From Prediction to Action: Grand Challenges in Personal Informatics and AI in Diabetes Self-Management

Pooja M. Desai

Columbia University Medical Center, pmd2137@cumc.columbia.edu

David J. Albers

University of Colorado Denver - Anschutz Medical Campus, david.albers@cuanschutz.edu

Lena Mamykina

Columbia University Medical Center, om2196@cumc.columbia.edu

CCS CONCEPTS • Human-centered computing • HCI Design and Evaluation Methods • Field Studies

Additional Keywords and Phrases: personal informatics, diabetes, forecasting, self-management

1 INTRODUCTION

Personal Informatics (PI) tools are a powerful class of digital solutions that help individuals gather data, identify trends, and adjust health behaviors. Emerging AI technologies have shown tremendous potential to augment PI by surfacing insights, facilitating decision support, and leveraging data to forecast future outcomes particularly in the context of chronic disease [2, 7, 10].

Our lab focuses on developing AI-based tools to support self-management of chronic diseases with a specific focus on type 2 diabetes (T2D). T2D is a condition that requires individuals to rigorously manage various lifestyle factors including nutrition, physical activity, and stress to maintain their health [11]. As part of this research, we have developed *GlucOracle* a mobile app that helps users with T2D predict their glycemic response to planned meals, and make in-the-moment adjustments. Insights from this work raise important considerations for the next era for PI tools leveraging AI. In this position paper we discuss two of these considerations and link them to the grand challenges in Personal Informatics and AI outlined in the workshop abstract:

1. Data sparsity is an ongoing challenge. PI tools can seek to leverage computational approaches that have lower data thresholds and design solutions that facilitate low-burden data capture.
2. While predictions help users reflect on consequences of planned behaviors, users struggle to act on insights. Personalized recommendations can make forecasts actionable, and support decision-making.

We expand on these lessons and discuss implications for the future development PI systems that leverage AI to support health management.

2 LESSONS FROM GLUCORACLE

In prior research, our team developed and conducted a feasibility pilot study of *GlucOracle*, a smartphone app to help individuals with T2D predict changes in their blood glucose levels in response to planned meals. *GlucOracle* users track meals by capturing photographs and short text descriptions as well as blood glucose levels before and after these meals to train a personalized prediction model. Once the model is trained with users self-tracking data, for each newly captured meal and pre-meal blood glucose reading *GlucOracle* presents them with forecast for how that meal is anticipated to affect their post-meal blood sugar, enabling them to make modifications to planned meals in real-time.

GlucOracle's prediction model leverages data assimilation with mechanistic physiological models to generate personalized real-time predictions of an individuals' evolving blood glucose levels. Unlike more traditional machine learning techniques, that typically require large volumes of data, incorporating expert knowledge in the form of mechanistic models allows to achieve high predictive accuracy with small datasets; in *GlucOracle*, users were required to record 25 meals with pre- and post-meal blood glucose readings in order to train the personal blood glucose prediction model [1].

We deployed *GlucOracle* in a 4-week feasibility study with 10 participants with T2D. Participants had varying levels of experience using technology for diabetes management (5 novice self-trackers unfamiliar with PI technologies, 5 experienced self-trackers). Following the 4-week pilot, participants were invited to take part in qualitative interviews to share their perceptions and experiences with receiving personalized forecasts [4]. Participants discussed the usefulness of AI-generated forecasts to support nutrition management and considerations for future personal informatics technologies to support health choices in everyday life.

Overall, *GlucOracle* users found the AI-generated forecasts interesting and informative. At the same time, users highlighted the need to couple predictions with other, more actionable forms of decision support, such as recommendations for specific ways to change meals to improve forecasts. However, interviews also suggested the importance of personalizing such recommendations not only to individuals' physiology and blood glucose regulation, but also to their preferences and lifestyle. Below we expand on insights from our deployment of *GlucOracle* and outline perspectives to consider for the future development of human-centered AI systems to support everyday health decision-making.

2.1 Designing for Data Limitations (Grand Challenge: Limitations of self-tracking data in AI models)

An important consideration that emerged in the development and evaluation of *GlucOracle* was with regards to challenges around tracking burdens and limited data available for generating personal predictions [4]. *GlucOracle'*s personal prediction engine required a minimum of 25 meal recording with pre- and post-meal BG for the model to generate forecasts with an error rate comparable to expert dieticians [1]. While this low number is a substantial improvement as comparing to a typical dataset required for traditional machine learning techniques, it still required users to track meals for approximately 1 week to enable personalized predictions. Furthermore, during this time and while the model was training, users could not benefit from personalized predictions, which led some users to tracking lapses and further delayed personal forecasts. Similarly, as users gained more experience with forecasts, they became better at anticipating impact of typical meals, and wished to only use GlucOracle for new dishes, or new contexts, which, again, led to fragmentation in their logs needed for training the model. Such tracking in "fits and spurts" poses considerable challenges for developing unbiased algorithms as it can result in biased datasets that only capture individuals in particular states of health [3].

Our findings highlight two important considerations for addressing data sparsity in PI tools. First, computationally, it is important to design and evaluate models with data sparsity and the cold-start problem in mind, "stress testing" algorithms on their ability to perform accurately with limited amounts of data. Second, it is crucial that systems are designed to facilitate low-burden data capture. Past work in HCI has highlighted ways data tracking can be incentivized through simplifying the data entry process (e.g., duplicating previous entries) and leveraging gamification techniques (e.g. points, badges, visualizations, etc.) for tracking [3]. Additionally, automated data collection through devices can also facilitate data collection. Finally, these findings highlight the need for new machine learning techniques that are able to achieve high predictive accuracy with small sparse and irregular datasets.

2.2 What Now?: From Forecasts to Actionable Guidance (Grand Challenge: Forms of support and interaction paradigms)

While both novice and expert *GlucOracle* users found forecasts helped prompt reflection on planned meals, participants identified several challenges in deciding how to make modifications in response to forecasts. First, users highlighted that since forecasts were delivered after a meal had been prepared, large modifications like full ingredient substitutions were seldom feasible. However, smaller modifications such as reducing portion sizes or adding food to the plate were more feasible if they could identify what changes were necessary. Second, while forecasts helped users understand if a given meal would cause a BG spike, they provided little information about what specific ingredients were predicted to cause BG spikes and would make good targets for intervention. Users wondered what parts of the meal could be modified and by how much in order to keep their BG in a healthy range. They desired new features that could help them identify actionable changes that could reduce a meal's glycemic impact.

These findings highlight that while predictions can facilitate reflection, potential opportunities to couple predictions with other forms of decision support such as recommendations can help users transition from reflecting on their health to deciding how to act to improve it. Additionally, they highlight a need for explainable models that can not only provide predictions but also outline factors that contributed to the predictions.

At the same time, our findings suggest that providing users with recommendations for changing meals need to be personalized not only to individuals' physiology, but also to their preferences, lifestyle, and context [9]. While preferences and context have been explored by recommender systems and context-aware computing systems, there remain many open questions about how these approaches can be integrated into one coherent system. For example, previous work in other domains of chronic disease management has shown the value of AI generated personalized recommendations that account for user's context and preferences (e.g. recognizing a user is at a restaurant, has certain allergies, or is unwilling to modify particular cultural staples in their diet) [10]. Similar opportunities and approaches could be examined in other chronic conditions and other lifestyle recommendations.

3 A WAY FORWARD: MAKING FORECASTS MORE ACTIONABLE

Our *GlucOracle* pilot study underscored several important considerations for the next era of personal informatics solutions, specifically those that rely on computational models and predictions in order to inform behaviors and choices. We believe these considerations are aligned with the grand challenges proposed for discussion at the workshop, and that our takeaways can contribute to outlining future directions for research in PI and AI.

First, it is important to consider data limitations and challenges with sporadic data collection when developing PI tools. This is an important consideration in training AI models to reduce errors and bias. In addition, from the HCI perspective it is crucial for PI tools to incorporate low friction ways to capture data.

Second, users seek to understand what factors impacted nutrition forecasts. This can facilitate reflection and help users identify potential triggers and where to focus modifications. Beyond *GlucOracle*, other PI tools for health prediction, particularly in mental health, have highlighted the importance of ethically communicating non-modifiable risk factors to users in ways that are informative and support proactive engagement with health.

Third, users seek personal recommendations to make forecasts more actionable. Users highlight important dimensions of personalization, including tailoring towards taste preferences, cultural dietary habits, context, and prior successes and failures.

4 CONCLUSION

In this workshop we hope to engage with fellow HCI researchers and explore new ways to design PI technologies to better support health management. We believe our insights will be a valuable contribution to the workshop. We look forward to discussing these grand challenges and learning from fellow researchers.

ACKNOWLEDGMENTS

We thank the organizers of the CHI 2022 "Grand Challenges in Personal Informatics and AI" workshop, participants, translators, and members of the ARCH Lab who made this work possible. This project was funded with support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) R01DK090372 and R56DK113189; and the National Library of Medicine (NLM) R01 LM012734-01.

REFERENCES

[1] David J Albers, Matthew Levine, Bruce Gluckman, Henry Ginsberg, George Hripcsak, and Lena Mamykina. 2017. Personalized glucose fore- casting for type 2 diabetes using data assimilation. PLoS computational biology 13, 4 (2017), e1005232.

[2] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. ACM Transactions on Computer-Human Interaction 20, 5: 1–27. https://doi.org/10.1145/2503823

[3] Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., ... & Smith, I. (2008, April). Activity sensing in the wild: a field trial of ubifit garden. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1797-1806)

[4] Pooja M. Desai, Elliot G. Mitchell, Maria L. Hwang, Matthew E. Levine, David J. Albers, and Lena Mamykina. 2019. Personal Health Oracle: Explorations of Personalized Predictions in Diabetes Self- Management. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/ 3290605.3300600

[5] Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. 2016. Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1109–1113. DOI:https://doi.org/10.1145/2858036.2858045

[6] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 731–742.

[7] Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker. 2017. What does all this data mean for my future mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. Human–Computer Interaction 32, 5-6 (2017), 208–267.

[8] Jochen Meyer, Merlin Wasmann, Wilko Heuten, Abdallah El Ali, and Susanne C.J. Boll. 2017. Identification and Classification of Usage Patterns in Long-Term Activity Tracking. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 667–678. DOI:https://doi.org/10.1145/3025453.3025690

[9] Shriti Raj, Kelsey Toporski, Ashley Garrity, Joyce M. Lee, and Mark W. Newman. 2019. "My blood sugar is higher on the weekends": Finding a Role for Context and Context-Awareness in the Design of Health Self-Management Technology. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 119, 1–13. DOI:https://doi.org/10.1145/3290605.3300349

[10] Darius A. Rohani, Andrea Quemada Lopategui, Nanna Tuxen, Maria Faurholt-Jepsen, Lars V. Kessing, and Jakob E. Bardram. 2020. MUBS: A Personalized Recommender System for Behavioral Activation in Mental Health. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376879

[11] Jaakko Tuomilehto, Jaana Lindström, Johan G Eriksson, Timo T Valle, Helena Hämäläinen, Pirjo Ilanne-Parikka, Sirkka Keinänen- Kiukaanniemi, Mauri Laakso, Anne Louheranta, Merja Rastas, et al. 2001. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine 344, 18 (2001), 1343–1350.