Dual hormone closed loop in type 1 diabetes: A randomized trial (DARE)

STATISTICAL ANALYSIS PLAN

Version 1.1, 30-09-2024

Revision history

Version	Date	Change description
1.1	30-09-2024	Sensitivity analyses added to evaluate impact of discontinuation of index treatment (section 7.3, p 16-17)

Index

Contributors

- M. Jancev, Department of Vascular Medicine, Diabetology and Endocrinology, UMC Utrecht, Utrecht, The Netherlands
- Dr. T.T. van Sloten, Department of Vascular Medicine, Diabetology and Endocrinology, UMC Utrecht, Utrecht, The Netherlands
- Dr. H.W. de Valk, Department of Vascular Medicine, Diabetology and Endocrinology, UMC Utrecht, Utrecht, The Netherlands
- Dr. N.P.A. Zuithoff, Department of Data Science & Biostatistics, Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht University, Utrecht, The Netherlands UMC Utrecht
- Dr. P.M. van de Ven, Department of Data Science & Biostatistics, Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht University, Utrecht, The Netherlands UMC Utrecht
- Prof. dr. K.G.M. Moons, Department of Epidemiology & Health Economics, Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht University, Utrecht, The Netherlands UMC Utrecht
- Dr. G.W.J. Frederix, Department of Epidemiology & Health Economics, Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht University, Utrecht, The Netherlands UMC Utrecht
- Dr. A.C. van Bon, Department of Internal Medicine, Rijnstate Hospital, Arnhem, The Netherlands
- Em. prof. dr. J. Hans DeVries, Department of Internal Medicine, Amsterdam UMC, Amsterdam, The Netherlands
- H. Knottnerus, Dutch Diabetes Association, Amersfoort, The Netherlands
- Dr. E.H. Serné, Department of Internal Medicine, Amsterdam UMC, Amsterdam, The Netherlands
- Em. prof. dr. F.J. Snoek, Medical Psychology, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

1. Overview

In this document, the approach to the sample size and statistical analyses are outlined for the Dual hormone closed loop in type 1 diabetes: a randomized trial (DARE). This statistical analysis plan was written in accordance with research protocol version 7.0 dated 29-05-2024.

The DARE trial is a 12 month open-label, two-arm randomised parallel-group trial aiming to determine in patients with type 1 diabetes (T1DM) the long-term effectiveness of treatment with a dual-hormone (insulin and glucagon) fully closed loop system (DHFCL) during 12 months compared to both the current most advanced technological care and to the currently most used care. The study consists of 240 patients in total, with study arm 1 including 170 patients randomised to either DHFCL (index) or a hybrid closed loop (HCL) as control, and study arm 2 with in total 70 patients randomised to either DHFCL (index) or multiple daily injections (MDI) in combination with continuous or flash glucose monitoring (CGM or FGM) as control. In both arms, patients are 1:1 randomised to either receiving the DHFCL (index group) or continuation of their current treatment (control group), stratified by participating study centre. Thus, from the 170 patients on HCL, 85 will be randomised to the DHFCL treatment and 85 will continue their HCL treatment. Likewise, from the 70 MDI patients, 35 will be randomised to DHFCL and 35 patients to continue their MDI treatment.

All patients will be asked to wear a blinded sensor (FSL Pro IQ, Abbott) to measure glucose levels at baseline, 3, 6, 9 and 12 months.

The primary endpoint is the time in range (TIR; % of time spent between 3.9-10 mmol/l) at 12 months. Secondary endpoints include other glycaemic endpoints such as time above and below range (TAR and TBR, respectively) and glycaemic variability, patient reported outcome measurements (PROMs), cost effectiveness outcomes and safety and device-related outcomes.

2. Statistical hypothesis

The primary outcome is:

- Superiority in the Time in Range (TIR; % of time spent in the 3.9-10 mmol/l target range) at 12 months

There is no key secondary outcome. All secondary outcomes are described in section 4 (P7-9).
The primary analysis will be performed for each of the study arms separately (i.e. DHFCL vs HCL, and DHFCL vs MDI).

Time in range outcome study arm 1 - HCL as control treatment and DHFCL as index treatment:

- Null hypothesis: There is no difference in the mean TIR at 12 months between DHFCL and HCL group
- Alternative hypothesis: there is a nonzero difference in the mean TIR at 12 months between DHFCL and HCL group

Time in range outcome study arm 2 - MDI+FGM/CGM as control treatment and DHFCL as index treatment:

- Null hypothesis: There is no difference in the mean TIR at 12 months between DHFCL and MDI+FGM/CGM group
- Alternative hypothesis: there is a nonzero difference in the mean TIR at 12 months between DHFCL and MDI+FGM/CGM group

3. Sample size

3.1 Study arm 1: HCL as control treatment and DHFCL as index treatment

A recent cross over study of the index intervention treatment (DHFCL) showed a median mean TIR of 86% (SD 4%).[1] Previous studies show a mean TIR under HCL treatment of 70 to 78% (SD 16% to 17%).[2–4] To be conservative, we assume a 7% increase in mean TIR from 78% under the HCL control treatment (SD 17%) to 85% under the index intervention treatment (SD 4%), using the two sample t-test with unequal variances, two-sided significance level of 0.05 and a power of 90%. Accordingly, the required sample size for this arm is 68 per group. To allow for potential loss to follow-up and to address the main secondary endpoints, we include a sample size of 85 per group in this arm of the trial: 85 in the HCL control treatment group and 85 in the DHFCL index treatment group.

3.2 Study arm 2: MDI+FGM/CGM as control treatment and DHFCL as index treatment.

Previous studies showed a mean TIR under MDI+FGM/CGM treatment of 50-59% (SD of 12 to 20%).[5] To be conservative again, we assume a mean TIR in the MDI+FGM/CGM control group of 70% (SD 20%) and for the DHFCL index intervention of 85% (SD 4%) (similarly as above), thus assuming a 15% increase in mean TIR due to the index intervention. Using the two sample t-test with unequal variances, two-sided significance level of 0.05 and a power of 90%, the required sample size for this arm is 22 per group. To allow for potential loss to follow-up and to address the main secondary endpoints we include a sample size of 35 per group in this arm of the trial: 35 in the MDI+FGM/CGM control treatment group and 35 in the DHFCL index treatment group.

4. Outcome measures

4.1 Primary endpoint

- The main study endpoint is the time in range (TIR; % of time spent in the 3.9-10 mmol/l target range) at 12 months.

4.2 Secondary endpoints

Secondary endpoints include other glycaemic outcomes, patient reported outcome measurements, and cost-effectiveness outcomes and safety and device-related outcomes.

4.2.1 Glycaemic control endpoints at 0, 3, 6, 9 and 12 months:

- Time in range (TIR) (3.9-10 mmol/l) (%; amount of time (hours/minutes)) at 3, 6 and 9 months;
- Time above range (TAR):
    - level 1 and 2 hyperglycaemia: % of time spent >10.0 mmol/l;
    - level 2 hyperglycaemia: % of time spent >13.9 mmol/l;
- Time below range (TBR):
    - level 1 and 2 hypoglycaemia: % of time spent <3.9 mmol/l;
    - level 2 hypoglycaemia: % of time spent <3.0 mmol/l;
- Number of hypoglycaemic events: defined as glucose <3.0 mmol/l for 15 consecutive minutes[6], when the time between two successive events is less than 30 minutes, they will be combined and counted as one event;
- Mean glucose (mmol/l) per patient per time point:
    - day and night;
    - day: from 06:00 to 23:59 hours;
    - night: from 00:00 to 05:59 hours;
- Glycaemic variability:
    - Coefficient of variation (%);
    - Standard deviation (mmol/l);
- HbA1c:
    - Mean per patient per time point (mmol/mol);
    - Percentage of patients achieving HbA1c ≤ 53 mmol/mol.
- Percentage patients with ≥5% points TIR improvement from baseline (%) at 12 months;
- Percentage patients achieving >70% TIR and/or <4% TBR at 12 months;
- Time in tight range (3.9-7.8 mmol/l) (%; amount of time (hours/minutes).

4.2.2 Patient reported outcome measures (PROMs), at 0, 3, 6, 9 and 12 months:

- World Health Organization-Five Well-Being Index (WHO-5) score;
- Health-related quality of life scores (EQ-5D-5L);
- Diabetes Treatment and Satisfaction Questionnaire status and change (DTSQ-s and DTSQ-c) scores;
- Problem Areas In Diabetes (PAID-5) score;
- Hypoglycaemia Fear Survey-II (HFS-II) Worry subscale score;
- Pittsburgh Sleep Quality Index (PSQI) score;

- Insulin delivery systems: perceptions, ideas, reflections and expectations (INSPIRE) scores; only for HCL and DHFCL groups;
- Hypoglycaemia unawareness (Gold-Clarke): only at 0 and 12 months.

These scores will be calculated as defined previously.[8–16]

4.2.3 Cost effectiveness endpoints:

- Medical Consumption Questionnaire (MCQ), at 0, 3, 6, 9 and 12 months;
- Productivity Cost Questionnaire (PCQ), at 0, 3, 6, 9 and 12 months;
- Detailed hospital health care consumption for each individual patient (collected from electronic patient files, including frequency of unplanned patient contact with the diabetes team);
- Cost effectiveness: cost per quality adjusted life year.

4.2.4 DHFCL endpoints, at 3, 6, 9 and 12 months:

- Daily glucagon use (mg/day);
- Percentage of time the control algorithm was active while wearing the DHFCL device. For this calculation, we will use the algorithm status output that is logged every 10 minutes while the DHFCL was in use.

4.3 Safety and other endpoints

- Adverse events and serious adverse events ((S)AEs);
- Device issues;
- Aspartate aminotransferase (ASAT) (U/L) and alanine aminotransferase (ALAT) (U/L) at 12 months;
- NSAID use (due to excluded daily acetaminophen use) and associated drug complications (e.g. stomach ulcers) rate;
- Daily insulin use at 3, 6, 9 and 12 months (UI/day);
- Weight (kg), blood pressure (mmHg) and pulse (beats/min) at 12 months;
- Frequency of unplanned patient contact with the diabetes team;
- Concomitant medication, at screening, baseline, 3, 6, 9 and 12 months;
- Continuation rate: expressed as the percentage of participants that continue DHFCL treatment after 1 year of use;
- Reasons for discontinuation of the DHFCL treatment;
- Glycaemic data for at least 70% of the collection period of 14 days (N, %: see section 4, 4.1, and 5.1 for details).
- Glycaemic data for at least 50% of the collection period of 14 days (N, %: see section 4, 4.1, and 5.1 for details).

4.4 Definitions safety parameters

Safety parameters include AEs, SAEs and device-related issues.

AEs are defined as any undesirable experience occurring to a subject during the study, whether or not considered related to the DHFCL. AEs of special interest include:

- Events possibly related to glucagon or fluctuating glucose levels: nausea/fatigue/headache/restlessness;
- Skin reaction tape infusion set: erythema/discomfort/irritation/rash;
- Reaction infusion set/site: discomfort/inflammation/pain/lipodystrophy;
- Skin reaction tape glucose sensor: erythema/discomfort/irritation/rash/ damage;
- Reaction glucose sensor (insertion site): bruising/bleeding/pain.
- New diabetes complications;
- Deterioration (from baseline) of diabetes complications;
- Ketoacidosis;
- Severe hypoglycaemia (hypoglycaemia that leads to a seizure or unconsciousness),

Hypoglycaemia is a disease specific endpoint, and not recorded as AE in this trial.

SAEs are defined as any untoward medical occurrence or effect that

- results in death;
- is life threatening (at the time of the event);
- requires hospitalisation or prolongation of existing inpatients' hospitalisation;
- results in persistent or significant disability or incapacity;
- is a congenital anomaly or birth defect; or
- any other important medical event that did not result in any of the outcomes listed above due to medical or surgical intervention but could have been based upon appropriate judgement by the investigator.

An elective hospital admission will not be considered as a SAE.

Device incidents are defined as any malfunction or deterioration in the characteristics or performance of the DHFCL system, including use-error due to ergonomic features, as well as any inadequacy in the information supplied by the manufacturer and any undesirable side-effect.

5. Analysis datasets and statistical analyses

The primary and secondary analyses will be performed on an intention-to-treat (ITT) principle and per-protocol (PP) principle. All randomized participants will be analyzed for the ITT analysis and all safety analyses.

The PP analysis will be limited to visits of patients that adhered to the randomised treatment. For example, a patient that adhered to the treatment regime up to 6 months of follow-up and subsequently changed the treatment regime will be included until the 6 month visit. All subsequent visits will be excluded from the PP analyses, also when patients returned to the randomized treatment later in the follow-up. The PP analysis will also be restricted to participants with:

- No glucose-lowering medications used other than those acceptable in the protocol;
- For participants in the DHCL group, closed loop mode is active for at least 80% of the time during the trial duration before drop-out (if any)

This PP analysis will only be done if >5% of the participants are excluded based on these criteria.

The primary outcome is measured with a blinded sensor with the aim to collect 14 consecutive days of glucose measurements. However, not all patients may be able or willing to provide continuous glucose measurements for 14 days. This will be further evaluated with sensitivity analyses.

5.1 Analysis of the primary endpoint

The primary analysis will be performed for each of the study arms separately (i.e. DHFCL vs HCL, and DHFCL vs MDI) and will follow the ITT principle with the data from each participant analysed according to the treatment assigned by randomization.

Included subjects

All subjects who had at least ≥168 hour of blinded sensor data (i.e. 50%) at baseline and at least one follow-up visit.[7] Valid measurements of primary outcomes at any visit will be included in the analyses: for example, a patient with a missing 6 month measurement (for any reason) and valid measurements at subsequent visits will be included in the analysis with all available measurements. If the subject has <168 hours of sensor data, then the baseline metrics will not be calculated and will be set to missing. Likewise, if a subject has <168 hours of sensor data, then post-randomization sensor metrics will not be calculated and will be set to missing.

5.2 Analysis of the secondary endpoints

Included subjects

All secondary outcomes assessed at follow-up visits (glycaemic control outcomes and PROMs) will be included in the analysis, similar to what is described for the primary outcome. The cost-effectiveness analysis is described separately in section 6.3.

5.3 Baseline descriptive statistics

Baseline characteristics of all included patients will be reported for each treatment (index and control) group separately, for both arms (see below). Continuous variables will be included as means with standard deviations, medians with 25th and 75th percentiles and number of non-missing values, categorical variables will be reported as numbers with percentages and number of non-missing

values. Patients, sex, age, center of inclusion as well as all other baseline measurements of primary and secondary outcomes (where applicable) will be included in the table. In line with current recommendations for clinical trials,[17] p-values for a comparison between treatment and control groups at baseline will not be included.

Baseline variables:

- Age;
- Sex;
- Body mass index (kg/m^2);
- Systolic and diastolic blood pressure (mmHg) and pulse (beats/min);
- HbA1c (mmol/mol);
- Duration of diabetes (years);
- Specification current treatment (sensor/pump and type);
- Duration of current treatment (years);
- Patient reported daily insulin use (units/day);
- Diabetes complications and treatment (yes/no):
- Retinopathy:
    - Retinopathy and treatment;
    - Maculopathy and treatment;
- Nephropathy:
    - eGFR (mL/min/1.73 m2);
    - Albuminuria (micro-/macroalbuminuria);
- Neuropathy:
    - Distal neuropathy;
    - Gastric autonomic neuropathy;
- Macro-angiopathy:
    - Cardiac;
    - Cerebral;
    - Peripheral;
    - Other;
- Non-diabetic disease;
- Smoking (yes/no);
- Alcohol (units/week);
- Education level;
- Professional status;
- Domestic situation.

6. Statistical methods

6.1 Primary endpoint

The primary endpoint (mean TIR at 12 months) will be compared between the DHFCL treatment and the two control treatments. The analysis will be performed for each of the study arms (see above) separately. To obtain this comparison, a mixed model analysis for continuous outcomes will be performed with TIR at 3, 6, 9 and 12 months (hence forward referred to as time) as dependent variable. This is a linear regression model that includes a random intercept for center of inclusion and a residual covariance matrix to correct for multiple measurements per patient. The initial analysis will be performed with a heterogeneous autoregressive matrix that assumes different variances for each measurement, a constant correlation between adjacent measurements and consistently lowering correlation when time between measurements is longer. The first analysis will include time, treatment and an interaction between time and treatment as fixed effects. In the second analysis, correction for the predefined prognostic factors, i.e. TIR at baseline, patients age and sex, will be included in the analysis as fixed effects. As the primary endpoint comparison is defined at 12 months, a specific contrast will be included in the analysis to compare mean TIR at 12 months between DHFCL control treatments. Results will be reported as difference in mean TIR with 95% confidence intervals (CIs). P-values will only be reported for the primary analysis in the ITT population. Study arm 1 and 2 will be considered two separate trials in different populations and as such no multiplicity correction is planned for performing these two primary comparisons under a single study protocol. The p-value for study arm 1 and study arm 2 will each be separately compared to the threshold 0.05 to declare significance.

The primary analysis may result in convergence problems when the residual covariance matrix is not correctly specified, given the data. In this case, the residual covariance matrix may be adjusted. Convergence problems may also arise for the estimation of the random intercept for center of inclusion when one or more centers include a low number of patients or when included patients from different centers show very similar outcome measurements. In this case, the random intercept may be excluded from the analysis and the addition of a variable representing academic centers versus peripheral centers may be added to the analyses models as a fixed effect. The regression model with residual covariance matrix is specifically chosen as the primary analysis, as it allows to incorporate all non-missing measurements of patients over time. Further details on missing outcomes are given in section 7 on missing data.

Validity of the distributional assumptions for the main analysis model (normal distribution, homoscedasticity) will be assessed with residual analysis. Raw residuals from a covariance pattern model may (compared to a 'standard' linear regression) show both correlation and unequal variances. Raw residuals will be transformed with Cholesky decomposition prior to the assessment described in this section.[18]

Histograms and QQ-plots will be created to assess normality of residuals. Residuals will be plotted against raw and normalized predicted values to assess homoscedasticity and linearity. Violation of the assumption of homoscedasticity will be corrected with robust (i.e. Hubert-White's) estimation of standard errors.[18]

When a serious violation of the normality assumption is observed in the residual plots, the main analysis will be changed to a mixed regression analysis based on beta distributions.[19] The steps for inclusion of predefined prognostic factors will be included as specified for the primary analysis (see above). Multiple measurements per patient will be corrected with a random intercept. The residual covariance matrix will be excluded from the model, as such models cannot be estimated under maximum likelihood.

6.2 Secondary endpoints

The analysis for continuous secondary outcomes follows the analysis specifications of the primary outcome (see above), except when distributional assumptions are not met. In this case, a log-transformation will first be used and subsequently assessed with residual analyses. When log-transformations still show deviations from distributional assumptions, alternative distributions or non-parametric analysis will be considered. As for the analysis of the primary outcome, all regression analyses will be performed by adjusting for the same a priori defined prognostic factors and repeated measurements.

Binary secondary outcomes will be analyzed with mixed logistic regression analysis. Secondary count outcomes (e.g. the number of hypoglycemic events) will be analyzed with a generalized linear mixed model: the distributions utilized will depend on the observed distribution. Initial analysis will be performed with a negative binomial distribution. The distribution may show a high number of patients with zero events. If so, the distribution will be adjusted to either a zero-inflated Poisson distribution or a zero-inflated negative binomial distribution.[20] For binary and count outcomes, analyses with a random intercept will initially be performed, as these analyses may be performed under maximum likelihood estimation. Assumption for linearity will be assessed and adjustments will be made when necessary.

Secondary outcomes measured only once during follow-up (i.e. at 12 months) will be analyzed with linear, logistic and negative binomial regression models (as applicable) without adjustment for repeated measurements (i.e. regular linear and logistic regression). Results from linear models will be reported as a difference in means (with 95% CIs). For binary outcomes, results will be reported as odds ratios with 95% CIs. For negative binomial models for count outcomes, results will be reported as rate ratios with 95% CIs.

Secondary endpoints will be considered supportive and are not included in the formal testing strategy. P-values will therefore not be provided.

6.3 Cost-effectiveness

The health economic evaluation will be performed according to the Dutch Health Care Institute pharmacoeconomic guidelines.[21]

6.4 Software

Analysis of primary and secondary endpoints will be performed with SAS (SAS Institute Inc., Cary, NC, USA) or R software (R Core Team, Vienna, Austria). Multiple imputation will be performed in R software with the MICE package.[22] Results from the imputation will be exported as imputed datasets in a format that will allow subsequent analysis in SAS. Some graphical representations and

descriptive tables may be constructed with SPSS (IBM Corp. IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp). For any software or R package, the most recent available version will be used.

7. Missing data

Missing data will be minimised as we will use centralised data acquisition (eCRF) for all participating centres. This will be coordinated by the UMC Utrecht with appropriate monitoring on data queries. Missing data may nevertheless be unavoidable and may subsequently result in biased results.

7.1 Baseline variables

Multiple imputation will be applied for missing values on baseline variables that may be used in any of the statistical analyses. Prior to the imputation, an analysis will be conducted to evaluate associations between the occurrence of missing values and available patient and disease characteristics.

Multiple imputation will be applied with a fully conditional specification and performed for control groups and intervention groups separately,[23] with a correction for clustering in centers of inclusion. Initial imputations will be performed on data in a so-called wide format, where all measurements over time are included as separate variables. The logistic regression approach will be used for categorical variables, continuous measurements will be imputed with predictive mean matching. The imputation of baseline variables will be performed on all variables that will be included in the statistical analysis including primary and secondary outcomes at baseline and at follow-up. Additionally, other observed variables considered to be associated with missing values will be included in the imputation. The number of imputations will initially be equal to the percentage of patients with missing data on variables to be included in any analysis plus 10. When this percentage is below 10%, a minimum of 20 imputations will be used. When the percentage of patients with missing patients is larger than 50%, the number of imputations will be increased to 100.[24,25]

7.2 Missing primary and secondary outcome measurements

A blinded sensor is used for measuring the primary endpoint. Patients with at least 168 hours of sensor data on a specific follow-up time will be included in the analyses. All valid TIR measurements at a specific time point during follow-up will be included in the analysis, also for patients with incomplete follow-up. Sensitivity analyses on the primary endpoint will be conducted to assess the impact of incomplete sensor measurements. Based on the consensus statement on glucose sensor measurement,[7] analyses will be repeated by calculating the TIR based on the first 235.2 hours (i.e. 70% of 14 days) of sensor measurements for all patients, excluding patients with less than 235.2 hours of sensor data.

Even though primary and secondary outcomes will be included in the imputation of baseline variables, the imputed values of outcomes will not be included in the primary analysis, as this may result in biased estimation of treatment effect or loss of efficiency.[24,26] For patients with missing outcomes due to intolerability of the DHFCL, a separate imputation and sensitivity analyses of primary and secondary outcomes is specified in the section on drop-outs (section 7.3). In these sensitivity analyses, imputed outcomes will be included. Patients with incomplete baseline variables that drop-out early after randomisation (i.e. prior to the measurement at 3 months) may be fully excluded.

7.3 Drop-outs due to intolerability of the DHFCL

7.3.1 Intolerability and lost to follow-up

A potential bias may occur in both ITT and PP analyses when a substantial number of patients do not follow the index treatment regime due to intolerability of the DHFCL. This is of special concern when these patients completely withdraw from the study and do not participate in follow-up measurements. When this occurs in more than 6 patients in arm 1 or more than 3 patients in arm 2 (to be decided separately for each arm), a separate sensitivity analysis will be performed. For these patients, we assume a return to the treatment-as-usual (comparable to the control group), missing outcomes of these patients will be imputed based on an imputation model that is estimated in the control group only.[27] Further specifications of the multiple imputation will be applied as described above. For this specific sensitivity analysis, missing outcomes of patients who drop out for other reasons will be included in the analysis, yet based on the imputation in the group they were randomised in.

7.3.2 Intolerability and continuation of study measurements

Intolerability of the DHFCL is also of concern for patients who discontinue treatment while continuing with the study. Follow-up measurements in these patients will be performed as planned, and these measurements will be included in the ITT analysis as described in section 6.1. The treatment effect derived from the ITT analysis, while being of primary interest for decision making, will not represent the full treatment potential of the DHFCL. Results of the PP analysis may be biased, as balance between treatment groups after randomisation may be lost. To further evaluate the beneficial effect of the DHFCL, two additional sensitivity analyses will be performed.

The first analysis aims to evaluate the treatment effect under a hypothetical scenario where all patients were treated with the DHFCL and were fully adherent. Initially, observations of the outcome after treatment discontinuation or non-adherence will be excluded from the analysis and set to missing. Subsequently, multiple imputation will be used to impute these outcomes. Multiple imputation will be applied as described in section 7.1, inclusion of additional variables in the imputation model will be considered based on clinical literature as well as observed patterns of missing data and discontinuation. As described in section 7.1, imputations will be performed for each arm and each treatment group separately. Analysis of the primary outcome will be performed as described in section 6.1. In an additional step, all follow-up outcomes, both prior to or after discontinuation or non-adherence, will be set to missing and imputed. Again, the resulting imputed data will be analysed as described in section 6.1. It should be noted that imputations described in this section will be performed separately from imputations described in sections 7.1.

The second analysis aims to evaluate the treatment effect in the principal stratum of patients that tolerate and fully adhere to DHFCL. This analysis is performed in two steps. In the first step, data from patients randomised to the intervention group are used to construct a propensity score (PS) model for discontinuation due to intolerability of the DHFCL. In the second step, the analysis as described in section 6.1 is performed with weights based on the propensity scores.
The propensity score will be constructed with logistic regression with intolerability (yes/no) as outcome variable in the intervention groups only. The model will include patient characteristics and

variables related to diabetes status measured at baseline. Patient characteristics include age, sex, educational level. Diabetes-related variables include TIR, duration of diabetes and patient reported daily insulin use. Variables to be included in the PS model may be subject to change based on available numbers (e.g. the number of discontinuing patients and valid observations for baseline variables) and clinical information on reasons for discontinuation. Outcomes and other variables measured during follow-up will not be included in the PS model.

The PS model will be used to estimate predicted probabilities of discontinuation due to intolerability for patients in the control group, these probabilities will be used as weights. Weights will only be applied in the control group. In the intervention group only two weights will be applied: a weight of 0 for patients who discontinued due to intolerability (i.e. effectively excluding these patients from the analysis) and a weight of 1 for all others. The weight will be applied for the analysis as described in section 6.1.[28]

8. Safety analyses

All safety outcomes will be presented for all events from baseline to final visit. Analyses will have a similar approach as with the secondary endpoints in the case of sufficient numbers of events.

9. Planned interim analyses

A formal interim analysis for primary and secondary outcomes is not planned. A Data Safety Monitoring Board will be established to monitor (S)AEs.

10. Subgroup analyses

In exploratory analyses, the primary and key secondary outcomes will be assessed separately for interaction with certain baseline variables as described below.

Subgroup analyses will assess the effectiveness of the DHFCL by the following patient characteristics:

- Age groups (18 to <30 years old, 30 to <50 years old, ≥50 to 75 years old)
- Sex
- Baseline HbA1c (<53 mmol/mol, 53 to <64 mmol/mol, 64 to <75 mmol/mol, ≥75 mmol/mol)
- Type 1 diabetes duration (<5 years; 5 to <10 years; 10 to <20 years; ≥20 years)
- Body mass index subgroups as follows:
    - Underweight – BMI <18.5 kg/m^2
    - Normal weight – BMI 18.5 to <24.9 kg/m^2
    - Overweight – BMI ≥25 kg/m^2
- Education (low [no education, primary education or lower vocational education], middle [general secondary education, general vocational education, higher secondary and pre-university education], and high [higher vocational education and university])
- rtCGM or FGM use prior to enrollment (MDI group only)

Effect sizes and 95% confidence intervals will be presented per subgroup in a forest plot.
Interpretation of the subgroup analyses will be made with caution, particularly if the primary analysis is not significant.

11. Exploratory analyses

No exploratory analyses are planned.

12. References

1. Blauw H, Onvlee AJ, Klaassen M, van Bon AC, DeVries JH. Fully Closed Loop Glucose Control With a Bihormonal Artificial Pancreas in Adults With Type 1 Diabetes: An Outpatient, Randomized, Crossover Trial. Diabetes Care. 2021;44(3):836–8.

2. Bergenstal, R. M., Nimri, R., Beck, R. W., Criego, A., Laffel, L., Schatz, D. FSG. A Comparison of Two Hybrid Closed-Loop Systems in Adolescents and Young Adults with Type 1 Diabetes: A Multicenter Randomised Crossover Trial. Lancet. 2021;397(10270):208–19.

3. Brown, S. A., Kovatchev, B. P., Raghinaru, D., Lum, J. W., Buckingham, B. A., Kudva, Y. C., ... & Beck RW. Six-Month Randomized, Multicenter Trial of Closed-Loop Control in Type 1 Diabetes. New England Journal of Medicine. 2019;381(18):1707–17.

4. Amadou C, Franc S, Benhamou PY, Lablanche S, Huneker E, Charpentier G, et al. Diabeloop DBLG1 Closed-Loop System Enables Patients With Type 1 Diabetes to Significantly Improve Their Glycemic Control in Real-Life Situations Without Serious Adverse Events: 6-Month Follow-up. Diabetes Care. 2021;44(3):844–6.

5. Visser MM, Charleer S, Fieuws S, De Block C, Hilbrands R, Van Huffel L, et al. Comparing real-time and intermittently scanned continuous glucose monitoring in adults with type 1 diabetes (ALERTT1): a 6-month, prospective, multicentre, randomised controlled trial. The Lancet. 2021;397(10291):2275–83.

6. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range. Diabetes Care. 2019;42(8):1593–603.

7. Battelino T, Alexander CM, Amiel SA, Arreaza-Rubin G, Beck RW, Bergenstal RM, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol. 2023;11(1):42–57.

8. EuroQolGroup T. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New York). 1990;16(3):199–208.

9. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 well-being index: A systematic review of the literature. Psychother Psychosom. 2015;84(3):167–76.

10. Welch GW, Jacobson AM, Polonsky WH. The problem areas in diabetes scale: An evaluation of its clinical utility. Diabetes Care. 1997;20(5):760–6.

11. Cox DJ, Irvine A, Gonder-Frederick L, Nowacek G, Butterfield J. Fear of hypoglycemia: quantification, validation, and utilization. Diabetes Care [Internet]. 1987;10(5):617–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3677982

12. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28:193–213.

13. Clarke WL, Cox DJ, Gonder-Frederick LA, Julian D, Schlundt D, Polonsky W. Reduced Awareness of Hypoglycemia in Adults With IDDM: A prospective study of hypoglycemic frequency and associated symptoms. Diabetes Care [Internet]. 1995 Apr 1;18(4):517–22. Available from: https://diabetesjournals.org/care/article/18/4/517/18500/Reduced-Awareness-of-Hypoglycemia-in-Adults-With

14. Gold AE, Macleod KM, Frier BM. Frequency of Severe Hypoglycemia in Patients With Type I Diabetes With Impaired Awareness of Hypoglycemia. Diabetes Care [Internet]. 1994 Jul 1;17(7):697–703. Available from: https://diabetesjournals.org/care/article/17/7/697/18251/Frequency-of-Severe-Hypoglycemia-in-Patients-With

15. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Roijen LH Van. The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. Value in Health [Internet]. 2015;18(6):753–8. Available from: http://dx.doi.org/10.1016/j.jval.2015.05.009

16. iMTA Productivity and Health Research Group. Manual iMTA Medical Cost Questionnaire (iMCQ). [Internet]. Rotterdam; 2018. Available from: https://www.imta.nl/questionnaires/imcq/

17. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340.

18. Fitzmaurice GM, Laird NM, Ware HJ. Applied Longitudinal Analysis. Wiley; 2011.

19. Ferrari SLP, Cribari-Neto F. Beta regression for modelling rates and proportions. J Appl Stat. 2004;31(7):799–815.

20. Payne EH, Gebregziabher M, Hardin JW, Egede LE. An empirical approach to determine a threshold for assessing overdispersion in Poisson and negative binomial models for count data. Commun Stat Simul Comput. 2018;47(6):1722–38.

21. Dutch Institute National Health Care (Zorginstituut Nederland). Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidzorg (Protocol for the execution of economic evaluation in healthcare). 29-02-2016 [Internet]. 2016;(november):120. Available from: https://www.ispor.org/PEguidelines/source/NL-Economic_Evaluation_Guidelines.pdf

22. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67.

23. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? Stat Methods Med Res. 2018;27(9):2610–26.

24. Austin PC, White IR, Lee DS, van Buuren S. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. Canadian Journal of Cardiology [Internet]. 2021;37(9):1322–31. Available from: https://doi.org/10.1016/j.cjca.2020.11.010

25. Von Hippel PT. How to impute interactions, squares, and other transformed variables. Sociol Methodol. 2009;39(1):265–91.

26. Von Hippel PT. Regression with missing data: An improved strategy for analyzing multiply imputed data. Sociol Methodol. 2007;37(1):83–117.

27. Carpenter, JR, Roger, JH, Kenward M. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23(6):1352–71.

28 . Stuart, EA, Booil, J, Assessing the sensitivity of methods for estimating principal causal effects, Stat Methods Med Res., 2015; Vol. 24(6): 657–674