8%
- Percentage of time spent in Automated mode
- Change from baseline in BMI (at 6 and 12 months)
- Insulin usage (units, units/kg)
- Patient-reported outcome questionnaires at 6 months and 12 months (or withdrawal from study):
  - DQOL-brief – used to assess the relative burden of an intensive diabetes treatment regimen
  - HCS – used to measure hypoglycemia awareness, hypoglycemia frequency, severity, and impact

4 ANALYSIS OF PRIMARY ENDPOINT

4.1 SAMPLE SIZE

The sample size estimation is based on the following assumptions:

- Power of 90% and two-sided significance level of 5%
- Difference between groups of 10% in TIR 70-180 mg/dL
- Standard deviation of 16.5% in each group
- 2:1 randomization (Intervention:Control)

Under these assumptions, a total sample size of 131 participants is required (87 Intervention, 44 Control). To allow for pre-randomization attrition of up to 15%, and post-randomization attrition of up to 22.9%, a total sample size of up to 200 will be enrolled to give 170 participants randomized and 131 participants completing the 13 weeks. The anticipated difference in TIR 70-180 mg/dL was projected from the OP5 pivotal study.

The sample size calculations were performed using PASS 2021.

4.2 PRIMARY ANALYSIS

The primary endpoint is per-participant percentage of time in range (TIR) 70-180 mg/dL as measured by study CGM and comparing the Intervention group to the Control group during the 13 weeks. The null hypothesis is that the mean percentage of TIR 70-180 mg/dL in the Intervention group is equal to the mean percentage of TIR 70-180 mg/dL in the Control group during the 13 weeks. The alternative hypothesis is that mean percentage of TIR 70-180 mg/dL in the Intervention group is not equal to the mean percentage of TIR 70-180 mg/dL in the Control group during the 13 weeks.

Formally, the null and alternative hypotheses for the primary endpoint are:

H_0: $\mu_I - \mu_C = 0$

H_A: $\mu_I - \mu_C \neq 0$

where μ_I and μ_C are the population means of per-participant percentage of TIR 70-180 mg/dL as measured by the study CGM during the 13 weeks in the Intervention and Control groups, respectively. The calculation of percentage of time in a specified range is described in Section 7.4.

The primary endpoint will be analyzed using a repeated measures linear mixed effects model with TIR 70-180 mg/dL at baseline and during the 13-week randomized period as the dependent variable. The correlation between baseline and follow-up will be modelled using an unstructured covariance matrix. The model will include randomized treatment group, age, sex, country, and duration of diagnosis as fixed effects and site as a random effect. This model adjusts for baseline TIR 70-180 mg/dL by forcing the treatment groups to have the same mean value at baseline. By including baseline in the response vector, the model allows participants to be included even if they only have data at one of the time points.

Missing data will be handled by using a direct likelihood approach which will allow participants to be included even if they only have data at one of the time points. The direct likelihood method approach assumes data are missing at random (MAR).

Standard residual diagnostics will be performed for all analyses. It is expected that the primary outcome will follow an approximately normal distribution, however, residual values will be examined to confirm an approximate normal distribution. If values are highly skewed, then robust regression using M-estimation will be used instead. In this case, the model will include TIR 70-180 mg/dL during the 13 weeks randomized period as the dependent variable, baseline TIR 70-180 mg/dL as a covariate, and missing data will be handled using multiple imputation.

The primary analysis will utilize the mITT analysis set and will be tested at a two-sided significance level of 5%.

4.3 SENSITIVITY ANALYSIS

Missing data handling in described in Section 7.6 of this document. The following sensitivity analysis will be performed for the primary endpoint.

4.3.1 Complete Cases Only

The primary analysis will be replicated including only participants with non-missing TIR 70-180 mg/dL at baseline and follow-up. This assumes missing completely at random (MCAR).

4.3.2 Multiple Imputation Assuming MAR

Consider a model with TIR 70-180 mg/dL during the 13-week randomized period as the dependent variable, randomized treatment group, baseline TIR 70-180 mg/dL, age, sex, and duration of diagnosis as fixed effects and site as a random effect.

Missing data will be handled using multiple imputation assuming missing at random (MAR). The imputation model will include treatment group, baseline TIR 70-180 mg/dL, age, sex, and duration of diagnosis. Multiple imputation will involve the following steps:

1. Generate 100 imputed data sets.
2. Fit the mixed effects linear regression model on each imputed dataset. Calculate the point estimate and standard error of the treatment group difference for each imputed dataset.
3. Pool the estimates across imputations to get an overall point estimate and confidence interval for the treatment group difference.

4.3.3 Two-way Tipping Point Analysis

A two-way tipping point sensitivity analysis will be performed to evaluate the possibility of missing not at random (MNAR).

First, multiple imputation will be applied to missing TIR 70-180 mg/dL assuming MAR, as described above. This procedure will then be repeated applying various shifts to the imputed values. Since the reasons for missing data may vary by treatment group, a separate shift will be applied for each treatment group. A shift of zero for both treatment groups is equivalent to the model assuming MAR. A positive shift indicates that we expect TIR 70-180 mg/dL to be worse in that treatment group for those with missing values than those with observed data. The range of shifts that are considered will include scenarios where the outcomes for patients missing data are worse among patients randomized to the treatment arm than those randomized to the control arm. The goal is to examine how far one needs to shift for the results of the study to tip from significant to not significant for between-treatment comparison. Tipping points that change the significance will be reported and clinically evaluated for whether such points are considered plausible.

4.3.4 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 model for the primary endpoint. The determination of a meaningful baseline imbalance will be based on clinical judgement and not a p-value.

4.4 POOLABILITY

As the study plans to enroll site in the US and France, the poolability will be assessed first by country, and then by site.

A repeated measures mixed effect model with treatment group and country as fixed effects and TIR 70-180 mg/dL at baseline and during the week 13 randomized period as the dependent variable will be fit for the comparison of treatment groups. The model will also include treatment group by country interaction term. A p-value for the interaction term of <0.15 will be considered to indicate heterogeneity. If the p-value for the interaction term from the model is <0.15, a repeated measures mixed effect model with treatment as fixed effect and country as random effect will be fit.

Further, within each country, a similar model with treatment and study site as fixed effects; TIR 70-180 mg/dL at baseline and during the week 13 randomized period as the dependent variable; and a treatment by site interaction term will be fit. A p-value for the interaction term of <0.15 will be considered to indicate heterogeneity within that country. If the p-value for the interaction term from the model is <0.15, a repeated measures mixed effect model with treatment as fixed effect and study site as random effect will be fit.

For the poolability analysis that assesses homogeneity within each country, sites with ≤5 participants contributing to analysis will be combined into pseudo-centers of at least 10 participants, or combined with existing centers to make sure that each site contributes at least 10 participants. If site is included in other analyses, e.g., random effect in the analysis of primary endpoint, the same approach of combining sites will be used.

4.5 SUBGROUP AND STRATIFIED ANALYSES

Regardless of results of poolability analysis of primary endpoint, the primary endpoint and any secondary endpoints found statistically significant in the hierarchical testing will be further explored to understand the treatment effect. The general approach for these exploratory analyses will be to add an interaction term for the subgroup factor by treatment to the primary and secondary endpoint models described above. Additional factors may be identified based on study results or those deemed of interest based on clinical assessment.

The endpoint results for the factors for which the interaction term was statistically significant (at a two-sided 5% significant level) will be presented in a stratified manner. Categorical variables will be stratified by category (although multiple categories may be combined for ease of interpretation). Continuous variables will be stratified based on mean/median value or clinical or scientific justification. At a minimum, additional stratifications will include:

- Country (USA vs. France)
- Sex
- Race/Ethnicity (White non-Hispanic vs. Other)
- Age
- BMI (≤25 vs >25 kg/m^2)
- Day and night (daytime: 6AM to <12AM; nighttime as 12AM to <6AM)
- Baseline value of the endpoint

5 ANALYSIS OF OTHER ENDPOINTS

5.1 ANALYSIS OF SECONDARY ENDPOINTS (PERIOD 1)

To preserve the family-wise type 1 error for selected key secondary endpoints, a fixed sequence hierarchical testing approach will be used. If the primary endpoint is found to be statistically significant at a two-sided significance level of 5%, testing for secondary endpoints can commence. Secondary endpoints will be tested in a hierarchical fashion at a one-sided significance level of 2.5% (or two-sided significance level of 5% as appropriate). Each endpoint with a two-sided p-value ≤0.05 (or one-sided p-value of ≤0.025, as appropriate) is considered to be met until a secondary endpoint with a two-sided p-value >0.05 (or one-sided p-value of >0.025, as appropriate) is encountered. This process continues iteratively moving to the next endpoint down on the list until a non-significant result is observed, or all endpoints 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 for each key secondary endpoint by treatment group. A 95% confidence interval for the treatment effect will also be calculated for all key secondary outcomes listed above. However, a confidence interval that is below zero (superiority) or that is below the non-inferiority margin (noninferiority) will not be considered a statistically significant result if an outcome variable higher on the hierarchical list failed to reach statistical significance.

The secondary endpoints will be analyzed using a repeated measures linear mixed effects model with the endpoint at baseline and at 13 weeks as the dependent variable. The correlation between baseline and follow-up will be modelled using an unstructured covariance matrix. The model will include randomized treatment group, age, sex, and duration of diagnosis as fixed effects and site as a random effect. This model adjusts for baseline value of the endpoint by forcing the treatment groups to have the same mean value at baseline. Missing data will be handled by using a direct likelihood approach which will allow participants to be included even if they only have data at one of the time points.

Binary PRO-derived endpoints will be presented as number and percent of participants achieving treatment success with corresponding 95% asymptotic confidence limits in each group. Binary endpoints will be analyzed using a logistic mixed effects model with the endpoints as the outcome. The following variables will also be included in the model as fixed effects: randomized treatment group, age, sex, duration of diagnosis, and baseline continuous value of the outcome being tested. In addition, the model will adjust for site as a random effect. A 95% confidence interval for the treatment group adjusted risk difference will be produced using parametric bootstrapping.

Standard residual diagnostics will be performed for all secondary analyses. Residual values will be examined to confirm an approximate normal distribution. If values are highly skewed, then robust regression using M-estimation will be used instead. In this case, the model will include the endpoint at 13 weeks as the dependent variable, baseline value of the endpoint as a covariate, and missing data will be handled using multiple imputation.

The secondary analyses will utilize the mITT analysis set, and will be tested at a two-sided significance level of 5% or a one-sided significance level of 2.5%, as appropriate. If there is evidence of heterogeneity either across countries or study sites based on the poolability analysis, a random effect for country or site, as appropriate, will also be included.

The secondary endpoints will be testing in the following order:

5.1.1 Percentage of Time <54 mg/dL (Non-Inferiority)

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C \geq 1$
- H_A: $\mu_I - \mu_C < 1$
  where μ_I and μ_C are the population means of per-participant percentage of time <54 mg/dL as measured by the study CGM during the 13 weeks in the Intervention and Control groups, respectively

The non-inferiority margin will be set to 1%. The calculation of percentage of time in a specified range is described in Section 7.4. If the upper bound of the two-sided 95% confidence interval (corresponding to 1-sided 97.5% CI) for the difference (μ_I minus μ_C) is <1%, non-inferiority will be established. If the upper bound of the two-sided 95% confidence interval for the difference (μ_I minus μ_C) is <0%, superiority of the Intervention group compared to the Control group will be established.

5.1.2 Percentage of Time >180 mg/dL

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C = 0$
- H_A: $\mu_I - \mu_C \neq 0$
  where μ_I and μ_C are the population means of per-participant percentage of time <180 mg/dL as measured by the study CGM during the 13 weeks in the Intervention and Control groups, respectively

The calculation of percentage of time in a specified range is described in Section 7.4.

5.1.3 Mean Glucose

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C = 0$
- H_A: $\mu_I - \mu_C \neq 0$
  where μ_I and μ_C are the population means of per-participant mean glucose as measured by the study CGM during the 13 weeks in the Intervention and Control groups, respectively

5.1.4 Change from Baseline in A1C

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C = 0$
- H_A: $\mu_I - \mu_C \neq 0$
  where μ_I and μ_C are the population means of change from baseline (Visit 1) in A1C after 13 weeks in the Intervention and Control groups, respectively; change from baseline is calculated as follows:

$$Change\ from\ Baseline\ in\ A1C = A1C\ (Visit\ 1) - A1C\ (13\ weeks)$$

5.1.5 Percentage of Time <70 mg/dL

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C = 0$
- H_A: $\mu_I - \mu_C \neq 0$

where μ_I and μ_C are the population means of per-participant percentage of time <70 mg/dL as measured by the study CGM during the 13 weeks in the Intervention and Control groups, respectively

The calculation of percentage of time in a specified range is described in Section 7.4.

5.1.6 Change from Baseline in T1-DDS Total Score

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C = 0$
- H_A: $\mu_I - \mu_C \neq 0$
  where μ_I and μ_C are the population means of change from Baseline (Visit 2) in T1-DDS total score after 13 weeks in the Intervention and Control groups, respectively; change from baseline is calculated at follows:
  Change from Baseline in T1-DDS = T1-DDS (13 weeks) – T1-DDS (Visit 2)

The Type-1 Diabetes Distress Scale is further described in Section 7.7.

5.1.7 Change from Baseline in HCS Total Score

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C = 0$
- H_A: $\mu_I - \mu_C \neq 0$
  where μ_I and μ_C are the population means of change from Baseline (Visit 2) in HCS total score after 13 weeks in the Intervention and Control groups, respectively; change from baseline is calculated as follows:
  Change from Baseline in HCS = HCS (13 weeks) – HCS (Visit 2)

The Hypoglycemia Confidence Scale (HCS) is further described in Section 7.7.

5.1.8 Change from Baseline in DQOL-brief Total Score

The hypotheses associated with this endpoint are:

- H_0: $\mu_I - \mu_C = 0$
- H_A: $\mu_I - \mu_C \neq 0$
  where μ_I and μ_C are the population means of change from Baseline (Visit 2) in DQOL-brief total score after 13 weeks in the Intervention and Control groups, respectively; change from baseline is calculated as follows:
  Change from Baseline in DQOL-brief = DQOL-brief (13 weeks) – DQOL-brief (Visit 2)

The Diabetes Quality of Life-brief is further described in Section 7.7.

5.1.9 Proportion of Participants with T1-DDS MCID (Total Score Improvement ≥0.19) after 13 Weeks

The proportion of participants achieving a minimally clinically important difference (MCID) of ≥0.19 points (improvement) on the T1-DDS questionnaire total score at 13 weeks will be compared between treatment groups. [1] Those hypotheses are stated as:

- H_0: $\pi_I - \pi_C = 0$
- H_A: $\pi_I - \pi_C \neq 0$
  where π_I and π_C are the proportion of participants achieving an improvement of ≥0.19 points after 13 weeks in the Intervention and Control groups, respectively.

This endpoint will be considered met if the treatment group fixed effect in the logistic regression is significant at a two-sided significance level of 5%.

5.1.10 Proportion of Participants with DQOL-brief Clinically Meaningful Improvement after 13 Weeks

The proportion of participants achieving a clinically meaningful improvement in the DQOL-brief total score after 13 weeks will be compared between groups [2, 3].

- H_0: $\pi_I - \pi_C = 0$
- H_A: $\pi_I - \pi_C \neq 0$
  where π_I and π_C are the proportion of participants achieving a clinically meaning improvement (defined as ≥0.5*standard deviation of change from baseline in the DQOL-brief total score at 13 weeks) in the DQOL-brief total score after 13 weeks in the Intervention and Control groups, respectively.

This endpoint will be considered met if the treatment group fixed effect in the logistic regression is significant at a two-sided significance level of 5%.

5.1.11 Proportion of Participants with HCS Total Score ≥3 after 13 Weeks

The proportion of participants in each treatment group achieving a score of ≥ 3 points in the HCS total score after 13 weeks will be compared. Those hypotheses are stated as:

- H_0: $\pi_I - \pi_C = 0$
- H_A: $\pi_I - \pi_C \neq 0$
  where π_I and π_C are the proportion of participants achieving a score of ≥3 points in the HCS total score after 13 weeks in the Intervention and Control groups, respectively.

This endpoint will be considered met if the treatment group fixed effect in the logistic regression is significant at a two-sided significance level of 5%.

5.2 ANALYSIS OF EXPLORATORY ENDPOINTS (PERIOD 1)

Various exploratory endpoints will be evaluated and compared in the Intervention group versus Control group at the end of the 13-week randomized period.

5.2.1 A1C at 13 Weeks and Improvement from Baseline in A1C

Corelab A1C will be collected at Visit 1 and Visit 8. A1C will be evaluated based on the following exploratory endpoints:

- Proportion of participants with A1C <7% at 13 weeks
- Proportion of participants with A1C <8% at 13 weeks
- Proportion of participants with ≥1% improvement from baseline (Visit 1) in A1C at 13 weeks, where improvement is calculated as follows:

$$Improvement\ in\ A1C = A1C\ (Visit\ 1) - A1C\ (13\ weeks)$$

- Proportion of participants with ≥10 relative percentage point improvement from baseline (Visit 1) in A1C at 13 weeks, where relative percentage point improvement is calculated as follows:

$$Relative\ improvement\ in\ A1C = \frac{A1C\ (Visit\ 1) - A1C\ (13\ weeks)}{A1C\ (Visit\ 1)} * 100\%$$

- Proportion of participants with either ≥1% improvement from baseline (Visit 1) in A1C at 13 weeks, ***or*** ≥10 relative percentage point improvement from baseline (Visit 1) in A1C at 13 weeks

The endpoints will be compared between groups. The denominator for the calculation of these endpoints includes all participants for whom an A1C at 13 weeks is available. Further, for the endpoints involving improvement in A1C, the A1C at baseline (Visit 1) will also need to be available.

5.2.2 Change from Baseline in Body Mass Index (BMI)

Change from Baseline in BMI at 13 weeks will be compared between groups. Height is collected at Visit 1; weight is collected at Visit 1 and at Visit 8. Change in BMI is calculated as follows:

$$Change\ from\ Baseline\ in\ BMI = BMI\ (Visit\ 8) - BMI\ (Visit\ 1)$$

where BMI at baseline (Visit 1) is calculated as:

$$BMI\ (Visit\ 1) = \frac{Weight\ at\ Visit\ 1\ (kg)}{(Height\ at\ Visit\ 1\ (m))^2}$$

and BMI at follow-up (Visit 8) is calculated as:

$$BMI\ (Visit\ 8) = \frac{Weight\ at\ Visit\ 8\ (kg)}{(Height\ at\ Visit\ 1\ (m))^2}$$

If a participant's weight is reported in pounds, the following conversion will be performed:

$$Weight\ (kg) = \frac{Weight\ (lbs)}{2.2}$$

If a participant's height is reported in inches, the following conversion will be performed:

$$Height\ (m) = \frac{Height\ (in) * 2.54}{100}$$

5.2.3 Change from Baseline in Total Daily Insulin (TDI)

Change from Baseline in TDI (units/kg) at the end of the 13-week randomized period will be compared between groups. The insulin requirements will be collected at Visit 3 and at Visit 8, and change in TDI will be calculated as follows:

$$TDI\ \left(change\ from\ Baseline, \frac{units}{kg}\right) = \frac{TDI\ (13\ weeks) - TDI\ (Visit\ 3)}{subject's\ weight\ at\ Visit\ 1}$$

5.2.4 Additional CGM-Derived Endpoints

The following additional CGM-derived endpoints, such as coefficient of variation (CV), percentage of time >250 mg/dL, percentage of time <54 mg/dL (superiority), TIR during daytime and nighttime hours will be compared between groups at the end of the 13-week randomized period.

CV is calculated per participant as follows, then summarized and compared between groups:

$$CV(\%) = \frac{SD\ of\ glucose}{mean\ glucose}$$

5.2.5 Episodes of Severe Hypoglycemia and DKA

As the number of episodes (events) of severe hypoglycemia (SH) and DKA are of significant interest, statistical testing will be performed for the rate of SH (and DKA) if there are ≥5 events (of SH or DKA, respectively) combined across both treatment groups during the randomized period. The rates will be calculated as the number of events per 100 person years. The rates will be compared using a Poisson regression with site as a random effect. The amount of follow-up will be included as an offset term. If a participant withdraws from the study prior to Visit 8 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 randomized period for the purpose of calculating follow-up time for safety endpoints.

If there are concerns with model convergence (e.g., due to zero events in one group), Fisher's exact test for the number of events will be used instead.

The number of severe hypoglycemia and DKA adverse events, and the number and percent of participants with at least one event in each group will be summarized.

5.2.6 Analysis Methods for Exploratory Endpoints

The mITT analysis set will be the main analysis set used to analyze these endpoints.

For continuous endpoints, mean and standard deviation or median and interquartile range, as appropriate, in each group will be presented. Continuous endpoints will be analyzed using a repeated measures linear mixed effects with the endpoint at baseline and at 13 weeks as the dependent variable. The correlation between baseline and follow-up will be modelled using an unstructured covariance matrix. The model will include randomized treatment group, age, sex, and duration of diagnosis as fixed effects and site as a random effect. This model adjusts for baseline value of the endpoint by forcing the treatment groups to have the same mean value at baseline. If values are highly skewed, then robust regression using M-estimation will be used instead, as described above for the secondary endpoints.

For binary endpoints, number and percent of participants in each group will be presented. Binary endpoints will be analyzed using a logistic mixed effects model with the endpoints as the outcome. The following variables will also be included in the model as fixed effects: randomized treatment group, age, sex, duration of diagnosis, and baseline continuous value of the outcome being tested. In addition, the model will adjust for site as a random effect. A 95% confidence interval for the treatment group adjusted risk difference will be produced using parametric bootstrapping. Analysis of binary endpoint will be based on complete cases only.

Since the results of exploratory endpoint analyses will not be used to support clinical claims, no adjustment for multiplicity will be performed.

5.3 ANALYSIS OF PERIOD 2 OUTCOMES

There are no pre-specified hypotheses associated with the safety or performance endpoints. The Period 2 outcomes will be summarized using descriptive statistics. P-values may be calculated for relevant comparisons but will be considered exploratory.

All available data from Period 2 will be included in analysis. Participants who have agreed to participate in Period 2, but provide minimal data, may be excluded from analysis at the discretion of the sponsor.

At the commencement of Period 2, in the event of logistic and/or site scheduling/approval or device delivery challenges occurring for some participants of the Intervention group preventing a continuous use, a switch-back to previous device is accepted for a maximum of 6 months, before activating therapy with the CE-marked commercial Omnipod 5 System with the clinical study site support. Similarly, such a delay in transitioning participants in the Control group to the Omnipod 5 System is also acceptable. Any data collected during that "study pause" will not be included in analysis.

6 ANALYSIS OF SAFETY

All adverse events reported during the study will be reviewed and adequately reported to comply with applicable regulations. All reportable AEs will be assessed by the investigator who will determine whether the event is related to the study procedures or related to the study device and whether the event meets any of the criteria for seriousness. The event will be considered serious if the event meets at least one criterion for seriousness.

6.1 MEDICAL MONITOR

A Medical Monitor will be responsible for individual and timely review of adverse events, including serious adverse events (SAEs) and adverse device events. The Medical Monitor will conduct adjudication of any events of DKA or severe hypoglycemia or any additional events as requested by the Sponsor. The specified events will be adjudicated to determine event relatedness to the study procedures and/or the devices (participant's own insulin pump, study CGM and Omnipod 5), and whether an adverse event is unanticipated. The adjudication decision of the Medical Monitor will be used for the final classification of events, including relatedness to the study procedures and/or the devices, for the determination of safety endpoints and for all regulatory reports, product labeling, and publications or presentations.

6.2 RELATIONSHIP TO STUDY DEVICE AND PROCEDURE

The causal relationship to the study procedures and the investigational device for each adverse event will be rated as follows:

- Unrelated: The event is not related to the procedures or the investigational device.
- Possibly Related: The temporal sequence is such that the relationship is not unlikely or there is no contradicting evidence that can reasonably explain the participant's

condition. There is a possibility of any relation between the event and the procedures or the devices.

- Related: The temporal sequence is relevant, or the event abates upon completion of the procedure/ use of devices, or the event cannot be reasonably explained by the participant's condition or comorbidities. The event is related or most likely associated with the procedures or the devices.
- Causal Relationship (for French sites only): The serious event is associated with the investigational device or with procedures beyond reasonable doubt.

For the purposes of dichotomizing the causal relationship in safety summaries, events that are "Related" or "Possibly Related" (or have a "Causal Relationship", French sites only) will be considered related.

6.3 SEVERITY

The severity of the adverse event will be rated based upon the following grades:

- Mild: Usually transient, requires no special treatment, and does not interfere with the participant's daily activities.
- Moderate: Usually causes a low level of inconvenience, discomfort or concern to the participant and may interfere with daily activities but is usually ameliorated by simple therapeutic measures and participant is able to continue in study.
- Severe: Interrupts a participant's usual daily activities, causes severe discomfort, may cause discontinuation of study device, and generally requires systemic drug therapy or other treatment.

6.4 OUTCOME

The outcome of each reportable adverse event will be classified by the investigator as follows:

- Recovered/Resolved: The participant recovered from the AE/SAE without sequelae.
- Recovered/Resolved with Sequelae: The event persisted and had stabilized without further anticipated change in the event status.
- Fatal: A fatal outcome is defined as the SAE that resulted in death. Only the event that was the cause of death should be reported as fatal. AEs/SAEs that were ongoing at the time of death; however, were not the cause of death, will be recorded as "resolved" at the time of death.
- Not Recovered/Not Resolved (Ongoing): An ongoing AE/SAE is defined as the event was ongoing with an undetermined outcome.

- Unknown: An unknown outcome is defined as an inability to access the participant or the participant's records to determine the outcome (for example, a participant that was lost to follow-up).

6.5 ANALYSIS OF ADVERSE EVENTS

Safety summaries and analyses will be based on all participants that are randomized into the study (i.e., modified Intention to Treat analysis set). All adverse events reported over the course of the study will be summarized and tabulated by study phase (standard therapy vs. randomized period), event category, seriousness, severity, and relationship to the study procedures and the investigational device. Except where indicated, a participant reporting the same adverse event more than once will be counted once when calculating the number and percentage of participants with that particular event.

Adverse events leading to death or to discontinuation from the study will be listed separately. A listing of all adverse events will be provided. Adverse events for participants excluded from mITT analysis set will be listed separately.

6.6 EVALUATION OF DEVICE DEFICIENCIES

Device deficiencies will be tabulated and listed in a manner similar to the methods described for adverse events. Any device deficiency leading to an adverse event or study termination will be listed separately.

7 GENERAL STATISTICAL CONSIDERATIONS

7.1 ANALYSIS WINDOWS

7.1.1 Sensor-Derived Endpoints

Participants are required to wear a Dexcom G6 sensor for 14 days prior to randomization to assess eligibility ("standard therapy" period). Current Dexcom G6 users at US sites can be exempt from this assessment period if they are willing to provide 14 days of data from their personal device from the 30 days prior to randomization.

During the 13-week randomized period, CGM data are collected continuously for both treatment arms. The period for CGM data collection is defined as follows:

Period	Start Date/Time	End Date Time
13-week randomized period	For Intervention group: • Earlier of the first non-error CGM reading using Omnipod 5, or Date/Time participant entered automated mode at Visit 3 For Control group: • Noon on the day of Visit 3	For Intervention group: • Earlier of (1) the last non-error CGM reading using Omnipod 5, on or prior to the date the Controller was returned to the site, or (2) the last non-error CGM reading using Omnipod 5 on the date of study exit visit For Control group: • The last CGM reading on the date of study exit visit

7.1.2 Other Endpoints

For A1C, PROs, and BMI, baseline is defined as the value collected prior to randomization (at Visit 1 or 3). The analysis window for the 13-week timepoint is as follows:

Timepoint	Target Date	Analysis Window	
		A1c/BMI	PROs
13 Week	Study day 90	± 2 weeks	± 5 weeks

If a measurement was collected within analysis window, and has been attributed to Visit 8 (13 weeks), then that measurement will be used in the analysis of the endpoint. Otherwise, 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.

7.2 ANALYSIS SETS

The following analysis sets are planned for the study (Period 1 only):

7.2.1 ITT (Intention to Treat) Analysis Set

The ITT analysis set includes all participants that are enrolled in the study.

7.2.2 mITT (modified Intention to Treat) Analysis Set

The modified Intention to Treat (mITT) analysis set is a subset of the ITT analysis set. The mITT analysis set will consist of all participants who have been randomized. The mITT analysis set will be used as the primary analysis for the primary and secondary endpoints, as well as safety analyses, and for other clinical outcome data.

7.2.3 PP (Per Protocol) Analysis Set

The Per-Protocol (PP) analysis set is a subset of the mITT analysis set. Participants will be included in the PP analysis set if they have a minimum of 80% system use as measured by the study CGM (control group) or Omnipod 5 (intervention group) during the 13 weeks over a minimum duration of 10 weeks and have completed the study without major protocol deviations. The PP analysis set will be used as supportive analysis for the endpoints. The following will be considered major protocol deviations:

- Major inclusion/exclusion criterion deviation
- Significant protocol non-compliance that may confound the study objective data (e.g., use of prohibited medications, not using control group pump for prolonged period)

The list of participants excluded from the PP analysis set will be determined prior to analysis. If the PP analysis set does not differ from the mITT analysis set, separate analyses will not be presented. If the results based on the PP analysis set different from mITT analysis differ, the per-protocol analysis should be interpreted with caution.

7.3 CONTROL OF SYSTEMATIC BIAS

Several measures are incorporated into the study design to help minimize study bias as follows:

1) This is a randomized, multi-center study to help ensure that investigator or site or participant enrollment bias is minimized. Selection of participants will be made from the Investigator's usual participant load. Consecutively eligible participants should be enrolled into the study.
2) This document specifies appropriate statistical methodology to ensure that bias is minimized.
3) The effectiveness measures will be based on the direct output from the device or are analyzed at a NGSP certified central laboratory, ensuring objective reporting.
4) A Medical Monitor will adjudicate all SAEs and device-related adverse events to ensure consistency across all evets; the Medical Monitor's assessment of adverse events will be used for regulatory reports, product labeling, and publications or presentation.

7.4 CALCULATION OF PERCENTAGE OF TIME

Several effectiveness endpoints involve calculation of percentage of time in a specific glycemic range. These endpoints will be based on the direct output from a device (either a CGM or Omnipod 5). The percentage of time in a specified range (e.g., TIR) will be calculated per participant as follows:

$$100 \times \frac{\#\ of\ CGM\ readings\ in\ a\ specified\ range}{\#\ of\ evaluable\ CGM\ readings}$$

The following CGM readings will be excluded from analysis, and therefore, from calculation of TIR endpoints:

- No glucose value is provided in the device output, such as due to an error or device deficiency during which the device does not record glucose readings
- CGM readings reported prior to start of the applicable period, or after the end of the applicable period (e.g., no readings prior to start of the randomized period will be included in the analysis of TIR during randomized period)

At least seven days' worth of sensor readings (168 hours, 2,016 readings) is required to calculate CGM metrics for a period. At least 126 and 42 hours' worth of sensor readings is required to calculate CGM metrics the daytime and nighttime hours, respectively. If there are insufficient data to calculate CGM metrics, the endpoints will be considered missing.

7.5 LENGTH OF RANDOMIZED PERIOD

For certain analyses, it may be necessary to calculate the length of the randomized period. The length of randomized period is calculated as the difference between the start of the randomized period and the end of the randomized period. The start and end of the randomized period for sensor-derived endpoints for each treatment group is described in Section 7.1.1 of this document. For other analyses and summaries, the length of randomized period is calculated (in days) as date of Visit 8 (Study Exit) minus date of Visit 3, +1.

7.6 MISSING DATA HANDLING (PERIOD 1 ONLY)

7.6.1 Missing Data Handling for Sensor-Derived Endpoints

The primary endpoint of TIR 70-180 mg/dL during the 13-week randomized period will be considered missing for any participant with less than 7 days' worth of sensor readings (equal to 2,016 readings).

Due to the nature of the data and study design, it is anticipated that all randomized participants will contribute at least some sensor data post-randomization. Hence, it is possible that there will be no need for imputation of missing values for these endpoints. All randomized participants (i.e., mITT analysis set) will provide an endpoint value for the primary endpoint and all secondary sensor-derived endpoints as follows:

- If a randomized participant provides at least seven days' worth of sensor readings (2,016 readings) during the randomized period, all the available sensor readings will be collapsed (averaged) for a per-participant value for a specific CGM-derived endpoint.

- If a randomized participant does not provide at least seven days' worth of sensor readings (2,016 readings) during the randomized period, the sensor readings for that participant will be excluded from analysis, and instead, a per-participant value for a specific CGM-derived endpoint will be imputed.

Gaps in sensor data, such as due to loss of signal, will not be imputed as imputing single sensor readings data points is clinically inappropriate and would render single CGM data imputation clinically not meaningful.

Because there are numerous reasons that data may be missing for patients in a clinical trial and the methods for handling missing data makes a number of untestable assumptions, sensitivity analyses will be performed to explore (1) whether results are similar for the primary analysis when using different methods and (2) to evaluate the impact of deviations from missing at random (MAR) and missing completely at random (MCAR) assumptions for the primary endpoint.

7.6.1.1 Missing Data Handling for A1C- and PRO-Derived Endpoints

It is likely that at least some participants do not contribute data for A1c, BMI, or Patient-Reported Outcomes (PROs) (e.g., due to early withdrawal). For the main analysis of such endpoints, missing values for the participants in the mITT analysis set will be imputed as follows:

- If a participant withdrew from the study early and provided an endpoint value collected within 13-week analysis window (Section 7.1), the last known value will be used in the analysis of the endpoint.
- If a participant is confirmed to have withdrawn early from the study due to unsatisfactory treatment or potential loss of efficacy outside of the analysis window (Section 7.1), the change from baseline in a specific endpoint will be assumed to be zero (e.g., no change).
- If a participant does not provide a value for the specific endpoint for other reason (e.g., lost sample, participant lost to follow-up), the endpoint data will be imputed.

This allows for all participants in the mITT analysis set to be included in the analysis of these endpoints.

7.7 PATIENT-REPORTED OUTCOMES

The following questionnaires will be used to evaluate general and disease and disease-specific quality of life, and device usability.

Type 1-Diabetes Distress Scale (T1-DDS)

The T1-DDS is a 28-item self-report instrument. Each item is rated on a 6-point scale from (1) "not a problem" to (6) "a very serious problem." The scale yields an overall distress score based on average responses along the 1-6 scale for all 28 items (range = 1-6). The

scale also yields a score for each of seven subscales based on the average response on all of the items in that subscale (range = 1-6).

Hypoglycemia Confidence Scale (HCS)

The HCS is a nine-item self-report scale that examines the degree to which people with diabetes feel able, secure, and comfortable regarding their ability to stay safe from hypoglycemic-related problems. Each item is rated on a 4-point scale from (1) “not confident at all” to (4) “very confident”. The total score is calculated as the sum of each item divided by the number of items completed.

Diabetes Quality of Life-brief (DQOL-brief)

The DQOL-brief is a 15-item instrument that provides a total quality-of-life score. Each item is rated on a 5-point scale from (1) “Very Dissatisfied” or “All the Time” to (5) “Very Satisfied” or “Never.” Respondents must answer at least ten (two thirds) questions to be included in the analysis. The DQOL total score is calculated by summing the scores for all items and dividing by the numbers of responses answered. Higher scores indicate a more positive quality of life.

Clarke Hypoglycemia Awareness

Clarke Questionnaire is an eight-item questionnaire to assess reduced awareness of hypoglycemia. Responses to each item are scored as “A” or “R” (or “U” which is equal to four “R” responses). Total of four or more “R” responses are considered to indicate reduced awareness. Generally, no missing data imputation will be performed, but it is possible in certain cases to determine whether or not a participant has reduced awareness, even in the presence of missing responses (e.g., participant is missing a single response, but if the participant only has two “R” responses for the remaining seven items, it can be concluded that the participant does not have reduced awareness since adding another “R” response would result in fewer than four “R”s).

EQ-5D-3L

The EQ-5D-3L is a short, generic, validated quality-of-life instrument used to measure health outcomes. It consists of five dimensions (questions) regarding mobility, self-care, usual activities, pain/discomfort and anxiety/depression, and one visual analog scale to measure the participant’s self-rated health. Each of the five dimensions is divided into three levels of perceived problems: Level 1 indicating no problem, Level 2 indicating some problems, and Level 3 indicating extreme problems. Responses to five dimensions can be converted into a single summary index score.

The EQ VAS records the participant's self-rated health on a 0 to 100 scale. Missing values in the EQ VAS should be coded as "999" and ambiguous values (e.g., the line crosses the scale twice) should be treated as missing values.

As the study is conducted in US and France, the US participants will be analyzed based on US value set, and the France participants will be analyzed based on France value set.

DAWN Impact of Diabetes Profile (DIDP)

The DIDP is a six-item questionnaire soliciting responses regarding the impact of diabetes on certain aspects of life. Seven response options range from left to right as "Very negative impact" (score=7) to "Very positive impact" (score=1). Response option "Not Applicable" will be treated as missing. Up to two missing or "N/A" responses can be tolerated. Item scores can be reported independently. The composite score is calculated by summing individual items and diving by the number of completed items (therefore, excluding "N/A" or missing responses). The composite score is converted into a percentage score by subtracting 1 from the composite score, dividing by 6, and multiplying by 100. Lower scores indicate greater positive impact across different aspects.

Pittsburgh Sleep Quality Index (PSQI)

The PSQI is used to measure the quality and patterns of sleep, by measuring seven domains: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction over the last month. Responses are converted to a 0 to 3 scale, with higher scores indicating more negative quality of sleep. A global sum of ≥5 indicates a "poor" sleeper. The PSQI global score is calculated as the sum of the seven component scores.

System Usability Scale (SUS)

The SUS is a 10-item questionnaire constructed on a 5-point Likert scale, with responses ranging from (1) "Strongly disagree" to (5) "Strongly agree". It is not recommended that the individual items be scored, but rather a single composite measure of the overall usability of the system should be reported. The SUS score ranges from 0 to 100, and is calculated as follows:

1. For items 1, 3, 5, 7, and 9, subtract 1 from the reported score
2. For items 2, 4, 6, 8, and 10, the contribution equals 5 minus the reported score
3. Sum the contributions (each will range from 0 to 4), multiply by 2.5

Insulin Dosing Systems: Perceptions, Ideas, Reflections, and Expectations (INSPIRE)

The INSPIRE for adults is a 22-item validated questionnaire to evaluate the impact of automated insulin dosing (AID) systems on the psychosocial functioning and quality of life (QoL) of individuals with Type 1 diabetes. Each item is rated on a six-point scale, from strongly disagree (score=0) or not applicable, to strongly agree (score=4). Mean score is calculated based on answered responses, then multiplied by 25. The final score will range from 0 to 100, with higher scores indicating more positive appraisal of the AID.

For validated questionnaires, the prescribed scoring algorithm will be followed. The results will be presented using summary statistics. Scores will be compared between treatment groups using the same methods described above for the secondary and exploratory endpoints.

7.8 OTHER DATA SUMMARIES

The distribution of each baseline characteristic or demographic parameter of interest (such as age, gender, medical history, etc.) will be presented. Data on all enrolled participants and participants included in mITT analysis set will be presented. Continuous variables will be summarized using count, mean, median, standard deviation, and range. Categorical variables will be summarized using counts and percentages.

7.9 STATISTICAL SOFTWARE

The statistical software package SAS® 9.4 or later will be used for all the data derivations, summarization, data listings and statistical analyses. Additional statistical software may be used for graphics or validation purposes as appropriate.

7.10 PERIODIC REVIEW OF CONTROLLER DATA

CGM and Omnipod 5 device data will be made available to Sponsor representatives for periodic reviews. Reviews of the controller data will be conducted to identify any anomalies and to confirm the algorithm appears to be functioning as expected. These reviews, including any findings, will be documented. Unless potential safety concerns are identified, the results of such reviews will not be shared with study sites or participants and thus, will not impact study endpoints. Should a potential safety concern be identified, sites may be notified as appropriate to ensure participant safety and measures may be taken to address the issue such as, but not limited to, changes to the protocol and/or device. Any resulting device deficiencies will be reported, if applicable. A brief summary of the controller data reviews will be included in the final clinical study report.

8 REFERENCES

[1] L. Fisher, D. Hessler, W. Polonsky, L. Strycker, U. Masharani and A. Peters, "Diabetes distress in adults with type 1 diabetes: Prevalence, incidence and change over time," *J Diabetes Complications,* vol. 30, no. 6, pp. 1123-8, 2016.

[2] Y. Mouelhi, E. Jouve, C. Castelli and S. Gentile, "How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods," *Health and Quality of Life Outcomes,* Vols. 18, 136, no. 136, pp. 1-17, 2020.

[3] T. H. P. Draak, B. T. A. de Greef, C. G. Faber and I. S. J. Merkies, "The minimum clinically important difference: which direction to take," *Eur J Neurol,* vol. 26, no. 6, pp. 850-855, 2019.