Supplementary Appendix

Effect of self-managed lifestyle treatment on glycemic control in patients with type 2 diabetes

TABLE OF CONTENTS

SUPPLEMENTARY DISCUSSION: COST-EFFECTIVENESS ANALYSIS

DETAILED RESULTS OF THE BASE CASE SCENARIO

The baseline characteristics and changes of relevant biomarkers were used to estimate the future risk of complications in patients using the tool biweekly, monthly and bimonthly and patients on usual care, respectively (Supplementary Figure 3). These data, together with information on costs for glucose-lowering treatment and diabetic complications, productivity loss and operating expenses was used to model the overall cost-effectiveness of the tool compared with usual care during a simulated 20-year time horizon (Supplementary Table 20).

BASECASE SCENARIO

The incremental gain of QALYs in users of the tool is mainly attributed to increased survival, improved quality-oflife, and reduced BMI. The changes in productivity loss had only a small effect on overall cost increments, which is a result of the mean baseline age being 62 years, i.e. close to the age of retirement.

Overall, the tool was dominant over usual care (more effective and less costly). In the base case scenario, the total estimated cost-savings amounted to \$4,116 per person over 20 years compared to usual care.

SENSITIVITY ANALYSIS

To study how sensitive the model results are to changes in key parameters and assumptions, we tested eight different scenarios in addition to the base case scenario. These include:

1,2) A 10-year and a 40-year time horizon

Time horizon is an important parameter, as diabetic complications develop over extended time. While a time horizon of 20 years was used in the base case scenario, we also examined the effects over 10 and 40 years, respectively. In the 40-year scenario, it was assumed that treatment was intensified with additional basal insulin of 0.1 IU per kg of body weight when HbA1c reached a threshold level of 10.1% (87 mmol/mol). This was based on observations of the average HbA1c levels for initiation of insulin therapy in the ANDIS registry. The anticipated effects of insulin treatment on HbA1c and BMI was based on previously reported data.³⁰ These considerations were not relevant for the other time horizons, as average HbA1c did not drift to the threshold level in this cohort during shorter time periods.

Compared with the base case scenario, the 10-year time horizon resulted in lower QALY gains and cost savings, because fewer long-term complications are taken into account. Increasing the time horizon from 20 to 40 years did not have a major impact on total costs and QALYs, mainly because of discounting effects and limited survival during the entire 40-year period.

3) Inclusion of changes in HbA1c but not secondary variables

As HbA1c was the primary study variable, we investigated a scenario where no changes of secondary variables were included in the model. In this scenario, cost savings were similar to the base case results, while the QALY gain was lower, mainly because changes in quality-of-life and BMI were not considered.

4) Linear increase of HbA1c from first year in both users and controls

Recent meta-analyses have shown that the average metabolic response to digital lifestyle interventions declines after six months.⁷ The continuous improvement of HbA1c observed during the entire follow-up period in the present study suggests that the tool leads to more sustainable effects. In the base case, a linear increase of HbA1c (0.15% per year) was assumed to start after three years in users of the tool and after one year in controls, an assumption that was based on the study observations. We analyzed also a more conservative scenario, in which a similar HbA1c increase over time, starting after the first year, was assumed in both groups. This resulted in a change of cost-savings from \$4,116 to \$3,208 per user over 20 years.

5) Inclusion of all secondary variables

The base case scenario included only variables with indicated differences between study participants and controls, based on 95% confidence intervals. This approach is often applied in cost-effectiveness analyses in order to remain conservative. It could, however, be argued that all point estimates should be used in the model. We therefore examined a scenario that also included the observed mean differences of total cholesterol, LDL cholesterol and triglycerides between study participants and controls. This did not have a noticeable impact on the results.

6) Exclusion of BMI-related effects on quality-of-life.

The base case scenario included both the direct effects on quality-of-life, as measured by EQ-5D-5L, and indirect effects related to decreased body weight (reduced body weight in patients with type-2 diabetes, as observed in the study, is generally associated with a positive change in quality-of-life).⁵⁵ As this could potentially lead to an overestimation of the gain in quality-of-life, we studied a scenario where the model was adjusted to exclude any indirect BMI-related quality-of-life effects and only consider the direct, measured effects. The model results were robust in this respect, and the gain of QALY was only slightly decreased.

7) Effects based on usage at least once per month including only HbA1c

We analyzed the cost-effectiveness when the tool was used at least once per month but included only changes in HbA1c, no secondary variables (corresponding to scenario 3 for recommended usage).

8) Effects based on usage at least once every other month including only HbA1c

We analyzed the cost-effectiveness when the tool was used at least every other month but included only changes in HbA1c, no secondary variables (corresponding to scenario 3 above).

In each of the scenarios of the sensitivity analysis, the tool resulted in cost savings and more QALYs compared to usual care (Supplementary Table 21). The hypothetical maximum yearly operating expenses of the tool in order to remain cost-saving ("break-even cost") was calculated as total cost savings divided by total remaining life expectancy (ranging from 7.8 to 12.2 years in the different scenarios). The break-even cost varied between \$110 and \$411 per patient (\$369 in the base case scenario), which is well above the actual cost of \$7.5. Non-users do not incur any costs and were therefore not included in the models.

The cost-effectiveness analysis was based on Swedish cost data. Given the low operating expenses and assuming that similar metabolic outcomes can be obtained in other contexts, the main conclusions should be applicable to other country-specific settings as well. Since the results indicated that the tool would be cost-saving up to yearly operating expenses of approximately \$300 per user, there is room for implementation costs and language adaptations (beyond the current English and Swedish versions) while still obtaining cost-saving outcomes.

CLINICAL CONTEXT

Cost-effectiveness models as the one used here are based on several assumptions that make exact estimations uncertain. In this case, however, a key driver of the results is the low marginal cost of offering the tool to patients. This means that also modest benefits in terms of reducing very costly diabetes complications will make the tool cost-effective. The cost per patient can be put into perspective by comparing it to the annual expenses of metformin, which are at least 10 times higher. This is particularly important considering the large unmet clinical need for scalable solutions to support lifestyle self-management, not least in low-income settings. The intention is to make the tool available to users via 1) healthcare providers (in conjunction with other lifestyle management activities) and 2) direct access by patients.

From the onset, the aim has been to develop a tool that is freely available to the individual patient. Many digital solutions that are technically scalable still fail to reach large patient groups because of associated costs. Some solutions require expensive coaching or enhanced healthcare support that limit broad applicability. Several tools also apply user fees or require patients to give back health data to be used for third party purposes.

For many drug and lifestyle interventions it is critical to compare the costs and/or side effects versus overall efficacy. In this case, the tool is provided at low cost per user and has no known negative side effects. It means that even if the tool is not used by all patients, it could still be of considerable clinical importance. A clinician or healthcare provider could choose to 1) provide the tool only to patients with MOD characteristics (or apply a BMI cut-off), which is likely to give substantial benefits in that population, analogous to a tailored drug, or 2) offer it to all patients with type 2 diabetes, as any additional patient incurs a minimal cost and those who use it regularly are likely to get overall improvements.

SUPPLEMENTARY FIGURES

Supplementary Figure 1*.* **User journey on the tool.**

Upper left: At each login, participants can choose one of 80 different themes. Themes are presented with a brief introductory text that is displayed when clicking on the theme title.

Upper right: A theme consists of tests in order to explore current habits, texts with health information and exercises to learn behavior change techniques.

Lower left: At the end of each theme, participants ask a question to themselves on how to implement the content and insights from the theme. They are encouraged to reflect on their question in daily life, explore different options and return to the tool within two weeks.

Lower right: When returning, participants choose a new theme. They can also follow their personal journey and the questions they asked, and revisit themes they have previously done and found valuable.

Supplementary Figure 2. Study profile as CONSORT diagram.

A total of 370 individuals were randomized to usual care or to access the tool. Twenty-six of the participants randomized to the tool and eight on usual care discontinued between first and second visit (reported reasons in the figure). Change of HbA1c from first to second visit was compared between randomization groups.

After the second visit, the randomization groups were merged to enable all participants to use the tool during an extended open-label period of 730 days on average. The change of HbA1c from baseline was compared between participants and matched controls on usual care. Baseline for study participants during the long-term assessment period was defined as HbA1c before accessing the tool, which was the first visit for those randomized to immediate access and the second visit for those initially randomized to wait. Twenty participants who had been on wait discontinued at the second visit and did therefore not provide any data beyond baseline, and 42 participants were excluded from analysis because of changed glucose-lowering medication just after baseline assessments. If participants changed glucose-lowering medicines during the follow-up period, then data from the last visit with unchanged medication was used for analysis.

Supplementary Figure 3. Estimated incidence of microvascular and macrovascular complications.

Baseline characteristics and changes of relevant biomarkers in participants using the tool and matched controls were used to model the future risk of complications. Estimated cumulative incidence data for first stroke, heart failure, background retinopathy and macroalbuminuria are shown for usage of the tool biweekly, at least monthly and at least every other month, respectively. Each cycle in the model corresponds to one year.

SUPPLEMENTARY TABLES

Supplementary Table 1. Demographic and baseline characteristics of patients from ANDIS in longitudinal analysis of psychological factors and glycemic control. *

* Data are n (%) or mean (SD). Data on socioeconomic status and education were not available from all patients.

**The body mass index is the weight in kilograms divided by the square of the height in meters.

***Glycated hemoglobin level (HbA1c) was analyzed according to International Federation of Clinical Chemistry (IFCC) standard.

****Basic level refers to up to 9 years of education; medium level is up 12 years of education.

Supplementary Table 2. Association between questionnaire scores and HbA1c across visits. *

*The association between questionnaire scores and HbA1c across time (average 32 months) was analyzed using a linear model in which values from each visit were included as discrete observations and grouped by study subject. The unstandardized beta coefficient from the linear model is reported with 95% confidence intervals (CI). **Negative beta coefficient implies that increased questionnaire score is associated with reduced HbA1c over time.

* Data are n (%) or mean (SD).

**The body mass index is the weight in kilograms divided by the square of the height in meters.

***Basic level refers to up to nine years of education; medium level is up 12 years of education.

*The behavior change techniques (BCT) included in themes on the tool are based on the taxonomy proposed by Michie and colleagues⁵⁰ and adapted to digital form.

**A total of 24 BCT were used altogether out of the 40 techniques⁵⁰ in the taxonomy.

*Up to four behavior change techniques (BCT) were incorporated into each theme. They are ranked according to predominance within the theme.

**Core questions refer to a set of themes that focus on the different aspects of existential health that the World Health Organization (WHO) has proposed.^{53,54} These themes aim to stimulate questions on overall life context and how it relates to current habits and disease coping. The themes are more essayistic in style compared with other themes on the tool and may include references to e.g. Socrates, Victor Frankl, Martin Buber, Dag Hammarskjöld and others as starting points to inspire further reflections by the individual user.

*Data are n (%) or mean (SD). Data on glucose-lowering medication, socioeconomic status and education were not available from all.

**The body mass index is the weight in kilograms divided by the square of the height in meters.

***Basic level refers to up to 9 years of education; medium level is up 12 years of education.

* Study participants were categorized as MOD and non-MOD, respectively, using the ANDIS clustering methodology.¹⁴ Data are mean (SD).

**The body mass index is the weight in kilograms divided by the square of the height in meters.

* Data are n (%) or mean (SD). Data on glucose-lowering medication, socioeconomic status and education were not available from all participants.

**A total of 34 participants (26 randomized to access and 8 randomized to wait) discontinued between first and second visit during the randomization period. Another 20 participants who had been on wait discontinued at second visit and did not provide any observational data beyond baseline. Participants who changed glucose-lowering medication immediately after baseline (n=42) were not included in the follow-up analyses but remained in the study and are therefore not reported as discontinued.

***The body mass index is the weight in kilograms divided by the square of the height in meters.

****Basic level refers to up to 9 years of education; medium level is up 12 years of education.

* The pattern of usage was observed in study participants from baseline to end of follow-up. Data are n (%) or mean (SD). Data on glucose-lowering medication, socioeconomic status and education were not available from all participants. Biweekly usage refers to biweekly usage during at least a one-year time frame in the follow-up period (a patient may e.g. use it less than biweekly during the first year and then use it biweekly during the second year, but also low-activity periods were included in the analyses to investigate the effects over the entire follow-up). ** Physical activity measured at baseline by IPAQ

Supplementary Table 10. Weighted descriptive statistics when adjusting for potential confounders related to frequent usage of the tool*

*Analysis weights based on propensity scores were used to adjust for potential confounders related to frequent usage of the tool. The weights are based on the probability of using the tool as recommended and are applied to statistically adjust the composition of both the group of non-users and the group using the tool biweekly to estimate the mean difference between groups if they were more comparable in terms of baseline characteristics. (The number of individuals in adjusted groups will not always be integers because of the applied weights). Data are n (%) or mean (SD).

**The weights were optimized to minimize the standard mean difference (SMD) between the groups (a smaller SMD indicates more comparable groups). The unweighted SMD and the SMD when applying weights are given for each baseline variable.

***The number of categories was reduced to enable weight optimization (one category refers to no treatment or oral only; the other refers to insulin with or without combined oral treatment).

****The number of categories was reduced to enable weight (one category refers to basic or medium level; the other refers to college/university education).

*Matching characteristics for study participants and control patients from ANDIS. Data are n (%) or mean (SD).

**Controls were matched exactly on gender and on Mahalanobis distance based on age, body mass index and glycated hemoglobin level.

***Balance after matching was evaluated using the standardized mean difference (SMD).

****The body mass index is the weight in kilograms divided by the square of the height in meters.

*Table presents baseline data that were not used as matching variables (see Supplementary Table 11) between study participants and controls. Data are means (SD).

**EQ-5D-5L data were only available for study participants. Score ranges from 0 to 1 with higher score indicating better quality-of-life.

*Trajectories of HbA1c progression were analyzed in study participants when the tool was used as recommended (biweekly), at least monthly and at least bimonthly and were also analyzed in patients with type-2 diabetes in ANDIS. Individuals who had a baseline HbA1c at 52 mmol/mol or above and at least three HbA1c measurements over a three-year time frame with no known medication changes were included. Average follow-up period was 632 days in ANDIS and 730 for users of the tool. Data are presented as number of patients (% of total number of patients eligible for analysis).

**Four common patterns of HbA1c trajectories were identified. The number of patients corresponding to each pattern was compared between ANDIS patients and users of the tool using Fisher's exact test.

***Fisher's exact test statistic <0.001.

****Fisher's exact test statistic <0.05.

*Table presents results from mediation analyses with usage of tool, defined as number of completed themes, as the independent variable, change of HbA1c as the outcome variable and change of body weight or change of HOMA2- IR as mediator. Analyses are decomposed into 1) effect of the independent variable on the outcome variable, 2) effect of the independent and mediator variable, respectively, on outcome, and 3) the total effect of the independent variable on the outcome, the indirect effect via the mediator and the fraction of the total effect that is estimated to be an indirect effect. The total effect is not entirely similar between the analyses since the number of participants with available data on body weight and HOMA2-IR at time points corresponding to those of HbA1c measures differed (n=258 and n=244, respectively).

**Point estimates and 95% confidence intervals of regression coefficients.

***Baseline BMI and sex were used as moderators in the model.

****Baseline HOMA2-IR and sex were used as moderators in the model.

*Overall scores of responders and non-responders for every BCT across all themes during one year (see also Methods and Supplementary Table 4-5).

**The distribution of responder and non-responder scores was analyzed by chi-square tests to assess whether a BCT was overrepresented in themes completed by participants who responded to the tool by reduced HbA1c. BCT in bold have chi-square statistics corresponding to a *P* value of 0.05 or less.

*At every login, participants were given the opportunity to comment on whether their questions had resulted in lifestyle changes or not. The number of reported lifestyle changes for participants asking more abstract than concrete questions are presented. These analyses focused on data during the first year.

The number of reported lifestyle changes in participants asking more concrete than abstract questions *The association between the number of questions and change of HbA1c from baseline to one year was analyzed by linear regression. The beta coefficients from the regression models are reported with 95% confidence intervals. ****The proportion of abstract to concrete questions in responders (lower HbA1c at one year compared to baseline) and non-responders (similar or higher HbA1c at one year compared to baseline), respectively, in participants using the tool as recommended.

Data are means (SD); n refers to the number of participants included in the analyses.

*Since the model is based on a cycle length of one year, the change of relevant biomarkers from baseline to one year in participants using the tool as recommended and matched controls on usual care, respectively, was included in the model

**The model equations are based on HbA1c measured in NGSP units (%). HbA1c in IFCC units (mmol/mol) was converted to NGSP (%) using the formula $HbA1c_{\text{NGSP}} = (0.915 * HbA1c_{\text{FCC}}) + 2.15$.

***The body mass index is the weight in kilograms divided by the square of the height in meters.

****Data on total cholesterol, LDL cholesterol and triglycerides were only used in the sensitivity analysis.

*Values are means of baseline data for participants using the tool and matched controls on usual care. Since a few biomarkers included in the model were not reported in detail in the ANDIS registry, we extracted corresponding baseline data from the UK Prospective Diabetes Study (UKPDS) or clinical registries for patients with similar age, disease duration and glycemic control, in order to obtain as precise measures as possible.

Estimated glomerular filtration rate is presented as ml per minute per 1.73 square meter of body surface area. *Assumed to be zero from data on cohorts with similar baseline characteristics.^{58,59}

*Based on data from study participants. Percentages add up to more than 100 because of co-treatment.

**Mean yearly cost per patient including consumables. Insulin costs include 1 needle, 1 test strip and 1 lancet per day. Glucagon-like peptide-1 analogue cost includes 1 needle per day.

***Approximately 50% using each compound.

****Basal neutral protamine Hagedorn (NPH) insulin was used for calculations to remain conservative.

*Table presents values per person during 20 years, disaggregated into components that contribute to total QALYs and costs in patients on usual care and patients using the tool in addition to usual care, respectively. Increment refers to the difference between the groups.

**Treatment costs include usual care plus tool.

***Incremental cost-effectiveness ratio is incremental total costs divided by incremental QALYs. Dominant outcome implies that the tool is more effective and cost-saving compared to usual care.

* Table presents QALYs and costs per person during 20 years in patients on usual care and patients using the tool in addition to usual care, respectively. Increment refers to the difference between the groups. Data are shown for the base case and eight other scenarios, as detailed in the Supplementary Discussion.

**The hypothetical maximum yearly operating costs for the tool in order to remain cost-saving ("break-even") was calculated for each scenario as total cost savings divided by total remaining life expectancy.