The effects of a one-year intensive lifestyle intervention on health-related quality of life and mental health in patients with type 2 diabetes: Statistical Analysis Plan for secondary analyses from the randomized U-TURN trial

Version 1.0 (May 13, 2019)

SAP authors:
Christopher Scott MacDonald, Robin Christensen, Sabrina Mai Nielsen, Julie Midtgaard.

Statistical analyst:
Sabrina M. Nielsen, MSc; Biostatistics fellow.

Statistical advisor:
Robin Christensen, MSc, PhD; Professor of Biostatistics and Clinical Epidemiology; Head of Musculoskeletal Statistics Unit.

Affiliations:
CSM: Department of CopenRehab, Faculty of Health Sciences, Copenhagen University, Copenhagen, Denmark AND Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
RC: Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen AND Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark
JM: The University Hospitals Centre for Health Research, Copenhagen University Hospital Rigshospitalet, Denmark; Department of Public Health, Section of Social Medicine, University of Copenhagen, Denmark
SMN: Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen AND Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark

Collaborators:
Robert Cummins, Henning Langberg, Mathias Ried-Larsen.

Trial registration: NCT02417012

Study population

The primary analyses will be based on the Intention to Treat (ITT) population, based on the Full Analysis Set. The ITT principle asserts the effect of a treatment policy (that is, the planned treatment regimen) rather than the actual treatment given (i.e. it is independent of adherence). This has the consequence that participants allocated to a treatment group should be followed up, assessed and analysed as members of that group irrespective of their compliance to the planned course of treatment (i.e. independent of withdrawals and cross-over phenomena). In general, it is advantageous to demonstrate a lack of sensitivity of the principal trial results to alternative choices of the set of participants analysed. In addition to the ITT-approaches, we will analyse the results as per protocol (described below).

We plan to conduct both an analysis of the full analysis set and a per protocol analysis, so that any differences between them can be explicitly discussed and interpreted. The data analysis based on the per-protocol population will still depend on the group allocation, but will have a stringent definition of whether participants adhered to the prescribed protocol.

We define the per protocol population as:

The particular outcome measure of interest is available both at baseline and at 12-month follow-up (i.e. complete case population). Further, specific for the individual groups:

U-TURN (intervention group)

- Attending at least four (of five [80%]) medical consultations
- Conducting ≥70% of all exercise sessions (supervised and un-supervised – as assessed by the exercise registration)
- Only gets the prescribed medications and/or the prescribed combination of medications according to the treatment algorithm

Standard care (comparator: control group):

- Attending at least four (of five [80%]) medical consultations
- Only gets the prescribed medications and/or the prescribed combination of medications according to the treatment algorithm

Primary analyses:

Our primary analyses will be based on the ITT population, including all randomized participants with available data at baseline. Missing data will be handled indirectly and statistically modeled using repeated-measures linear mixed models (see below). These models will be valid if data are 'Missing at Random' (MAR): "*Any systematic difference between the missing values and the observed values can be explained by differences in observed data*". Contrasts between groups will be estimated based on repeated-measures analysis of covariance applied in mixed linear models.

Statistical methods

Sample size considerations: The sample size in this study (according to the original protocol) was based on what was considered feasible (1) within the local context, and it enabled up to 120 participants to be enrolled in the trial period (29th of April 2015 to 17th of August 2017). The sample size of the main-study was truncated at 120 participants or the *N* reached at the end of the recruitment period. At the end of the recruitment period, 64 patients were randomized to the intervention group and 34 patients to the standard care group (3).

No formal power analysis was performed for the present secondary analysis study. However, assuming that 98 participants will be available from the ITT population (64 vs 34) we should have a sufficient statistical power (>90%) to detect a statistically significant difference between the groups corresponding to a moderate to large clinical effect size in any of the HRQoL domains. According to Cohen's guidelines for interpreting effect sizes a small effect size is 0.2, a moderate effect size is 0.5, whereas as large effectsize is 0.8 (4). For a two-sample pooled t-test of a normal mean difference with a two-sided significance level of 0.05, a total sample size of 96 (allocation ratio of 2 to 1) has a power of 95.5% to statistically detect a standardized mean difference (SMD) of 0.8 (i.e. Cohen's index). With the expected sample size in the ITT population, the secondary analyses should have a reasonable power to detect even a moderate clinical effect size (Cohen's index of 0.60 would correspond to a statistical power around 80%).

Statistical analyses: The primary statistical model will consist of repeated measures linear mixed models, which state that observed data consist of two parts; fixed effects and random effects. Fixed effects define the expected values of the observations, and random effects define the variance and covariances of the observations. In this study (with secondary analyses) from the U-turn trial with repeated measures, participants were randomly assigned to treatment groups (U-turn vs Standard of

Care), and observations were made at five time points (0, 6, 8, 10, 12) for each participant. Basically, there are two fixed-effect factors: group and time. Random effects result from variation between and within participants. We anticipate that measures on the same patient at different times are correlated, with measures taken close together in time being more highly correlated than measures taken far apart in time; observations on different participants will be assumed independent (5).

The objectives of repeated measure designs are to make inferences about the expected values of the observations, that is, about the means of the populations from which participants are sampled. This is done in terms of treatment and time effects in the model. Data will be analyzed using R, with the particular outcome variable at baseline level as a covariable - using a multilevel repeated measures random effects model with participants as the random effect factor based on arestricted maximum likelihood (REML) model.

The change in the SF36-PCS value will be the (primary) response variable, and the baseline value (one for each participant), treatment group (two levels), and time (five levels) will be included as covariates, as well as the interaction between treatment group and time, and Patient ID as a random effect. This statistical model holds all between-group comparisons at all assessment points (incl. baseline) and allows for evaluation of the average effect, as well as the trajectory over the time period from baseline to 12-month follow-up.

Categorical changes for dichotomous end points will be analyzed with the use of logistic regression with the same fixed effects and covariates as the respective analysis of covariance. Since Odds Ratios (ORs) for outcomes of common incidence either over- or under estimate the corresponding risk estimate, we will convert all the calculated OR values and 95% confidence intervals into approximate Risk Ratios (6). Following these analyses, the proportion of patients experiencing the outcome of interest will be reported for the groups and interpreted based on the number needed to treat (NNT) - communicating the effect size in absolute terms (7). The NNT is computed by taking the reciprocal of the absolute risk difference (RD); the NNT will indicate how many diabetes patients must be managed on average with U-turn intervention rather than standard care to achieve 1 additional good outcome.

For the statistical analyses, we will primarily use the statistical software R (version 3.3.3 or newer) (8) with the packages *lme4* (9) and *nlme* (10). The following codes will be used for the main analyses:

```
#Primary analysis of primary outcome:
```

```
lme(SF36PCS_change ~ factor(TIME) + GROUP + factor(TIME)*GROUP + SF36PCS_0, random=~1|PtID,
                  corr=corGaus(form=~TIME|PtID,nugget=TRUE), data=d)

#Analysis of secondary outcomes:
lme(OutcomeX_change ~ factor(TIME) + GROUP + factor(TIME)*GROUP + OutcomeX_0, random=~1|PtID,
                  corr=corGaus(form=~TIME|PtID,nugget=TRUE), data=d)
lm(OutcomeY_change ~ GROUP + OutcomeY _0, data = d)
glmer(OutcomeZ ~ TIMEfac + GROUP + TIMEfac*GROUP + OutcomeZ_0 + (1 | PtID), data = d, family = binomial)
```

Analysis populations: Handling of missing data and sensitivity analyses

As explained above, we plan to conduct both an analysis of the full analysis set (ITT population) and a per protocol analysis, so that any differences between them can be explicitly discussed and interpreted. Robustness is a concept that refers to the sensitivity of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis. Robustness implies that the treatment effect and primary conclusions of the trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches.

Sensitivity analyses: Loss to follow-up and missing data for various reasons is hard to avoid in randomized trials and in particular in trials like the U-turn trial where data collection has been finalized. We will apply the analysis framework suggested by White et al (2011) where missing data related to the ITT approach depends on making plausible assumptions about the missing data and including all participants in subsequent sensitivity analyses (11):

1. Attempt to follow up all randomized participants, even if they withdraw from allocated treatment (i.e., U-turn data collection already finalized)
2. Perform a main analysis of all observed data that are valid under a plausible assumption about the missing data (i.e. Model-based: data as observed; using linear mixed models assumes that data are 'Missing at Random' [MAR])
3. Perform sensitivity analyses to explore the effect of departures from the assumption made in the main analysis (i.e. A non-responder-imputation: using the value at baseline to replace missing data will correspond to a non-responder imputation; these models will potentially be valid even if data are 'Missing Not At Random' [MNAR])
4. Account for all randomized participants, at least in the sensitivity analyses (covered by #2 and #3 above plus the corresponding analyses based on the Per protocol population).

The interpretation of the corresponding statistical measures of uncertainty of the treatment effect and treatment comparisons will involve consideration of the potential contribution of bias to the *p*-value, 95% confidence interval, and inference in general (12).

#1+2: Our primary analysis population will be all participants with available data at baseline statistically modelled using repeated-measures linear mixed models (see above). These models will be valid if data are 'MAR'.

#3+4 Sensitivity: We will analyze all variables with missing data being replaced by imputation of the baseline level; i.e. interpreted as assuming that those who dropped out returned to their baseline level (13). These estimates could potentially be valid even if data are 'Missing Not At Random'.

Because the degree of potential confounding, when outcomes are missing, cannot be determined (the required data are by definition missing) the only statistically sound approach is sensitivity analyses, which involve assessing the robustness of the result to a range of plausible mechanisms responsible for the missing data (14).

Multiplicity Considerations

All applicable statistical tests will be 2-sided and will be performed using a 5% significance level. All confidence intervals presented will be 95% and two-sided. The selection of an appropriate statistical strategy for dealing with multiplicity (repeated statistical tests) is critical for performing reliable inferences and maximizing the probability of an apparent success in a clinical trial like U-Turn. Multiplicity considerations play a central role in the assessment of efficacy evidence in the presence of competing clinical objectives; i.e. the more comparisons that are made, the more likely it is that a comparison that appears to be significant will be falsely so. In order to preserve the family wise error rate of the multiple analyses, the multiplicity of the analyses of the primary and selected secondary efficacy outcomes will be adjusted using a gatekeeping procedure (15). The analyses will be performed in sequence until one of the analyses has failed to show the significant difference or all analyses have been completed at a statistical significance level of 0.05. The sequence of the analyses for the selected secondary efficacy outcomes are listed below.

Rank of the outcomes (including SF36-PCS) in the Hierarchy of Multiplicity Adjustment	
Outcomes assessed after 12 months	**Rank#**
Change in SF36-PCS, points	1
SF36-PCS responders, no (%)	2
Change in SF36-MCS, points	3
SF36-MCS responders, no (%)	4
Change in general health (SF36), points	5
Change in physical functioning (SF36), points	6
Change in physical role functioning (SF36), points	7
Change in bodily Pain (SF36), points	8
Change in vitality (SF36), points	9
Change in emotional role functioning (SF36), points	10
Change in mental health (SF36), points	11
Change in social functioning (SF36), points	12
Change in postiive affect derived from Global Mood Scale, points	13
Change in negative affect derived from Global Mood Scale, points	14
Change in mental Health continuum, points	15
Flourishing (MHC-SF), no (%)	16
Moderately mentally healthy (MHC-SF), no (%)	17
Languishing (MHC-SF), no (%)	18
Change in Warwick Edinboruogh Mental Well-being scale, points	19

A result of the this "gate keeping approach" is that formal comparison with respect to the first secondary outcome to be tested (change from baseline in the SF36-PCS responders measured by SF-36 at month 12) will be conducted conditional on the test results for primary efficacy outcome. Thus, if the test of change in SF36-PCS is statistically significant (i.e., two sided p-value ≤ 0.05), then the comparison for change from baseline in the SF36-PCS responders measured by SF-36 at month 12 will be performed at $\alpha = 0.05$. If the test result of SF36-PCS is not statistically significant, the formal statistical tests will not be interpreted as statistically significant for change in SF36-PCS responders at month 12 and for the remaining of secondary/exploratory endpoints.

For planned statistical tests that are not formally considered statistically significant as a result of aforementioned "gate keeping" multiplicity adjustment strategy (15), nominal 2-sided p-values (without adjustment for multiplicity) will still be computed as a measure of the strength of the association between the outcomes and the U-turn effect rather than formal test of hypotheses.

References:

1. Ried-Larsen M, Christensen R, Hansen KB, Johansen MY, Pedersen M, Zacho M, et al. Head-to-head comparison of intensive lifestyle intervention (U-TURN) versus conventional multifactorial care in patients with type 2 diabetes: protocol and rationale for an assessor-blinded, parallel group and randomised trial. BMJ Open. 2015;5(12):e009764.
2. Ware J, Snoww K, Kosinski M, Gandel B. SF36 Health Survey: Manual and Interpretation Guide. Lincoln, RI: Quality Metric, Inc, 1993. 1993.
3. Johansen MY, MacDonald CS, Hansen KB, Karstoft K, Christensen R, Pedersen M, et al. Effect of an Intensive Lifestyle Intervention on Glycemic Control in Patients With Type 2 Diabetes: A Randomized Clinical Trial. JAMA. 2017;318(7):637-46.
4. Cohen J. A power primer. Psychol Bull. 1992;112(1):155-9.
5. Diggle PJ. An approach to the analysis of repeated measurements. Biometrics. 1988;44(4):959-71.
6. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA. 1998;280(19):1690-1.
7. Saver JL, Lewis RJ. Number Needed to Treat: Conveying the Likelihood of a Therapeutic Effect. JAMA. 2019.
8. Team RC. R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria.; 2013 [Available from: http://www.R-project.org/.
9. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1):1-48.
10. Pinheiro J, Bates D, DebRoy S, Sarkar D, Heisterkamp S, Van Willigen B, et al. Linear and Nonlinear Mixed Effects Models 2018 [
11. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ. 2011;342:d40.
12. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012;367(14):1355-60.
13. Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med. 2003;348(21):2136-7.
14. DeMets DL, Cook T. Challenges of Non-Intention-to-Treat Analyses. JAMA. 2019;321(2):145-6.
15. Dmitrienko A, D'Agostino RB, Sr. Multiplicity Considerations in Clinical Trials. N Engl J Med. 2018;378(22):2115-22.