nature portfolio

Peer Review File

Personalized Insulin Dosing using Reinforcement Learning for High-Fat Meals and Aerobic Exercises in Type 1 Diabetes: a Proof-of-Concept Trial

Editorial Note: This manuscript has been previously reviewed at another journal that is not operating a transparent peer review scheme. This document only contains reviewer comments and rebuttal letters for versions considered at *Nature Communications*.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In recognition of the opposing effects of high fat meals vs aerobic exercise on glucose levels the authors have developed a computerised personalised decision support system using machine learning which advises insulin dosing for people with type 1 diabetes managing their glucose levels on multiple daily injections. They tested feasibility in a single-arm 16-week outpatient a study in 15 adults with type 1 diabetes in a before and after study design. They report benefits in hypoglycaemia reduction and incremental areas under the curve post-prandially and post-exercise with less insulin administered at mealtimes. The authors have an excellent track record. The manuscript reads well. I am not able to provide expert comment on the algebraic formula. My comments below focus on the clinical aspects and study design which are not as well detailed as the mathematical proofs of the algorithm.

Major Comments

1/ The greatest weakness in the protocol is the before and after study design without randomisation in a small number of participants (n=15). It is possible that a learning effect contributed to the improvements observed in the second stage. Also, it is possible that in operating the insulin dosing advisor the participants' attention was drawn to their food intake with the timing and dosing of the rapid acting insulin independent of the algorithm. As the authors state "However, as this is an uncontrolled study, findings would need to be confirmed in randomized controlled trials." We can only conclude that the intervention is feasible but I would have reservations regarding any conclusions about efficacy.

2/ Great detail is provided regarding the algorithm development. Very limited information is provided regarding the study design itself. How were data on meals and exercise documented? Was there any way of confirming that exercise was performed during the post prandial period? How soon after the meal bolus did exercise occur, and was it indeed aerobic? How were mealtimes confirmed so that CGM metrics during the five-hour post-prandial period could be calculated?

3/ Participants announced if they were eating a high fat (≥20g) meal. Given that the sentence before implies that not all were able to assess carbohydrate content, how reliable was their estimation of the fat content? Was there any confirmation that they did actually eat a high fat meal?

4/ While insulin dose is provided no descriptors are included regarding the meals apart from the observation that there were approximately 50% more high fat meals in the intervention stage suggesting that meals were not controlled in the two stages.

5/ While one accepts that the numbers would be small it would have been helpful to see what happened with high fat meals that did and did not precede exercise, and non-high fat meals that did and did not precede exercise.

6/ The closer that exercise is performed to the meal-time insulin bolus the greater the risk for hypoglycaemia. How close to the meal boluses did the bouts of exercise occur? Unless a substantial proportion of exercise bouts were within 2 hours of the meal bolus, ultimately, the person with type 1 diabetes still must forward plan when exercising.

7/ Exercise will also impact insulin dosing for the meal that follows. Were there any observations relating to algorithm performance with this?

8/ What did the users think of the insulin dose calculator? What proportion of meals did they use it for? Did they always follow the advice provided, and did its use tail off with time?

Minor Comments

1/ The abstract includes the following statement:

"The 5-hour postprandial incremental area under the curve of glucose levels was significantly

improved in the evaluation period compared to the baseline period of the study for both high-fat meals (378±222 vs 38±223 mmol/L/min, p=0.03) and meals followed by exercises (-395±192 vs 132±181 mmol/L/min, p=0.007).”
My interpretation is that the sentence suggests evaluation comes before baseline but the results sequence show that the baseline levels are more stable. One suspect that this is an error as it reverses what is documented in the results section.
2/ The authors may wish to consider including the detailed proofs of their equations as an appendix or supplementary material rather than the main text. Many readers may skip over this part which constitutes a substantial proportion (10 pages) of the methods section.
3/ Given comparisons are being made between an intervention and run-in, was the trial registered?

Reviewer #2 (Remarks to the Author):

The authors introduce reinforcement learning algorithms to optimize dosing adjustments for high-fat meals and postprandial aerobic exercises, as well as for carbohydrate ratios and basal insulin doses. The validation of this algorithm was conducted through a 16-week clinical study. While this study is pioneering in its application of reinforcement learning algorithms in real-world debates care using a smartphone app and cloud server, I have several concerns:
* The motivation behind the algorithm design is not sufficiently described. A more detailed explanation regarding the composition of the state space, action space, reward function, and the distinctions between basal and bolus insulin settings (multi-agent vs. single-agent) would be beneficial.
* The approach to the exploration-exploitation trade-off in reinforcement learning is not clearly delineated. From the presented details, it appears that no exploration was undertaken for both Q-learning algorithms.
* Given that the algorithms were updated on a weekly basis, it raises the question of whether they truly converged to an optimal policy. A comparison of Q-values (from baseline to evaluation) would be insightful. In terms of evaluation on Q values, it would be better for the authors to confirm their findings with clinical experts.
* The clinical study has several limitations that need addressing. The sample size is relatively small, and the study design is single-armed. This makes it challenging to determine if the observed improvements are attributable to the decision support algorithms or merely the utilization of the app (digital systems). Furthermore, many results lack statistical significance. The study does not provide clarity on patient adherence in the outpatient setting, specifically regarding whether patients consistently administered the recommended insulin dosages.
* Evaluating the proposed models using type 1 diabetes simulators would significantly enhance the study's robustness. Such simulators would allow for the testing of various algorithm configurations, providing a more comprehensive understanding of their efficacy.

Reviewer #3 (Remarks to the Author):

The authors used reinforcement learning to develop a personalised decision support system for insulin dosage following meals and exercise. The proposed system was evaluated in a small, single-arm trial (pilot study) with 15 Type I diabetic patients and conducted over a period of 16 weeks. The results disclosed by the authors indicate promising potential for the more effective management of blood glucose levels. The study conscientiously acknowledges its limitations, notably the modest sample size, which consisted exclusively of adult participants, and the absence of a control group.

This manuscript contributes to the burgeoning field of AI applications in the medical domain. Notably, it is commendably well-written, with only a handful of minor grammatical issues that warrant attention. Following are minor comments aimed at improving its quality.

The authors mention that the clinical team reviews therapy parameter values before inputting them into the iBolusV2 app but do not provide information on whether adjustments or modifications to these values were necessary or how frequently this occurred. This information is crucial as it pertains to the app's autonomy from human intervention and the reliability of its decision-making process.

The manuscript could benefit from an explanation regarding why comparisons between baseline and evaluation periods were not adjusted for age-related effects, especially given the wide age range of the patients in the study (ranging from 22 to 63 years). It would also be valuable to comment on the influence of gender on these comparisons. Furthermore, the reference to the baseline and evaluation periods as "confounding variables" in the first sentence of the statistical analysis section may require revision. The mixed model analysis compares the outcomes at baseline and evaluation periods and so these two time periods are levels of the variable of interest.

We are grateful for the helpful feedback by the Reviewers. We responded to all points and have modified the manuscript accordingly. The page numbers referenced below are with respect to the highlighted revised manuscript. We also included new supplementary material in response to the Reviewers questions including 5 new figures, 4 new tables, participants' feedback from interviews, and the details of the learning algorithm.

Response to Reviewer 1

In recognition of the opposing effects of high fat meals vs aerobic exercise on glucose levels the authors have developed a computerised personalised decision support system using machine learning which advises insulin dosing for people with type 1 diabetes managing their glucose levels on multiple daily injections. They tested feasibility in a single-arm 16-week outpatient study in 15 adults with type 1 diabetes in a before and after study design. They report benefits in hypoglycaemia reduction and incremental areas under the curve post-prandially and post-exercise with less insulin administered at mealtimes. The authors have an excellent track record. The manuscript reads well. I am not able to provide expert comment on the algebraic formula. My comments below focus on the clinical aspects and study design which are not as well detailed as the mathematical proofs of the algorithm.

Major Comments

1. The greatest weakness in the protocol is the before and after study design without randomisation in a small number of participants (n=15). It is possible that a learning effect contributed to the improvements observed in the second stage. Also, it is possible that in operating the insulin dosing advisor the participants' attention was drawn to their food intake with the timing and dosing of the rapid acting insulin independent of the algorithm. As the authors state "However, as this is an uncontrolled study, findings would need to be confirmed in randomized controlled trials." We can only conclude that the intervention is feasible but I would have reservations regarding any conclusions about efficacy.

We agree with the Reviewer that the study design does not support any efficacy claims. The manuscript has been revised and every wording that may suggest efficacy has been modified.

2. Great detail is provided regarding the algorithm development. Very limited information is provided regarding the study design itself. How were data on meals and exercise documented?

Meals and exercise data were logged into the study app when the user used it to calculate prandial boluses. The meal data included the carbohydrate content of the meals (if the user counts the carbohydrate) and whether they were high-fat meals. The exercise data included the exercise description (e.g., soccer) and whether they will be performed less than 90 minutes or between 90 minutes and 4 hours from the mealtime. The app transmitted all data to the cloud whenever connected to Wi-Fi. This has now been clarified in the revised manuscript (page 17, paragraph 2nd last).

3. Was there any way of confirming that exercise was performed during the post prandial period?

There was no way for us to confirm that the exercises were indeed performed after their announcement to the system. In future studies, this could be mitigated by providing participants with a fitness tracker (e.g., Fitbit) and crosscheck their data with the study app. This has now been clarified in the revised manuscript (page 17, paragraph 1).

4. How soon after the meal bolus did exercise occur, and was it indeed aerobic?

64% of the exercises were announced to be performed within 90 minutes of the mealtime while 36% to be performed between 90 minutes and 4 hours from the mealtime. This has now been added to the revised manuscript (page 20, paragraph 2nd).

The exercises' descriptions (Extended Data Table 3) indicated that they were aerobic.

5. How were mealtimes confirmed so that CGM metrics during the five-hour post-prandial period could be calculated?

The mealtimes were logged in the study app whenever the user used the bolus calculator. These times were uploaded to the cloud and used to define the postprandial periods. This has now been clarified in the revised manuscript (page 17, paragraph 2nd last).

6. Participants announced if they were eating a high fat (≥20g) meal. Given that the sentence before implies that not all were able to assess carbohydrate content, how reliable was their estimation of the fat content? Was there any confirmation that they did actually eat a high fat meal?

At the admission visit, participants received training on how to identify high-fat meals. Our team introduced participants to the "Diabetes Exchange List[1]" from the American Diabetes Association as a guide for assessing fat content. Additionally, our team reviewed participants recalls of their common meals and determined whether these meals qualified as high-fat or not. Importantly, the high-fat meal assessment in our study is relatively straightforward compared to carbohydrate assessment, as it involves making a binary decision regarding whether the fat content exceeds 20 grams or not. This is now clarified in the manuscript (page 116, paragraph last).

There was no way for us to confirm that participants did indeed eat high-fat meals after announcing them to the system. This has been now clarified in the manuscript (page 17, paragraph 1).

7. While insulin dose is provided no descriptors are included regarding the meals apart from the observation that there were approximately 50% more high fat meals in the intervention stage suggesting that meals were not controlled in the two stages.

[1] http://www.diabetesed.net/page/_files/THE-DIABETIC-EXCHANGE-LIST.pdf

We thank the reviewer for this comment. The number of high-fat meals was actually comparable between the baseline and evaluation periods (111 vs 120 meals). This is indicated in Table 1 and Figure 2.a.

8. While one accepts that the numbers would be small it would have been helpful to see what happened with high fat meals that did and did not precede exercise, and non-high fat meals that did and did not precede exercise.

When analyzing meals with both high-fat content and postprandial exercise, the 5-hour postprandial incremental area under the glucose curve was numerically improved from the baseline to the evaluation period (-560±456 vs 128±461, p=0.06) and the 5-hour postprandial percentage time spent below 3.9 mmol/L was also numerically improved (5.2±3.3 vs 3.5±3.2, p=0.65; Table S1), though the differences did not reach statistical significance.

Similar trends were observed when conducting the same analysis but for non-high-fat meals and exercise. The 5-hour postprandial incremental area under the glucose curve was numerically improved from the baseline to the evaluation period (-272±258 vs 158±238, p=0.052) and the percentage time spent below 3.9 mmol/L was reduced (5.4±1.1 vs 1.0±1.0, p<0.001).

These outcomes are now reported in Table S1 and Figure S1.

9. The closer that exercise is performed to the meal-time insulin bolus the greater the risk for hypoglycaemia. How close to the meal boluses did the bouts of exercise occur? Unless a substantial proportion of exercise bouts were within 2 hours of the meal bolus, ultimately, the person with type 1 diabetes still must forward plan when exercising.

Whenever the user announced an upcoming postprandial exercise, the app also asked them to indicate whether the exercise will be performed less than 90 minutes or between 90 minutes and 4 hours from the mealtime. 64% of the exercises were announced to be performed within 90 minutes while 36% to be performed between 90 minutes and 4 hours. This has been now discussed in the manuscript (page 17, paragraph 2).

We agree with the Reviewer that these number suggest that forward planning is indeed needed. If the majority of the exercises were performed in the late postprandial period, pre-exercise carbohydrate intake could be an option to prevent hypoglycemia during exercise. However, given the proximity of the exercise to mealtime, a substantial carbohydrate intake would be needed unless pre-prandial insulin is reduced. This has been now discussed in the manuscript (page 22, paragraph 3).

10. Exercise will also impact insulin dosing for the meal that follows. Were there any observations relating to algorithm performance with this?

We thank the Reviewer for raising this interesting question. We conducted additional analysis in which we compared meals that were preceded, but not followed, by exercise between the baseline and the evaluation periods. The 5-hour postprandial incremental area under the glucose curve was numerically reduced (358±152 vs 89±160, p=0.20) and hypoglycemia was also

numerically reduced (5.8±2.8 vs 4.0±2.8, p=0.34; Table S2), though none of the outcomes were statistically significant.

These outcomes are now reported in Table S2 and Figure S2.

11. What did the users think of the insulin dose calculator? What proportion of meals did they use it for? Did they always follow the advice provided, and did its use tail off with time?

We have conducted semi-structured interviews at the end of the interventions. We received informative feedback that will guide future iterations of the app and the algorithm.

Some users felt that the need to announce the exercise up to four hours in advance is not always practical:

> *"Sometimes I didn't plan for hours ahead. So sometimes I would only know, one or two hours in advance." (Participant 15)*

> *"Unless you have a real schedule that you're going to follow, it's a little bit less practical." (Participant 9)*

We therefore plan to add a feature to announce exercise at any time (as opposed to only at the time of the meal). If the exercise is announced outside the mealtime, then, instead of a prandial insulin reduction, the app will recommend a personalized carbohydrate intake based on the exercise, insulin-on-board, and the current glucose level.

Users also appreciated the ability to announce the time of the exercise (early vs late postprandial) and having the insulin dose recommendations adjusted accordingly:

> *"It definitely helped first of all having the within 90 minutes and after." (Participant 13)*

> *"If I was doing exercise within 90 minutes obviously it would give me a little bit less-lesser of an insulin bolus than it would if I wasn't…so I did it and it worked out." (Participant 10)*

We plan to keep this important feature, but to further increase flexibility, we plan replace the discrete timing option (early vs late postprandial) with a sliding scale of 30-min or 60-min increments.

Interestingly, several participants would have liked to be able to log snacks in the app:

> *"I think what I most would have liked to see is somewhere where I can log what I'm eating that I'm not getting insulin for." (Participant 8)*

> *"It would have been nice to be able to add like a snack, you know, like if I was eating something, there was nowhere to log it." (Participant 9)*

> *"If there was like a feature for me to enter instead of a meal if I'm having just like a small snack, because sometimes I would only have like a bowl of chips, or maybe like a granola*

bar or something, just a quick snack, to maybe log that instead of like a meal" (Participant 15)

We plan to modify he app to allow manual entry of carbohydrate at any time, and to use that information in the learning algorithm calculations.

Participants also appreciated the reminder to deliver the second bolus to manage high-fat meals:

> *"It helps because it reminds you to check again in two hours. So that was great." (Participant 3)*

> *"In the past, when I had like, let's say pizza, um, I would calculate just for the meal and take that amount, um I would drop into hypo and then immediately spike into hyper and stay really high for a long time. So that part of the app was really useful and helpful." (Participant 6)*

These participants' feedback is now included in the **supplement**.

Participants were instructed to use the app for all meals that required insulin boluses. Participants used the app for a median 2.4 times per day, which was relatively stable across the study duration. We have included a new Figure S3 in the revised manuscript illustrating the app usage over the study duration.

Participants followed the app's insulin dose recommendations for high-fat meals and meals followed by an exercise 88% and 87% of the time, respectively. This has been included in the revised manuscript (page 19, paragraph 2).

Minor Comments

1. The abstract includes the following statement: "The 5-hour postprandial incremental area under the curve of glucose levels was significantly improved in the evaluation period compared to the baseline period of the study for both high-fat meals (378±222 vs 38±223 mmol/L/min, p=0.03) and meals followed by exercises (-395±192 vs 132±181 mmol/L/min, p=0.007)." My interpretation is that the sentence suggests evaluation comes before baseline but the results sequence show that the baseline levels are more stable. One suspect that this is an error as it reverses what is documented in the results section.

This is indeed an error. We thank the Reviewer for pointing this out. We have corrected this in the revised manuscript.

2. The authors may wish to consider including the detailed proofs of their equations as an appendix or supplementary material rather than the main text. Many readers may skip over this part which constitutes a substantial proportion (10 pages) of the methods section.

We thank the Reviewer for this suggestion. We have moved the details of the algorithms to the supplementary material.

3. Given comparisons are being made between an intervention and run-in, was the trial registered?

Yes, the trial was registered at ClinicalTrials.gov (NCT05041621). This is mentioned in the Methods section of the manuscript (page 16, paragraph 2nd last).

Response to Reviewer 2

The authors introduce reinforcement learning algorithms to optimize dosing adjustments for high-fat meals and postprandial aerobic exercises, as well as for carbohydrate ratios and basal insulin doses. The validation of this algorithm was conducted through a 16-week clinical study. While this study is pioneering in its application of reinforcement learning algorithms in real-world debates care using a smartphone app and cloud server, I have several concerns:

1. The motivation behind the algorithm design is not sufficiently described. A more detailed explanation regarding the composition of the state space, action space, reward function, and the distinctions between basal and bolus insulin settings (multi-agent vs. single-agent) would be beneficial.

We thank the Reviewer for this question. We selected features in the state vector of the multi-agent algorithms to steer the late postprandial glucose levels into the desired target range while avoiding hypoglycemia. This was achieved by using the postprandial glucose error and the time spent below 3.9 mmol/L in the state vector. For high-fat meals, we additionally included the incremental area under the curve for both the early and late postprandial periods, considering the impact of high-fat meals on both time periods. For postprandial exercise, we incorporated estimated hypoglycemia treatments during exercise and the rate of glucose level change at the minimum glucose level value to minimize the risk of hypoglycemia beyond the observation period of the state features.

The reward function structure in the multi-agent algorithms was shaped with an exponential form so the agents receive rewards and penalties that increase exponentially in response to their actions, promoting faster convergence. The action space was chosen to be in the discrete domain to reduce the computational burden of running the algorithm. Moreover, the discrete state space accommodates the scarcity of the training and validation data as continuous function approximators like neural networks are impossible to use.

For the single-agent algorithm, the features in the state vector were chosen to steer the nighttime glucose level to the target range while avoiding hypoglycemia. To achieve this, we incorporated in the state vector the nighttime mean glucose levels, nighttime spent below 3.9 mmol/L and above 10 mmol/L, the early morning (4am to 7am) rate of glucose levels change, and the estimated hypoglycemia treatments (10pm to 2am). The rationale behind the structure of the reward function and action space is similar to that of the multi-agent algorithms.

Owing to the limited data available for high-fat meals and postprandial exercise in each main meal (breakfast, lunch, and dinner), we employed multi-agent reinforcement learning. In multi-agent reinforcement learning, each agent learns not only from its own experiences but also benefits from the collective knowledge gained by other agents (Equation. 6; supplementary

material). This collaborative learning strategy mitigates against the scarcity of data. In contrast, for basal insulin, we applied single-agent reinforcement learning, as we have daily data available, unlike the scenarios for high-fat meals and postprandial exercises, where data is less frequent.

These details are now mentioned in the supplement (pages 14-15, 17, and 19) and in the revised manuscript (page 8).

2. The approach to the exploration-exploitation trade-off in reinforcement learning is not clearly delineated. From the presented details, it appears that no exploration was undertaken for both Q-learning algorithms.

We thank the reviewer for this comment. To attain safer and faster convergence in the clinical study, we implemented two strategies for exploration. First, based on common clinical guidelines and exploratory analysis of data collected in another study (NCT04123054), we initialized the Q-values at baseline such that higher values are assigned to better state-action pairs. This approach bears similarity to the concept of reward shaping[2] and helps the agent to explore and learn optimal behaviour. To understand this, consider a scenario where each action incurs a negative reward (e.g., current dosing is too aggressive and leads to hypoglycemia, whether the action increases or decreases the dose by 10%). If we initialize all Q(s, a) to 0, the agent will quickly recognize that all its actions result in similarly unfavorable outcomes, causing it to stuck near its starting state. Conversely, when we adopt an optimistic initialization of Q(s, a) with large positive values for some state-action pairs, the agent will maintain hopeful expectations during the learning process, allowing it to explore further regions within the state-action space (e.g., continue to decrease doses until hypoglycemia is either eliminated or until enough negative rewards accumulate that forces the agent to explore another state-action pair).

Secondly, in the clinical study, our algorithms continued learning while exploring 10% of the time to fine-tune the Q-values in the presence of continuous variability between and within individuals ($\varepsilon = 0.1$ in Algorithm 1 and Algorithm 2; supplement).

This is now clarified in Algorithm 1 and Algorithm 2 in the supplement (pages 16 and 20).

3. Given that the algorithms were updated on a weekly basis, it raises the question of whether they truly converged to an optimal policy. A comparison of Q-values (from baseline to evaluation) would be insightful. In terms of evaluation on Q values, it would be better for the authors to confirm their findings with clinical experts.

The learning algorithms are run on each day's data to produce daily recommendations that are not sent the user. The weekly recommendations that are sent to the user are determined using the median of the daily recommendations. This is now clarified in the Methods section of the revised manuscript (page 15, paragraph last).

In the revised manuscript, we have included Figures S4 and S5 to illustrate the convergence of the Q-tables from the baseline to the evaluation period. The figures show the norm of the

[2]Wiewiora E. Potential-based shaping and Q-value initialization are equivalent. Journal of Artificial Intelligence Research. 2003 Sep 1;19:205-8.

difference matrix of consecutive Q-tables, which decreased over time and stabilizes around week 10. Importantly, the policy on each state of the Q-tables did not change after week 10 for any study participant.

In our study, the clinical team had to review and approve the algorithms' recommendations whenever they deviated by more than 30% from the baseline values or from the last approved values (page 17, paragraph 2nd last). Out of a total of 485 algorithm runs, 32 required medical reviews, and only 3 of the recommendations were overridden by the clinical team, corresponding to 0.6% override rate. This suggests a strong alignment between the algorithm's recommendations and those of the clinical team. This is now stated in the revised manuscript (page 19, paragraph 2).

4. The clinical study has several limitations that need addressing. The sample size is relatively small, and the study design is single-armed. This makes it challenging to determine if the observed improvements are attributable to the decision support algorithms or merely the utilization of the app (digital systems). Furthermore, many results lack statistical significance. The study does not provide clarity on patient adherence in the outpatient setting, specifically regarding whether patients consistently administered the recommended insulin dosages.

We agree with the Reviewer that the single-arm design raises the possibility that the observed benefits, or part of it, may be the results of the study effect (including the use of the app). The manuscript has been revised to remove any efficacy claim. The discussion has been amended to emphasize the limitation of the single-arm design (page 22, paragraph 3).

We agree that many comparisons lacked statistical significance. However, the main outcomes of interest (AUC of glucose for high fat meals and time in hypoglycemia for exercise meals) showed strong statistical significance, precluding chance as a factor in these differences.

The reviewer is correct that we are not able to confirm that the recommended doses were indeed delivered. However, we note that study participants used the app for a median 2.4 times per day (Figure S3) and that they overrode the recommendations of high-fat meals doses and postprandial exercise doses (indicating that they injected a different dose than the algorithm recommendation) 12% and 13% of the time, respectively, suggesting active engagement and usage of the app. This is now stated in the revised manuscript (page 19, paragraph 2).

5. Evaluating the proposed models using type 1 diabetes simulators would significantly enhance the study's robustness. Such simulators would allow for the testing of various algorithm configurations, providing a more comprehensive understanding of their efficacy.

We have performed preliminary simulations to test our algorithms, but we chose not to include the results in the manuscript as the simulation models are not fully validated and to focus our manuscript on the clinical assessment of the algorithms. For the Reviewer's interest, we include the results here.

Our simulator[3] for individuals with type 1 diabetes is based on the Hovorka model[4]. Each in-silico individual possesses unique and optimal carbohydrate ratios and long-acting insulin dose. To simulate exercise, the parameters of glucose uptake and insulin sensitivity are varied based on the exercise duration and intensity, as outlined in a previously published work from our group[5]. To simulate high-fat meals, we varied the parameters of peak time to meal-glucose appearance and insulin sensitivity[6].

We performed 60-day simulations in 40 in-silico individuals. Each in-silico individual consumed 40g, 60g, and 80g of carbohydrates for breakfast, lunch, and dinner, respectively, and consumed 15g of carbohydrates to treat hypoglycemia below 3.9 mmol/L. In-silico individuals consumed one high-fat meal every day in either breakfast, lunch, or dinner, randomly. We also introduced postprandial exercises every other day of 60 minutes at 50% of VO_{2max}, occurring 60 minutes after either breakfast, lunch, or dinner, randomly. There was no high-fat content and postprandial exercise for the same meal.

The Figures and Tables below compare the outcomes from the last 14 days to the first 14 days in all in-silico individuals. The outcomes for both high-fat meals and meals followed by exercise are comparable qualitatively to the outcomes observed in our clinical trial.

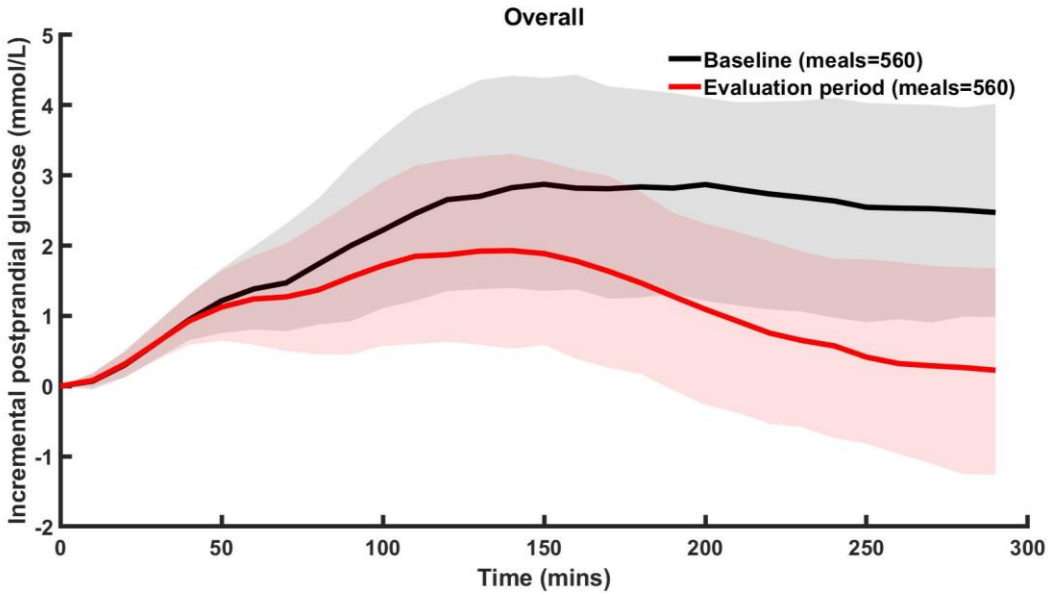

Figure. Simulated postprandial incremental glucose (median and interquartile range) after high-fat meals.

[3]El Fathi A, Kearney RE, Palisaitis E, Boulet B, Haidar A. A model-based insulin dose optimization algorithm for people with type 1 diabetes on multiple daily injections therapy. IEEE Transactions on Biomedical Engineering. 2020 Sep 11;68(4):1208-19.

[4]Hovorka R, Shojaee-Moradie F, Carroll PV, Chassin LJ, Gowrie IJ, Jackson NC, Tudor RS, Umpleby AM, Jones RH. Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT. American Journal of Physiology-Endocrinology and Metabolism. 2002 May 1;282(5):E992-1007.

[5]Alkhateeb H, El Fathi A, Ghanbari M, Haidar A. Modelling glucose dynamics during moderate exercise in individuals with type 1 diabetes. Plos one. 2021 Mar 26;16(3):e0248280.

[6] Laxminarayan S, Reifman J, Edwards SS, Wolpert H, Steil GM. Bolus estimation—rethinking the effect of meal fat content. Diabetes technology & therapeutics. 2015 Dec 1;17(12):860-6

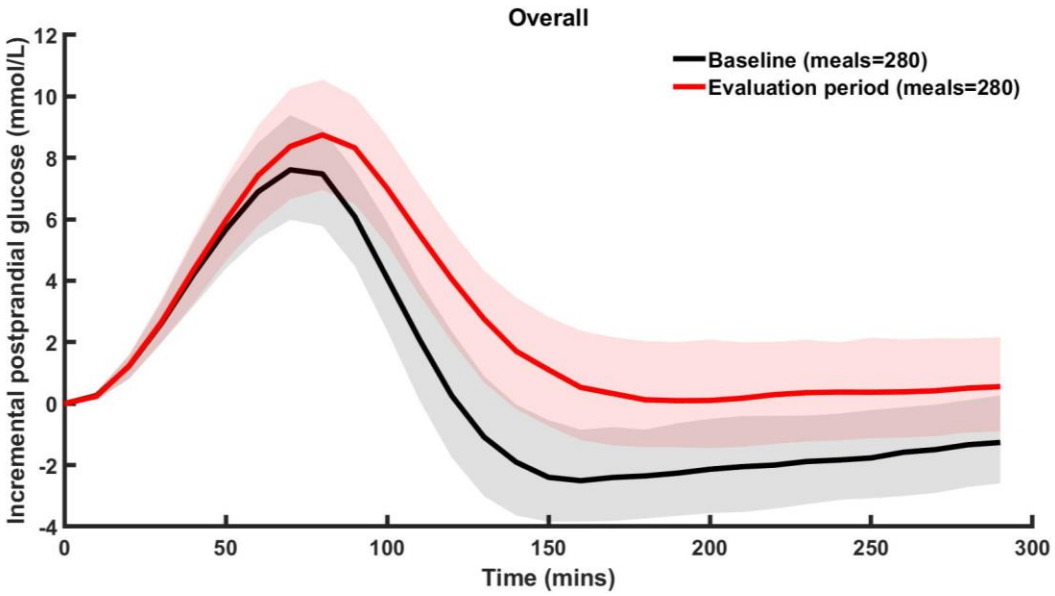

Figure. Simulated postprandial incremental glucose (median and interquartile range) after meals with exercise.

Table. Comparison of the baseline with the evaluation period in postprandial outcomes after high-fat meals with no postprandial exercises.

Overall	Baseline	Evaluation	P value
Number of High-Fat Meals	560	560	-
iAUC 0-300 min (mmol/L/min)	623±31	311±27	**<0.001**
TBR < 3.9 mmol/L	0.0±0.0	0.34±0.0	0.23
Mealtime Insulin (U)	6.1±0.4	5.2±0.5	**<0.001**
Later Bolus Insulin (U)	0.0±0.0	1.0±0.2	**<0.001**

Outcomes are reported as mean ± standard error.

Table. Comparison of the baseline with the evaluation period in postprandial outcomes after non high-fat meals with postprandial exercises.

Overall	Baseline period	Evaluation period	P value
Number of Exercise Sessions	280	280	-
iAUC 0-300 min (mmol/L/min)	150±46	350±49	**<0.001**
TBR < 3.9 mmol/L	5.0±0.1	0.3±0.0	**<0.001**
Mealtime Insulin (U)	6.3±0.1	2.0±0.6	**<0.001**

Outcomes are reported as mean ± standard error.

Response to Reviewer 3

The authors used reinforcement learning to develop a personalised decision support system for insulin dosage following meals and exercise. The proposed system was evaluated in a small, single-arm trial (pilot study) with 15 Type I diabetic patients and conducted over a period of 16 weeks. The results disclosed by the authors indicate promising potential for the more effective management of blood glucose levels. The study conscientiously acknowledges its limitations, notably the modest sample size, which consisted exclusively of adult participants, and the absence of a control group.

1. This manuscript contributes to the burgeoning field of AI applications in the medical domain. Notably, it is commendably well-written, with only a handful of minor grammatical issues that warrant attention. Following are minor comments aimed at improving its quality.

We thank the Reviewer for the positive assessment of our work.

2. The authors mention that the clinical team reviews therapy parameter values before inputting them into the iBolusV2 app but do not provide information on whether adjustments or modifications to these values were necessary or how frequently this occurred. This information is crucial as it pertains to the app's autonomy from human intervention and the reliability of its decision-making process.

We thank the reviewer for this comment. In our study, the clinical team reviewed the algorithms' recommendations whenever they deviated by more than 30% from the baseline values or from the last approved values (page 17, paragraph 2nd last). Out of a total of 485 algorithm runs, 32 required medical reviews, and only 3 of the recommendations were overridden by the clinical research team, corresponding to 0.6% override rate. This data supports the algorithm's autonomy from human intervention. These numbers are now stated in the revised manuscript (page 19, paragraph 2).

3. The manuscript could benefit from an explanation regarding why comparisons between baseline and evaluation periods were not adjusted for age-related effects, especially given the wide age range of the patients in the study (ranging from 22 to 63 years). It would also be valuable to comment on the influence of gender on these comparisons. Furthermore, the reference to the baseline and evaluation periods as "confounding variables" in the first sentence of the statistical analysis section may require revision. The mixed model analysis compares the outcomes at baseline and evaluation periods and so these two time periods are levels of the variable of interest.

We thank the Reviewer for this comment. We ran additional linear mixed model analysis (shown in Tables S3 and S4) in which we incorporated two more variables, age and sex (unfortunately, we did not collect gender data), to adjust the outcomes of both the baseline and evaluation periods. This supplementary analysis did not alter the conclusions of the previously observed results without adjustments.

The Reviewer is right that the baseline and evaluation periods are levels of the variable of interest and not confounding variables. This has been corrected in the revised manuscript (page 18, paragraph 2nd last). We thank the Reviewer for pointing this out to our attention.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I thank the authors for their detailed responses. I have no further comments regarding content. There are some minor editorial matters that I would recommend they address
1/ Randomized controlled trial to establish the efficacy of this personalized insulin dosing system are warranted." Should read as the plural "randomised controlled trials."
2/ Please ensure that references to all of the Supplementary material and Extended Data are made in the body of the manuscript at the relevant junctures with numbers in sequence and uniform nomenclature. I do not believe that this is the case at present. For example reference in the text to "Figure S1" comes after reference ti "supplement Figure S3" and (forgive me if I have missed it) I cannot find reference to Figure S2 which I think results from a typographical error:
"Separate outcomes for high-fat meals and non-high-fat meals with postprandial exercises are reported in the supplement (Table S1 and Figure S1). Outcomes for meals that were preceded, but not followed, by exercise are also reported in the supplement (Table S1 and Figure S1).
3/ In the early part of the Discussion the statement "This uncontrolled pilot clinical study suggested that with the use of an advanced decision support system based on reinforcement learning…"
To
"This uncontrolled pilot clinical study demonstrated feasibility of an advanced decision support system based on reinforcement learning, and suggested…"

Reviewer #2 (Remarks to the Author):

While the authors have addressed some of my previous concerns and provided additional details about their study, there are still significant issues that seem unresolvable within the scope of the current work. I believe that these issues prevent the paper from meeting the journal's standards, particularly in the following areas:

Concerns Regarding Glucose Level Measurements in Figures 2 and 3:

Upon closer examination of Figures 2 and 3, I noticed that the starting glucose levels during the evaluation period are higher than the baseline (overall plot). This discrepancy might be a crucial factor in the reported significant iAUC. Aligning the curves to match the initial glucose levels reveals that the postprandial glucose curves between the baseline and evaluation periods are quite similar. I suggest that the authors also present other metrics, such as AUC and TAR, for a more comprehensive analysis. However, I am concerned these metrics might not show any significant differences. Specifically, in Figure 2, the starting glucose level of 9.9 mmol/L during the evaluation period and the resulting low iAUC do not necessarily indicate effective glycemic control, as the glucose levels predominantly remain in the hyperglycemic zone. Presenting actual glucose levels, instead of incremental postprandial values starting from zero, would provide more clarity.

Rationale Behind Algorithm Design:

The authors assert that their use of multi-agent reinforcement learning is due to its lesser data requirements during training. However, this claim is questionable for several reasons:
The complexity of multi-agent environments, larger state and action spaces, and inherent non-stationarity in such settings generally make the learning process more data-intensive, not less. In particular, each additional agent in the environment exponentially increases the number of possible states and actions (in combination), making the learning process more data-intensive.

Multi-agent environments are non-stationary from the perspective of any individual agent because the strategies and actions of other agents are constantly changing. These factors challenge the authors' rationale behind choosing a multi-agent approach for their algorithm.

Safety Concerns with Greedy Policy Implementation:

The application of a greedy policy with a 0.1 factor implies that approximately 10% of the time, the agent selects actions randomly. This approach raises serious concerns about the algorithm's safety, especially if deployed in real-life scenarios without human oversight. I question the practicality and safety of using this application or algorithm for self-managing Type 1 Diabetes (T1D) without further validation and safeguards.

Inadequacies in Clinical Study Design:

The single-arm nature of the clinical study and its small sample size limit the ability to conclusively demonstrate the algorithm's effectiveness. It is plausible that any observed improvement is attributable to the app usage rather than the algorithm itself. The number of meal announcements on the app do not sufficiently represent this, particularly concerning basal insulin recommendations.

Minor Differences in Overall Glycemic Control and Lack of Technical Details:

Extended Data Table 4 shows minimal differences in overall glycemic control, with almost no change in all metrics between the first 10 days and the last 10 days of the study period. This observation casts doubt on the efficacy of the intervention. Additionally, the paper lacks detailed information on the engineering aspects of the app development and algorithm implementation, which are crucial for evaluating the study's technical rigor.

Reviewer #3 (Remarks to the Author):

The authors have addressed all my concerns.

We are grateful for the helpful reviews. We responded to all points and modified the manuscript accordingly. The page numbers referenced below are with respect to the highlighted revised manuscript.

Response to Reviewer 1

I thank the authors for their detailed responses. I have no further comments regarding content. There are some minor editorial matters that I would recommend they address

We thank the Reviewer for the additional editorial comments.

1. Randomized controlled trial to establish the efficacy of this personalized insulin dosing system are warranted." Should read as the plural "randomised controlled trials."

We have corrected this.

2. Please ensure that references to all of the Supplementary material and Extended Data are made in the body of the manuscript at the relevant junctures with numbers in sequence and uniform nomenclature. I do not believe that this is the case at present. For example reference in the text to "Figure S1" comes after reference ti "supplement Figure S3" and (forgive me if I have missed it) I cannot find reference to Figure S2 which I think results from a typographical error:

"Separate outcomes for high-fat meals and non-high-fat meals with postprandial exercises are reported in the supplement (Table S1 and Figure S1). Outcomes for meals that were preceded, but not followed, by exercise are also reported in the supplement (Table S1 and Figure S1).

We have corrected this. All tables and figures now are cited in the text with a consistent and a sequentially numbered nomenclature.

3. In the early part of the Discussion the statement "This uncontrolled pilot clinical study suggested that with the use of an advanced decision support system based on reinforcement learning…"
To
"This uncontrolled pilot clinical study demonstrated feasibility of an advanced decision support system based on reinforcement learning, and suggested…

We have rephrased this sentence (page 12).

Response to Reviewer 2

While the authors have addressed some of my previous concerns and provided additional details about their study, there are still significant issues that seem unresolvable within the scope of the current work. I believe that these issues prevent the paper from meeting the journal's standards, particularly in the following areas:

1. Concerns Regarding Glucose Level Measurements in Figures 2 and 3:

Upon closer examination of Figures 2 and 3, I noticed that the starting glucose levels during the evaluation period are higher than the baseline (overall plot). This discrepancy might be a crucial factor in the reported significant iAUC. Aligning the curves to match the initial glucose levels reveals that the postprandial glucose curves between the baseline and evaluation periods are quite similar. I suggest that the authors also present other metrics, such as AUC and TAR, for a more comprehensive analysis. However, I am concerned these metrics might not show any significant differences. Specifically, in Figure 2, the starting glucose level of 9.9 mmol/L during the evaluation period and the resulting low iAUC do not necessarily indicate effective glycemic control, as the glucose levels predominantly remain in the hyperglycemic zone. Presenting actual glucose levels, instead of incremental postprandial values starting from zero, would provide more clarity.

We thank the Reviewer for this comment. It is customary to use iAUC rather than AUC to assess postprandial glucose control as iAUC is more robust to variations in starting glucose levels. Nevertheless, in the revised manuscript, we have included AUC and time spent above 10 mmol/L outcomes (Tables 1 and 2) along with the AUC graphs for both high-fat meals and postprandial exercises (Figure S2).

For high-fat meals, overall AUC (0 to 300 minutes) and time spent above 10 mmol/L did not differ between the evaluation and baseline periods (p=0.38 and p=0.44; Table 1). However, AUC for the late postprandial period (240 to 300 minutes) was significantly lower (611±43 mmol/L/min vs 750±43 mmol/L/min, p=0.02; Table 1) in the evaluation period compared to the baseline period.

High-fat meals pose two main challenges: i) hypoglycemia in the early postprandial period[1], likely due to slowing gastric emptying, and ii) hyperglycemia in the late postprandial period[1], likely due to increased gluconeogenesis. In our study, early postprandial (< 2 hours) hypoglycemia was reduced by 56% (4.3±1.5% vs 1.9±1.4%, p=0.038; Table 1) in the evaluation period compared to the baseline period, combined with improvement in the AUC from 240 to 300 minutes. This was achieved while the algorithm reduced participants' insulin doses at mealtimes (9.3±1.2U vs 10.5±1.1U, p=0.09; Table 1) and increase it 2 hours after the meals (1.5±1.5U vs 0.3±1.0U, p<0.001; Table 1). Note that these adjustments to the insulin doses were not pre-defined but resulted from reactive real-time continuous adaptations for each individual based on their glucose levels. This has been discussed in the revised manuscript (Discussion, paragraph 3).

For meals followed by exercise, AUC was increased in the evaluation compared to the baseline period (p<0.001; Table 2), reflecting less drop in postprandial glucose levels, without increasing time spent above 10 mmol/L (p=0.16; Table 2). The main challenge with aerobic exercise is hypoglycemia. In our study, hypoglycemia was reduced by 73% (1.4±1.1% vs 5.3±1.2%, p=0.003; Table 2) in the evaluation period compared to the baseline period, with no increase in hyperglycemia. This was achieved with participants needing less insulin at mealtimes (8.0±1.4U vs 10.0±1.3U, p<0.001; Table 2) in the evaluation period compared to the baseline period. As

[1] Bell KJ, Fio CZ, Twigg S, Duke SA, Fulcher G, Alexander K, McGill M, Wong J, Brand-Miller J, Steil GM. Amount and type of dietary fat, postprandial glycemia, and insulin requirements in type 1 diabetes: a randomized within-subject trial. Diabetes Care. 2020 Jan 1;43(1):59-66.

with high-fat meals, these insulin doses were not pre-defined but resulted from reactive real-time continuous adaptations for each individual based on their glucose levels.

2. Rationale Behind Algorithm Design:
The authors assert that their use of multi-agent reinforcement learning is due to its lesser data requirements during training. However, this claim is questionable for several reasons:
The complexity of multi-agent environments, larger state and action spaces, and inherent non-stationarity in such settings generally make the learning process more data-intensive, not less. In particular, each additional agent in the environment exponentially increases the number of possible states and actions (in combination), making the learning process more data-intensive. Multi-agent environments are non-stationary from the perspective of any individual agent because the strategies and actions of other agents are constantly changing. These factors challenge the authors' rationale behind choosing a multi-agent approach for their algorithm.

Our aim was to provide personalized recommendations for high-fat meals and postprandial exercise for each main meal (breakfast, lunch, and dinner). For this task, we required three separate single-agents for high-fat meals and six agents for exercise (2 per each meal; one for exercises within 90 minutes of each meal and another for later than 90 minutes). Therefore, our problem inherently involves multiple agents.

We agree with the Reviewer that, in multi-agent learning, the state-action space grows exponentially with the number of agents, which would necessitate more data for training if each agent kept track of all other agents (i.e., centralized training and execution). However, a common method to mitigate this is to employ a decentralized learning and execution scheme, in which each agent learns and acts independently or with limited information from other agents. In our study, we used multi-agent learning with agents configuration in which each agent interacts with the environment using its own local state, action, and reward. To promote faster convergence, each agent additionally leverages the collective experience gained by other agents (Equation. 6; supplementary material). Consequently, in this multi-agent setting, we needed less data compared to using completely independent agents for each meal (i.e., several parallel single-agent reinforcement learnings). The language in the supplement (page 13-14) were slightly revised to clarify this point.

To illustrate the previous point, consider an agent for the breakfast meal that observed 100 meals, and that these meals were sufficient for the agent to be trained and to recommend optimal breakfast dose recommendations. With our algorithm, when the first lunch meal is consumed, the lunch agent will not need 100 new meals to learn its optimal recommendations. Instead, it will leverage the experiences gathered from the breakfast meals, until it becomes self-sufficient (Equation. 6-8; supplementary material).

We also agree with the Reviewer that this is a non-stationarity problem (i.e, the reward resulting from an action taken by a specific agent may vary depending on the actions taken by the other agents). In our study, this was addressed in four ways. First, we used a global Q-table to update each agent's individual Q-table during the initial learning period. That global Q-table represents an aggregation holding all agents' interactions with the environment in a weighted-average manner and partly alleviates the non-stationary problem as this global Q-table is known to all agents. Second, as each agent accumulates sufficient visitations in its own state-action pair, it

increasingly relies on its Q-value for that state-action pair and eventually transitions to a single-agent setting after the visitation for a particular state-action pair exceeds eight, as outlined in Equation 7. Third, in our study, only 17% and 33% of the days had more than one high-fat meal and postprandial exercise, suggesting minimal non-stationarity. Fourth, the weekly recommendations sent to the user are calculated based on the median of the daily recommendations, which additionally addresses the non-stationarity problem by smoothing out recommendations from days with one or more than one agent active.

Note also that the non-stationarity is an issue, albeit to a lesser extent, even when using only one agent for all meals (assuming there are no physiological variations across meals). For instance, variations in nighttime glucose control, whether improved or worsened, can affect daytime outcomes. Similarly, factors like missed boluses, snacks, errors in carbohydrate counting leading to suboptimal boluses, and the residual effects of exercise from the previous day can all contribute to changes in the dynamics of the environment.

3. Safety Concerns with Greedy Policy Implementation:
The application of a greedy policy with a 0.1 factor implies that approximately 10% of the time, the agent selects actions randomly. This approach raises serious concerns about the algorithm's safety, especially if deployed in real-life scenarios without human oversight. I question the practicality and safety of using this application or algorithm for self-managing Type 1 Diabetes (T1D) without further validation and safeguards.

We thank the Reviewer for this comment. The agent did not select 10% of the actions randomly. Instead, we implemented a 10% exploration while running the algorithms on the daily data, but the final weekly recommendations that were sent to participants were based on the median of the daily recommendations. Additionally, we implemented two additional safeguards. First, if four days or more have hypoglycemia related to a parameter, then the median is taken only using those hypoglycemia days. Second, new parameter recommendations were confined to be within 20% of the previous week's recommendations. These details are now reported in the revised manuscript (page 7, under *Final recommendations* section).

4. Inadequacies in Clinical Study Design:
The single-arm nature of the clinical study and its small sample size limit the ability to conclusively demonstrate the algorithm's effectiveness. It is plausible that any observed improvement is attributable to the app usage rather than the algorithm itself. The number of meal announcements on the app do not sufficiently represent this, particularly concerning basal insulin recommendations.

We agree with the reviewer that the uncontrolled study design does not demonstrate effectiveness. The manuscript has been revised carefully to avoid such a claim. A controlled study to demonstrate effectiveness would require over 100 participants, based on our initial power calculations.

We are not sure what the Reviewer meant by referring to the number of announcements on the app usage. We do not claim that the number of announcements indicates efficacy of the algorithm but indicate active engagement with the app during the study.

5. Minor Differences in Overall Glycemic Control and Lack of Technical Details: Extended Data Table 4 shows minimal differences in overall glycemic control, with almost no change in all metrics between the first 10 days and the last 10 days of the study period. This observation casts doubt on the efficacy of the intervention. Additionally, the paper lacks detailed information on the engineering aspects of the app development and algorithm implementation, which are crucial for evaluating the study's technical rigor.

The Reviewer is right that there were minimal differences in the overall glucose outcomes despite the observed improvements in the postprandial outcomes of high-fat meals and meals with exercise. This is expected, since there were only 0.7 high-fat meals per day and 0.45 meals with exercise per day, and thus the overall glucose outcomes are mainly driven by the time after other meals, time between meals, and nighttime. This is stated in the revised manuscript (Discussion, paragraph 5). As an analogy, single-meal studies with the newest insulin Lyumjev showed significant improvements in the postprandial state, and even though these improvements are theoretically observed with every meal in outpatient studies, overall glucose control in free living conditions is limited to only 2.4% in time in range due to times other than postprandial[2].

The mobile app was developed using JavaScript in React Native framework, allowing cross-platform compatibility with a single code. Firebase served as the back-end service, leveraging its encryption for enhanced security measures. The mobile app was developed for Android versions 6.0 and above (API level 23) as well as iOS versions 9.0 and higher. The mobile app was tested using Xray in Jira. Throughout the app development, GitHub was used for version control. TestFlight was used to distribute the mobile app to participants using iOS devices, while Firebase App Distribution was used to distribute the app to participants using Android devices. The reinforcement learning algorithms were implemented in MATLAB and ran on a Nitro 5 Core i7 computer. The weekly recommendations were sent to study participants using a cloud server. The app received Health Canada's Investigational Testing Authorization as a CLASS II medical device. These details are now stated in the revised manuscript (supplement material, page 22).

Response to Reviewer 3

The authors have addressed all my concerns.

We thank the Reviewer for reviewing our manuscript.

[2] Bergenstal RM, Bode BW, Bhargava A, Wang Q, Knights AW, Chang AM. Assessing Time in Range with Postprandial Glucose-Focused Titration of Ultra Rapid Lispro (URLi) in People with Type 1 Diabetes. Diabetes Therapy. 2023 Sep 23:1-3.

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

The detailed discussion and elaboration added in the manuscript further enhance the understanding of the research's context and methodology.

However, as shown in Figure S2, the differences in overall glucose outcomes between the baseline and evaluation periods are minor, which is agreed by the authors. While improvements in postprandial outcomes for high-fat meals and meals with exercise are noted, the small number of these scenarios seems to limit the impact of the findings and to weaken the claim. These aspects, combined with the nature of a single-arm study, significantly weaken the overall conclusions about the effectiveness of the proposed approach.

In terms of result presentation, Figure 2 appears potentially misleading. The delta curve starting from zero, regardless of varying initial glucose levels, might give an overly optimistic view of the results. A more accurate and transparent representation could be achieved using Figure S2, which shows postprandial glucose curves with their actual starting points. This adjustment would provide a clearer understanding of the data's true variability and impact.

The effectiveness of transferring learning from one agent to another in significantly different contexts (e.g., breakfast vs. dinner meals) might not be as straightforward as implied. A detailed comparison involving different numbers of agents and varying thresholds in Supplementary Equation 7, would be beneficial. If it is difficult to evaluate the efficacy, the authors might consider the use of FDA-approved simulation tools in diabetes.

The algorithm's design predominantly relies on empirical evidence. To strengthen the study, incorporating more case studies or corroborating evidence from existing literature on reinforcement learning would be advantageous.

Response to Reviewer 2

The detailed discussion and elaboration added in the manuscript further enhance the understanding of the research's context and methodology.

However, as shown in Figure S2, the differences in overall glucose outcomes between the baseline and evaluation periods are minor, which is agreed by the authors. While improvements in postprandial outcomes for high-fat meals and meals with exercise are noted, the small number of these scenarios seems to limit the impact of the findings and to weaken the claim. These aspects, combined with the nature of a single-arm study, significantly weaken the overall conclusions about the effectiveness of the proposed approach.

We thank the Reviewer for re-iterating this point. We have reviewed the manuscript and confirm that no efficacy statements were made. Moreover, we softened any sentence that may suggest efficacy. We also shortened the abstract to improve readability.

In terms of result presentation, Figure 2 appears potentially misleading. The delta curve starting from zero, regardless of varying initial glucose levels, might give an overly optimistic view of the results. A more accurate and transparent representation could be achieved using Figure S2, which shows postprandial glucose curves with their actual starting points. This adjustment would provide a clearer understanding of the data's true variability and impact.

We thank the reviewer for this comment. We have combined Figure S2 with Figures 2 and 3 in the revised manuscript, so the reader may readily find the incremental glucose levels as well as the actual glucose levels.

The effectiveness of transferring learning from one agent to another in significantly different contexts (e.g., breakfast vs. dinner meals) might not be as straightforward as implied. A detailed comparison involving different numbers of agents and varying thresholds in Supplementary Equation 7, would be beneficial. If it is difficult to evaluate the efficacy, the authors might consider the use of FDA-approved simulation tools in diabetes.

We thank the Reviewer for this comment. A detailed efficacy comparisons of different number of agents and thresholds is not possible with the retrospective clinical data that we have. Performing simulations to justify design parameters (equation 7, reward functions, state definitions, etc.) is out of the scope of this paper. The scope and the novelty of our work is to test, for the first time, an adaptive and a personalized dosing scheme to manage glucose levels after high fat meals and exercises. Thus, the paper included the description of the dosing scheme followed by the results of the clinical study. The scope is not focused on testing various designs of the dosing scheme nor to optimize it. We have modified the title accordingly to put less emphasis on the particular dosing algorithm used and more on the personalization of the insulin dosing for meals with high fat meals and exercise.

We acknowledge that the effectiveness of our multi-agent design compared to a single-agent design can not be quantified with our study design. An explicit statement about this is now added to the revised manuscript (page 14, paragraph 3).

4. The algorithm's design predominantly relies on empirical evidence. To strengthen the study, incorporating more case studies or corroborating evidence from existing literature on reinforcement learning would be advantageous.

We thank the Reviewer for this comment. Our multi-agent reinforcement learning algorithm bears resemblance to the coordinated reinforcement learning approaches[1-3] where the global Q-function (Equation 6), being a target for the algorithm, is calculated using the weighted sum of local Q-functions. The main difference between these approaches and ours is that they calculate the global Q-function using equal weights from all agents in contrast to our method where we weight each agent based on its visitation to a specific state-action pair divided by the total visitations of all agents to that same state-action pair. This has been stated in the revised manuscript (supplement material, page 23).

[1] Kok JR, Vlassis N. Sparse tabular multiagent Q-learning. *In Annual Machine Learning Conference of Belgium and the Netherlands* 2004; 65-71. Enschede: *Universiteit Twente Press*.
[2] Guestrin C, Venkataraman S, Koller D. Context-specific multiagent coordination and planning with factored MDPs. *In Proceedings of the Eighteenth National Conference on Artificial Intelligence* 2002; 253-259.
[3] Kok JR, Vlassis N. Collaborative multiagent reinforcement learning by payoff propagation. *Journal of Machine Learning Research* 2006; **7**: 1789–1828.