

**Medtronic****Statistical Analysis Plan**

Clinical Investigation Plan Title	SMILE: S tudy of M iniMed™ 640G I nsulin Pump with SmartGuard™ in prevention of L ow Glucose E vents in adults with Type 1 diabetes
Clinical Investigation Plan Identifier	CEP 311
Sponsor/Local Sponsor	Medtronic International Trading Sàrl. ("Medtronic") Route du Molliat 31 1131 Tolochenaz, Switzerland Medtronic of Canada 99 Hereford St. Brampton ON L6Y 0R3, Canada
Document Version	version 2.0, 20 February 2018
Confidentiality Statement The information contained in this document is confidential and the proprietary property of Medtronic. Any distribution, copying, or disclosure without the prior written authorization of Medtronic is strictly prohibited. Persons to whom the information is disclosed must know that it is confidential and that it may not be further disclosed by them.	

Table of Contents

1.	Version History	4
2.	List of Abbreviations and Definition of Terms	5
3.	Introduction.....	6
4.	Study Objectives.....	7
5.	Investigation Plan	7
6.	Determination of Sample Size	8
6.1.	Statistical Hypothesis	8
6.2.	Sample size.....	9
7.	Statistical Methods	9
7.1.	Study Subjects	9
7.1.1.	Disposition of Subjects.....	9
7.1.2.	Per Protocol Criteria	9
7.1.3.	Analysis Sets	10
7.1.3.1.	Efficacy analysis set.....	10
7.1.3.2.	Per Protocol set.....	11
7.1.3.3.	Safety set.....	11
7.2.	General Methodology	11
7.3.	Center Pooling.....	11
7.4.	Handling of Missing Data and Dropouts.....	12
7.5.	Adjustments for Multiple Comparisons.....	13
7.6.	Demographic and Other Baseline Characteristics	13
7.7.	Treatment Characteristics	14
7.7.1.	Sensor compliance information for TREATMENT Arm: Sensor wear/usage from Visit 6 to visit 15	14
7.7.2.	Selection of sensor reading and sensor wear/usage calculation during the 3 periods of 2 weeks for between arm comparisons.	15
7.7.3.	Pump wear (Treatment and Control Arm)	16
7.8.	Interim Analyses.....	17
7.9.	Evaluation of Objectives	17
7.9.1.	Analysis of primary endpoint	17
7.9.2.	Analysis of secondary endpoints	19
7.9.2.1.	Number of sensor glucose hypoglycemic events for several thresholds	19
7.9.2.2.	Mean duration of hypoglycemic excursions ≤ 55 mg/dL and < 54 mg/dL.....	19

7.9.2.3.	Time spent per day and daily area under the curve (AUC) of sensor glucose values below several thresholds	19
7.9.2.4.	Time spent per day and daily AUC of sensor glucose values above several thresholds 20	
7.9.2.5.	Time spent of sensor glucose values within several ranges.....	20
7.9.2.6.	Glycemic variability (measure from CGM)	21
7.9.2.7.	HbA1c.....	21
7.9.2.8.	Sensor glucose values surrounding SmartGuard triggered insulin suspensions (Treatment arm only).....	21
7.9.2.9.	Proportion of Suspend before low events that reached the pre-set Low Limit (Treatment arm)	21
7.9.3.	Analysis of tertiary endpoints.....	22
7.9.3.1.	Mean number of SMBG	22
7.9.3.2.	Questionnaires' scores	22
	<i>Gold Questionnaire</i>	22
	<i>Hypoglycemia Fear Survey (HFS)</i>	22
7.9.3.3.	Total insulin dose	24
7.9.3.4.	HbA1c level and number of sensor hypoglycemia events per week by subgroups 24	
7.9.3.5.	Mean weight.....	24
7.10.	Safety Evaluation.....	25
7.10.1.	Number of severe hypoglycemic events	25
7.10.2.	Number of severe hyperglycemic events	26
7.10.3.	Number of diabetic ketoacidosis events	26
7.10.4.	Number of Serious Adverse Events, Serious Adverse Device Effects and Device Deficiencies 26	
7.11.	Health Outcomes Analyses.....	27
7.12.	Changes to Planned Analysis.....	27
8.	Validation Requirements	28
9.	References	28

1. Version History

Version	Summary of Changes	Author(s)/Title
1.0	<ul style="list-style-type: none"> New Document 	<div>████████████████████</div> <div>████████</div>
2.0	<ul style="list-style-type: none"> For secondary endpoints related to hypoglycemic regions using CGM values, the StAP now only uses '<' consistently Adds explanation on how to handle dates collected at baseline with missing day and/or month Adds descriptive analysis of A1c by type of insulin (group 1: Lispro (Humalog), group 2: Aspart (Novolog), group 3: Glulisine (Apidra) and group 4: Other). This was prespecified in CIP. Clarifies how the descriptive analysis of the PLGM will be performed Per protocol criteria section is aligns with CIP recommended versus mandatory criteria. 	<div>████████████████████</div> <div>████████</div>

2. List of Abbreviations and Definition of Terms

Abbreviation	Definition
AE	Adverse Event
AUC	Area under the curve
CIP	Clinical Investigation Plan
CGM	Continuous Glucose Monitoring
CSII	Continuous Subcutaneous Insulin Infusion
DMC	Data Monitoring Committee
DTSQ	Diabetes Treatment Satisfaction Questionnaire
eCRF	electronic Case Report Forms
HFS	Hypoglycemia Fear Survey
ITT	Intent To Treat
LGS	Low Glucose Suspend
MAGE	Mean amplitude of glycemic excursions
MNSHE	Mean Number of Sensor Glucose Hypoglycemic Events
PP	Per Protocol
PLGM	Predictive Low Glucose Management
SAE	Severe Adverse Event
SAP	Sensor-Augmented Pump
SMBG	Self Monitoring Blood Glucose
StAP	Statistical Analysis Plan

3. Introduction

Managing type 1 diabetes requires constant vigilance and attention to diet, exercise, and insulin regimens, and depends on consistently delivering the right amount of insulin at the right time. Medtronic's insulin delivery systems have been helping type 1 diabetes subjects with their diabetes management for more than 30 years. The insulin pump system is indicated for the continuous delivery of insulin, at set and variable rates, for the management of diabetes mellitus in persons requiring insulin. Technological advances in insulin pump delivery systems and key clinical trials have established the value of this therapy in helping subjects with their daily diabetes management.

First improvement was real-time continuous glucose monitoring (CGM) data, used in conjunction with insulin pump therapy, allowing subjects to safely reduce their HbA1c values in comparison with multiple daily injection (MDI) therapy and continuous subcutaneous insulin infusion (CSII) alone by using CGM data to guide them in their insulin dosing regimen. Thereafter a feature was introduced, called Low Glucose Suspend (LGS), allowing the pump to automatically suspend insulin delivery when a pre-set sensor glucose threshold was reached. When this LGS feature is enabled in a Paradigm® Veo™ pump and a sensor glucose value at or below the pre-specified value is detected, the pump will stop insulin delivery for up to 2 hours, if not restarted earlier by the user. The benefits of sensor-augmented pump (SAP) therapy using this feature have been proven in several clinical trials, showing reduction in the incidence, severity and duration of hypoglycemic events. In addition, also data analyses based on routine real world use of pumps with the LGS feature confirm significant reductions in hypoglycemia.

The most recent advancement in SAP therapy is SmartGuard™. Medtronic's SmartGuard technology is implemented in the recently commercialized MiniMed™ 640G system and consists of the algorithms called, suspend on low, and suspend before low. The suspend on low is similar to the LGS algorithm in the Paradigm Veo pump and the Suspend before low feature, also called Predictive Low Glucose Management (PLGM), is the newest sensor algorithm that can suspend insulin delivery when the sensor glucose value is predicted to reach or fall below a pre-set low glucose limit within 30 minutes. In addition, the Suspend before low algorithm allows for automatic resumption of basal insulin delivery after 30-120 minutes, if the system predicts subject's recovery from hypoglycemia. During the suspension, insulin delivery can be manually resumed by the user at any time, and settings and alerts are customizable within multiple time blocks depending on the subject's needs in the MiniMed™640G system.

Hypoglycemia is an important barrier in achieving tight glycemic control. Even mild events of hypoglycemia have been shown to disrupt sleep, impair memory, and cause inflammatory reactions, with impacts on driving, work and daily living. Approximately one-third of subjects with type 1 diabetes have evidence of impaired hypoglycemia awareness, further increasing their risk of severe hypoglycemia. However, avoidance of hypoglycemia in these subjects has been shown to restore awareness in small-scale proof of concept studies. Until now, most studies of CSII and CGM therapy have excluded these subjects, and therefore, there are few studies to demonstrate if these technologies can provide an improvement specifically in this population.

The SMILE study will specifically recruit subjects with impaired awareness of hypoglycemia and aim to investigate if this technology can reduce the incidence of hypoglycemia in this group of high-risk subjects.

Despite the available clinical evidence, as described above, SAP therapy is not widely adopted, partly because CGM is not yet reimbursed in many countries. The current clinical evidence on the ability of SAP to reduce the risk of hypoglycemic events could be considered as not robust enough (e.g. short study duration, small number of subjects, not the right comparator, sub-optimal randomization process). Additional clinical evidence on the ability of SAP with SmartGuard technology to reduce the risk of hypoglycemic events can potentially lead to wider adoption and reimbursement of CGM technology to benefit subjects' needs.

This Statistical Analysis Plan (StAP) documents, before data is analyzed, the planning of final analyses that will minimally be included in study reports for the SMILE Study. This StAP is based on Version 1.0 (14 March 2016) of the SMILE Study Clinical Investigational Plan (CIP).

4. Study Objectives

Primary Objective: The primary study objective is to demonstrate a reduction in the mean number of hypoglycemic events when using the MiniMed 640G system with the SmartGuard algorithm:

The between arm difference in the incidence of hypoglycemic events below or equal to 55 mg/dL (3.0 mmol/L) during 6 months of SAP therapy (MiniMed 640G with SmartGuard), as compared to subjects on CSII therapy (MiniMed 604G) over the same period of time, in a population of T1 diabetic subjects with an increased risk of hypoglycemia, will be evaluated. A reduction in the mean number of hypoglycemic events below or equal to the threshold of 55 mg/dL (3.0 mmol/L) is estimated to be of clinical value. The evaluation will be made by comparing subject sensor data collected in both arms. The primary endpoint is the mean number of sensor glucose hypoglycemic events (MNSHE) per subject per week.

5. Investigation Plan

This is an international, multi-center, prospective, open-label, adaptive, randomized, controlled, pre-market clinical investigation with parallel arms. The study is designed to compare CGM-based hypoglycemic events in a treatment arm versus a control arm.

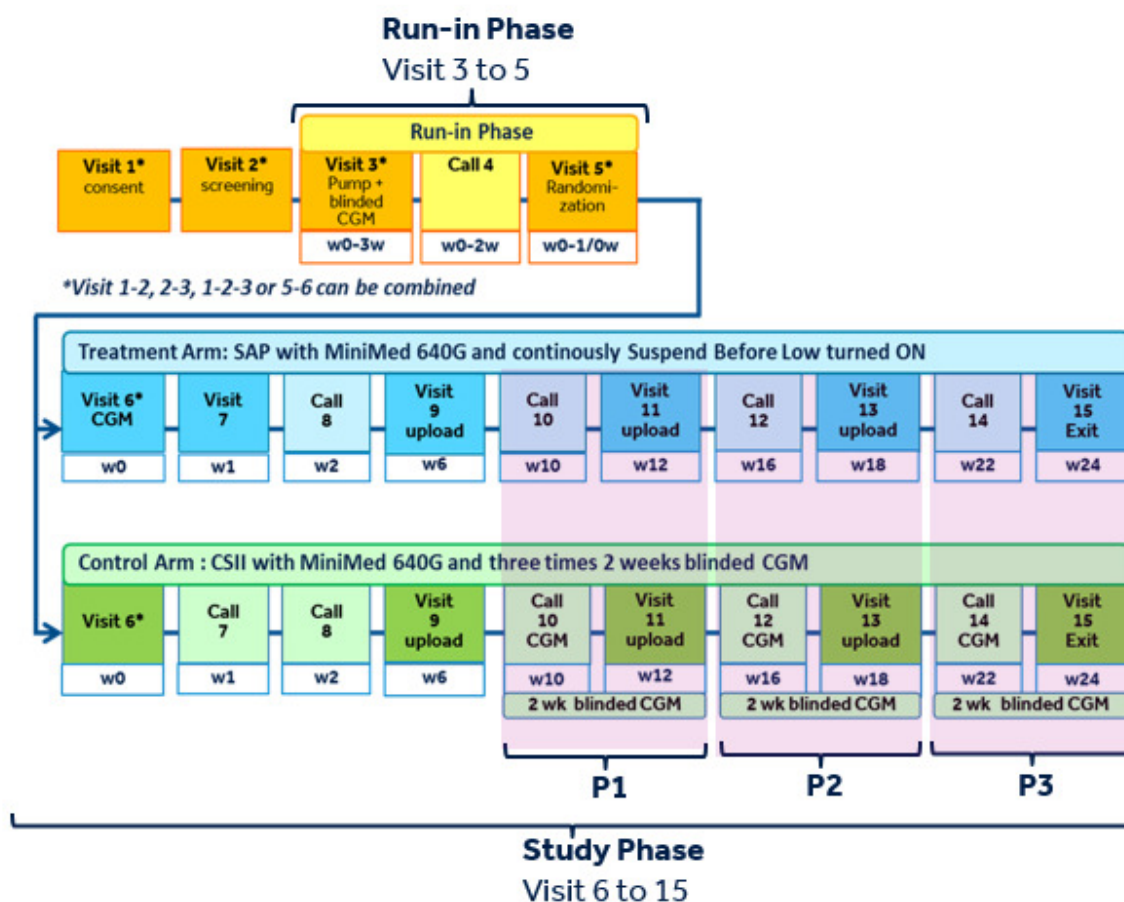
Enrolled subjects will enter a 2 weeks' run-in period, receive training and start CSII therapy with the MiniMed 640G insulin pump and usage of blinded Continuous Glucose Monitoring (CGM). At the end of the run-in period, eligible subjects that meet the randomization criteria will be randomized into the treatment phase of six (Figure 1) months in a 1:1 ratio to either:

- Treatment Arm: training and start of Sensor Augmented Pump therapy with Suspend before Low feature of SmartGuard turned ON.
- Control Arm: continuation of CSII therapy with blinded CGM usage for a total of 6 weeks during the treatment phase.

Sensor data is collected at 3 periods of 2 weeks. Period 1 starting at week 10 corresponds to 3 months after randomization, Period 2 starting at week 16 corresponds to 4.5 months after randomization and Period

3 starting at week 22 corresponds to 6 months after randomization. This study has an adaptive design. Due to the uncertainty about the hypoglycemic event rate and the magnitude of the true effect of treatment, the sample size will be re-estimated by an independent Data Monitoring Committee (DMC) in an interim assessment based on conditional power after a total of at least 60 subjects have been randomized and followed up for 6 months. Details of the interim analysis and adaptive design are available in the DMC Analysis Plan.

Figure 1: Study design overview



6. Determination of Sample Size

6.1. Statistical Hypothesis

There are 3 periods of 2 weeks in which sensor data is collected for both Treatment and Control arms. The mean number of sensor glucose hypoglycemic events (MNSHE) below or equal to 55 mg/dL (3.0 mmol/L) per subject/week will be calculated at each period using the sensor data collected in the TREATMENT arm (SAP with MiniMed 640G and SmartGuard) and the CONTROL arm (CSII with MiniMed 640G).

The overall MNSHE will be calculated as the average of P1, P2 and P3. The overall between group difference in MNSHE per week is of primary interest and will be compared using the following hypotheses:

Null-hypothesis: $H_0: \text{MNSHE}_{\text{TREATMENT}} = \text{MNSHE}_{\text{CONTROL}}$

Alternative hypothesis: $H_A: \text{MNSHE}_{\text{TREATMENT}} < \text{MNSHE}_{\text{CONTROL}}$

6.2. Sample size

The sample size calculation was performed based on assumptions from the results of the *ASPIRE study* (Bergental, et al., 2013). In particular, the mean number of sensor glucose hypoglycemic events below or equal to 55 mg/dL (3.0 mmol/L) per subject/week is expected to be 2 with standard deviation of 1.65 in the Control arm (CSII therapy alone with MiniMed 640G).

Assuming a two-sided two-sample t-test, $\alpha=0.05$, power=80% and 40% reduction in the mean number of sensor glucose hypoglycemic events per subject/week in the Treatment arm (SAP with MiniMed 640G and SmartGuard), as compared to the control arm, the required sample size is 136 subjects (68 in each arm) completing the 6 month follow up of the study phase.

The dropout rate during the 6-month follow-up after randomization has been estimated to be around 15%. Incorporating this dropout rate, a total of 160 subjects will be randomized in a 1:1 ratio either to the SAP therapy arm (MiniMed 640G with SmartGuard) or to the CSII therapy arm (MiniMed 640G).

7. Statistical Methods

7.1. Study Subjects

7.1.1. Disposition of Subjects

Number of enrolled subjects and number of completed visits will be reported. An overview of study discontinuations will be provided. A description of the total numbers of subjects screened, randomized, per study arm, and a list of reasons for failed screening, dropouts number and corresponding reasons for dropping out will be reported.

A CONSORT flow diagram (Begg, Cho, & Eastwood, 1996), displaying the progress of all participants through the trial (Run-In phase: from Visit 1 to Visit 5 included, and Study phase: after Visit 5 to Visit 15 included) will report the number of subjects that contributed to primary objective analysis.

7.1.2. Per Protocol Criteria

Below there is a list of criteria that will be used to identify subject's data for the per protocol sensitivity analysis of the primary endpoint as described in section 7.1.3.2. Occurrence of each deviation will be

reported and occurrence of any deviation will result in exclusion of the subject from the per protocol analysis.

- Only for CONTROL arm subjects:
 - SmartGuard features should be turned OFF 100% of the time. To account for possible short time setting errors, a maximum of 1 day (0.55%) of the study period with SmartGuard features turned ON will be allowed, otherwise the subject's data will be excluded from PP analysis.
- Only for TREATMENT arm subjects:
 - Continuous unblinded sensor use is expected > 70% of the 6 months of the study phase, otherwise subject's data is excluded from the per protocol analysis.
 - While in sensor use, it is expected to have > 85% of time with Suspend before Low turned ON, otherwise subject's data is excluded from the per protocol analysis.
 - While in sensor use, exclude from Per Protocol if the low limit is outside the range (< 55mg/dL or >70mg/dL) for >50% of time.
- For both CONTROL and TREATMENT arms:
 - Pump usage/wear: In both arms, pump wear is expected to be a minimum of 80% of the 6 months study phase time. Exclude from PP analysis if pump wear is less than 80% of the time (this includes both, pump data loss and days with Total Daily Dose =0).
 - Average number of SMBG per day should be >2; otherwise exclude from PP analysis.
 - Sensor usage during the 3 periods of 2 weeks (starting at weeks 10, 16 and 22), for the assessment of the primary endpoint, will be monitored for each subject. Combining the three periods, a total of 6 weeks of sensor data is expected but a subject's data is excluded from PP analysis if a subject has less than 3 weeks of sensor data when the 3 periods are combined (this is equivalent to <50% of the time for P1+P2+P3). The total number of subjects with less than 3 weeks of sensor will be reported.
 - Exclude from PP analysis if a subject drops out of the study phase of 6 months. Number of subjects dropping out of the study phase will be reported
 - The length of follow-up, from visit 6 to visit 15, should be within a window of 161 days to 182 days (calculated from the date of visit 6, 168-7 days study phase duration or 168+14 days study phase duration). If the length of follow-up is less than 161 days, exclude from PP analysis.

7.1.3. Analysis Sets

7.1.3.1. Efficacy analysis set

For the primary endpoint, efficacy analyses will be performed in the Intent to Treat (ITT) basis. The ITT set will be composed of all randomized subjects. These subjects will be assessed and analyzed as members of the intended randomized arm, irrespective of their compliance to the planned course of treatment or deviations from protocol.

7.1.3.2. Per Protocol set

For sensitivity purposes, efficacy analysis of the primary endpoint will also be performed on the Per Protocol set. The Per Protocol (PP) set is defined as the set of successfully randomized subjects with no violation of entry criteria and who are compliant with the protocol. The PP set will include subjects who do not have any deviations as listed in section 7.1.2.

7.1.3.3. Safety set

Safety analysis will include all subjects with signed informed consent and safety data will be presented by phase (run-in phase and study phase). Safety reporting will be done separately for the run-in period and the study phase period. Any AE and SAE occurring from the date of Visit 1 to before the date of Visit 5 will be reported as happening in the run-in phase. Any AE and SAE occurring from the date of visit 5 to the date of Visit 15 will be reported as happening in the study phase and will be reported according to the randomization arm that the subject was assigned.

For comparison of safety in the two arms, the safety set is composed of all randomized subjects, and assessed and analyzed as members of the intended randomized arm.

7.2. General Methodology

Number of enrolled patients and number of completed visits will be reported. An overview of study discontinuations will be provided. By phase (run-in phase and study phase), a description of the total number of patients screened, randomized, per study arm, and a list of reasons for failed screening, dropouts number and corresponding reasons for dropping out will be reported.

Summary statistics such as n (number of subjects), mean, standard deviation, median, minimum and maximum will be used for continuous variables. For categorical variables, summary statistics such as n, frequency counts and percentages will be used.

The templates for Tables, Listings and Figures (TLFs) will be found in the TLFs document.

7.3. Center Pooling

At all investigational centers, the same clinical investigational plan and training is followed and standardized data collection methodology and electronic case report forms (eCRF) are used. Results for the primary endpoint will be reported by center and country using a format similar to the table below. In particular, values in the sixth column (between groups difference of averages) will allow for examination of centers with outlying results. Grouping of sites based in geographies may also be reported. Available data will be used to populate the table and in case of missing data, no imputation will be performed. Centers will be pooled for analyses. No statistical test will be performed.

Country	Center	Number of patients	Treatment Arm Average (STD), of average of P1, P2 and P3 number of sensor glucose hypoglycemic events per subject per week	Control Arm Average (STD), of average of P1, P2 and P3 number of sensor glucose hypoglycemic events per subject per week	Between group difference Average of average of P1, P2 and P3 number of sensor glucose hypoglycemic events per subject per week (Treatment Arm – Control Arm)
xxxx	1	xx	xx.x (xx.xx)	xx.x (xx.xx)	xx.xx
xxxx	2	xx	xx.x (xx.xx)	xx.x (xx.xx)	xx.xx
	...				
Overall for country xxxx		xx	xx.x (xx.xx)	xx.x (xx.xx)	xx.xx
...					
Total		xxx	xx.x (xx.xx)	xx.x (xx.xx)	xx.xx

7.4. Handling of Missing Data and Dropouts

Efficacy analyses of the primary endpoint will be performed in the Intent to Treat (ITT) basis. The analysis will be based on a repeated measures random effects model (section 7.9.1) with PROC MIXED that uses all available data and accounts for possible missing at random data.

Efficacy analysis of other secondary endpoints where hypothesis testing may be performed will use random mixed models. For exploratory analysis of other endpoints without hypothesis testing, available data without imputation of missing values will be used.

In case of dates collected at baseline with missing day and/or month the next procedure will be applied. A missing month will be imputed using the month 'July' and a missing day will be imputed with the day '15'. This applies only to dates related to medical history and baseline information with missing month and/or day.

For tables and listings of safety data a conservative/worst case scenario approach will be taken in case of partially missing dates. For date of discharge with missing day, the day will be set to last day of that month. For date of admission with missing day, the day will be set to first day of that month. If that is before day of start of study phase (within the same month), then day will be set equal to day of start of study phase. And if it is before day of start of run-in, then it will be set to day of start of run-in. Thus, duration is set to be as long as possible and by preference after intervention.

7.5. Adjustments for Multiple Comparisons

The following hierarchical test procedure reflects the relative importance of the endpoints and controls for multiplicity.

- Primary study endpoint

The mean number of sensor glucose hypoglycemic events ≤ 55 mg/dL will be tested for superiority as described in section 7.9.1 and a p-value < 0.05 will be considered statistically significant.

- Fixed sequential testing of selected secondary endpoints

For the following endpoints, the procedure test hierarchically the ordered hypotheses in sequence at level $\alpha=0.05$ until a first hypothesis is non-rejected.

1. Mean Time spend ≤ 55 mg/dL
Superiority test, if p-value ≤ 0.05 reject null hypothesis and continue, else stop
2. Mean Time in target 70-180mg/dL
Non-inferiority test with non-inferiority margin of 40 minutes, if p-value ≤ 0.05 reject null hypothesis and continue, else stop
3. Mean Time in target 70-180mg/dL
Superiority test, if p-value ≤ 0.05 reject null hypothesis.

- Exploratory analysis

For the endpoints listed below and for all other endpoints, hypothesis testing may be performed and p-values will be reported but may not be claimed

- The mean number of sensor glucose hypoglycemic events < 54 mg/dL
- Mean Time spend < 54 mg/dL
- HFS score (Superiority test, exploratory only)
- DTSQs and DTSQc (Superiority test, exploratory only)
- Mean Time spend > 180 mg/dL (Non-inferiority test with non-inferiority margin of 40 minutes, exploratory only)
- Mean duration of hypoglycemic excursions < 54 mg/dL (Superiority test, exploratory only)

7.6. Demographic and Other Baseline Characteristics

A summary of basic subject demographics and baseline characteristics will be included in the study report using appropriate summary statistics (n, mean, standard deviation, median, minimum and maximum for continuous variables; n, frequency counts and percentages for categorical variables).

Baseline data will be obtained from visit 2 (CRF: demographics, medical history, HbA1c, Creatinine, Clark and Gold score) and visit 5 (baseline sensor data: CareLink clinical upload).

Although this is a randomized trial, there may be baseline differences between the two groups just by chance. Baseline characteristics for all patients and by treatment group will be reported.

7.7. Treatment Characteristics

7.7.1. Sensor compliance information for TREATMENT Arm: Sensor wear/usage from Visit 6 to visit 15

For the Treatment Arm, the amount of sensor wear (%) will be reported. The maximum sensor wear per subject is defined as the number of days from the start of CGM until end of study (from Visit 6 to Visit 15~168 days [24 weeks]), multiplied by the maximum number of sensor readings per day (288 readings per day). Even with continuous sensor wear, we expect that no sensor reading will be obtained due to replacement/changes of the sensors, setting up of transmitter and so on. To account for this, we assume that a maximum of 6 hours per week will have no sensor readings, which means that for the 168 days of CGM use there will be 6 days of expected no sensor readings.

When sensor is wear continuously, the total expected number of sensor reading for subject i is calculated as:

(Number of days from Visit 6 to Visit 15 - 6 days of expected no sensor readings)*288

As each subject will have different number of days in the study phase due to the allowable visit window in the protocol (-7 days to +14 days), the total expected number of sensor readings could be as low as 44640 sensor readings (for 168 - 7 = 161 days study phase duration, with (168-7-6)*288 sensor readings), or as high as 50688 sensor readings (for 168+14=182 days study phase duration, with (168+14-6)*288 sensor readings) when sensor is wear continuously.

The amount of sensor wear (%) for subject i is calculates as:

Sensor wear (%) for subject i =

$$[(\text{Total observed number of sensor reading for subject } i) / (\text{total expected number of sensor reading for subject } i)] * 100$$

There may be subjects that continuously wear sensor and spend less than 6 hours per week replacing/changing sensors, setting up transmitters and so on, and would end up with a calculated sensor wear (%)>100%. In those cases, the sensor wear (%) will take value = 100%.

The total observed number of sensor reading for subject i is calculated as the number of sensor reading in CareLink from the date of Visit 6 to the date of Visit 15 (end of study visit date)

In case a subject drops out of the study before Visit 15, the total observed number of sensor reading for that subject is calculated as the number of sensor reading in CareLink from the date of Visit 6 to date of study drop out.

In case a subject drops out of the study before Visit 15, total expected number of sensor reading for that subject is calculated as: $(168-6)*288= 46656$

7.7.2. Selection of sensor reading and sensor wear/usage calculation during the 3 periods of 2 weeks for between arm comparisons.

The blinded (for control arm) and unblinded (for treatment arm) sensor data collected during the 3 periods of 2 weeks (each period starting at weeks 10, 16 and 22) will be used to assess the between arm difference in the mean number of sensor glucose hypoglycemic events per subject per week.

Sensor wear/usage during the 3 periods of 2 weeks will be monitored for each subject in each period.

The first period (P1) starts the date of Visit 10 and finishes at Visit 11. During this period, the total expected number of sensor readings in CareLink is $14\text{days} * 288 = 4032$. *Only the 4032 sensor readings closer to Visit 11 will be used for analysis, starting from the time of the upload at Visit 11 and going backwards up to the time of the call at Visit 10.*

From Visit 10 to Visit 11 there is a maximum of 21 days (14 days + additional 7 days window). If during this period there are 4032 or more sensor readings for subject i , then the sensor usage in P1 for subject i is equal to 100%. If during this period there are less than 4032 sensor readings, then the sensor usage in P1 for subject i is equal to $(\text{observed number of sensor reading in P1 for subject } i \div 4032) * 100\%$.

Using a similar logic, the second period (P2) starts the date of Visit 12 and finishes at Visit 13. During this period, the total expected number of sensor readings is $14\text{days} * 288 = 4032$. *Only the 4032 sensor readings closer to Visit 13 will be used for analysis, starting from the time of the upload at Visit 13 and going backwards up to the time of the call at Visit 12.*

From Visit 12 to Visit 13 there is a maximum of 21 days (14 days + additional 7 days window). If during this period there are 4032 or more sensor readings for subject i , then the sensor usage in P2 for subject i is equal to 100%. If during this period there are less than 4032 sensor readings, then the sensor usage in P2 for subject i is equal to $(\text{observed number of sensor reading in P2 for subject } i \div 4032) * 100\%$.

For the last 2-weeks period, a similar logic is used. The Third period (P3) starts the date of Visit 14 and finishes at Visit 15. During this period, the total expected number of sensor readings is $14\text{days} * 288 = 4032$. *Only the 4032 sensor readings closer to Visit 15 will be used for analysis, starting from the time of the upload at Visit 15 and going backwards up to the time of the call at Visit 14.*

From Visit 14 to Visit 15 there is a maximum of 21 days (14 days + additional 7 days window). If during this period there are 4032 or more sensor readings for subject i , then the sensor usage in P3 for subject i is equal to 100%. If during this period there are less than 4032 sensor readings, then the sensor usage in P3 for subject i is equal to $(\text{observed number of sensor reading in P3 for subject } i \div 4032) * 100\%$.

In case a subject has less than 3 days of sensor data (< 3*288 sensor points) during one of the periods of 2 weeks, then the sensor data for that subject in that period will be missing.

The mean sensor usage (%) for P1, P2 and P3 will be reported separately for the treatment and Control arms.

The accumulated sensor wear (adding up the 3 periods: P1 + P2 + P3) will also be calculated as:

Accumulated (P1 + P2 + P3) Sensor wear (%) for subject i =

$$100 \left(\frac{a}{(4032 * 3)} \right)$$

Where:

a = (Total observed number of sensor reading for subject i in P1) + (Total observed number of sensor reading for subject i in P2) + (Total observed number of sensor reading for subject i in P3)

If the total observed number of sensor readings (adding up the 3 periods: P1 + P2 + P3) is larger than 4032*3, then the accumulated sensor wear is equal to 100%.

The mean accumulated sensor wear (%) (adding up the 3 periods: P1 + P2 + P3) will be reported separately for the treatment and Control arms.

If a subject drops out of the study before any of the periods, the Total observed number of sensor reading for the missing period will be = 0 and the total expected number of sensor readings will be = 4032

7.7.3. Pump wear (Treatment and Control Arm)

A day of pump wear is a day with any total daily insulin amount (total daily insulin amount > 0), otherwise it is a day of no pump wear (total daily insulin amount = 0).

Given the visit windows, the number of days for a subject completing the study phase can be between 168 - 7 = 161 days and 168+14=182 days.

The pump wear (%) for subject i is calculated from CareLink as:

Pump wear (%) for subject i =

$[(\text{Number of days with total daily insulin amount} > 0, \text{ for subject } i) / (\text{number of days in the study phase for subject } i)] * 100$

This formula will be used for all randomized subjects to assess compliance with pump wear while in the study.

In addition, the compliance with pump wear with respect to the complete expected study phase will be assessed. For this the formula above will be used for subjects completing the study phase and for subjects dropping out during the study phase, the denominator will be = 161 days.

Summary statistics such as n, mean, standard deviation, median, minimum and maximum will be reported for pump wear (%) in both, the Treatment and Control arm.

7.8. Interim Analyses

An interim analysis for sample size re-estimation will be based on the conditional power approach by Li et al. (2002) (Li, Shin, Xie, & Lu, 2002) and a method by Chen et al. (2004) (Chen, DeMets, & Lan, 2004) as extended by Mehta and Pocock (2011) (Mehta & Pocock, 2011). All the details regarding the interim analysis can be found in the Statistical Analysis Plan of the Data Monitoring Committee.

7.9. Evaluation of Objectives

7.9.1. Analysis of primary endpoint

A sensor hypoglycemic event is defined as sensor glucose values less or equal to 55 mg/dL (3.0 mmol/L) for more than 20 consecutive minutes. When the time between two successive events is less than 30 minutes, they will be combined and counted as one event.

The primary endpoint is the mean number of sensor glucose hypoglycemic events (MNSHE) per subject per week computed using the sensor data collected during each of the 3 periods of 2 weeks. As described in section 7.7.2, the last 4032 sensor readings from each period are used for analysis.

The goal is to compare the MNSHE per week in the two study arms. The between group difference in MNSHE for the overall at 3, 4.5 and 6 months (using Periods 1, 2 and 3) is of primary interest and will be compared using the following hypotheses:

Null-hypothesis: $H_0: \text{MNSHE}_{\text{TREATMENT}} = \text{MNSHE}_{\text{CONTROL}}$

Alternative hypothesis: $H_A: \text{MNSHE}_{\text{TREATMENT}} < \text{MNSHE}_{\text{CONTROL}}$

The null-hypothesis will be rejected in favor of the alternative hypothesis if the mean of the estimated regression coefficient β_2, β_4 and β_6 is <0 and the associated p-value for this contrast is less than 0.05 using the mixed model:

$$\begin{aligned} \text{MNSHE}_{ij} = & (\beta_1 + \beta_2 * \text{group}_i) * \text{period1} + \\ & (\beta_3 + \beta_4 * \text{group}_i) * \text{period2} + \\ & (\beta_5 + \beta_6 * \text{group}_i) * \text{period3} + \beta_7 * \text{BLMNSHE}_i + b_{0i} + \epsilon_{ij} \end{aligned}$$

Where:

MNSHE_{ij} is the mean number of sensor glucose hypoglycemic events per week for subject i at period $j=1, 2$ or 3 , BLMNSHE_i is the baseline mean number of sensor glucose hypoglycemic events per week for subject i , and b_{0i} are random effects.

If period = 3 months, then period1= 1, 0 otherwise

If period = 4.5 months, then period2=1, 0 otherwise

If period = 6 months, then period3=1, 0 otherwise

If arm=Treatment, then group=1, 0 otherwise

In case of partially missing baseline mean number of sensor glucose hypoglycemic events (BLMNSHE_i), the method proposed by White and Thompson will be use (Adjusting for partially missing baseline measurements in randomized trials, 2005, Statistics in Medicine).

In each period, it is expected to have 4032 sensor readings, which are equivalent to 14 days of sensor data and represents 100% sensor usage. The mean number of hypoglycemic events per subject per week is calculated for each of the subjects in each of the periods (period1, period2 and period3) using the blinded (for control arm) or unblinded (for treatment arm) sensor data collected as indicated in section 7.7.2. With the available sensor readings, the number of hypoglycemic events will be calculated and expressed as events per week for each period for each subject. For example: If a subject has 1 sensor hypoglycemic events in 14 days (2 weeks) in a particular period, then the number of sensor glucose hypoglycemic events per week for that subject in that period is: (1 hypoglycemic event/ 2 weeks) = 0.5 hypoglycemic events/week. If a subject in a particular period has 2016 sensor readings (representing 50% sensor usage), which are equivalent to 1 week and 1 sensor hypoglycaemic event, then the number of sensor glucose hypoglycemic events per week for that subject in that period is: (1 hypoglycemic events / 1 week)= 1 hypoglycemic event/week.

Selection of sensor data: In case a subject has less than 3 days of sensor data ($< 3*288$ sensor points) during one of the periods of 2 weeks, then the mean number of hypoglycemic events/week for that subject in that period will be missing.

In case the MNSHE shows a Poisson like distribution, the transformation \sqrt{MNSHE} (and $\sqrt{BLMNSHE_i}$) may be considered for the mixed model as a sensitivity analysis.

7.9.2. Analysis of secondary endpoints

The blinded (for the control arm) or unblinded (for the treatment arm) sensor data collected during the 3 periods of 2 weeks will be used. Difference between the treatment arm and the control arm will be analyzed in a similar way as the primary endpoint using a random effects model when hypothesis testing may be performed. If

7.9.2.1. Number of sensor glucose hypoglycemic events for several thresholds

Mean number of sensor glucose hypoglycemic events < 40 mg/dL (2.2 mmol/L), ≤ 55 mg/dL (3.0 mmol/L), < 54 mg/dL and < 70 mg/dL (3.9 mmol/L) per subject per week will be calculated in a similar way as the primary endpoint (see section 7.9.1). Subgroup analysis will be conducted according to the time the events started, subgroup 1: starting at day-time (08:01 to 21:59) and subgroup 2: starting at night-time (22:00 to 08:00).

A hypoglycaemic event is defined as sensor glucose values less or equal (for \leq), or less than (for $<$) the predefined threshold for more than 20 consecutive minutes. When the time between two successive events is less than 30 minutes, they will be combined and counted as one event.

7.9.2.2. Mean duration of hypoglycemic excursions ≤ 55 mg/dL and < 54 mg/dL

A hypoglycemic excursion happens when sensor glucose values are ≤ 55 (3.0 mmol/L) for more than 20 consecutive minutes. For subjects with hypoglycemic excursions, the mean duration of all the excursions in each period will be calculated and used for analysis in a similar way as the primary endpoint. This will also be analyzed for sensor glucose values < 54 mg/dL.

7.9.2.3. Time spent per day and daily area under the curve (AUC) of sensor glucose values below several thresholds

Mean time spent per day (in minutes/day) and mean daily AUC of sensor glucose values < 40 mg/dL (2.2 mmol/L), < 54 mg/dL, ≤ 55 mg/dL (3.0 mmol/L) and < 70 mg/dL (3.9 mmol/L) will be analyzed.

For a given time (in days) with available sensor readings, the time spent per day and daily AUC below a particular threshold will be calculated for subject i in period j as:

- Time spent per day below a particular threshold in in period j =

(Total number of minutes below threshold for subject i in period j) / (number days with CGM data for subject i in period j)

- Daily AUC below a particular threshold in period j =

(Total AUC below threshold for subject i in period j) / (number days with CGM data for subject i in period j)

These will be calculated for each subject and for each period ($j=1, 2$ and 3) for analysis and the mean reported per arm and period

For example, if for a given period a study subject has 10 days of available sensor data (2880 sensor points) of which 0.4 days were spent below a particular threshold, the time spent per day below the threshold will be calculated for that subject in that period as: $0.4/10=0.04$ days 57.6 minutes/day.

Subgroup analysis will be conducted according to the time the excursion started, subgroup 1: starting at day-time (08:01 to 21:59) and subgroup 2: starting at night-time (22:00 to 08:00), and calculated respectively as:

(Total number of minutes below threshold during day-time for subject i in period j) / (number of day-time periods with CGM data for subject i in period j) or

(Total number of minutes below threshold during night-time for subject i in period j) / (number of night-time periods with CGM data for subject i in period j)

7.9.2.4. Time spent per day and daily AUC of sensor glucose values above several thresholds

Mean time spent per day (in minutes/day) and mean daily AUC of sensor glucose values above 180 mg/dL (10.0 mmol/L), above 240 mg/dL (13.3 mmol/L), and above 300 mg/dL (16.7 mmol/L) will be calculated and analyzed in a similar way as described above in 'Time spent per day (in minutes/day) and daily AUC of sensor glucose values below several thresholds' in each period. Subgroup analysis will be conducted according to the time the excursion started, subgroup 1: starting at daytime (08:01 to 21:59) and subgroup 2: starting at night-time (22:00 to 08:00)

For mean time spend > 180mg/dL, also a non-inferiority test will be performed with non-inferiority margin of 40 minutes as pre-specified in section 7.5.

7.9.2.5. Time spent of sensor glucose values within several ranges

Mean time spent of sensor glucose values within range and including 70-140 mg/dL (3.9 -7.8 mmol/L) and 70-180 mg/dL (3.9-10.0 mmol/L) will be calculated for subject i in period j as:

- Time spent per day within particular range =

(Total number of minutes within range for subject i in period j) / (number of days with CGM data for subject i in period j)

For mean time in target 70-180mg/dL also a non-inferiority test will be performed with non-inferiority margin of 40 minutes as pre-specified in section 7.5.

Similar analysis as the described above will be conducted also categorized by starting at day-time (08:01 to 21:59) and night-time (22:00 to 08:00).

7.9.2.6. Glycemic variability (measure from CGM)

CGM measurements will be used to calculate the excursion amplitudes of the glucose values measured by mean amplitude of glycemic excursions (MAGE) and mean 24 hours standard deviation (SD) of glucose values in mg/dL in each period. Between groups difference in the mean MAGE and the mean 24 hours SD will be performed.

7.9.2.7. HbA1c

Mean HbA1c change from baseline to 6 months in each arm will be reported (both for % and mmol/mol) and a between group comparison will be performed. In addition, analysis of change from baseline to 6 months in HbA1c will be performed separately for two groups: a. baseline HbA1c level $\leq 7.5\%$ and b. baseline HbA1c level $> 7.5\%$.

7.9.2.8. Sensor glucose values surrounding SmartGuard triggered insulin suspensions (Treatment arm only)

Mean sensor glucose values within 60 min before SmartGuard triggered insulin suspension (suspend before low) and within 360 min after suspend before low will be calculated for each subject in the Treatment arm and summarized with mean, SD, median, interquartile range, min and max, and number of suspensions for suspensions lasting: a. < 30 , b. ≥ 30 to < 90 , c. ≥ 90 minutes, and d. all suspensions. A graphical representation will be prepared.

Summary statistics will also be reported for two subgroups: starting at day-time (08:01 to 21:59) and starting at night-time (22:00 to 08:00).

7.9.2.9. Proportion of Suspend before low events that reached the pre-set Low Limit (Treatment arm)

The proportion of Suspend before low events that reached the pre-set Low Limit as per the individualized Low Limit settings at any given time point will be calculated. For the Treatment arm it will be summarized with number of subjects contributing data, average number of suspensions per subject, average number of suspends before low events that reached the pre-set Low Limit and mean percentage of suspends before low events that reached the pre-set Low Limit.

7.9.3. Analysis of tertiary endpoints

Descriptive summary statistics such as n, mean, SD, median, interquartile range, min and max, and 95% confidence intervals for quantitative variables and n, percentage for categorical variables will be reported per study arm.

Additional ad-hoc analysis may be performed and will be described as such in the report of study results.

7.9.3.1. Mean number of SMBG

The mean number of self-monitored blood glucose (SMBG) readings and mean of SMBG values will be calculated in each study arm and reported using summary statistics for quantitative variables from the pump data uploaded in CareLink Clinical at visits: 5 (randomization, week=0-1), 6 (week=0), 9 (week=6), 11 (week=12), 13 (week=18) and 15 (week=24).

7.9.3.2. Questionnaires' scores

The Clarke score, the Gold score, the HFS score and Diabetes Treatment Satisfaction Questionnaire (status version, DTSQs and change version DTSQc) scores will be calculated in each study arm and reported using summary statistics for quantitative variables.

In case the subject withdraws from the study early, summary statistics will be reported for the study exit visit.

Clarke Questionnaire

The Clarke method comprises eight questions characterizing the participant's exposure to events of moderate and severe hypoglycemia. It also examines the glycemic threshold for, and symptomatic responses to, hypoglycemia. The score is calculated as the sum of the answers rated with [R]. A score of four or more implies impaired awareness of hypoglycemia.

Summary statistics for the Clarke questionnaire score will be reported at screening (Visit 2), at visit 11 (3 months of follow-up) and at visit 15 (6 months of follow-up), which is also the end of study. In addition, the proportion (%) of subjects with a score ≥ 4 will be reported.

Gold Questionnaire

The Gold questionnaire is a 7-point visual analogue scale to characterize the participant's impaired awareness of hypoglycemia as confirmed by a score of ≥ 4 in the Gold score.

Summary statistics for the Gold question score will be reported at screening (Visit 2), at visit 11 (3 months of follow-up) and at visit 15 (6 months of follow-up), which is also the end of study. In addition, the proportion (%) of subjects with a score ≥ 4 will be reported.

Hypoglycemia Fear Survey (HFS)

Hypoglycemia can lead to various aversive symptomatic, affective, cognitive, physiological, and social consequences, which in turn can lead to the development of possible phobic avoidance behaviours associated with hypoglycemia. The hypoglycemia fear survey (HFS) is a psychometric instrument designed to quantify this fear. The instrument has internal consistency and test-retest stability, and varies with elevated glycosylated haemoglobin. The HFS has in most translations two subscales, the behaviour subscale and the worry subscale and has a recollection period of 6 months.

The two subscales of the HFS are scored as: behavior subscale (first 15 items) items are added together, the 18 Worry subscale items are added together. Additionally, a total score, adding all 33 items, will be computed and reported using summary statistics (n, mean, Std, median).

Summary statistics for the HFS score will be reported at screening (Visit 2), at visit 11 (3 months of follow-up) and at visit 15 (6 months of follow-up), which is also the end of study.

DTSQs and DTSQc

The Diabetes Treatment Satisfaction Questionnaire (DTSQ) has been specifically designed to measure satisfaction with diabetes treatment regimens in people with diabetes.

The DTSQ [status version (DTSQs)] is an eight-item questionnaire, in which six questions assess treatment satisfaction and the other two assess perceived frequency of hyper- and hypoglycemia.

Each item is scored from 6 (very satisfied) to 0 (very dissatisfied) such that the Treatment Satisfaction scale can range from 36 (very satisfied) to 0 (very dissatisfied) and the perceived frequency of hyper- and hypoglycemia scores range from 6 (most of the time) to 0 (none of the time).

The DTSQs has 8 items and produces the following measures:

- Treatment satisfaction: Items 1, 4, 5, 6, 7 and 8 are summed to produce a Treatment Satisfaction score (range 0-36). The higher the score the greater the satisfaction with treatment.
- Individual satisfaction with treatment items (items 1, 4, 5, 6, 7 and 8) can be considered separately. All are rated from 6 to 0. The higher the score the greater the satisfaction with each aspect of treatment

Although the DTSQs has proved highly sensitive to change, in many studies where patients are very satisfied with treatment used at baseline, the DTSQs cannot show improvements when they switch to a new treatment, even though they might be even more satisfied with the new treatment.

To overcome the limitation of the DTSQs, a change version (DTSQc) has also been developed, which asks participants to rate how their current treatment compared with their previous treatment.

The DTSQc instrument contains the same 8 items as the DTSQs version. The difference lies in the wording of the response options and instructions, which, in the DTSQc, direct the respondent to compare their experience of treatment before the study began. All items are rated from +3 to -3.

The DTSQc instructions and response options differ from those of the DTSQs to produce measures of relative change in satisfaction rather than measures of absolute satisfaction:

- Treatment satisfaction (change): items 1, 4, 5, 6, 7 and 8 are summed to produce a Treatment Satisfaction (change) score (range: +18 to -18). The higher the score, the greater the improvement in satisfaction with treatment; the lower the score, the greater the deterioration in satisfaction with treatment. A score of 0 represents no change.
- Individual satisfaction with treatment change items (items 1, 4, 5, 6, 7 and 8) can be considered separately. The higher the score, the greater the improvement in satisfaction with each aspect of treatment, the lower the score, the greater the deterioration in satisfaction with treatment.

A major advantage of the DTSQs and DTSQc is that they have been developed to be suitable for people with type 1 or type 2 diabetes using a wide range of treatments, including various methods of insulin delivery, oral medications and diet alone, and is, therefore, appropriate for use before and after patients switch between very different treatment regimens.

Treatment Satisfaction scores for the DTSQs and individual satisfaction with treatment items will be reported for visit 2 and visit 14.

Treatment Satisfaction scores for the DTSQc and individual satisfaction with treatment items will be reported for visit 14.

7.9.3.3. Total insulin dose

The mean insulin total daily delivery dose will be calculated from pump data uploaded in CareLink Clinical at visits: 5 (randomization, week=0-1), 6 (week=0), 7 (week=1), 9 (week=6), 11 (week=12), 13 (week=18) and 15 (week=24). Summary statistics for quantitative variables will be reported at every visit.

7.9.3.4. HbA1c level and number of sensor hypoglycemia events per week by subgroups

Summary statistics for (quantitative variables) will be reported for HbA1c level and for the mean number of sensor glucose hypoglycemic events per subject per week (sensor glucose values of ≤ 55 mg/dL (3.0 mmol/L)) by study arm and by:

- Age group
 - o Grouping A. group 1: age <65, group 2: Age ≥ 65
 - o Grouping B. group 1: age < median age, group 2 \geq median age.
- Duration of diabetes
 - o Grouping A. group 1: <10 years, group 2 ≥ 10 years
 - o Grouping B. group 1: <20 years, group 2 ≥ 20 years
 - o Grouping C. group 1: <30 years, group 2 ≥ 30 years
- Duration of pump therapy at time of screening
 - o Grouping A. group 1: < median duration of pump therapy, group 2 \geq median duration of pump therapy
- Baseline HbA1c
 - o Grouping A. group 1: Baseline HbA1c < 6%, group 2: Baseline HbA1c $\geq 6\%$
 - o Grouping B. group 1: Baseline HbA1c < 7%, group 2: Baseline HbA1c $\geq 7\%$
 - o Grouping C. group 1: Baseline HbA1c $\leq 7.5\%$, group 2: Baseline HbA1c > 7.5%
 - o Grouping D. group 1: Baseline HbA1c < 8%, group 2: Baseline HbA1c $\geq 8\%$
 - o Grouping E. group 1: Baseline HbA1c < 9%, group 2: Baseline HbA1c $\geq 9\%$
- Baseline type of insulin
 - o group 1: Lispro (Humalog), group 2: Aspart (Novolog), group 3: Glulisine (Apidra) and group 4: Other

7.9.3.5. Mean weight

The mean weight will be reported for screening (Visit 2) and end of study phase (Visit 15). Summary statistics for quantitative variables will be reported at every visit.

7.10. Safety Evaluation

Safety analysis will include Number of severe hypoglycemic events, diabetic ketoacidosis events, (Serious) Adverse Events, (Serious) Adverse Device Effects and Device Deficiencies.

Analysis will include: hypoglycemia-related number and mean duration of hospitalizations, number and mean duration ICU (intensive care unit) care, number of emergency room admissions, number of events requiring ambulance assistance and number of lost days from school or work.

7.10.1. Number of severe hypoglycemic events

A severe hypoglycemic event is an event *requiring* assistance of another person due to altered consciousness to actively administer carbohydrate, glucagon, or other resuscitative actions. This means that the subject was impaired cognitively to the point that he/she was unable to treat his or her self, was unable to verbalize his or her needs, and was incoherent, disoriented and/or combative. These episodes may be associated with sufficient neuroglycopenia to induce seizure or coma. Plasma glucose measurements may not be available during such an event, but neurological recovery attributable to the restoration of plasma glucose to normal is considered sufficient evidence that the event was induced by a low plasma glucose concentration.

Total number of severe hypoglycemic events, reported as SAEs in the eCRF, during Run-In (Visit 1 to before Visit 5) and Study Phase will be reported. Total number of severe hypoglycemic events during the Study Phase (from date of Visit 5 to date of Visit 15) will be reported by treatment arm.

The number of severe hypoglycemic events per year will be computed for each subject based on the entire study phase duration (happening after Visit 5 to and including Visit 15). If a subject is followed up for x months and the number of observed severe hypoglycemic events during that follow up is m , the number of severe hypoglycemic events per year (SHE_{year}) for that patient will be estimated as:

$$SHE_{year} = \frac{12}{x} m$$

Annualized crude incidence rates will be expressed as number of severe hypoglycemic events per 100 patients' year and will be computed following these steps: First, the average of the number of severe hypoglycemic events per year is calculated as

$Mean_{(SHE_{year})} = \sum_{i=1}^n SHE_{year(i)} / n$, where $i = 1, 2, \dots, n$ and n total number of subjects. This represents the mean number of severe hypoglycemic events per year in a subject. Second, the number of severe hypoglycemic events per 100 patients' year is calculated as $100 * Mean_{(SHE_{year})}$.

The number of severe hypoglycemic events per 100 patients' year will be calculated and reported by study arm.

7.10.2. Number of severe hyperglycemic events

A severe hyperglycemic event is defined as hyperglycemia (blood glucose >300 mg/dL) with blood glucose ketones >0.6mmol/L or accompanied by symptoms of nausea, vomiting or abdominal pain

Total number of severe hyperglycemic events, reported as SAEs in the eCRF, during Run-In (Visit 1 to before date of Visit 5) and Study Phase will be reported. Total number of severe hyperglycemic events during the Study Phase (from date of Visit 5 to date of Visit 15) will be reported by treatment arm.

7.10.3. Number of diabetic ketoacidosis events

A diabetic ketoacidosis events is defined as an event of blood glucose greater than 250 mg/dL (or greater than 13.9 mmol/L), arterial pH less than 7.3, bicarbonate less than 15mEq/l, Anion gap greater than 12, moderate ketonuria or ketonemia and requiring treatment within a health care facility.

Number of diabetic ketoacidosis events (NKeto) will be reported using summary statistics by treatment arm.

7.10.4. Number of Serious Adverse Events, Serious Adverse Device Effects and Device Deficiencies

Serious adverse events are events that

- a) Led to death,
- b) Led to serious deterioration in the health of the subject, that either resulted in
 - 1) A life-threatening illness or injury, or
 - 2) A permanent impairment of a body structure or a body function, or
 - 3) In-patient or prolonged hospitalization, or
 - 4) Medical or surgical intervention to prevent life-threatening illness or injury or permanent impairment to a body structure or a body function,
- c) Led to fetal distress, fetal death or a congenital abnormality or birth defect.

Serious adverse device effect is an adverse device effect that has resulted in any of the consequences characteristic of a Serious Adverse Event.

Device deficiency is an inadequacy of a medical device with respect to its identity, quality, durability, reliability, safety or performance (NOTE: Device deficiencies include malfunctions, use errors, and inadequate labelling)

The number of each of these events will be reported for run-in and study phase separately and per study arm. A listing describing these events will be prepared.

7.11. Health Outcomes Analyses

Not applicable

7.12. Changes to Planned Analysis

This (StAP) has been written for internal use to document the planned analyses that will be conducted on the SMILE study with the possibility of including these analyses in reports or publications. This StAP will be executed in full (deviations will be explained in the final report) but does not limit the analysis in reports, and additional analysis of the study data beyond this plan is expected. Analyses beyond the StAP will be identified as such and referred to as not pre-specified.

In this StAP some minor changes have been made to the analysis of some endpoints as compared with the Study CIP. The following are the changes and the reason for these changes:

- Secondary endpoints for sensor glucose values < 54 mg/dL have been added to this StAP (not included in the study CIP). The reason of this addition is to comply with the newly released guideline from ADA and EASD (December 2016) recommending using glucose concentration <3.0 mmol/L (<54 mg/dL), which it considered to be clinically significant biochemical hypoglycemia. This guideline was released after the release of the SMILE study protocol.

- The study CIP defines hypoglycemia for sensor values \leq than several thresholds (Mean number of sensor glucose hypoglycemic events \leq 40 mg/dL (2.2 mmol/L), \leq 55 mg/dL (3.0mmol/L) and \leq 70 mg/dL (3.9mmol/L)). For values other than 55 mg/dL the analysis will be performed for sensor values < than the thresholds (Mean number of sensor glucose hypoglycemic events < 40 mg/dL (2.2 mmol/L), < 54 mg/dL (3.0mmol/L) and < 70 mg/dL (3.9mmol/L)). The reason for this change is to comply with the newly released guideline from ADA and EASD, while making no changes to the primary endpoint threshold of \leq 55 mg/dL.

- This StAP includes a new secondary endpoint "Mean duration of hypoglycemic excursions < 54 mg/dL (3.0mmol/L). The reason for this change is that the evidence points toward duration of severe hypoglycemia as a major component of severe consequence as seizure or death. This endpoint has been added in the StAP to be analyzed in an exploratory way.

- The method for handling missing data described in this StAP is different from the multiple imputation method mentioned in the study CIP. The analysis will be based on a repeated measures random effects ANOVA model (section 7.9.1) with PROC MIXED that uses all available data and accounts for possible missing at random data. Multiple imputation would require making assumption on the longitudinal profile over time for the endpoint while the updated ANOVA method in this StAP does not require making such assumptions while accounting for correlated data and possible missingness.

8. Validation Requirements

Level I validation is required for Statistical and SAS programming of primary endpoint. Level I requires that the peer reviewer independently programs output and then compares the output with that generated by the original Statistical Programmer.

For the other endpoints, minimally a Level II validation will be used. Level II requires that the peer reviewer reviews the code; where appropriate, performs manual calculations or simple programming checks to verify the output.

9. References

- Begg, C., Cho, M., & Eastwood, S. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*, 276:637-9.
- Bergental, R. M., Klonoff, D. C., Garg, S. K., Bode, B. W., Meredith, M., Slover, R. H., et al. (2013). Threshold-based insulin-pump interruption for reduction of hypoglycemia. *N Engl J Med*.
- Chen, Y., DeMets, D., & Lan, K. (2004). Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine*, 1023-1038.
- Clarke. (1995). Reduced awareness of hypoglycemia in adults with IDDM: a prospective study of hypoglycemic frequency and associated symptoms. *Diabetes Care*.
- Cox. (1987). Fear of Hypoglycemia: Quantification, Validation, and Utilization. *Diabetes Care*.
- International Hypoglycemia Study Group. (2017). Glucose Concentrations of Less Than 3.0 mmol/L (54 mg/dL) Should Be Reported in Clinical Trials: A Joint Position Statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*, 40:155-157.
- Li, G., Shin, w., Xie, T., & Lu, J. (2002). A sample size adjustment procedure for clinical trials based on conditional power. *Biostatistics*, 277-287.
- Mehta, C., & Pocock, S. (2011). Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Statistics in Medicine*, 3267-3284.
- Rubin, D. B. (1987). *Multiple Imputatiuon for Nonresponse in Surveys*. New York: Wiley.
- White, I., Thompson, S. (2005). Adjusting for partially missing baseline measurements in randomized trials. *Statistics in Medicine*, 24: 993-1007.