**A. SIGNIFICANCE**

**A-1. The Multilayered Challenge of Type 1 Diabetes Management and Limitations of Current Interventions**

Individuals with type 1 diabetes (T1D) must constantly self-manage their diet and insulin to mitigate the blood glucose fluctuations associated with common occurrences such as snacking, exercise, and mental stress. In addition, many of them have modifiable disease risk factors including overweight (60%), hypertension (40%), dyslipidemia (60%), and inadequate exercise (67%-82%)48 which exacerbate their already elevated risk of cardiovascular mortality47. Existing T1D self-management interventions (psychoeducation13-15, diabetes devices8-12, digital platforms2-7) have improved blood glucose control, but none showed treatment effect on cardiometabolic cardiovascular disease risk factors (body mass index, blood pressure, lipids) or lifestyle behaviors (exercise, diet) when measured with reliable instruments20,57. Thus, T1D self-management interventions that address a broader range of health targets than glycemic control alone are needed.

**A-2. The Benefits and Challenges of Exercise**

Exercise interventions could provide a novel solution for effective diabetes self-management and have been linked in dozens of studies with positive health outcomes in T1D (aerobic capacity, glycemic control, lipid profile)58. Unfortunately, most people with T1D do not regularly exercise59,60 as it poses unique challenges61. First, exercise can dysregulate blood glucose, leading to hypo- and/or hyperglycemia during and up to 24hr after. The direction and magnitude of these changes depend upon many factors, including exercise characteristics (intensity, muscle groups), individual characteristics (e.g., endogenous insulin sensitivity), and physiological context (pre-exercise blood glucose level, insulin- and carbohydrates-on-board, and concentration of counter-regulatory hormones)62. This glycemic dysregulation poses uncertainty regarding *adjusting diet and insulin around exercise to avoid dysregulation* and *tailoring the exercise prescription to optimize long-term impact upon glycemic control*. Second, this unpredictability leads to *fear of hypoglycemia* which is the most reported barrier to exercise among people with T1D49. Nearly all adults with T1D report some fear of hypoglycemia, and limited available data indicate that this fear varies daily50. Therefore, it may interfere not only with initiation but also consistency of exercise. Third, *other emotional factors* including stress, negative affect, and fatigue correlate with exercise behavior at a daily level among many adult populations50,53,63-67 including those with T1D52. Other common correlates of exercise among adults are *adequacy of prior night’s sleep*68*, weather conditions*69, and *competing time demands*70. Innovative strategies are needed to better understand and overcome these potential barriers for those with T1D.

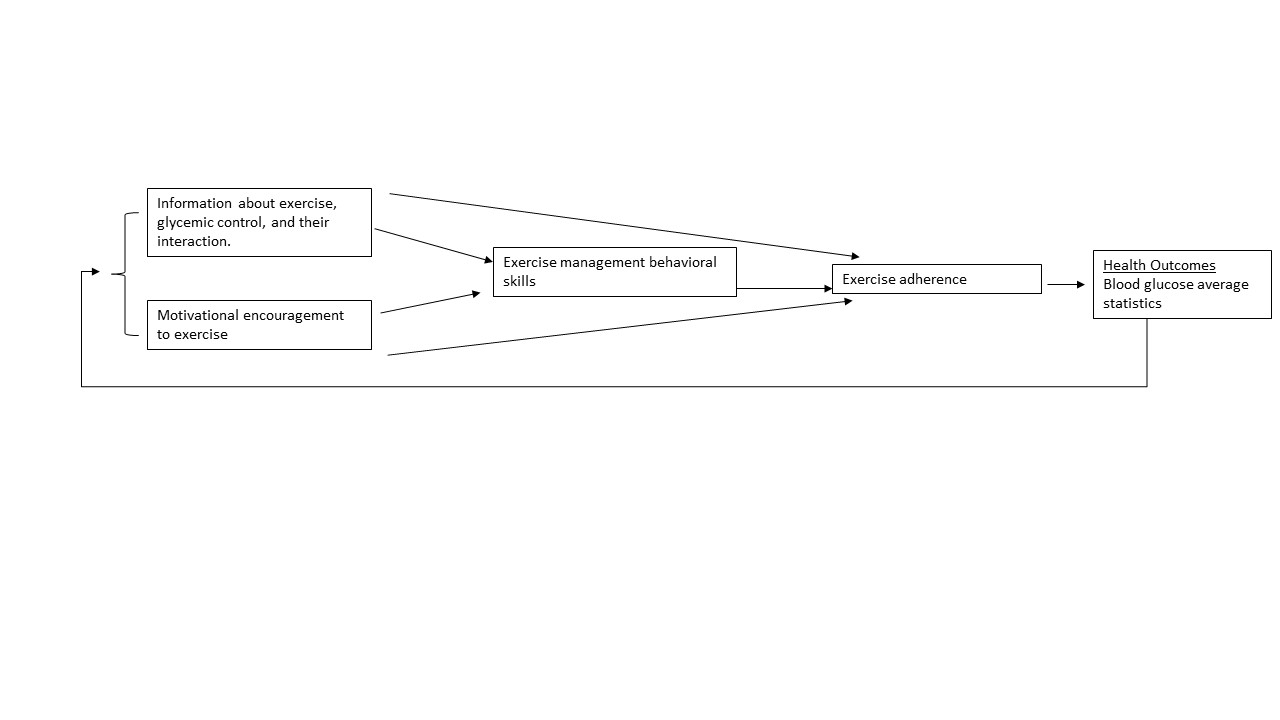
**A-3. Knowledge and Technology Supporting Exercise with Type 1 Diabetes**

Patients and their providers are encouraged to follow consensus statement guidelines for monitoring blood glucose levels and adjusting diet and insulin before, during, and after exercise to limit safety hazards and optimize benefits61. In addition, continuous glucose monitoring (CGM) is a now widely used (30%)71 technology among people with T1D that can support them to follow these exercise recommendations. First, by giving glucose trends rather than “snapshot” values, CGM helps the patient identify risk of exercise-induced hypo- or hyper-glycemia and adjust insulin and diet accordingly and mitigate fear of hypoglycemia. Second, CGM presents users with time-series feedback of their glucose levels, allowing them to infer whether the timing of behaviors such as exercise (e.g., morning vs. afternoon, after vs. before insulin bolus) contributes to achievement of the target range54. Patients with *type 2* diabetes have shown greater motivation to increase exercise when presented with feedback on the intersection of exercise and CGM data72. Third, patients using CGM report better confidence in glycemic control and motivation to more carefully consider lifestyle choices73. This combination of knowledge and technology holds great potential to help people with T1D overcome barriers to exercise, although research on CGM and exercise promotion is limited74. Yet, evidence-based interventions to translate CGM feedback into sustainable adherence to exercise-related behaviors are lacking74.

**A-4. Translating this Knowledge and Technology to Practice**

Most CGM users regularly view the real-time display of current blood glucose values and trending direction but few use mobile medical applications to comprehensively view CGM data (17%)71. These metrics are discussed with providers at clinic appointments71, but provider discussions often do not translate to daily self-management decisions by patients75. *Studies working with people with T1D to develop and test how they could independently use in-depth CGM information to promote safe, consistent, and effective exercise represent an underexplored, promising area for intervention development.*

We propose to address the gap between the available relevant medical information and translation to successful exercise by people with T1D by applying the Information-Motivation-Behavioral Skills (IMBS) model (Figure 1)55. The IMBS model has been successfully applied to self-management interventions for T1D76,77 and other chronic diseases78-80 and explained variance in self-management outcomes55,81,82. In the present application, we propose to develop mobile tools that leverage the diabetes literature, CGM data, and informatics to derive the needed components of IMBS for exercise (information about exercise, motivational encouragement to exercise, and feedback on long-term exercise outcomes) and deliver it to people with T1D.

**Figure 1.** Information-motivation-behavioral skills (IMBS) model for exercise with type 1 diabetes.

**A-5. Preliminary Data (human-delivered IMBS-based intervention)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2.** Components of human-delivered intervention MOVE-CGM ranked by satisfaction rating. | | | |
| **Component** | **Informational Aspects** | **Motivational Aspects** | **Rating (/ 5)** |
| Monthly motivational enhancement therapy session including biosensor feedback | \*Reports of CGM and exercise data  \*Advice to change | \*Presentation of data to draw inferences about CGM-exercise relationship  \*Person-centered therapy | 4.5±0.6 |
| Exercise videos | \*Exercise techniques,  \*Exercise benefits for T1D | \*Visual display of expected health impact† | 3.9±0.9 |
| Text-based exercise coach | \*Exercise advice | \*Personalization of messages | 3.3±0.8 |
| Self-monitoring of exercise and barriers | \*Common barriers to exercise | \*Self-monitoring | 2.2±0.6 |
| Sample was N=18, 56% women, 42±14 years old, 7.1±1.2% A1c, 30±5 kg/m2., previously sedentary.  Exercised 2.5±1.7 times per week (62.2±9.5% age-predicted maximum heart rate). Overall intervention rating was 3.6±0.9 out of 5. CGM adherence 90.4±8.8% of possible readings, daily diary adherence 91.4±14.1%  †Eg, diagram of areas where muscle insulin sensitivity improves for a specific exercise. | | | |

Participants gave the highest rating to monthly calls (Table 2) because they improved exercise management behavioral skills and motivation but stated a need for more frequent and sustained contact that was not satisfied by the texting and self-monitoring. *Together, these data suggest that our human-delivered IMBS intervention (MOVE-CGM v1.0) holds promise to promote exercise adherence, but a new version with improved automated tools (MOVE-CGM v2.0) is required to deliver more intensive, sustained support.*

**A-6. Digital subtyping for a future precision medicine approach (Exploratory Aim)**

Precision medicine means finding the right interventions for the right individuals at the right time. Machine learning upon biobehavioral data can support precision medicine by digitally phenotyping individuals’ unique behaviors and bioprocesses to help target interventions. For example, mathematical features of hip and wrist movement (e.g., mean, skewness, variability, consistency across various lag times) predict momentary symptoms and long-term onset of social anxiety disorder, depression83-86, and type 2 diabetes87. Further value to these predictions has been added by mathematical features of CGM (beyond the standard clinical summary statistics)88. These are “supervised” analyses meaning the dependent variable is a predefined clinical diagnosis such as type 2 diabetes.

An underexplored opportunity is conducting such analyses “unsupervised”: segregating and identifying individuals with similar biobehavioral data profiles, then profiling these clusters for characteristics that may be clinically meaningful or generate useful hypotheses for future analyses. For example, machine learning using polysomnography data among individuals with insomnia identified three intriguing polysomnography profile clusters, that did not conform to standard diagnostic insomnia subtypes. First, they were predictive of other sleep traits such as cortical arousal which could flag possible abnormalities warranting targeted intervention89. Second, they predicted the responsiveness of insomnia symptoms to a behavioral sleep treatment intervention90. We hypothesize that the biobehavioral data in our clinical trial (exercise video completion, other physical activity registered by Fitbit, CGM, insulin, mood, sleep) will yield biobehavioral subtypes within the population of sedentary adults with T1D that similarly flag intervention targets or predict intervention responsiveness.

**B. INNOVATION**

*The proposal is innovative in three ways:*

1. Current mobile physical activity applications encourage short, low-intensity bouts by addressing momentary barriers (e.g., “take another 500 steps before the end of the hour even if you feel fatigued”)91. This approach does not support the intensive, multimodal exercise recommended for T1D because the condition requires planning around typical daily commitments and T1D-specific management needs. Each pilot participant reported “not enough time” as their top barrier to exercise (p’s <0.05), indicating the need for intervention at the start of the day when there is time to plan. Our just-in-time adaptive intervention (Sect. C-3e2) is therefore targeted to days rather than moments of high vulnerability to missing exercise.
2. Current Bayesian models forecast blood glucose behavior based on insulin, physical activity, and diet92-94. However, there is a need for tools to efficiently evaluate whether an intervention (e.g., exercise) coincides with significant blood glucose deviation from the forecast after capturing and adjusting for blood glucose’s myriad of covariates (insulin, active energy expenditure outside of structured exercise, carbohydrates, glucagon, plasma cortisol, epinephrine) in a non-laboratory setting. To fill this need, we have developed a customized wrapper software that can process, prepare, and register a high number of high-frequency biomedical outcomes and covariates24. The proposed work will configure this software for communication with multiple biosensors to capture blood glucose and covariates (estimated insulin-on-board, active energy expenditure outside of structured exercise, carbohydrates). We have demonstrated that this software could accommodate a greater number of covariates24, anticipating the future when sensors will capture additional covariates of blood glucose such as directly sensed insulin, diet, glucagon, cortisol, and epinephrine.

3) Prior reports used supervised machine learning to identify digital phenotype predictors of established clinical outcomes83-87, whereas our exploratory aim proposes unsupervised machine learning to identify digital phenotype clusters that we will evaluate for characteristics, such as intervention responsiveness, that may be clinically meaningful (Sect. A-6). Our approach, variational autoencoding, is suitable for this goal because it evaluates performance based upon the ability to reproducibly reconstruct data, rather than comparing data to ground truth or external labels.

**C. APPROACH**

**C-1. Proposal Rigor and Reproducibility**

Exercise promotion for T1D is underexplored, and interventions to date are of limited efficacy. We will account for this weakness and achieve a robust and unbiased approach by developing tools based upon rigorous informatics, consensus exercise guidelines for T1D, stakeholder input, evaluation, and refinement. In addition, CGM has been rigorously documented to improve glycemic control8-10,12 and we will use it for exercise as a new target outcome. Finally, we are positioned to recruit equal numbers of men and women (see Inclusion of Women and Minorities), and sex will be accounted for as a between-subject factor in statistical analysis.

**C-2. Alpha Version**

**C-2a. Research Design**. We will conduct a single-group 4-week feasibility trial (N=24). All participants will be provided with biosensors–fitness watch (Fitbit Charge 4), smartphone diary software (PiLR mobile), CGM (Dexcom G6) if not currently using their own, insulin smartpen (Companion Medical Inpen) if not a pump user – and oriented to their respective commercial smartphone dashboards. They will use the devices during two weeks of normal daily living, our standard window to eliminate familiarization effects of diabetes technology1. Then, they will use the biosensors with the mobile app for 4 weeks, with weekly user satisfaction interviews.

**C-2b.** **Enrollment.** Inclusion criteria will be 30-65yr old with ≥6mo T1D diagnosis, sedentary (<1.0 exercise sessions/wk)95,96, smartphone ownership, and English literacy. Exclusion criteria will be chronic disease or physical disability that would influence exercise intervention, diabetic ketoacidosis not clearly related to pump site failure in past 6 months, >1 episode of severe hypoglycemia (defined as requiring assistance from another person to treat or complication such as physical injury or coma) in past 6 months, home blood pressure >145 mmHg systolic or >90 mmHg diastolic, myocardial infarction or angina in past 6 months, renal failure, pregnancy, cognitive impairment, severe retinopathy or neuropathy. We will recruit nationally using scalable strategies established from my pilot27. Eligible volunteers will complete a medical history phone screening, and those deemed eligible by Drs. Weinzimer or Nally will be invited to a televideo consenting visit, after which they will schedule a televideo intake visit to be fitted for biosensors (mailed in advance). Two weeks later, they will receive the mobile app and human-guided setup (Sect. C-2c1). *No in-person visits will occur.*

**C-3c. Mobile App**

|  |  |  |
| --- | --- | --- |
| **Table 3.** App components and fit with Information-Motivation-Behavioral Skills (IMBS) model | | |
| **Component** | **Informational Aspects** | **Motivational Aspects** |
| **Tool 1.** Exercise videos | \*Exercise techniques,  \*Exercise benefits for T1D | \*Visual display of expected health impact† |
| **Tool 2.** Daily just-in-time adaptive text messages to overcome exercise barriers at times of vulnerability | \*Advice about overcoming barriers | \*Personalization of message timing and content |
| **Tool 3.** Weekly personalized review of short-term exercise safety hazards with tips to avoid them | \*Glycemic patterns that result in safety hazards  \*Tips to avoid them | \*Personalization of tips according to occurrence of safety hazards |
| **Tool 4.** Monthly personalized feedback of long-term impact of exercise on blood glucose control | \*Exercise achievement metrics (frequency, intensity, duration, type)  \*Causal impact of exercise to blood glucose control | \*Personalized encouragement based on causal impact results |

**C-2c1. Access to a library of 200 exercise videos following brief human orientation consultation (Tool #1).** Our ~200 exercise videos (GlucoseZone™, New Haven, CT) include a combination of aerobic, resistance, flexibility, and balance exercises, in accord with American Diabetes Association recommendations. They can be filtered by difficulty level, duration, type of exercise (aerobic vs. resistance), muscle groups, mobility limitations (knee-friendly, shoulder-friendly, chair-based), indoor/outdoor, and home equipment availability. Using these filters, participants can select single sessions or 30-session series that progress in difficulty and duration. Participants self-schedule exercise time in their home environment or outdoor guided walking routes. Participants may select routines requiring no equipment, use a $10 set of resistance bands we recommend, use their dumbbells or household items like frozen water bottles. A fitness professional certified by GlucoseZone in exercise for diabetes schedules a 20min phone call to help download the app, discuss exercise levels and goals, and recommend starting videos and series.

**C-2c2. Daily just-in-time adaptive text messages to overcome exercise barriers on vulnerable days (Tool #2).** The MEI PiLR™ Health cloud will generate data about exercise adherence by tracking usage of exercise guidance (videos or guided outdoor walking) through GlucoseZone’s API. Fitbit heart rate readings (mean error ±3%97) verify exercise timing, duration, and intensity. A Borg Rating of Perceived Exertion98 prompt captures intensity for non-aerobic exercises such as weightlifting. It will also record contextual variables that may predict vulnerability to missing exercise: 1) 24hr blood glucose derived from interstitial vales sampled every 5min by the CGM (mean error 9%-10%99 even during exercise100) and expressed by consensus metrics (coefficient of variation, mean, % time >180, 70-180, <70mg/dL101. 2) Mood states (positive affect, negative affect, energy, fatigue) previously correlated with exercise within-participants53,63-67 assessed by 1- to 4-item subscales (ɑ=0.76 to 0.95102,103) each morning at self-expected waketime and immediately pre- and post-exercise using prompted surveys on the PiLR mobile diary. 3) Fear of hypoglycemia assessed by a single item each morning “considering your schedule today, how worried are you that you may become hypoglycemic (low blood sugar) the rest of today?” and evening “how worried are you that you might become hypoglycemic overnight?” These items were developed with stakeholder input and had significant correlations with other common factors related to fear of hypoglycemia (e.g., glycemic variability)43. 4) Sleep Quality assessed each morning using a single 10-point numerical scale by the well-validated Pittsburgh Sleep Diary104. 5) Competing demands assessed each morning at self-reported waketime by a single 4-point scale “How busy is today going to be for you?”70 6) Weather (temperature, humidity, windchill) captured using GPS location and the National Weather Service API. 7) Day of the week (weekday vs. weekend). 8) Insulin dosing (including entered carbohydrates) from the participant’s device (InPen, Tandem, Omnipod, or Medtronic) by Apple Health API. 9) Physical activity outside of exercise videos captured by the proprietary Fitbit algorithm (intensity classification accuracy 85%105, step counts ±3%97).

The above markers of individual context, day of the week, and between-participant characteristics, including demographics and clinical characteristics that can impede exercise (body composition, mobility limitations), will be transmitted by PiLR to our Amazon Virtual Machine for processing by our machine-learning system. The system is programmed to identify predictors of exercise “lapse,” defined as the 3rd missed day of the week or 2nd consecutive missed day given that the goal is 150 min/week spread over ≥5 days106. Thereafter, when the Virtual Machine detects these predictors, it signals the PiLR cloud to push a specific encouragement message. E.g., if high blood glucose variability is a vulnerable state predictor, the just-in-time message meeting its occurrence would read: “Sometimes blood glucose can go up and down. You can still be active using the strategies you and your provider have discussed.”

The mathematical basis of the code is a binary classification model of exercise lapse using Python scikit-learn package v. 0.19.056, including random forests and gradient boosted tree ensembles (such as XgBoost107). Before the study with funding support awarded by American Heart Association, I will run analyses that train the model at the group level using the previous pilot dataset (Sect. A-5; k=1,512 person-days). The resulting models will be pruned to reduce the complexity of the final classifiers and avoid overfitting the training data. This group-level model will be the starting point for training a person-level model for each participant via supervised transfer learning methods during the study to give the participant tailored prompts. This strategy mitigates the shortness of the timeframe (4 weeks) to train each person’s model. At the end of the study, the data will first be used as an independent sample to validate the pilot group-level model. Second, they will be combined with the pilot data for a larger sample size to assess potential covariates, including pump vs. injection users, prior CGM usage, and age. If needed, separate models for subgroups can be established in the subsequent beta version of the app. Models are evaluated by the number of lapses required to train person-level models to reach acceptable accuracy (receiver-operator characteristic AUC .80, or precision-recall AUC .80 if <20% of days are lapses), thus estimating a “learning window” before users can expect adaptive feedback. We can anticipate this learning window by programming the app to utilize a simplified algorithm (e.g., chosen from message bank according to an *a priori* hierarchy of barriers present) until adaptive feedback starts.

**C-2c3. Personalized review of safety hazard occurrences around exercise and tips to avoid them (Tool #3).** The MEI PiLR™ Health cloud will also apply heuristic codes to the exercise and CGM data to recognize

common patterns that contribute to glycemic regulation (Table 4). These patterns were selected using a list in international consensus guidelines for the entire T1D population61 pared down to ones observed occurring in our pilot study. The PiLR cloud will send participants weekly reports to summarize their patterns, corresponding follow-up tips (Table 5), daily CGM tracings with indicated exercise times, and weekly summary statistics (Sect. C-2e2). The above patterns may occur less commonly than expected (≥1 of the above patterns occurring in ≥80% of person-weeks) or differ in occurrence by covariates noted above like pump vs. injection users, prior

CGM usage, and age. In that case, we will examine the data for other more commonly occurring patterns to include in this tool in the next version. CGM tracking of people with T1D in the earliest stages of exercise uptake is novel, so some new patterns are expected.

**C-2c4. Health outcome feedback regarding the causal impact of exercise upon blood glucose (Tool #4).**

The PiLR cloud will draw CGM metrics (Sect. C-2c2) and covariates (insulin-on-board, carbohydrates self-reported into insulin device, Fitbit physical activity at sub-exercise intensity) and transmit them to the Virtual Machine for processing by our Bayesian R-code24. The results will be sent to the PiLR cloud to document a positive, neutral, or negative effect of exercise which PiLR relays to the GlucoseZone digital coach once per month. The coach will message the client relaying this information and any recommended adjustments (e.g., “You have done purely resistance exercise for the past 28 days, and it has had the causal impact of increasing blood glucose time above target range. Consider more aerobic exercise.”)

The mathematical basis of the R-code is our Bayesian time series model24. Briefly, it compares CGM metrics over each 4-week period against a counterfactual predicted based upon the CGM during the 4 weeks before the period and the covariates over the 4-week period. The full formalism and an example of this analysis is available in our manuscript24. The advantage over linear modeling frameworks is the flexibility to evaluate the intervention’s effect strength at all points in the intervention period. If p-values do not reach significance (α=.05) then the next version of the app will extend the length of the testing window beyond 4 weeks before providing feedback, though even 4 weeks of CGM data at minimal adherence (70%) yields k=5,645 observations resulting in 99% power to detect a small effect size (d≥0.10) at α=.05. The duration of 4 weeks was chosen as a typical frequency of health coaching sessions for long-term maintenance of lifestyle change143.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4.** Example summary of exercise safety hazards with tips to avoid them. | | | |
| **Pattern** | **Definition** | **Times of occurrence** | **Follow-up Tip** |
| Failure to take adequate carbohydrate supplementation relative to start values | (Blood glucose <70 mg/dL at the start of exercise)  OR  (blood glucose 70-100 mg/dL at the start of exercise AND <70 mg/dL during or within 1hr after exercise) | Monday 9:43am  Wednesday 11:16am | A |
| Nocturnal hypoglycemia following exercise | Blood glucose <70 mg/dL for 30min of consecutive nocturnal readings | Wednesday 2:00-3:15am | A |
| Failure to adequately reduce pre-exercise bolus insulin | Insulin bolus <120min before start of exercise AND blood glucose <70 mg/dL during or within 1hr after exercise | Thursday 7:14pm  Saturday 3:44pm | A |
| Starting exercise with elevated blood glucose values | Blood glucose >270 mg/dL at the start of exercise | Friday 4:45pm  Sunday 4:17pm | B |
| A, Based on this blood glucose data, it may be a good idea to review the recommendations from your healthcare professional about how you would adjust your food intake or insulin dosing around exercise. If you do not remember, it would be a good idea to reach out. B, If your blood sugar is high (more than 270 mg/dL) before starting exercise, check your blood or urine for ketones. If you test positive for ketones, avoid vigorous activity. If you do not have ketones in your blood or urine and you feel well, it should be fine to exercise. | | | |

**C-2d. Interviews and Analysis.** Participants will complete weekly audiotaped 30min semi-structured interviews during the 4 weeks of app usage. They will be transcribed, reviewed, and analyzed by content analysis108.

Exercise videos (Tool #1). Interviews will focus minimally on this component because the videos have been refined through 5 years of commercial popularity and highly rated by our pilot users (Table 2).

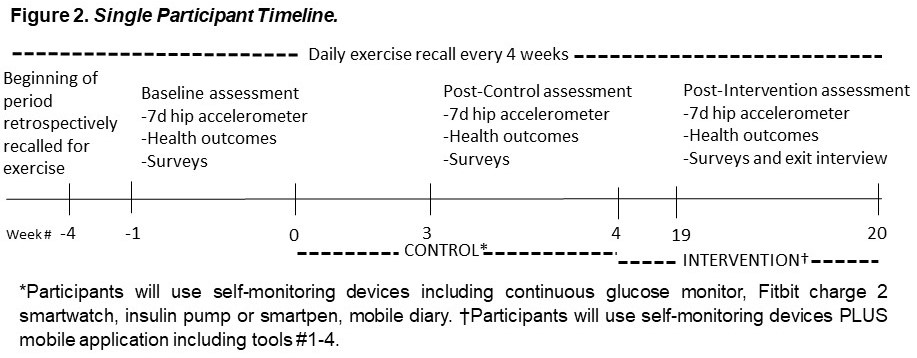
Just-in-time adaptive text messaging (Tool #2). 1) On the days you missed exercise, what was the barrier? 2) How did you find the messages addressed that barrier? 3) What else would you like to see in these messages? 4) For each barrier, what would help you in the moment besides texting? *(Content analysis will be used to identify themes of desirable attributes to include in the next version of the app.)*

Safety Hazard Review (Tool #3). The PiLR cloud will time the weekly report (Table 4) to arrive at the start of the interview time. Participants will review the report at a self-selected pace while the researcher administers structured questions via speakerphone: 1) What patterns do you see in your data highlighted by the tool? 2) What patterns do you see in your data that are not highlighted by the tool? 3) Why do you think these patterns are happening? 4) Were you aware of these patterns already? 5) What are your reactions to this information? 6) What does this information mean to you? 7) How might this information change your future interactions with our app? *(Content analysis will identify common themes to guide the personalized reports for the next version. For example, if participants consider specific patterns in their data more meaningful than the ones highlighted by the tool, we will program those patterns to be recognized in the next version.)*

Health Outcome Feedback (Tool #4). This tool will only be queried after the 4th week which is when it starts generating feedback. 1) Are you curious about the causal impact of exercise upon your blood glucose? 2) If you could learn even more about the causal impact of exercise upon your blood glucose, what would you use that information for? 3) Was the message at the end of the month helpful for refining your exercise prescription? 4) What message wording would help you refine your exercise prescription? 5) What message wording would increase your motivation to exercise? *(Content analysis will identify common desirable themes to include in the messages in the next version.)*

The above weekly semi-structured interviews will be supplemented by a quantitative survey at the end of the 4 weeks. It will include the system usability scale109 (10 Likert-style items assessing aspects such as ease of use and degree of technical support needed, α=.91) for each of the above tools and each biosensor, as well as Likert-style survey items specially designed for each of the novel mobile tools based upon the above questions. In addition, the study description page linked from our online advertisements will include a single question after the study description asking refusers to indicate their reason for refusal, so we can incorporate feedback from refusers into intervention refinement. We will evaluate participant use metrics to determine intervention component feasibility (i.e., diary and biosensor adherence, use of videos, receipt of text messages).

**C-3. Beta Version (Aim 2)**

**C-3a. Research Design**. After incorporating the refinements derived from Aim 1 into the PiLR tool, we will conduct a single-group within-participant trial (N=60) with weekly repeated measures primary outcomes (Figure 2). All participants will complete one week of baseline assessment (primary and secondary exercise behavior outcomes also retrospectively recalled day-by-day over the prior 4 weeks), a 4-week wait-list CONTROL (use biosensor devices only) followed by a 16-week INTERVENTION (use biosensor devices plus mobile app containing tools #1-4 described in Sect. C-2c). There will be a two-week adaptation period for participants not already using CGMs before the baseline assessment (week -3)1. The two-week adaptation will not count in the baseline exercise recall unless statistically identical to other recalled weeks.

Enrollment and the mobile app intervention will function as in the alpha version study (Sect. C-2b, C-2c). *Timetable is given in the Candidate section table and Sect. 2.7.* Clinical trial endpoints are given below.

**C-3b. Exercise behavior primary outcomes** 1) Hip accelerometer (Actigraph GT9X) will be worn for 7 days at each timepoint (Figure 2). We will classify ≥2020 counts/min as exercise110 and calculate total activity volume (kcals) by validated tri-axial vector magnitude equations111. 2) Timeline Followback for Exercise. Every month participants will be asked to recall exercise (type, duration, Borg Rating of Perceived Exertion scale98) for each calendar day going back 4 weeks (Figure 2) using calendar prompts and memory aids (e.g., holidays). I have demonstrated this method to be reliable (r= .79 - .97) and convergent valid with weekly exercise logs (r = .65 - .80)112. It captures more activities than accelerometry (e.g., resistance exercise) and dates outside the scope of app usage (i.e., up to 4 weeks before baseline). We will calculate intensity by the Ainsworth compendium113 and convert to kcals.

**C-3c. Critical health outcomes for T1D** *1) Glycemic control via HbA1c* (AccuBase A1c Home Test Kit (DTI Laboratories, Thomasville, GA))*.* *2) Resting blood pressure*via standard home monitoring procedures (Omron BP7350, Lake Forest, IL)114*. 3) Height and Weight*(height indicator tape ruler and Withings Body Weight Scale, Issy-les-Moulineaux, France) in light clothing without shoes to calculate body mass index.

**C-3d. Exercise facilitators and barriers** 1) Diabetes self-management115 queries how closely participants follow their prescribed regimen for 23 tasks that are part of standard diabetes care (ɑ=.78 to .87). 2) Fear of Hypoglycemia Worry Subscale includes 18 items related to concerns about hypoglycemia (ɑ=.89)116. 3) Exercise self-efficacy includes 9 items participants rate in terms of how much they would hinder them from successfully exercising (α=.92)117. 4) Exercise Self-Regulation Questionnaire includes 16 sources of self-regulation participants rate in terms of their influence upon their exercise (α=.73-.88)118.

**C-3e. Statistical Analyses****.** Statistical analyses will utilize an ITT approach and mixed models, which are one of two gold standards (together with multiple imputation) for handling outcome variable missing data in longitudinal studies. For exercise behavior outcomes (weekly minutes, frequency, volume) we will examine the effect of condition over time over week 4 or 20, controlling for baseline exercise if any. Specifically, we will evaluate changes in exercise scores and their time effects as estimated by a Generalized Linear Mixed Model (GLMM) and compare these effects between the 4-week CONTROL and the first 4 weeks of INTERVENTION. We will then perform extension models including exercise scores over the full 16 weeks of INTERVENTION. Using all repeated measures on individuals (i.e., each week of recorded exercise) in the context of a GLMM will allow us to assess temporal patterns of change over time and to use all available data on an individual. This approach, therefore, helps to avoid imputing missing data. The GLMM does not have reduced power compared to simpler end-point analysis, rather it allows us to obtain unbiased and efficient estimates of change over time and between group differences. GLMM incorporates correlations within participants. Within the GLMM, our primary hypothesis is on change from baseline to end-point and we will perform focused comparisons to assess these effects. We will select the best-fitting variance-covariance structure using Schwartz-Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). Normality of residuals from models will be assessed. When the normality assumption is not met, we have several options including applying transformations or categorizing an exercise outcome at a binary or ordinal level. For this case, GLMM will be employed for a binary or ordinal outcome. If the weeks of CGM familiarization (weeks -3 & -2) differ from the other baseline weeks (weeks -4 & -1), they will be excluded; the remaining 2 weeks is an adequate duration to capture exercise behavior119-121. Expected Results: We hypothesize that *INTERVENTION* will yield greater increase in exercise minutes than *CONTROL* which will correspond to significant condition by time effects with significant between-condition differences in change from baseline to end-point. A similar analysis will be used for secondary glycemic control outcomes from CGM (coefficient of variation, mean, % time >180, 70-180, <70mg/dL101). Finally, a similar analysis but reduced to three levels of the time factor (baseline, week 4, week 20) will be used for all other variables including the primary accelerometer-measured exercise outcome (weekly minutes), secondary accelerometer-measured exercise outcome (total weekly physical activity), and secondary health outcomes (HbA1c, systolic and diastolic blood pressure, body mass index, waist circumference) and secondary psychosocial outcomes (exercise-related diabetes self-management, fear of hypoglycemia, exercise self-efficacy). The study is not powered to test between-participant moderators (sex, race, ethnicity) but these will nonetheless be examined to generate hypotheses regarding groups most likely to benefit.

**C-3f. Sample Size and Power Considerations.** Aim 1. The sample size for a qualitative study is driven by the need for theoretical saturation which is typically reached with 10-12 participants in a homogenous population109; thus, we will purposively sample n=10-12 of each of the two insulin therapy modalities until theoretical saturation is achieved. Aim 2. Sample size (N=60) was chosen to accumulate a similar number of observations (k=112 days \* 60 participants = 6,720) to previous reports that have developed robust just-in-time adaptive algorithms122-124 while also being powered to detect a clinically significant, medium effect on exercise outcomes (Cohen’s d=0.5) after attrition (estimated 10% based on pilot study). The power is 94% for 5% significance and 85% for False Discovery Rate-adjusted 1% significance (for 5 outcomes: weekly exercise minutes, frequency, volume, accelerometer-measured weekly moderate to vigorous physical activity, accelerometer-measured total weekly physical activity). This power was calculated based on one pair of delta scores (i.e., CONTROL and INTERVENTION at endpoint). For the exercise variables measured weekly, the greater number of pairs (i.e., 4 weeks per condition plus extension models for the 16 weeks of INTERVENTION) should yield even greater power. Between-participant covariates do not affect power, since the analysis is within-person between the CONTROL and INTERVENTION conditions.

**C-4. Digital subtyping for a future precision medicine approach (Exploratory Aim)**

**C-4a. Overview.** Using the databank generated in the above clinical trial, we will use a variational autoencoder to cluster individuals according to a) their biobehavioral data during the pre-intervention control and b) the trajectory of their biobehavioral data during the first 4 weeks of the intervention (i.e., short-term behavioral changes which often predict longer-term changes in exercise behavioral studies125). Cluster membership will then be evaluated as a predictor of longer-term, important intervention outcomes including success or failure to meet clinical targets (≥150min weekly exercise, d≥0.5 change in mood or sleep states, ≥70% CGM time in target range), number of weeks to first reach these targets, and number of weeks sustained once reached.

**C-4b. Clustering method and preliminary results.** We have piloted the variational autoencoder clustering using our previously collected pilot data (Sect. A-5), in which 18 participants were observed for 12 weeks (k=1,512 total person-days) and we extracted 30 biobehavioral features from the biosensors and mobile diaries (Appendix 1). This timeframe included a 2-week CONTROL and 10-week INTERVENTION period, which have been combined in the below visualizations since they looked similar.

A variational autoencoder144 was trained on these data with batch size 32 and patience (i.e., tolerated number of iterations with no improvement) of 10. The decay of the autoencoder loss function over training iterations was highly aligned between randomly selected training and validation subsets (Appendix 2), suggesting that the model did not underfit or overfit the data. The autoencoder was set to reduce the data to 10 dimensions and showed clustering when any two of the 10 distinct dimensions are plotted against each other. We have randomly selected two dimensions for presentation purposes (Figure 3). Subsequent decoding reconstructed the original data with high accuracy (loss function output started 22, ended 8), indicating that the 10 reduced dimensions captured important features. In addition, the variational autoencoder clearly added analytic value because its resulting clusters were cleaner than a parallel analysis by linear reduction (i.e., principal component analysis).

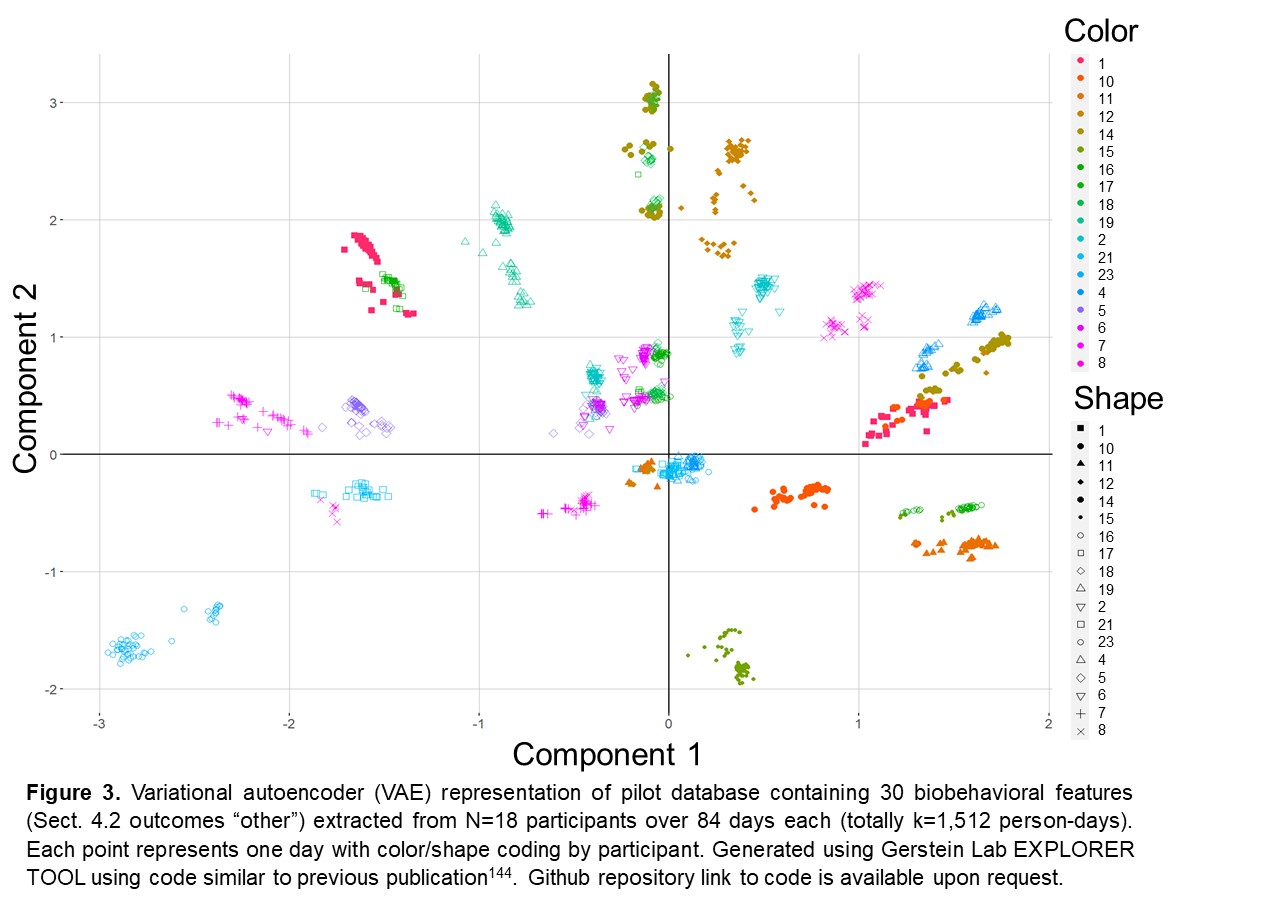
Visually inspecting the dimensions (Figure 3) revealed that days were heavily segregated by participant, suggesting a unique biobehavioral signature for each individual. Nonetheless, some individuals clustered with 1-2 other individuals. As the larger clinical trial in the present grant (Sect. C-3) enrolls individuals, we will add them to this plot. When clusters reach larger numbers of individuals, we can perform several analyses:

**1)** Test the clusters for differences in long-term intervention outcomes using GLMM adjusted for covariates that may influence outcomes, including age, sex, diabetes duration, comorbidities, with Bonferroni correction in case of > 2 clusters. These tests will consider the average magnitude of interpoint distances in each cluster (i.e., “spread”) as an additional fixed factor.

**2)** Tabulate the highest-ranking features of the machine learning models (i.e., those most determinative of long-term trajectories) which would be considered desirable health goals, especially if amenable to intervention.

**3)** Calculate vector trajectory (magnitude and direction) from CONTROL period datapoints to INTERVENTION wk1-4 datapoints and upon the vectors, perform variational autoencoder clustering and tests of differences in long-term intervention outcomes (Sect. C-4a) between the clusters. This analysis will test the hypothesis that early intervention response can predict longer-term intervention response.

We expect long-term intervention outcomes to have inter-individual heterogeneity. As one example of several from our pilot study, average daily sleep quality improved by 0.5 standard deviations among a majority but not all participants (13 / 18, 72%), this improvement took 3.7±2.8 weeks to occur and was sustained for 3.2±2.2 weeks. It would be beneficial if clusters could predict this interindividual heterogeneity of outcomes in early stages of tracking the participants, so that precision medicine could tailor long-term intervention accordingly.

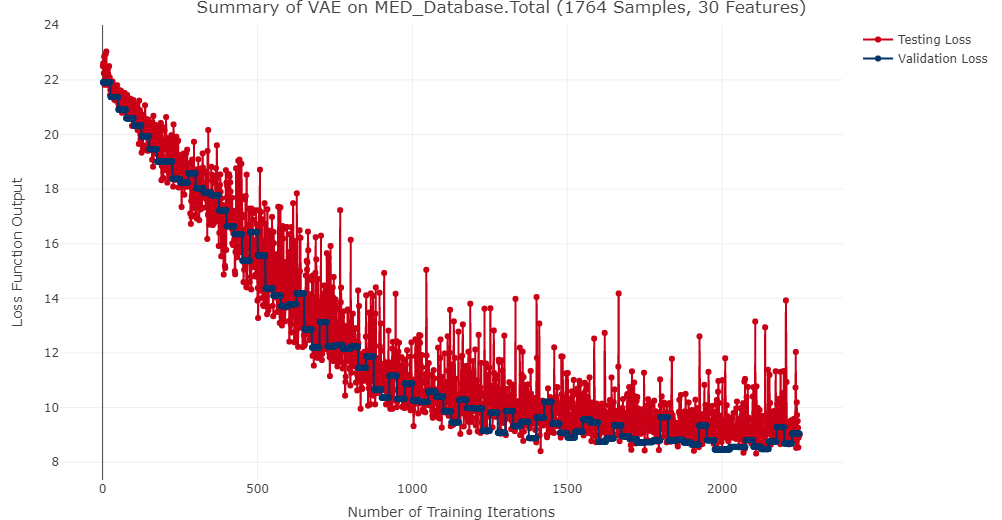
v

**C-4c. Power.** The sample size of N=60 with 10% dropout will detect large differences of pre-post intervention changes (d=0.8, 82% power, .05 significance) for k=2 clusters. To explore smaller effects or more clusters, another option is incorporating biobehavioral data from each month of the intervention into the baseline phenotype after adjusting for dependency by estimating the correlation of repeated measures, giving power between N=60-342 depending on the degree of this correlation. Effect sizes below this additional threshold are not testable but will inform calculations for future precision medicine studies.

**C-5. Potential Problems and Alternative Strategies**

**1)** While we could design a fully automated app, brief human elements -- orientation to exercise videos (Tool #1), relaying of monthly personalized information (Tool #4) – are included since these services are infrequent (15 min/month), disseminatable by GlucoseZone’s internal certification training, and cost-effective ($12.99/month retail value). **2)** We used run-in control design which prevents randomization but maximizes the number of participants assigned to intervention, thus proving the concepts of just-in-time adaptive machine learning and digital phenotyping to predict intervention response. **3)** We considered supervised machine learning for the exploratory aim based on predefined classifications of intervention responders and non-responders, but no clinical precedent exists for such classification. Hence, we opted to use unsupervised clustering.

|  |
| --- |
| **Appendix 1.** Biobehavioral metrics used as features for variational autoencoder. |
| **Exercise at the day level** |
| Exercise duration completed (minutes) |
| Exercise volume completed (minutes weighted by caloric expenditure intensity) |
| **Mood and Sleep at the day level** |
| Plans not to exercise (binary) |
| Not Enough Time to exercise (binary) |
| Lacked Positive Feelings (binary) |
| Had Negative Feelings (binary) |
| Lacked Energy (binary) |
| Felt Fatigued (binary) |
| Blood sugar management issue (binary) |
| Sick (binary) |
| Morning Fear of Hypoglycemia (5-point scale) |
| Evening Fear of Hypoglycemia (5-point scale) |
| Sleep Quality (5-point scale) |
| **Mood Immediately Before Exercising** |
| Fear Of Hypoglycemia (5-point scale) |
| Positive Feelings (5-point scale) |
| Negative Feelings (5-point scale) |
| Energy (5-point scale) |
| Fatigue (5-point scale) |
| **Mood Immediately Post-Exercise** |
| Fear Of Hypoglycemia (5-point scale) |
| Positive Feelings (5-point scale) |
| Negative Feelings (5-point scale) |
| Energy (5-point scale) |
| Fatigue (5-point scale) |
| **Glucose\*** |
| Daily 24hr mean (mg/dL) |
| Daily 24hr variance (coefficient of variation, %) CV\_total |
| Daily 24hr Time above target range (%) |
| Daily 24hr Time in target range (%) |
| Daily 24hr Time below target range (%) |
| Daily Nighttime mean (mg/dL) |
| Daily Nighttime time below target range (%) |
| \*Further work could enter these data at the raw time series level rather than the summary features shown here. |



**Appendix 2.** Decay of autoencoder loss function with increasing training iterations for randomly selected testing subset (red) and validation subset (blue). The high overlap indicates the model did not overfit or underfit the data.