Journal Pre-proof

Quantifying the impact of physical activity on future glucose trends using machine learning

Nichole S. Tyler, Clara Mosquera-Lopez, Gavin M. Young, Joseph El Youssef, Jessica R. Castle, Peter G. Jacobs

PII: S2589-0042(22)00158-4

DOI: https://doi.org/10.1016/j.isci.2022.103888

Reference: ISCI 103888

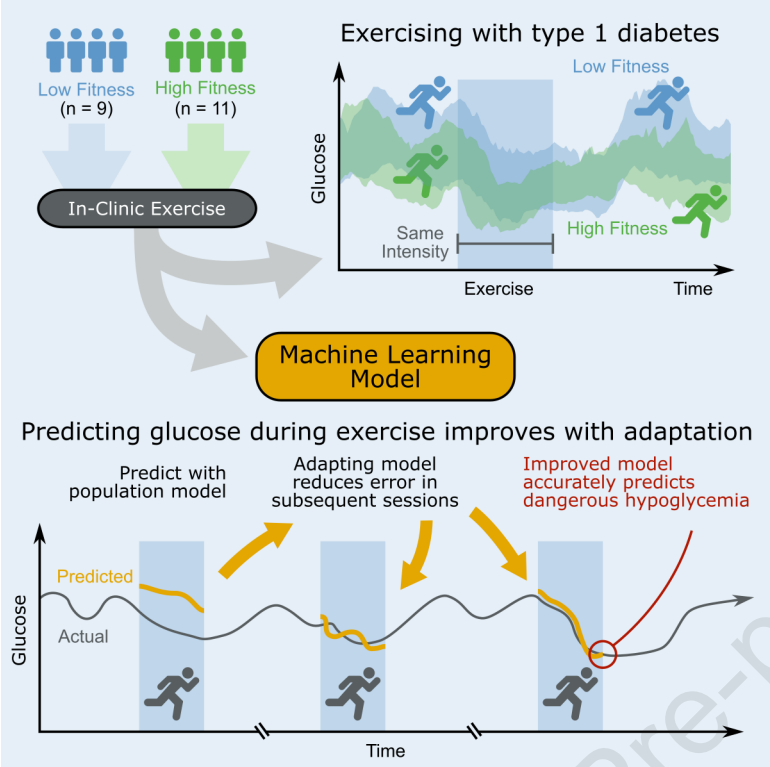
To appear in: *ISCIENCE*

Received Date: 24 June 2021

Revised Date: 19 November 2021

Accepted Date: 4 February 2022

Please cite this article as: Tyler, N.S., Mosquera-Lopez, C., Young, G.M., El Youssef, J., Castle, J.R., Jacobs, P.G., Quantifying the impact of physical activity on future glucose trends using machine learning, *ISCIENCE* (2022), doi: https://doi.org/10.1016/j.isci.2022.103888.



Quantifying the impact of physical activity on future glucose trends using machine learning

Nichole S. Tyler[1]*, Clara Mosquera-Lopez[1], Gavin M. Young[1], Joseph El Youssef[2], Jessica R. Castle[2], Peter G. Jacobs[1**]

Affiliations:

[1] Artificial Intelligence for Medical Systems (AIMS) Lab, Department of Biomedical Engineering; Oregon Health & Science University; Portland, OR, 97232; USA

[2]Harold Schnitzer Diabetes Health Center, Division of Endocrinology; Oregon Health & Science University; Portland, OR, 97239; USA

** **Lead Contact**: Peter G. Jacobs, Email address: jacobsp@ohsu.edu

*** Corresponding Author:** Nichole S. Tyler, Email address: tylern@ohsu.edu

Summary (148/150 words)

Prevention of hypoglycemia (glucose <70 mg/dL) during aerobic exercise is a major challenge in type 1 diabetes. Providing predictions of glycemic changes during and following exercise can help people with type 1 diabetes avoid hypoglycemia. A unique dataset representing 320 days, and 50,000+ time points of glycemic measurements was collected in adults with type 1 diabetes who participated in a 4-arm crossover study evaluating insulin-pump therapies, whereby each participant performed 8 identically designed in-clinic exercise studies. We demonstrate that even under highly controlled conditions, there is considerable intra- and inter-participant variability in glucose outcomes during and following exercise. Participants with higher aerobic fitness exhibited significantly lower minimum glucose and steeper glucose declines during exercise. Adaptive, personalized machine learning (ML) algorithms were designed to predict exercise-related glucose changes. These algorithms achieved high accuracy in predicting the minimum glucose and hypoglycemia during and following exercise sessions, for all fitness levels.

Keywords: Type 1 Diabetes, Diabetes, Aerobic Exercise, Decision Support, Artificial Intelligence, Machine Learning, Medicine, Endocrinology

INTRODUCTION

Physical activity has been shown to reduce cardiovascular risk factors in people with type 1 diabetes (Bohn et al., 2015) and regular physical exercise has recently been shown to result in improved time in target glucose range (70-180 mg/dL)(Riddell et al., 2020a). However, exercise is also known to cause substantial changes in glucose. These changes in glucose vary per exercise modality (Bussau et al., 2006; Colberg et al., 2016; Lascar et al., 2014; Moniotte et al., 2017; Reddy et al., 2018; Yardley et al., 2013) and are most dramatic during steady aerobic exercise (Riddell et al., 2020b). There is an increased risk of hypoglycemia during exercise that occurs due to altered muscular uptake of glucose during exercise, and delayed hypoglycemia that can occur on nights following exercise due to changes in insulin-sensitivity (Man et al., 2009; McMahon et al., 2007; Reddy et al., 2019; Wahren, 1977). These dynamic processes underlying glucose uptake are compounded by regular bouts of exercise (Boulé et al., 2005; Steenberg et al., 2019). While regular exercise can improve overall health, avoiding hypoglycemia during exercise is a known challenge for people with type 1 diabetes (Wilson et al., 2020b).

Continuous glucose monitoring technologies (CGM) can provide real-time alerts to the occurrence of hypoglycemia (< 70 mg/dL) or hyperglycemia (> 180 mg/dL) during exercise. And while certain commercial CGM systems like the Dexcom CGM have recently been reported to achieve 13.3% mean absolute relative error (MARE) during aerobic activity (Guillot et al., 2020), use of CGM alone is not sufficient to prevent

hypoglycemia. Commercially available automated insulin delivery (AID) systems have been shown to improve time in glucose target range across real-world daily activities (Brown et al., 2019; Garg et al., 2017), but the exercise modalities of these systems are limited to user-selected modifications to basal insulin and target glucose during announced physical activity (MiniMed 670G System User Guide, Medtronic, 2017; t:slim X2 Insulin Pump with Control-IQ Technology User Guide, Tandem Diabetes Care, 2020) (Wilson et al., 2022). AID algorithms that incorporate real-time physical activity data to prevent hypoglycemia typically reduce automated insulin and, in the case of dual-hormone systems, increase glucagon in anticipation of glucose drops during aerobic exercise (Castle et al., 2018; Jacobs et al., 2016; Wilson et al., 2020a). Furthermore, adaptive AID algorithms that incorporate activity data have been developed to estimate an individual's plasma insulin and future glucose concentrations for the purpose of personalizing insulin delivery (Hajizadeh et al., 2018a; Hajizadeh et al., 2018b). However, even these systems do not completely eliminate exercise-induced hypoglycemia. Consensus statement guidelines have been developed to help people with type 1 diabetes make decisions regarding modification of insulin dosages and carbohydrate intake prior to and during exercise (Moser et al., 2020; Riddell et al., 2017), but people with type 1 diabetes will oftentimes need to use trial-and-error approaches to learn how to avoid hypoglycemia during exercise. Both automated hormone delivery and decision support systems currently lack the ability to accurately predict exercise-induced changes in glucose. In addition, there can be significant inter and intra-person variability in glucose changes during exercise. Exercise-related glucose changes in people with type 1 diabetes have not yet been precisely quantified

in individuals and across populations when considering different insulin therapies, or baseline fitness levels.

Machine learning is a powerful tool whereby machines are designed to solve problems or perform sophisticated tasks and can even help to make medical decisions, or provide decision support, for diabetes management. Machine learning approaches have been used in disease detection (Li et al., 2018), insulin dose modification through decision support (Tyler and Jacobs, 2020; Tyler et al., 2020), and can be expanded to provide exercise decision support directly to a person living with type 1 diabetes, or to AID systems in order to adjust insulin during physical activity (Reddy *et al.*, 2019; Wilson *et al.*, 2020a). While algorithms that have been designed to predict future glucose exhibit relatively low root mean squared error (RMSE) during non-exercise periods (14.0 mg/dL-18.0 mg/dL) (Mosquera-Lopez and Jacobs, 2021; Pérez-Gandía et al., 2010; Zecchin et al., 2012; Zhu et al., 2020), recent studies have indicated that the accuracy of these algorithms is oftentimes far worse during exercise (46.16 mg/dL) (Hobbs et al., 2019). Machine learning models have already been developed to predict changes in glucose immediately following aerobic exercise (Ben Brahim et al., 2015; Hobbs *et al.*, 2019; Reddy *et al.*, 2019; Romero-Ugalde et al., 2019), and, when integrated with a decision support system, increase the minimum glucose measured during in-clinic exercise sessions (Breton et al., 2018). Still, these algorithms oftentimes have poor accuracy during real-world scenarios (Hobbs *et al.*, 2019), demonstrate large variability in performance between individuals (Xie and Wang, 2020) and have not been evaluated across varying physical fitness levels.

Population machine learning models are trained on a group of people and are designed to provide predictions for all people. Whereas a *personalized* model learns an individual's unique physiology in order to improve prediction accuracy for an individual. Personalized models can be designed by training machine learning models specifically on an individual's data (Romero-Ugalde *et al.*, 2019), by clustering a number of similar people into groups prior to model training and then training a model on that cluster (Contreras et al., 2017; Montaser et al., 2019), or by adapting a model in real-time using newly observed data in order to improve glucose predictions (Hajizadeh *et al.*, 2018b; Hobbs *et al.*, 2019). It is not yet clear how personalization impacts the prediction accuracy of exercise-related changes in glucose.

Herein we characterize the impact of aerobic exercise on glucose changes using a unique dataset collected during highly controlled, aerobic exercise sessions in adults with type 1 diabetes. Glucose variations are characterized per participant, insulin therapy, and are further explored with respect to baseline physical fitness. Personalized machine learning models were then designed to estimate the minimum glucose during aerobic exercise and four hours following the start of exercise, and to quantify the impact of personalization on model accuracy. We considered three machine learning algorithms, including a multivariate adaptive regression spline (MARS) model (Friedman, 1991), a previously described logistic regression model (Breton *et al.*, 2018), and an autoregressive (AR) model based on a previously described autoregressive model with exogenous inputs (ARX) (Romero-Ugalde *et al.*, 2019). The dataset used to

train and benchmark the approach was collected in a previously published study whereby aerobic exercise was performed 8 times per study participant under identical exercise intensity and duration, meal content and timing conditions, and across multiple diabetes management strategies including automated insulin delivery, automated insulin and glucagon delivery, insulin pump therapy with predictive low-glucose suspend, and standard insulin pump therapy (Castle *et al.*, 2018). The findings obtained from this unique dataset can serve as a benchmark for comparison with other adaptive prediction algorithms, since we anticipate that the repeatability of the changes in glucose will be substantially reduced under free-living exercise conditions compared with these controlled conditions.

RESULTS

Variations in blood glucose dynamics during identically designed exercise scenarios

To evaluate the repeatability of exercise-related glucose changes, participant glucose outcomes were obtained from 20 adults with type 1 diabetes who each performed 8 identically-designed aerobic exercise sessions at 70% VO_2max for 43.2 minutes on average (N = 160 observations). To control for additional variability in glucose trends that can impact exercise-related glucose changes, the in-clinic exercise sessions were designed such that participants consumed a self-selected breakfast at 8 am, daily activities at 10 am, lunch at 12 pm, and performed exercise at 2 pm. Meals of identical nutritional content were consumed at the same time, and aerobic treadmill exercise was performed at the same time for each of the 8 in-clinic visits. Figure 1 shows the variability in the changes in blood glucose during exercise for each participant across

the entire study (Figure 1A) and also organized by insulin therapy (Figure 1B-E). The difference in exercise-related blood glucose changes measured during highly controlled exercise sessions (Figure 1B-E, connecting dashed and solid lines) are reported as the difference averaged across all study arms, per participant in Table 1. Glucose dropped during exercise for nearly every exercise session, and glucose dropped further in the 4 hour period after exercise was concluded for some subjects (Figure 1B-E, circles). Despite highly repeatable exercise conditions, food intake, and glucose management strategies, there was still substantial intra-participant variability of the change in glucose during exercise across all 8 identical exercise scenarios, ranging across participants from 23.1 mg/dL (participant 13) to 56.4 mg/dL (participant 9) (Table 1). While variability is smaller for some participants when looking at the two exercise sessions performed under a given diabetes management strategy, substantial variability in glucose changes during exercise is still observable for other study participants (Figure 1B-E). The average change in blood glucose during exercise and variability in this change is reported per therapy arm and per participant in Table 1.

Physical fitness impacts changes in glucose observed during physical activity

Baseline aerobic fitness was assessed by VO_2max norms for men and women using a rating scale from the American College of Sports Medicine (American College of Sports Medicine's Complete Guide to Fitness & Health by Barbara Bushman, 2017) that ranks individuals on a scale of very poor, poor, fair, good, excellent, and superior. We found that participants with higher aerobic fitness (rated as good, excellent, and superior VO_2max) exhibited significantly lower minimum glucose during aerobic exercise than

those with lower aerobic fitness (rated as very poor, poor, and fair VO_2max) (average minimum glucose 75.9 mg/dL vs 103.1 mg/dL, p <0.001). Participants with higher aerobic fitness also exhibited lower CGM-measured minimum glucose compared with participants with lower aerobic fitness in the 4-hours following the start of exercise (70.4 mg/dL vs 85.4 mg/dL, p <0.001). And, the higher aerobic fitness participants had significantly steeper glucose drops during exercise (-2.2 mg/dL/min vs -1.8 mg/dL/min, p <0.05) (Figure 2A-C). Participants with higher aerobic fitness exhibited lower glucose values across the in-clinic study days (Figure 2D-E), with significantly lower glucose during activities of daily living when they were physically active ($p < 0.05$), during the aerobic exercise, and in the overnight period following in-clinic aerobic exercise.

Population model predictions achieve good prediction accuracy

Three types of population machine learning models were designed: a MARS model to predict minimum glucose following exercise, a logistic regression model to predict hypoglycemia following exercise, and an AR model to predict CGM values at the end of exercise. Features used to model minimum glucose during and after exercise were extracted from the data collected during each of the in-clinic exercise sessions (N = 160 exercise sessions) and are defined in Table S1. Leave-one-participant-out cross-validation was used during algorithm training to develop generalizable predictive models (Figure S1). Accuracy of the three machine learning models to predict minimum blood glucose at the end of exercise and also CGM-measured minimum glucose during the 4 hours following the start of exercise are reported in Table 2. The population MARS model estimated minimum glucose during exercise with an MAE of 20.0 mg/dL; a

sensitivity of 63%, and an accuracy of 67% to predict hypoglycemia when cross-validated across all 20 participants with each participant left out during the training. The population logistic regression model achieved a sensitivity of 64% and accuracy of 61% in predicting hypoglycemia during exercise when cross-validated on all 20 participants. The population AR model exhibited worse MAE than the MARS model, 23.8 mg/dL, and achieved the highest sensitivity (71%) and accuracy (81%) to predict CGM-measured glucose < 70 mg/dL 40 minutes after the start of exercise, when cross-validated across all 20 participants.

For longer prediction horizons of 4 hours after the start of exercise, the population MARS model exhibited a MAE of 20.1 mg/dL, and a sensitivity of 62% and an accuracy of 56% to detect CGM-measured hypoglycemia when cross-validated across all 20 participants. The results of the logistic regression model to predict hypoglycemia during exercise and 4 hours following the start of exercise were similar both during exercise and 4-hours after exercise. The logistic regression model achieved a sensitivity of 63% and accuracy of 58% to detect CGM-measured hypoglycemia when cross-validated across all 20 participants. The AR model was not designed for the 4-hour predictive window and therefore results are not shown.

Prior exercise-related changes in glucose help to predict future nadir glucose

The benefit of personalization was evaluated by first considering whether the inclusion of participant exercise history, or data collected during previous exercise sessions, can improve accuracy to predict the minimum glucose during exercise. To do this, a second

MARS model was designed that also incorporates participant exercise history features (Table S2). Exercise data features that were found to be predictive of future glucose trends included (1) the participant's average metabolic expenditure measured during other aerobic exercise sessions, and (2) the average change in glucose measured during other aerobic exercise sessions by the participant. When evaluated on the holdout set, the MARS model that included exercise history reduced MAE by 39%, from 23.4 mg/dL to 14.3 mg/dL, improved sensitivity to predict hypoglycemia during exercise from 50% to 73%, and improved accuracy from 75% to 81% (Table 2). Cross-validation across all 20 participants showed that the inclusion of participants' exercise history into the MARS model reduced MAE from 20.0 mg/dL to 17.6 mg/dL, improved sensitivity from 63% to 66% to detect hypoglycemia, and improved accuracy from 67% to 70%.

For longer prediction horizons of 4 hours, the MARS model that included exercise history performed similarly to the MARS model that was designed without exercise history, when cross-validated across all 20 participants (Table 2).

Adaptive personalization improves the accuracy of predictive models

The benefit of personalization was also investigated through adaptation of the machine learning models to better predict individual participants' exercise-related glucose changes. Stochastic gradient descent (An overview of gradient descent optimization algorithms, Ruder, 2016) was used to incorporate the exercise information obtained from a participants exercise session (e.g., data collected during their first study visit) in order to update the population model parameters. The adapted model was then used to

predict the same participant's outcomes for a separate, held-out exercise session (e.g., their second study visit). This adaptation procedure was repeated for each held-out exercise session, enabling the machine learning model parameters to adapt to an individual's data over time as more exercise sessions were observed. Personalization of the model coefficients through stochastic gradient descent adaptation improved the accuracy of all of the predictive algorithms (see Table 2) to estimate glucose during exercise and 4 hours after the start of exercise. The improvement from adaptation was not influenced by the order of the observed exercise sessions, and we report the results from the original order prior to shuffling. Gradient descent adaptation of model coefficients reduced the predictive error of the MARS model from an MAE of 20.0 mg/dL to an MAE of 18.1 mg/dL, reduced sensitivity from 63% to 61%, and significantly improved the 20-fold cross-validation accuracy of the MARS model in predicting hypoglycemia during exercise from 67% to 78% ($p<0.05$). The predictive error per-participant can be seen in Table 1. Adaptation of the logistic regression parameters improved the sensitivity to predict hypoglycemia during exercise from 64% to 68%, and significantly improved the accuracy from 61% to 70% ($p<0.05$), when cross-validated across all 20 participants. Adaptation of the AR model improved the cross-validation MAE from 23.8 mg/dL to 22.0 mg/dL, and improved the sensitivity to detect hypoglycemia during exercise from 71% to 76% and accuracy from 81% to 83%.

For longer prediction horizons of 4 hours following the start of exercise, adaptation reduced the predictive error and improved the accuracy of all of the models. The personalization through adaptation of the MARS model coefficients significantly reduced

the MAE from 20.1 mg/dL to 18.3 mg/dL, reduced sensitivity from 62% to 56%, and significantly increased the accuracy to predict CGM-measured hypoglycemia 4 hours following exercise from 56% to 68% ($p<0.05$). The adaptation of the MARS model designed to include prior exercise session metrics reduced the MAE from 21.1 mg/dL to 18.2 mg/dL, reduced sensitivity from 74% to 57% and increased the accuracy to detect CGM-measured hypoglycemia 4 hours following exercise from 57% to 69% when cross-validated across all 20 participants. Adaptation of the logistic regression model increased sensitivity from 63% to 64%, and significantly improved the accuracy from 58% to 70% ($p<0.05$) to predict CGM-measured hypoglycemia in the 4 hours following exercise when cross-validated across all 20 participants.

Figure 3 shows the Parkes consensus grid of the MARS model cross-validation across all 20 participants in predicting glucose at the end of exercise. Personalization of the population MARS model increased the number of observations in the consensus error grid region A from 110 observations to 115 observations, with no changes in regions C, D, or E. When exercise history was included in the design of the MARS model, adaptation increased the values in region A to 118 observations, with no observations in regions D and E and 99.4% of observations in the combined A + B regions (Figure 3C).

Physical fitness impacts predictive performance

The MARS models performed equivalently for higher fitness vs. lower fitness study participants in terms of mean absolute relative error (Table 3). The AR performed worse for the higher fitness participants than the lower fitness participants. The

accuracy to detect hypoglycemia during exercise, and in the 4 hours following start of exercise, was nominally lower in all machine learning models when evaluated on participants with higher aerobic fitness. Adaptation improved the accuracy to predict hypoglycemia for participants with higher and lower aerobic fitness, and across both prediction horizons (Table 3).

DISCUSSION

Herein we demonstrate that there is substantial variability in glucose changes during aerobic exercise in people with T1D even under highly repeatable food intake and exercise conditions, and that these changes are impacted by baseline physical fitness levels. We also present adaptive glucose-forecasting algorithms and demonstrate how personalization and prior history can improve the accuracy to predict minimum glucose during and following aerobic exercise. To our knowledge, this is the first analysis of exercise-related glucose changes and prediction strategies using an ideal dataset of highly regimented, identical study exercise visits and across multiple insulin therapies. In the published clinical study data set used to train the proposed predictive algorithms (Castle *et al.*, 2018), the specific variations in glucose during exercise were not presented and a demonstration of differences between individuals with varying aerobic fitness was not presented. The data demonstrate that individuals living with type 1 diabetes will experience considerable variability during exercise, even when exercise occurs in the context of identical meals, exercise intensity and duration, insulin therapy, and scheduled daily activities. For some participants, the magnitude of this variability was diminished when examined within the context of an individual insulin therapy. From

a clinical perspective, this highlights the challenge and uncertainty that individuals face during aerobic exercise; even if someone could undertake the exact same daily activities, meals, and exercise practices, there will be differences in their glucose outcomes during exercise. Part of this variability is explained by insulin therapy and insulin-on-board, but there are many other factors such as activity level in the days preceding exercise, and stressors such as sleep quality, illness, or timing of menstrual cycle that affect insulin sensitivity and glycaemia following exercise. And, baseline physical fitness can also have a significant impact on glycemic outcomes during exercise. The high intra- and inter-participant variability in glucose trends during exercise presents an opportunity for adaptive machine learning approaches to help people with type 1 diabetes avoid acute and long-term complications related to hypoglycemia.

The impact of exercise on glucose trends during exercise, and across participants with varying physical fitness levels, is still an open question (Moser *et al.*, 2020; Yardley and Sigal, 2021). While an inverse relationship has previously been observed between the regularity of exercise and the rate of severe hypoglycemia(Bohn *et al.*, 2015), it has also been reported that participants with higher aerobic fitness exhibit a greater risk of hypoglycemia (Al Khalifah et al., 2016). We contribute definitive findings that participants with higher aerobic fitness exhibit significantly steeper glucose trends during exercise, experienced significantly lower glucose at the end of exercise, and exhibit nominally lower variability in their glycemic outcomes. This may be due to physiologic differences; regular exercise impacts muscle fiber content (Yan et al., 2010),

and a single bout of exercise can prime muscle for future glucose uptake (Steenberg *et al.*, 2019). Behavioral differences are also a factor, as participants with higher aerobic fitness may sustain physical activity and metabolic expenditure longer and more consistently than participants with lower aerobic fitness. And although participants with varying aerobic fitness exhibited significantly different glucose outcomes following exercise, personalized metrics such as VO_2max and fitness ranking require in-clinic evaluation and are not yet feasible features for incorporation into the design of accessible predictive algorithms. It was also observed that participants with higher aerobic fitness were shown to have significantly lower CGM across the entirety of the 4-arm clinical study; sensor readings for these participants were significantly lower during activities of daily living, exercise, and in the nighttime and 48-hrs following aerobic exercise. This precise knowledge can help to inform new strategies to help people of different fitness levels avoid exercise-related hypoglycemia.

Other groups have presented various methods to predict glucose during exercise. Reddy et al. (Reddy *et al.*, 2019) developed a hypoglycemia prediction algorithm during exercise using a decision tree and random forest algorithm. This random forest model utilized data within first 10 minutes of aerobic exercise to form predictions, and achieved an 86% sensitivity and 87% specificity to hypoglycemia. This approach does not describe adaptation or personalization of models or utilize exercise history. It was also limited in that it required data during the first 10 minutes of exercise to estimate hypoglycemia which makes it impossible for the algorithm to provide automated hormone dosing or decision support prior to the start of exercise. The algorithms

proposed in this manuscript do not use data during the exercise event. The proposed algorithms were designed for use prior to the start of exercise, for the purpose of modifying hormone doses and/or carbohydrate intake. The AR model that we evaluated in this paper was presented originally in Romero-Ugalde et al. as an ARX model, where the model was designed to predict CGM values at 30 minutes following aerobic stair-step exercise, and achieved an RMSE of 7.75 mg/dL (Romero-Ugalde *et al.*, 2019). We repeated the methods described in Romero-Ugalde et al., and while we discovered this method achieves fair accuracy to predict CGM < 70 mg/dL, we were unable to achieve the performance that was previously reported. While the AR model, based on the ARX model described by Romero-Ugalde et al., did not achieve the same predictive error as the MARS model, the adaptation methods presented herein improved the accuracy of the AR model to predict CGM < 70 mg/dL and reduced the RMSE. Since the AR model only included the 0, 10, and 20-minute CGM data points as feature inputs, we explored whether including the 5 and 15-minute CGM data points would improve the accuracy of the AR model. However, we found that when including these data points, there was no statistically significant improvement in the accuracy. This was likely because the CGM data points at 0, 10, and 20 minutes were smoothed, and so they included information from the 5 and 15 minute data points. Breton et al. developed a hypoglycemia prediction algorithm utilizing the contextual physical activity predictors identified by Ben Brahim et al. (Ben Brahim *et al.*, 2015). The accuracy of this model was not reported and does not describe personalization (Breton *et al.*, 2018). In the current paper, we used identical features described by Breton et al. and demonstrated the performance of the model. We additionally showed that adaptation can significantly improve the

performance in predicting hypoglycemia during exercise. Each of the prior publications as well as our findings identified the importance of CGM or SMBG measurements at the start of exercise as a critical predictive feature. The current manuscript extends the work done previously by emphasizing the importance of personalization and physical fitness considerations when designing glucose forecasting algorithms during exercise.

Personalization of the population-based machine learning models was shown to improve the accuracy in almost every model-framework, across both short-term and long-term prediction horizons, and across all validation scenarios. Adaptation of model parameters using stochastic gradient descent was shown to significantly improve the accuracy of detecting hypoglycemia during exercise for the MARS and logistic regression models. And adaptation of the MARS and AR models improved overall accuracy of predictions in terms of MAE. Personalization of the MARS framework that included exercise history as an input feature significantly improved predictive accuracy to detect hypoglycemia during exercise. The personalized MARS models exhibited similar RMSE values for both short-term and long-term prediction horizons. This is likely due to the study design whereby participants were most active during exercise, and were instructed to rest until dinner. And, for some participants, the nadir glucose occurred during exercise and was equivalent for both prediction horizons. In real-world scenarios, predictive RMSE may be higher when people do activities that introduce variability in glucose in the 4-hour period. Taken together across all of the models and validation strategies presented in Table 2, personalization resulted in an average reduction in minimum glucose error estimations by 12.9%, and an average increase in

hypoglycemia prediction accuracy of 21.0%. A strength of the personalization methods presented in this manuscript is the simplicity of the gradient descent approach, which is computationally inexpensive and can be implemented easily in other predictive frameworks with just a few lines of code.

In summary, individuals on insulin pump therapy who perform aerobic exercise under highly regimented, nearly identical conditions and intensities will experience day-to-day variations in exercise-related glucose changes during and following exercise. Baseline physical fitness significantly impacts changes in glucose during exercise. Under these controlled conditions, glucose data at the start of exercise, as well as data from prior exercise sessions are informative of anticipated changes in glucose during future exercise sessions across participants of varying physical fitness levels. And while machine learning models can predict the expected changes in glucose during exercise and can be personalized to provide more accurate predictions, further work is needed to accurately predict hypoglycemia in participants with higher baseline physical fitness. Further studies are forthcoming to determine the performance of our adaptation strategy on at-home exercise session data across participants with varying physical fitness. The scientific community is invited to apply this benchmarking dataset in their research by contacting the lead author for access to the data.

LIMITATIONS

As a limitation, the candidate model structures described here must be compatible with gradient-based optimization procedures, and further evaluation is required before being

implemented in other non-linear model frameworks such as neural networks or decision tree structures. The models in this paper were designed and evaluated on in-clinic exercise data; future studies examining at-home exercise sessions will be required to develop algorithms for real-world use. Our analysis utilized data from 20 participants, and accounts for 320 cumulative days of real-world data, 160 days of which represent in-clinic exercise data, with over 50,000 data time points. While the sample size is small, we propose that this analysis reflects an ideal scenario, and that these results reflect the upper bound of adaptation performance and glucose variability.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award Number 1DP3DK101044-01, F31DK121436 and 1 R01DK120367-01, R01DK1225833-01, R01DK120367-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

Conceptualization, N.S.T and P.G.J.; Methodology, N.S.T., P.G.J., C.M.L., and G.Y.; Software, N.S.T.; Formal Analysis, N.S.T.; Resources, P.G.J., J.R.C., and J.E.Y.; Data Curation, N.S.T., P.G.J, J.R.C., and J.E.Y.; Writing – Original Draft, N.S.T and P.G.J.; Writing – Review and Editing, N.S.T., P.G.J., C.M.L., G.Y., J.R.C., and J.E.Y.; Visualization, N.S.T.; Funding Acquisition, N.S.T., P.G.J., J.R.C., and J.E.Y.;

DECLARATION OF INTERESTS

N.S.T. has nothing to disclose. P.G.J. and J. R. C. have a financial interest in Pacific Diabetes Technologies, Inc. a company that may have a commercial interest in this type of research. No other potential conflicts of interest relevant to the article were reported.

INCLUSION AND DIVERSITY

This analysis utilized previously published human subjects data. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to

ensure gender balance in the recruitment of human subjects. The final participant population demographics reflected the population of the city of Portland, OR. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper received support from a program designed to increase minority representation in science.

PRIOR PUBLICATION

Parts of this study were presented as a poster at the American Diabetes Association 78th Scientific Sessions, Orlando, FL, June 22-26 2018

MAIN FIGURE TITLES AND LEGEND

Figure 1. Change in blood glucose measured during identical aerobic exercise sessions.

(A) The change in glucose measured during 8 identical exercise sessions across a 4-arm clinical study. The box plot represents the median and interquartile range of the change in glucose measured during exercise, cross symbols represent outlier values and each whisker extends to the most extreme data point that is not an outlier (n = 158 observations of SMBG data from 20 participants, whereby each participant is represented by 8 SMBG observations, and participant 18 is represented by 6 SMBG observations).

(B-E) the change in glucose measured during aerobic exercise within a given insulin therapy. The black x symbol represents the change in glucose measured during an exercise session, and there are two x symbols per participant per study arm. The line drawn between two black x symbols represents the difference in glucose outcomes measured between the two identically-designed exercise sessions (n = 158 observations of SMBG data from 20 participants across 4 study arms, whereby participants are represented by 2 observations per study arm, and data is not available for participant 18 in the standard of care study arm). The open black circle represents the change glucose measured from the start of exercise, to the minimum glucose measured 4 hours after exercise, and these outcomes are connected by a dotted black line (n = 160 observations of CGM data from 20 participants across 4 study arms, whereby participants are each represented by two observations per study arm).

Figure 2. Differences in glycemic response across baseline physical fitness. Box plots represent the median and interquartile range of the data, cross symbols represent outlier values and each whisker extends to the most extreme data point that is not an outlier.

* represents significant differences p<0.05 between boxplot groups as determined by an independent t-test. ** represents significant differences p<0.05 between boxplot groups as determined by Wilcoxon rank-sum test. ○ represents significant differences p<0.05 between sensor glucose as determined by a Wilcoxon rank-sum test.

(A) The slope of glucose during aerobic exercise is significantly steeper in participants with higher aerobic fitness (n = 88 observations collected from 11 participants) than participants with lower aerobic fitness (n = 70 observations collected from 9 participants) (average trend -2.2 mg/dL/min vs -1.8 mg/dL/min, p = 0.03).

(B) The minimum glucose measured during aerobic exercise is significantly lower in participants with higher aerobic fitness (n = 88 observations collected from 11 participants) than in participants with lower aerobic fitness (n = 70 observations collected from 9 participants) (average minimum glucose 75.9 mg/dL vs 103.1 mg/dL, p = 4.7E-9)

(C) The minimum glucose measured by CGM in the 4-hrs following the start of aerobic exercise is significantly lower in participants with higher aerobic fitness (n = 88 observations collected from 11 participants) than in participants with lower aerobic fitness (n = 70 observations collected from 9 participants) and (average minimum glucose 70.4 mg/dL vs 85.4 mg/dL, p = 3.3E-5)

(D) Interquartile range of sensor glucose obtained from participants during in-clinic study days 1 and 4. Participants with higher aerobic fitness exhibit significantly lower glucose during activities of daily living and aerobic exercise, and in the nighttime following exercise ($p < 0.05$). The lower aerobic fitness group is represented by grey area (n = 72 sensor traces collected from 9 participants). The higher aerobic fitness group is represented by magenta area (n = 88 sensor traces collected from 11 participants). During the in-clinic exercise study visits, activities of daily living were performed starting at 10 am, and exercise at 70% VO_2max was performed at 2 pm. The number of sensor traces from 9 pm – 12 am is lower for both groups (lower fitness, n = 36, higher fitness, n = 44), representing data only from study day 1, whereas participants exited the clinical study on day 4 and overnight sensor data is therefore not available.

(E) Interquartile range of sensor glucose across the entire 4-day study. The lower aerobic fitness group is represented by grey area (n = 36 sensor traces collected from 9 participants). The higher aerobic fitness group is represented by magenta area (n = 44 sensor traces collected from 11 participants).

Figure 3. Consensus Error Grid for models predicting minimum glucose at the end of exercise. The regions of the consensus error grid indicate the clinical impact of prediction errors. Observations that land in regions A and B indicate safe predictions. Observations that lay in regions C, D, and E may result in clinical errors such as missed hypoglycemia, or false positive hypoglycemia that results in excessive carbohydrate

intake. The percentage of observations falling within each region is listed below each figure.

(A) Population MARS model validation (n = 158 observations of exercise data collected from 20 participants) without including prior exercise history.

(B) The MARS model predictions after personalization of population model coefficients (n = 158 observations of exercise data collected from 20 participants).

(C) The predictions of the MARS model that incorporates exercise history features, with additional personalization of the model coefficients (n = 158 observations of exercise data collected from 20 participants).

546 Tables
547

Participant ID	Mean Glucose Drop During Exercise [mg/dL]	Standard Care Arm	Predictive Low-Glucose Suspend Arm	Single-hormone AP Arm	Dual-hormone AP Arm	Average difference measured during identical exercise [mg/dL]	Ex model[b] MARE [%]	Ex model[b], RMSE [mg/dL]	4-hr model[c] MARE [%]	4-hr model[c] RMSE [mg/dL]
1[a]	-92.4 ± 24.0	-74.5 ± 10.6	-79.5 ± 3.5	-122.0 ± 12.7	-93.5 ± 31.8	20.8	11.9	8.5	13.8	10.0
2[a]	-100.3 ± 23.4	-107.0 ± 1.4	-122.0 ± 1.4	-106.5 ± 12.0	-65.5 ± 12.0	9.5	17.6	13.8	24.9	18.5
3	-41.4 ± 37.2	-61.0 ± 39.6	-54.0 ± 25.5	11.0 ± 9.9	-61.5 ± 3.5	27.8	11.7	15.9	26.1	33.5
4[a]	-91.1 ± 47.6	-132.0 ± 58.0	-93.0 ± 35.4	-93.5 ± 57.3	-46.0 ± 22.6	61.3	19.1	16.9	17.9	13.1
5[a]	-104.1 ± 28.4	-74.0 ± 25.5	-105.0 ± 25.5	-110.0 ± 36.8	-127.5 ± 4.9	32.8	22.7	21.4	25.3	18.8
6	-118.5 ± 48.1	-119.5 ± 13.4	-127.5 ± 91.2	-96.5 ± 57.3	-130.5 ± 54.4	76.5	24.2	41.9	21.9	30.6
7[a]	-83.6 ± 54.4	-79.5 ± 46.0	-109.5 ±102.5	-43.5 ± 10.6	-101.8 ± 52.0	74.6	10.2	8.0	10.5	8.4
8[a]	-101.1 ± 39.1	-145.5 ± 48.8	-106.5 ± 20.5	-86.0 ± 5.7	-66.5 ± 31.8	37.8	23.5	20.1	18.3	15.2
9[a]	-86.6 ± 56.4	-37.3 ± 34.3	-65.0 ± 56.6	-111.5 ± 61.5	-132.5 ± 53.0	72.6	17.1	17.2	33.1	43.2
10[a]	-97.8 ± 50.0	-71.0 ± 2.8	-74.5 ± 101.1	-126.5 ± 29.0	-119.0 ± 36.8	60.0	28.9	35.0	15.7	28.9
11	-94.4 ± 33.9	-95.5 ± 14.8	-93.0 ± 63.6	-106.5 ± 54.4	-82.5 ± 14.8	52.3	23.5	24.6	22.8	23.0
12[a]	-55.4 ± 26.7	-20.8 ± 15.2	-66.0 ± 22.6	-83.0 ± 5.7	-52.0 ± 7.1	17.9	22.1	41.0	17.8	21.3
13	-36.0 ± 23.1	-9.0 ± 2.8	-48.0 ± 15.6	-27.0 ± 18.4	-60.0 ± 9.9	16.5	16.3	15.1	13.0	12.4
14	-112.6 ± 42.8	-98.0 ± 46.7	-102.0 ± 65.1	-97.8 ± 35.7	-152.5 ± 29.0	62.4	19.5	24.6	19.3	18.5
15	-88.1 ± 36.7	-104.5 ± 82.7	-102.5 ± 17.7	-73.0 ± 11.3	-72.5 ± 16.3	45.3	16.1	15.9	16.1	16.8
16[a]	-77.1 ± 33.9	-57.5 ± 17.7	-45.0 ± 25.5	-91.5 ± 27.6	-114.5 ± 16.3	30.8	23.4	20.4	19.8	21.0
17	-69.4 ± 38.0	-23.0 ± 19.8	-111.5 ± 0.7	-79.5 ± 17.7	-63.5 ± 36.1	26.3	22.0	28.1	22.8	32.0
18	-97.0 ± 48.7	N/A	-115.5 ± 2.1	-135.0 ± 4.2	-40.5 ± 43.1	23.3	19.4	23.5	39.0	53.5
19	-74.0 ± 23.6	-78.0 ± 25.5	-68.0 ± 12.7	-95.0 ± 15.6	-55.0 ± 33.9	31.0	20.0	30.1	25.3	29.0
20[a]	-63.0 ± 54.0	-140.0 ± 21.2	-54.5 ± 20.5	-13.0 ± 33.9	-44.5 ± 26.2	36.0	20.1	20.1	22.5	19.0
Mean ± Std	-84.2 ± 43.26	-80.4 ± 46.6	-87.1 ± 42.9	-84.3 ± 44.2	-84.1 ± 40.7	40.8 ± 20.9	19.5 ± 4.7	22.1 ± 9.4	21.3 ± 6.8	23.3 ± 11.3

548
549 **Table 1 Changes in glucose during exercise, and results of the best performing ML models to predict minimum**
550 **glucose in the 4 hours following exercise**. [a] indicates participants with higher aerobic fitness. [b] indicates the
551 performance of the model designed to predict minimum glucose at the end of exercise, specifically, the MARS model
552 designed with exercise history and adaptive personalization. [c] indicates the performance of the model designed to predict

minimum glucose within 4-hours following the start of exercise, specifically, the MARS model that underwent adaptive personalization. Participant 18 SMBG data was not available for the standard-care arm, and is not reported.

	Population Model		Personalized Model, Coefficient Adaptation		Comparison between population model and personalized model	
	RMSE *(MAE)* [mg/dL]	[Sensitivity, Specificity] *(Accuracy)* [%]	RMSE *(MAE)* [mg/dL]	[Sensitivity, Specificity] *(Accuracy)* [%]	Δ MAE [%]	Δ Accuracy [%]
Predicting Minimum Glucose at the end of Exercise						
MARS Model						
Training, 16-fold CV	24.1 *(19.2)*	[73, 67] *(69)*	--	--	**--**	**--**
Validation, Holdout Set	26.5 *(23.4)*	[50, 86] *(75)*	23.1 *(19.6)*	[70, 86] *(81)*	-16.2	+ 8.3
Validation, 20-fold CV	24.6 *(20.0)*	[63, 63] *(67)*	23.0 *(18.1)*	[61, 78] *(78)*	- 9.5	+ 16.1 [a]
MARS Model + Exercise History Features						
Training, 16-fold CV	23.1 *(18.2)*	[75, 65] *(68)*	--	--	**--**	**--**
Validation, Holdout Set	18.7 *(14.3)*	[73, 86] *(81)*	19.7 *(15.8)*	[73, 95] *(88)*	+ 10.1	+ 7.7
Validation, 20-fold CV	22.6 *(17.6)*	[66, 69] *(70)*	22.1 *(17.5)*	[51, 83] *(77)*	- 0.6	+ 10.1 [a]
AR Model: Population Model [b]						
Training, 16-fold CV	28.8 *(22.7)*	[71, 94] *(83)*	--	--	**--**	**--**
Validation, Holdout Set	32.8 *(28.6)*	[59, 87] *(72)*	27.6 *(23.3)*	[59, 87] *(72)*	-18.7	+ 0
Validation, 20-fold CV	29.6 *(23.8)*	[71, 91] *(81)*	27.7 *(22.0)*	[76, 90] *(83)*	- 7.4	+ 3.1
Logistic Regression						
Training, 16-fold CV	--	[66, 67] *(66)*	--	--	--	--
Validation, Holdout Set	--	[73, 76] *(75)*	--	[73, 90] *(84)*	--	+ 12.5
Validation, 20-fold CV	--	[64, 56] *(61)*	--	[68, 61] *(70)*	--	+ 15.5 [a]
Predicting Minimum Glucose 4 hours after exercise						
MARS Model [b]						
Training, 16-fold CV	25.8 *(19.7)*	[67, 68] *(68)*	--	--	**--**	**--**
Validation, Holdout Set	25.7 *(21.6)*	[18, 76] *(56)*	21.5 *(16.3)*	[33, 96] *(78)*	- 24.8	+ 38.9
Validation, 20-fold CV	25.1 *(20.1)*	[62, 51] *(56)*	23.3 *(18.3)*	[56, 70] *(68)*	- 9.0 *	+ 21.4 [a]
MARS Model + Exercise History Features [b]						
Training, 16-fold CV	24.8 *(18.6)*	[79, 61] *(69)*	--	--	**--**	**--**
Validation, Holdout Set	30.7 *(26.1)*	[29, 61] *(47)*	23.0 *(16.0)*	[56, 96] *(84)*	-38.8	+ 80.0
Validation, 20-fold CV	26.3 *(21.1)*	[74, 52] *(57)*	23.9 *(18.2)*	[57, 70] *(69)*	- 13.8	+ 20.0
Logistic Regression [b]						
Training, 16-fold CV	--	[57, 72] *(65)*	--	--	--	**--**
Validation, Holdout Set	--	[32, 77] *(50)*	--	[53, 92] *(69)*	--	**+** 37.5
Validation, 20-fold CV	--	[63, 50] *(58)*	--	[64, 74] *(70)*	--	+ 20.4 [a]

Table 2 Comparing the effect of adaptation on the performance of models designed to predict exercise-related changes in glucose. Values represent the mean performance across participants. Training is performed with data from

n = 16 participants, while the holdout set includes data from n = 4 participants. The 20-fold validation includes data from all n = 20 participants. [a] indicates that the significance $p < 0.05$ determined Wilcoxon signed-rank test for paired, non-parametric data comparing the change in error or accuracy on a per-participant basis. [b] These models return predicted CGM, not SMBG values. The AR model is only designed to predict glucose approximately 43.2 minutes after the start of exercise, and the results for a 4 hour prediction horizon are not shown.

Model	Personalization	Accuracy [%]		MARE [%]	
		Lower VO_2max n = 70 obs	*Higher VO_2max* n = 88 obs	*Lower VO_2max* n = 70 obs	*Higher VO_2max* n = 88 obs
Predicting minimum glucose during exercise					
MARS Model	Population	73	65	23	20
	Adaptation	*84*	*72*	*20*	*20*
MARS Model + Exercise History	Population	74	65	19	19
	Adaptation	*86*	*70*	*19*	*20*
AR Model	Population	88	75 [a]	16	38 [a]
	Adaptation	*85*	*82*	*16*	*34* [a]
Logistic Regression	Population	65	55	--	--
	Adaptation	*72*	*65*	--	--
Predicting minimum glucose in the 4 hours following exercise					
MARS Model	Population	57	55	25	21
	Adaptation	*70*	*69*	*22*	*20*
MARS Model+ Exercise History	Population	63	52	25	24
	Adaptation	*69*	*68*	*23*	*20* [a]
Logistic Regression	Population	58	56	--	--
	Adaptation	*72*	*63*	--	--

Table 3 Comparing the effect of aerobic fitness on the performance of models designed to predict exercise-related changes in glucose. [a] indicates that the significance $p < 0.05$ as determined by Wilcoxon rank-sum test for non-parametric data, comparing algorithm performance on participants with higher aerobic and lower aerobic fitness rankings.

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for data and code should be directed and will be fulfilled by the lead contact, Peter G. Jacobs (jacobsp@ohsu.edu).

Materials Availability

This study did not generate unique reagents or materials.

Data and Code Availability

- De-identified human participant research data used in this analysis was granted for this analysis, and further data sharing is restricted and is not publically available. No standardized data types are reported in this manuscript. Data requests can be submitted to the lead contact. These requests are assessed on a case-by-case basis and require completion and signature of a sharing agreement, as defined by the Oregon Health & Science University Institutional Review Board (OHSU IRB). Summary statistics have been reported in the main manuscript.
- The algorithms designed in this manuscript are listed in the key resources table. The code used to perform the formal analysis of restricted participant data is available from the lead contact upon reasonable request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects and Study Setting

This analysis was performed upon approval of OHSU Institutional Review Board, study number 00019659. This analysis utilized data obtained during a previous clinical study (Castle *et al.*, 2018). The data was collected from 20 adults with type 1 diabetes (N = 20, 14 F, Age 34.5 ± 4.7y, duration diabetes 19.7 ± 8.6 y, BMI 26 ± 5.7, HbA1C 7.5 ± 0.8, VO_2max 37.1 ± 9.6) who participated in a 4-arm study. Each study arm consisted of 4 days of either (1) single-hormone automated insulin therapy, (2) dual-hormone (insulin and glucagon) automated therapy, (3) predictive low glucose suspend CGM-augmented pump therapy, or (4) standard of care CGM-augmented pump therapy. Participants visited the clinic on days 1 and 4 of each study arm. During in-clinic study visits, participants consumed a self-selected breakfast, lunch and dinner and performed aerobic exercise in the afternoon. Each participant consumed the same meals at the same time and performed the same physical activity at the same time for each of the 8 in-clinic visits (4 arms x 2 days). Participants underwent VO_2max testing using a modified Bruce protocol while wearing a gas-collecting mask. They performed aerobic exercise with graded work intensity every 3 minutes until volitional exhaustion or plateau of oxygen consumption. During the study, aerobic exercise was performed at 70% VO2max and was designed to last for 40 minutes. Participants sometimes exercised for less than 40 minutes if for example, their glucose dropped below 70 mg/dL. Participants sometimes exercised for longer than 40 minutes if they needed to stop in the middle of exercise for some reason. The average length of exercise across all participants was 43.2 ± 14 minutes. Participant accelerometer and heartrate data were obtained using ZephyrLife BioPatch devices (Zephyr, Annapolis, MD). The automated insulin and glucagon delivery systems were controlled using a custom exercise-aware

algorithm (Jacobs et al., 2015) installed on a Google Nexus smart phone. This automated delivery system wirelessly communicated the t:slim pumps (Tandem, San Diego, CA) and G5 CGM sensors (Dexcom, San Diego, CA) via Bluetooth. During the control arm, participants used their own insulin pumps. The insulin pumps in this study were filled with aspart insulin (Novo Nordisk, Plainsboro, NJ). This secondary analysis utilized participant data obtained from G5 devices, t:slim devices, ZephyrLife BioPatch devices and self-monitored blood glucose (SMBG) Contour Next devices (Bayer, Whippany, NJ).

METHOD DETAILS

Model Input Features and Outcome Measures

The participant data collected by the study devices during each of the in-clinic exercise sessions was processed for predictive exercise features and glucose outcomes following exercise (N = 160 exercise sessions). No observations were excluded from analysis on the basis of artifacts in the time series data, such as noise in CGM data due to calibration or movement, or signal dropout. The input features derived from the clinical data are defined in Table S1. Additional features describing participant exercise history are defined in Table S2. The final input features for each model were determined from *Greedy sequential variable selection (Whitney, 1971)*, or reproduced as described in previous publications (Breton *et al.*, 2018; Romero-Ugalde *et al.*, 2019). The algorithms were trained to predict (1) the minimum glucose from the start of exercise to the end of exercise as measured using self-monitored blood glucose

(SMBG) or continuous glucose monitor (CGM), and (2) minimum glucose 4 hours following the start of exercise as measured by CGM. SMBG measurements were measured by all participants at the start and end of exercise per study protocol, however SMBG was not always measured in the 4 hour period following exercise therefore CGM is used for the 4-hour prediction model. Participant age, sex, and VO_2max were used to classify each participant into categories of higher (including good, excellent, and superior VO_2max) aerobic fitness or lower (including very poor, poor and fair VO_2max) aerobic fitness, as defined by the American Society of Sports Medicine VO_2max aerobic fitness norms (American College of Sports Medicine's Complete Guide to Fitness & Health by Barbara Bushman, 2017).

Development of the Population Models

Three machine learning models were investigated to predict glucose outcomes during aerobic exercise. The first model is a MARS model was designed to predict minimum blood glucose during exercise, and minimum CGM-measured glucose in the 4 hours following exercise. The second model is a logistic regression model designed to predict hypoglycemia during exercise, and in the 4 hours following exercise. The third model developed was an AR model to predict CGM values approximately 43.2 minutes after the start of exercise. To investigate if exercise history is predictive of future glucose trends, a fourth model, a personalized MARS model was designed that incorporates participant exercise history features as inputs to the model (Table S2). Each population model was designed using a training set, which consisted of data from 16 participants. The population machine learning models were trained using leave-one-participant-out

667 cross-validation, meaning the input features and model parameters were selected using
668 fifteen of the participants in the training set, and then performance was evaluated on the
669 sixteenth held-out participant. The machine learning models were then evaluated on
670 data from a holdout set, which consisted of data from the 4 participants who were not
671 used in the training set. These 4 holdout participants were sampled to ensure that they
672 were representative of the population and had the same frequency of hypoglycemia and
673 minimum glucose as the training set. The general predictive accuracy of the models
674 were also evaluated using a 20-fold leave-one-participant-out cross-validation, where
675 the model parameters were retrained on 19 participants and the model performance
676 evaluated on 1 held-out participant (Figure S1).
677
678 *MARS Model to Predict Low SMBG after Exercise*
679 A MARS model implements a linear regression framework that also considers the
680 numerical range of the predictors. Each input feature (Table S1) was processed into
681 paired hinge-functions, representing the feature values above and below a specific
682 hinge point (i.e., SMBG values above and below a hinge point of 150 mg/dL are
683 considered separate variables with separate model coefficients). Candidate hinge
684 points for a given feature were determined by sorting observations within a feature and
685 selecting every 5th value for efficiency. The optimal hinge points were determined from
686 the set of candidate hinge points during supervised training of the algorithm. Next,
687 *Greedy sequential variable selection (Whitney, 1971)* was used to iteratively identify
688 optimal hinge-functions to predict minimum glucose during exercise. The MARS model
689 coefficients were designed using a weighted regression; this approach places a penalty

on MARS model misestimation of observations with glucose < 70 mg/dL. This essentially minimizes predictive error as well as improves sensitivity and specificity of the algorithm to detect hypoglycemia. The final model structure used to predict the minimum glucose during aerobic exercise is shown in equation (1). The model coefficients (in this case, β_0, β_1, β_2, and β_3) along with the hinge points are solved for each short-term and long-term prediction horizon model separately during model training.

$$\begin{aligned} &Minimum\ glucose\ during\ exercise \\ &= \beta_0 + \beta_1 * \max\left(0, CGM_{Start\ of\ Exercise} - 254\right) \left[\frac{mg}{dL}\right] \\ &+ \beta_2 \max\left(0, CBG_{Start\ of\ Exercise} - 124\right) \left[\frac{mg}{dL}\right] \\ &+ \beta_3 \max\left(0, HR_{10\ minutes\ prior\ to\ exercise} - 97.15\right) [BPM] \\ &+ \beta_4\ CGM\ Trend_{25\ minutes\ prior} \left[\frac{\frac{mg}{dL}}{min}\right] \end{aligned}$$

Equation 1

AR Model to Predict CGM Following Exercise

Romero-Ugalde et al. developed predictive models to forecast CGM measurements during aerobic exercise (Romero-Ugalde *et al.*, 2019). We used the methods and features described by Romero-Ugalde et al. to reproduce the population AR model that utilizes CGM data. In this approach, the CGM data is smoothed using a 1st-order simple moving average, whereby data-at time *t* is averaged with the preceding data at time *t*-5.

We found that the AR model using only CGM data and no exogenous inputs performed better than when including exogenous inputs. The AR with exogenous inputs (ARX) described in Romero-Ugalde et al. utilized raw activity data metrics from a different activity sensor than the one used in our study, and this may explain why they got better performance using an ARX model than using an AR model. We present the design and results of the AR model that achieved the highest accuracy during model validation. The exercise sessions in our dataset lasted on average for 43.2 ±14 minutes, therefore the AR model was designed to predict CGM at approximately 43.2 minutes following the start of exercise. The final model structure is shown below in equation (2) where the coefficients β_0, β_1, β_2, and β_3 are solved for during model training.

$$CGM\ 40\ minutes\ after\ the\ start\ of\ exercise = \beta_0 + \beta_1 SmoothCGM_{Start\ of\ Exercise}\left[\frac{mg}{dL}\right] + \beta_2 SmoothCGM_{10\ minutes\ preceding\ exercise}\left[\frac{mg}{dL}\right] + \beta_3 SmoothCGM_{20\ minutes\ preceding\ exercise}\left[\frac{mg}{dL}\right]$$

Equation 2

Logistic Regression to Predict Hypoglycemia

Breton et al. published a logistic regression model to predict hypoglycemia during exercise. We used the identical variables described by Breton et al. (Breton *et al.*, 2018) to train a population logistic regression model to predict the occurrence of hypoglycemia during aerobic exercise and in the 4 hours following exercise. The inputs

to this model were the CGM at the start of exercise, the average CGM trend in the hour preceding exercise, and the ratio of the active insulin (IOB) at the start of exercise to the participant's total daily insulin requirement (TDIR). The participant TDIR is defined as the total insulin dosed per day on average. The model is shown in equation (3) where the coefficients β_0, β_1, β_2, and β_3 are solved for during model training.

$$Probability\ of\ Hypoglycemia = logit\ (\beta_0 + \beta_1 CGM_{Start\ of\ Exercise} \left[\frac{mg}{dL}\right] + \beta_2 Average\ CGM\ Trend_{Prior\ Hour} \left[\frac{\frac{mg}{dL}}{min}\right] + \beta_3 \frac{IOB_{Start\ of\ Exercise}}{TDIR} \left[\frac{Units}{Units}\right])$$

Equation 3

MARS Model Personalized with Exercise History

The methods described above were used create a second personalized MARS model that incorporates exercise history from a given participant. The model was designed by identifying the optimal features included in Table S1, and also exercise history features included in Table S2 that describe participants' glucose dynamics during prior exercise sessions. The population model to detect minimum CGM-measured glucose during exercise is shown below in equation (4) whereby the coefficients β_0-β_6 were solved for each short-term and long-term prediction horizon model separately during training of the model.

$$
\begin{aligned}
Minimum\ glucose\ during\ exercise \\
&= \beta_0 + \beta_1 \max\left(0, CGM_{Start\ of\ Exercise} - 254\right) \left[\frac{mg}{dL}\right] \\
&+ \beta_2 \max\left(0, 254 - CGM_{Start\ of\ Exercise}\right) \left[\frac{mg}{dL}\right] \\
&+ \beta_3 \max\left(0, HR_{10\ minutes\ prior\ to\ exercise} - 97.15\right) [BPM] \\
&+ \beta_4 \max(0, Average\ \Delta CGM_{other\ exercise\ sessions} + 84.92) \left[\frac{mg}{dL}\right] \\
&+ \beta_5 \max(0, -84.92 - Average\ \Delta CGM_{other\ exercise\ sessions}) \left[\frac{mg}{dL}\right] \\
&+ \beta_6 \max(0, 5.97 - Average\ MET_{other\ exercise\ sessions}) [MET]
\end{aligned}
$$

Equation 4

Real-time Model Adaptation

To determine the impact of adaptation on prediction accuracy, the population model parameters were adapted to each participant left-out of model training using data from the participant's exercise observations. Stochastic gradient descent (An overview of gradient descent optimization algorithms, Ruder, 2016) was used to update the population model parameters using the participant's most recent observed exercise session, and the adapted model was then used to predict the same participant's outcomes of the next exercise session. This adaptation procedure was repeated successively for each held-out exercise observation, updating the population model parameters over time to better reflect a held-out participant's glucose dynamics as each exercise session was observed. In order to determine if the order of the exercise sessions impacted prediction accuracy, the order of the 8 identical exercise sessions were shuffled four times and the adaptation procedure was repeated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameter details are included here; additional parameters, including the number of observations and the precise statistical tests, are included in the figure and table legends. Significance testing was performed on a per-participant level, df = 19, to compare the change in error and accuracy to detect hypoglycemia before and after personalization. Normality of data was assessed using the Kolmogorov-Smirnov test to determine the appropriate statistical tests. The differences in model error before and after personalization were evaluated using a two-tailed paired t-test for parametric data, and a two-tailed Wilcoxon signed-rank test for non-parametric data, significance level of alpha = 0.05. The differences in glucose outcomes for participants in different physical fitness categories were evaluated using a two-tailed students t-test for parametric data, and a two-tailed Wilcoxon rank-sum test for non-parametric data, significance level alpha = 0.05. Glucose outcomes measured during exercise for each participant was explored with a boxplot, whereby the centerline of the boxplot indicates the median measurement and box edges represent the 25^{th} and 75^{th} percentiles, cross symbols represent outlier values and each whisker extends to the most extreme data point that is not an outlier. Model performance was assessed using root mean squared error (RMSE), mean absolute error (MAE), as well as the sensitivity, specificity and accuracy to detect observations with level 1 hypoglycemia (< 70 mg/dL). Leave-4-participant-out cross-validation was used to create a receiver operating curve for each algorithm to determine the optimal predictive threshold to detect hypoglycemia. The optimal threshold for each algorithm was then used to evaluate algorithm sensitivity, specificity, and accuracy to detect hypoglycemia for left-out participant data (Figure S1). The

Parkes consensus error grid analysis (Parkes et al., 2000) was used to determine the clinical impact of the algorithm predictions. Model design and assessment, and statistical analysis were performed in Matlab 2019b (MathWorks, Natick, MA). A power analysis was performed previously for the published clinical study; a study size of 20 participants was sufficient to detect a -3.3% change in % time-in-hypoglycemia and a 16.3% change in % time-in-target glucose (70-180 mg/dL), for >80% power and an alpha = 0.0125 (Castle *et al.*, 2018).

807

808 REFERENCES

809

810 Al Khalifah, R.A., Suppère, C., Haidar, A., Rabasa-Lhoret, R., Ladouceur, M., and
811 Legault, L. (2016). Association of aerobic fitness level with exercise-induced
812 hypoglycaemia in Type 1 diabetes. Diabetic Medicine *33*, 1686-1690.
813 https://doi.org/10.1111/dme.13070.
814 Ben Brahim, N., Place, J., Renard, E., and Breton, M.D. (2015). Identification of Main
815 Factors Explaining Glucose Dynamics During and Immediately After Moderate Exercise
816 in Patients With Type 1 Diabetes. J Diabetes Sci Technol *9*, 1185-1191.
817 10.1177/1932296815607864.
818 Bohn, B., Herbst, A., Pfeifer, M., Krakow, D., Zimny, S., Kopp, F., Melmer, A.,
819 Steinacker, J.M., and Holl, R.W. (2015). Impact of Physical Activity on Glycemic Control
820 and Prevalence of Cardiovascular Risk Factors in Adults With Type 1 Diabetes: A
821 Cross-sectional Multicenter Study of 18,028 Patients. Diabetes Care *38*, 1536-1543.
822 10.2337/dc15-0030.
823 Boulé, N.G., Weisnagel, S.J., Lakka, T.A., Tremblay, A., Bergman, R.N., Rankinen, T.,
824 Leon, A.S., Skinner, J.S., Wilmore, J.H., Rao, D.C., and Bouchard, C. (2005). Effects of
825 Exercise Training on Glucose Homeostasis. Diabetes Care *28*, 108.
826 10.2337/diacare.28.1.108.
827 Breton, M.D., Patek, S.D., Lv, D., Schertz, E., Robic, J., Pinnata, J., Kollar, L., Barnett,
828 C., Wakeman, C., Oliveri, M., et al. (2018). Continuous Glucose Monitoring and Insulin
829 Informed Advisory System with Automated Titration and Dosing of Insulin Reduces
830 Glucose Variability in Type 1 Diabetes Mellitus. Diabetes Technol Ther *20*, 531-540.
831 10.1089/dia.2018.0079.
832 Brown, S.A., Kovatchev, B.P., Raghinaru, D., Lum, J.W., Buckingham, B.A., Kudva,
833 Y.C., Laffel, L.M., Levy, C.J., Pinsker, J.E., Wadwa, R.P., et al. (2019). Six-Month
834 Randomized, Multicenter Trial of Closed-Loop Control in Type 1 Diabetes. The New
835 England journal of medicine *381*, 1707-1717. 10.1056/NEJMoa1907863.
836 Bussau, V.A., Ferreira, L.D., Jones, T.W., and Fournier, P.A. (2006). The 10-s maximal
837 sprint: a novel approach to counter an exercise-mediated fall in glycemia in individuals
838 with type 1 diabetes. Diabetes Care *29*, 601-606. 10.2337/diacare.29.03.06.dc05-1764.
839 Castle, J.R., El Youssef, J., Wilson, L.M., Reddy, R., Resalat, N., Branigan, D.,
840 Ramsey, K., Leitschuh, J., Rajhbeharrysingh, U., Senf, B., et al. (2018). Randomized
841 Outpatient Trial of Single- and Dual-Hormone Closed-Loop Systems That Adapt to
842 Exercise Using Wearable Sensors. Diabetes Care *41*, 1471-1477. 10.2337/dc18-0228.
843 Colberg, S.R., Sigal, R.J., Yardley, J.E., Riddell, M.C., Dunstan, D.W., Dempsey, P.C.,
844 Horton, E.S., Castorino, K., and Tate, D.F. (2016). Physical Activity/Exercise and
845 Diabetes: A Position Statement of the American Diabetes Association. Diabetes Care
846 *39*, 2065-2079. 10.2337/dc16-1728.
847 Contreras, I., Oviedo, S., Vettoretti, M., Visentin, R., and Vehí, J. (2017). Personalized
848 blood glucose prediction: A hybrid approach using grammatical evolution and
849 physiological models. PLOS ONE *12*, e0187754. 10.1371/journal.pone.0187754.
850 Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. The Annals of
851 Statistics *19*, 1-67. 10.1214/aos/1176347963.

Garg, S.K., Weinzimer, S.A., Tamborlane, W.V., Buckingham, B.A., Bode, B.W., Bailey, T.S., Brazg, R.L., Ilany, J., Slover, R.H., Anderson, S.M., et al. (2017). Glucose Outcomes with the In-Home Use of a Hybrid Closed-Loop Insulin Delivery System in Adolescents and Adults with Type 1 Diabetes. Diabetes technology & therapeutics *19*, 155-163. 10.1089/dia.2016.0421.

Guillot, F.H., Jacobs, P.G., Wilson, L.M., Youssef, J.E., Gabo, V.B., Branigan, D.L., Tyler, N.S., Ramsey, K., Riddell, M.C., and Castle, J.R. (2020). Accuracy of the Dexcom G6 Glucose Sensor during Aerobic, Resistance, and Interval Exercise in Adults with Type 1 Diabetes. Biosensors (Basel) *10*. 10.3390/bios10100138.

Hajizadeh, I., Rashid, M., Samadi, S., Feng, J., Sevil, M., Hobbs, N., Lazaro, C., Maloney, Z., Brandt, R., Yu, X., et al. (2018a). Adaptive and Personalized Plasma Insulin Concentration Estimation for Artificial Pancreas Systems. Journal of Diabetes Science and Technology *12*, 639-649. 10.1177/1932296818763959.

Hajizadeh, I., Rashid, M., Turksoy, K., Samadi, S., Feng, J., Sevil, M., Hobbs, N., Lazaro, C., Maloney, Z., Littlejohn, E., and Cinar, A. (2018b). Incorporating Unannounced Meals and Exercise in Adaptive Learning of Personalized Models for Multivariable Artificial Pancreas Systems. Journal of Diabetes Science and Technology *12*, 953-966. 10.1177/1932296818789951.

Hobbs, N., Hajizadeh, I., Rashid, M., Turksoy, K., Breton, M., and Cinar, A. (2019). Improving Glucose Prediction Accuracy in Physically Active Adolescents With Type 1 Diabetes. J Diabetes Sci Technol *13*, 718-727. 10.1177/1932296818820550.

Jacobs, P.G., El Youssef, J., Reddy, R., Resalat, N., Branigan, D., Condon, J., Preiser, N., Ramsey, K., Jones, M., Edwards, C., et al. (2016). Randomized trial of a dual-hormone artificial pancreas with dosing adjustment during exercise compared with no adjustment and sensor-augmented pump therapy. Diabetes, Obesity and Metabolism *18*, 1110-1119. 10.1111/dom.12707.

Jacobs, P.G., Resalat, N., El Youssef, J., Reddy, R., Branigan, D., Preiser, N., Condon, J., and Castle, J. (2015). Incorporating an Exercise Detection, Grading, and Hormone Dosing Algorithm Into the Artificial Pancreas Using Accelerometry and Heart Rate. J Diabetes Sci Technol *9*, 1175-1184. 10.1177/1932296815609371.

Lascar, N., Kennedy, A., Hancock, B., Jenkins, D., Andrews, R.C., Greenfield, S., and Narendran, P. (2014). Attitudes and barriers to exercise in adults with type 1 diabetes (T1DM) and how best to address them: a qualitative study. PLoS One *9*, e108019. 10.1371/journal.pone.0108019.

Li, Z., Keel, S., Liu, C., He, Y., Meng, W., Scheetz, J., Lee, P.Y., Shaw, J., Ting, D., Wong, T.Y., et al. (2018). An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs. Diabetes Care *41*, 2509. 10.2337/dc18-0147.

Man, C.D., Breton, M.D., and Cobelli, C. (2009). Physical activity into the meal glucose-insulin model of type 1 diabetes: in silico studies. J Diabetes Sci Technol *3*, 56-67. 10.1177/193229680900300107.

McMahon, S.K., Ferreira, L.D., Ratnam, N., Davey, R.J., Youngs, L.M., Davis, E.A., Fournier, P.A., and Jones, T.W. (2007). Glucose requirements to maintain euglycemia after moderate-intensity afternoon exercise in adolescents with type 1 diabetes are increased in a biphasic manner. The Journal of clinical endocrinology and metabolism *92*, 963-968. 10.1210/jc.2006-2263.

Moniotte, S., Owen, M., Barrea, T., Robert, A., and Lysy, P.A. (2017). Outcomes of algorithm-based modifications of insulinotherapy during exercise in MDI vs insulin pump-treated children with type 1 diabetes: Results from the TREAD-DIAB study. Pediatric diabetes *18*, 925-933. 10.1111/pedi.12509.

Montaser, E., Diez, J.L., Rossetti, P., Rashid, M., Cinar, A., and Bondia, J. (2019). Seasonal Local Models for Glucose Prediction in Type 1 Diabetes. IEEE Journal of Biomedical and Health Informatics, 1-1. 10.1109/JBHI.2019.2956704.

Moser, O., Riddell, M.C., Eckstein, M.L., Adolfsson, P., Rabasa-Lhoret, R., van den Boom, L., Gillard, P., Nørgaard, K., Oliver, N.S., Zaharieva, D.P., et al. (2020). Glucose management for exercise using continuous glucose monitoring (CGM) and intermittently scanned CGM (isCGM) systems in type 1 diabetes: position statement of the European Association for the Study of Diabetes (EASD) and of the International Society for Pediatric and Adolescent Diabetes (ISPAD) endorsed by JDRF and supported by the American Diabetes Association (ADA). Diabetologia *63*, 2501-2520. 10.1007/s00125-020-05263-9.

Mosquera-Lopez, C., and Jacobs, P.G. (2021). Incorporating Glucose Variability into Glucose Forecasting Accuracy Assessment Using the New Glucose Variability Impact Index and the Prediction Consistency Index: An LSTM Case Example. Journal of Diabetes Science and Technology, 19322968211042621. 10.1177/19322968211042621.

Parkes, J.L., Slatin, S.L., Pardo, S., and Ginsberg, B.H. (2000). A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. Diabetes Care *23*, 1143-1148. 10.2337/diacare.23.8.1143.

Pérez-Gandía, C., Facchinetti, A., Sparacino, G., Cobelli, C., Gómez, E.J., Rigla, M., de Leiva, A., and Hernando, M.E. (2010). Artificial Neural Network Algorithm for Online Glucose Prediction from Continuous Glucose Monitoring. Diabetes Technology & Therapeutics *12*, 81-88. 10.1089/dia.2009.0076.

Reddy, R., El Youssef, J., Winters-Stone, K., Branigan, D., Leitschuh, J., Castle, J., and Jacobs, P.G. (2018). The effect of exercise on sleep in adults with type 1 diabetes. Diabetes Obes Metab *20*, 443-447. 10.1111/dom.13065.

Reddy, R., Resalat, N., Wilson, L.M., Castle, J.R., El Youssef, J., and Jacobs, P.G. (2019). Prediction of Hypoglycemia During Aerobic Exercise in Adults With Type 1 Diabetes. J Diabetes Sci Technol *13*, 919-927. 10.1177/1932296818823792.

Riddell, M., Li, Z., Beck, R.W., Gal, R., Jacobs, P.G., Castle, J.R., Gillingham, M., Clements, M.A., Patton, S.R., Dassau, E., et al. (2020a). More time in glucose range during exercise days than sedentary days in adults living with type 1 diabetes. Diabetes Technology & Therapeutics. 10.1089/dia.2020.0495.

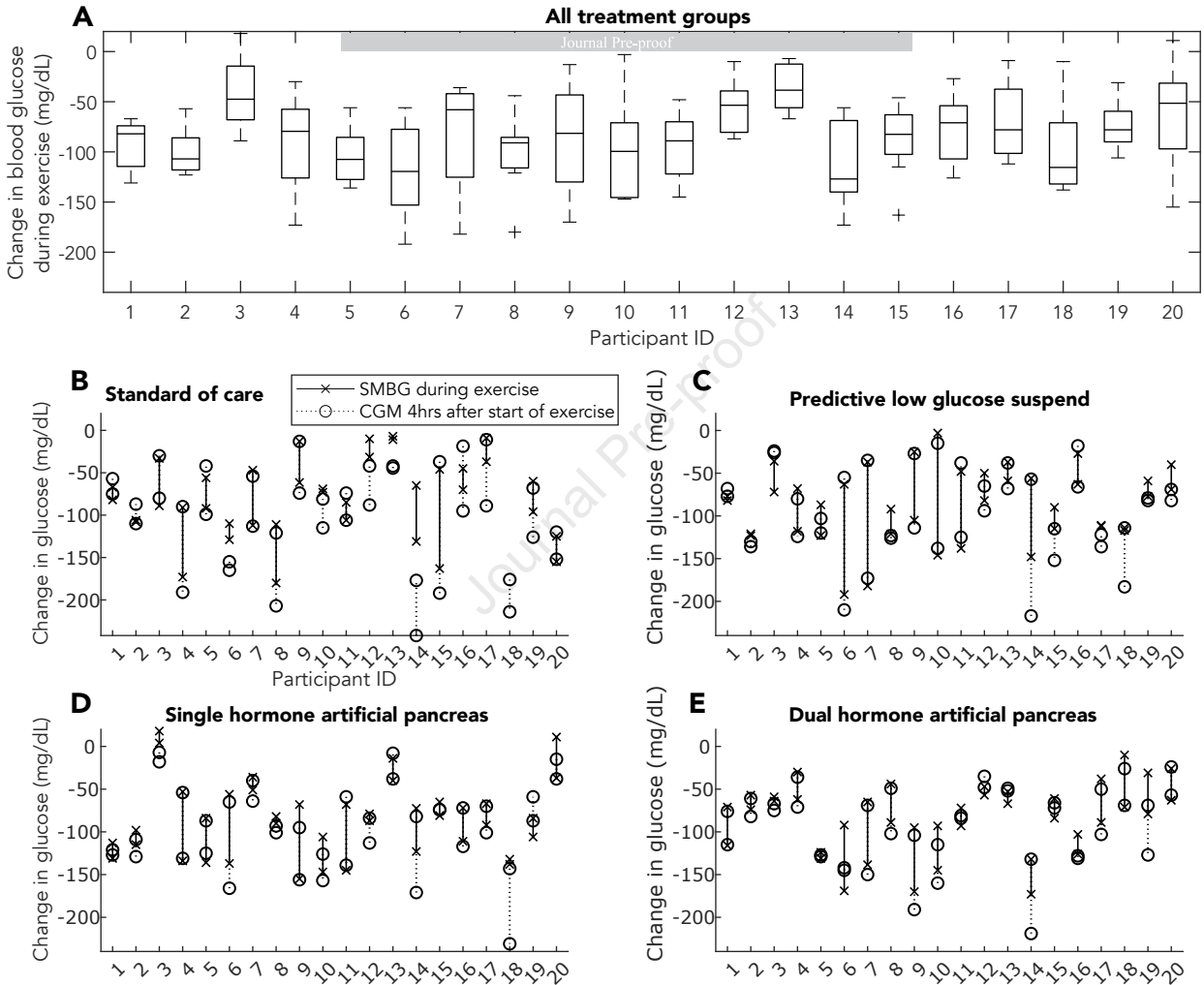
Riddell, M.C., Gallen, I.W., Smart, C.E., Taplin, C.E., Adolfsson, P., Lumb, A.N., Kowalski, A., Rabasa-Lhoret, R., McCrimmon, R.J., Hume, C., et al. (2017). Exercise management in type 1 diabetes: a consensus statement. Lancet Diabetes Endocrinol *5*, 377-390. 10.1016/s2213-8587(17)30014-1.

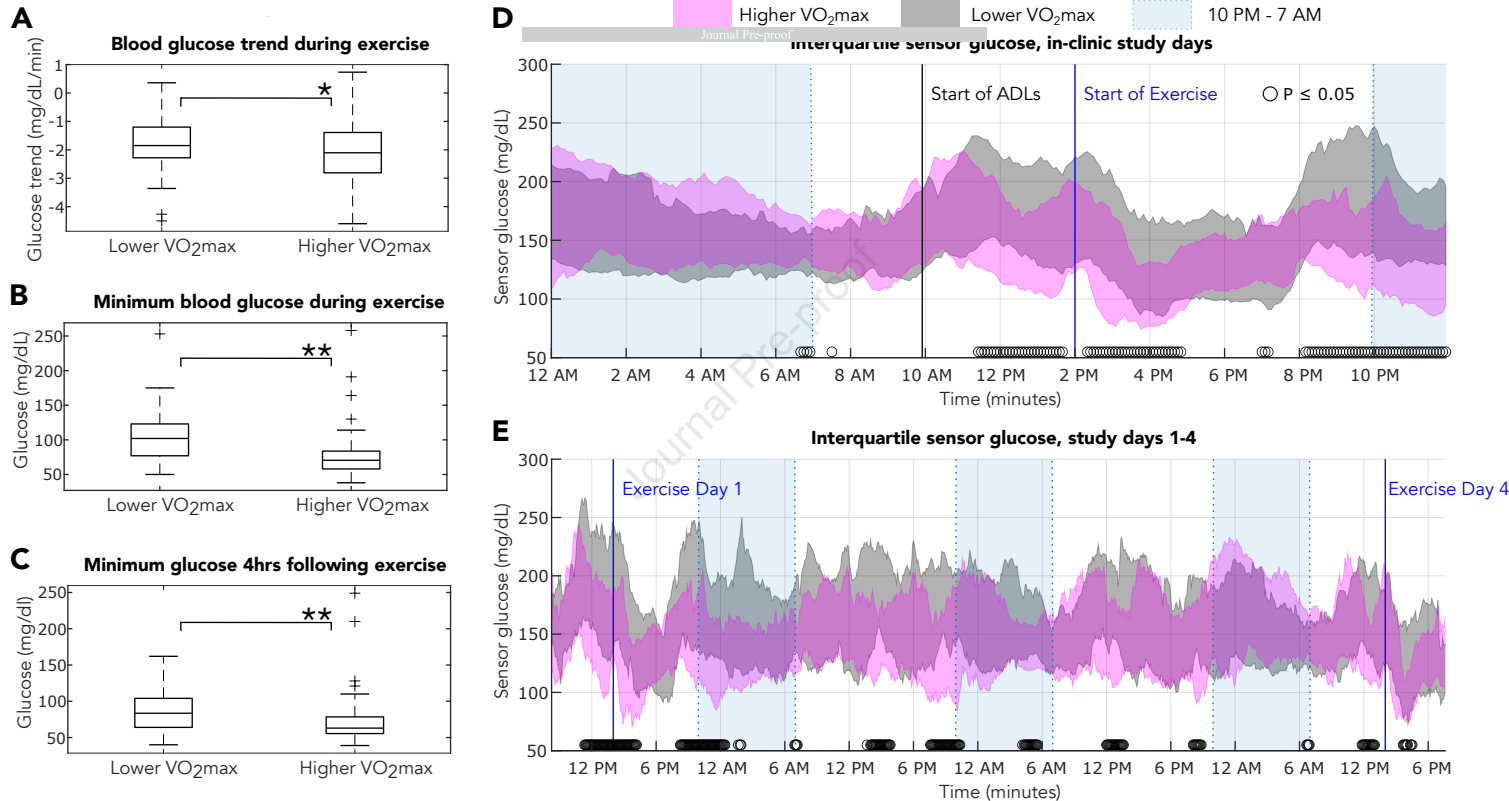
Riddell, M.C., Pooni, R., Fontana, F.Y., and Scott, S.N. (2020b). Diabetes Technology and Exercise. Endocrinology and Metabolism Clinics of North America *49*, 109-125. https://doi.org/10.1016/j.ecl.2019.10.011.

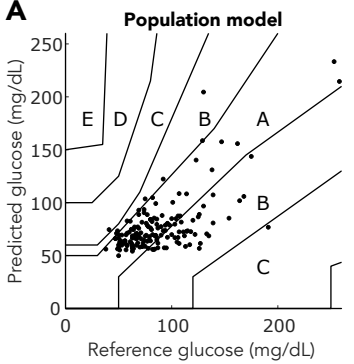
Romero-Ugalde, H.M., Garnotel, M., Doron, M., Jallon, P., Charpentier, G., Franc, S., Huneker, E., Simon, C., and Bonnet, S. (2019). ARX model for interstitial glucose

prediction during and after physical activities. Control Engineering Practice *90*, 321-330. https://doi.org/10.1016/j.conengprac.2019.07.013.
Steenberg, D.E., Jørgensen, N.B., Birk, J.B., Sjøberg, K.A., Kiens, B., Richter, E.A., and Wojtaszewski, J.F.P. (2019). Exercise training reduces the insulin-sensitizing effect of a single bout of exercise in human skeletal muscle. J Physiol *597*, 89-103. 10.1113/JP276735.
Tyler, N.S., and Jacobs, P.G. (2020). Artificial Intelligence in Decision Support Systems for Type 1 Diabetes. Sensors (Basel, Switzerland) *20*, 3214. 10.3390/s20113214.
Tyler, N.S., Mosquera-Lopez, C.M., Wilson, L.M., Dodier, R.H., Branigan, D.L., Gabo, V.B., Guillot, F.H., Hilts, W.W., El Youssef, J., Castle, J.R., and Jacobs, P.G. (2020). An artificial intelligence decision support system for the management of type 1 diabetes. Nature Metabolism *2*, 612-619. 10.1038/s42255-020-0212-y.
Wahren, J. (1977). Glucose turnover during exercise in man. Annals of the New York Academy of Sciences *301*, 45-55. 10.1111/j.1749-6632.1977.tb38184.x.
Whitney, A.W. (1971). A Direct Method of Nonparametric Measurement Selection. IEEE Transactions on Computers *C-20*, 1100-1103. 10.1109/T-C.1971.223410.
Wilson, L.M., Jacobs, P.G., Ramsey, K.L., Resalat, N., Reddy, R., Branigan, D., Leitschuh, J., Gabo, V., Guillot, F., Senf, B., et al. (2020a). Dual-Hormone Closed-Loop System Using a Liquid Stable Glucagon Formulation Versus Insulin-Only Closed-Loop System Compared With a Predictive Low Glucose Suspend System: An Open-Label, Outpatient, Single-Center, Crossover, Randomized Controlled Trial. Diabetes Care *43*, 2721. 10.2337/dc19-2267.
Wilson, L.M., Jacobs, P.G., Riddell, M.C., Zaharieva, D.P., and Castle, J.R. (2022). Opportunities and challenges in closed-loop systems in type 1 diabetes. The Lancet Diabetes & Endocrinology *10*, 6-8. https://doi.org/10.1016/S2213-8587(21)00289-8.
Wilson, L.M., Tyler, N., Jacobs, P.G., Gabo, V., Senf, B., Reddy, R., and Castle, J.R. (2020b). Patient Input for Design of a Decision Support Smartphone Application for Type 1 Diabetes. Journal of diabetes science and technology *14*, 1081-1087. 10.1177/1932296819870231.
Xie, J., and Wang, Q. (2020). Benchmarking Machine Learning Algorithms on Blood Glucose Prediction for Type I Diabetes in Comparison With Classical Time-Series Models. IEEE Transactions on Biomedical Engineering *67*, 3101-3124. 10.1109/TBME.2020.2975959.
Yan, Z., Okutsu, M., Akhtar, Y.N., and Lira, V.A. (2010). Regulation of exercise-induced fiber type transformation, mitochondrial biogenesis, and angiogenesis in skeletal muscle. Journal of Applied Physiology *110*, 264-274. 10.1152/japplphysiol.00993.2010.
Yardley, J.E., Kenny, G.P., Perkins, B.A., Riddell, M.C., Balaa, N., Malcolm, J., Boulay, P., Khandwala, F., and Sigal, R.J. (2013). Resistance versus aerobic exercise: acute effects on glycemia in type 1 diabetes. Diabetes Care *36*, 537-542. 10.2337/dc12-0963.
Yardley, J.E., and Sigal, R.J. (2021). Glucose management for exercise using continuous glucose monitoring: should sex and prandial state be additional considerations? Diabetologia *64*, 932-934. 10.1007/s00125-020-05373-4.
Zecchin, C., Facchinetti, A., Sparacino, G., Nicolao, G.D., and Cobelli, C. (2012). Neural Network Incorporating Meal Information Improves Accuracy of Short-Time Prediction of Glucose Concentration. IEEE Transactions on Biomedical Engineering *59*, 1550-1560. 10.1109/TBME.2012.2188893.

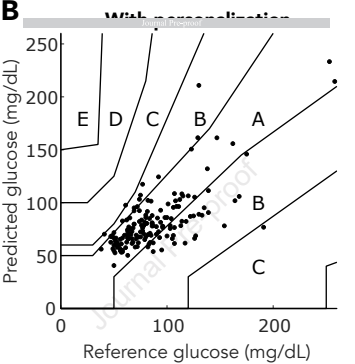
990 Zhu, T., Li, K., Chen, J., Herrero, P., and Georgiou, P. (2020). Dilated Recurrent Neural
991 Networks for Glucose Forecasting in Type 1 Diabetes. Journal of Healthcare Informatics
992 Research *4*, 308-324. 10.1007/s41666-020-00068-2.
993
994



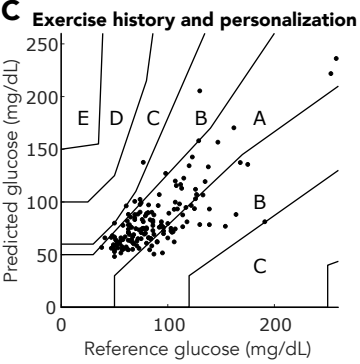

A
Blood glucose trend during exercise
Glucose trend (mg/dL/min)
1
0
-1
-2
-3
-4
*
Lower VO2max
Higher VO2max
B
Minimum blood glucose during exercise
Glucose (mg/dL)
250
200
150
100
50
**
Lower VO2max
Higher VO2max
C
Minimum glucose 4hrs following exercise
Glucose (mg/dl)
250
200
150
100
50
**
Lower VO2max
Higher VO2max
D
Higher VO2max
Lower VO2max
10 PM - 7 AM
Journal Pre-proof
Interquartile sensor glucose, in-clinic study days
Sensor glucose (mg/dL)
300
250
200
150
100
50
Start of ADLs
Start of Exercise
P ≤ 0.05
12 AM
2 AM
4 AM
6 AM
8 AM
10 AM
12 PM
2 PM
4 PM
6 PM
8 PM
10 PM
Time (minutes)
Journal Pre-proof
E
Interquartile sensor glucose, study days 1-4
Sensor glucose (mg/dL)
300
250
200
150
100
50
Exercise Day 1
Exercise Day 4
12 PM
6 PM
12 AM
6 AM
12 PM
6 PM
12 AM
6 AM
12 PM
6 PM
12 AM
6 AM
12 PM
6 PM
Time (minutes)



A = 69.6%
B = 29.1%
C = 1.3%
D = 0%
E = 0%
A+B = 98.7%



A = 72.8%
B = 25.9%
C = 1.3%
D = 0%
E = 0%
A+B = 98.7%



A = 74.7%
B = 24.7%
C = 0.6%
D = 0%
E = 0%
A+B = 99.4%

Highlights (3-4 with no more than 85 characters, including spaces)

- People with type 1 diabetes exercised in 8 identically-designed treadmill sessions
- Intra-person glycemic response varies even under controlled and repeated conditions
- Glucose trends downward more quickly in people with higher aerobic fitness
- Adaptive ML algorithms predict exercise-related nadir glucose with high accuracy

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Human Participant Data	Castle, J.R., El Youssef, J., Wilson, L.M., Reddy, R., Resalat, N., Branigan, D., Ramsey, K., Leitschuh, J., Rajhbeharrysingh, U., Senf, B., et al. (2018). Randomized Outpatient Trial of Single- and Dual-Hormone Closed-Loop Systems That Adapt to Exercise Using Wearable Sensors. Diabetes Care *41*, 1471-1477. 10.2337/dc18-0228	N/A
Software and Algorithms		
Matlab 2019b	Mathworks, Natick, MA. www.mathworks.com	N/A
Multivariate Adaptive Regression Spline	Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistics *19*, 1-67. 10.1214/aos/1176347963	N/A
Logistic Regression Model for Hypoglycemia Prediction	Breton, M.D., Patek, S.D., Lv, D., Schertz, E., Robic, J., Pinnata, J., Kollar, L., Barnett, C., Wakeman, C., Oliveri, M., et al. (2018). Continuous Glucose Monitoring and Insulin Informed Advisory System with Automated Titration and Dosing of Insulin Reduces Glucose Variability in Type 1 Diabetes Mellitus. Diabetes Technol Ther *20*, 531-540. 10.1089/dia.2018.0079	N/A
AR Model for Exercise Glucose Prediction	Romero-Ugalde, H.M., Garnotel, M., Doron, M., Jallon, P., Charpentier, G., Franc, S., Huneker, E., Simon, C., and Bonnet, S. (2019). ARX model for interstitial glucose prediction during and after physical activities. Control Engineering Practice *90*, 321-330. https://doi.org/10.1016/j.conengprac.2019.07.013	N/A
Stochastic Online Gradient Descent Algorithm	Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747	N/A