A Critical Reflection on Designing and Evaluating a Personalized Self-Care Intervention for Care Partners: Lessons from a Randomized Controlled Trial

Rongqi Bei
University of Michigan
Ann Arbor, USA
rbei@umich.edu

Christopher Graves
University of Michigan
Ann Arbor, USA
gravesch@med.umich.edu

Mark W. Newman
University of Michigan
Ann Arbor, USA
mwnewman@umich.edu

Predrag Klasnja
University of Michigan
Ann Arbor, USA
klasnja@umich.edu

Jennifer Miner
University of Michigan
Ann Arbor, USA
jenminer@med.umich.edu

Sung Won Choi
University of Michigan
Ann Arbor, USA
sungchoi@med.umich.edu

Angelle M. Sander
TIRR Memorial Hermann
Ann Arbor, USA
Angelle.Sander@memorialhermann.org

Noelle E. Carlozzi
University of Michigan
Ann Arbor, USA
carlozzi@med.umich.edu

Abstract

Care partners often face significant mental strain, but are rarely able to prioritize their own well-being. To support this vulnerable group, we designed the CareQOL app, an adaptive intervention combining daily self-monitoring of caregiver strain, depression, and anxiety with Fitbit-tracked health behaviors to deliver personalized push notifications targeting self-care. We evaluated this app in a six-month randomized controlled trial with care partners (N=254). Despite high compliance, adding personalized notifications to self-monitoring did not significantly improve participants' mental health or other health-related quality of life outcomes, though participants who rated the app more usable demonstrated a greater likelihood of improvement. This prompted reflection on intervention design and study process, facilitated by follow-up interviews (N=36). We found that personalization based solely on passive sensing and self-reports did not always match participants' dynamic needs and preferences, while dashboard visualizations without contextual details hindered in-depth self-reflection. Additionally, extended study durations made it hard for participants to recall experiences and left systems vulnerable to technical breakdowns. We offer practical recommendations for research at the intersection of ubiquitous computing and mental healthcare: enable user-initiated personalization through lightweight interactions, support contextual annotations in self-monitoring dashboards, adopt data-prompted interview approaches, and implement batch enrollment with technical fallback strategies.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

mobile health; personalized intervention; self-management; care partners

ACM Reference Format:
Rongqi Bei, Christopher Graves, Mark W. Newman, Predrag Klasnja, Jennifer Miner, Sung Won Choi, Angelle M. Sander, and Noelle E. Carlozzi. 2025. A Critical Reflection on Designing and Evaluating a Personalized Self-Care Intervention for Care Partners: Lessons from a Randomized Controlled Trial. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '25).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Care partners (i.e., informal family caregivers) of individuals with traumatic brain injury (TBI) often experience considerable mental strain [2, 13], manifesting in increased anxiety, depression, fatigue, and social isolation over time [4, 12, 20]. Compounding this burden, care partners often prioritize their care recipients' needs over their own health, creating a cycle where declining caregiver health-related quality of life (HRQOL) ultimately impacts both the caregiver and the individual receiving care [17, 22].

Recognizing that mobile health (mHealth) interventions offer greater accessibility, flexibility, and scalability than traditional in-person medical care for self-care improvement [10, 11], we designed the CareQOL app [3] to provide low-burden, personalized self-management support for care partners (Figure 1-A-C). Each day, users complete an ecological momentary assessment (EMA) consisting of single-item measures of caregiver strain [5, 6], depression [8], and anxiety [8] administered through a weekly reset computer adaptive test (CAT), and wear a Fitbit® to track their health behaviors (e.g., sleep and steps). The app provides visualizations of these self-monitoring data (Figure 1-C) and uses these data as tailoring variables to classify users' mental health, sleep, and activity into high, medium, or low levels. The app randomizes the delivery of a tailored push notification (Figure 1-B), with a

---

UbiComp '25, Espoo, Finland
2025. ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

50% chance of delivery daily. When delivery is selected, a message that targets one of six health aspects (i.e., caregiver stress, sadness, worry, sleep, mood and mindfulness, and physical activity) and is tailored to the user's level for that category (i.e., high, medium, or low) is randomly selected for delivery. Figure 1-E shows the workflow from data collection to notification delivery. Figure 1-F provides sample scenarios illustrating how users receive personalized notifications tailored to their self-reports and Fitbit data.

To evaluate the efficacy of this personalized intervention, we conducted a six-month randomized controlled trial (RCT) with 254 care partners of individuals living with TBI using the CareQOL app in their daily lives. Participants were randomized to: Study Arm 1, self-monitoring only (no notifications), or Study Arm 2, self-monitoring plus personalized notifications (receiving tailored push notifications with a 50% daily chance). We assessed their mental health and other HRQOL outcomes monthly, as well as 3 months and 6 months post-intervention. Methodology has been detailed in prior work [7]. Contrary to our expectations, we did not observe significant differences in mental health or other HRQOL outcomes between the two study arms, despite high protocol compliance across both arms [7]. We also observed varied outcomes: approximately one-third of participants improved, one-third remained unchanged, and one-third worsened, with participants who rated the app as more usable demonstrating a greater likelihood of improvement compared to their counterparts [7]. These findings prompted us to critically reflect on the research process and explore potential explanations. We revisited the intervention design and the clinical trial evaluation processes. To ground this reflection in participants' experiences, we also conducted follow-up interviews with a subset of participants (N = 36, including 16 from Study Arm 2) to understand their perspectives on the intervention. For interview participants from Study Arm 2, we created individualized slides that visualized their health data and personalized notifications they had received (Figure 1-D) to support recall and elaboration during interviews. The remainder of this paper will focus on the experiences of participants in Study Arm 2.

In this workshop paper, we share lessons learned from our critical reflection and offer practical recommendations. We organized the lessons learned around two key aspects: the personalized mHealth intervention design and the study methodology. Our goal is to contribute actionable insights to guide future designs and evaluations at the intersection of ubiquitous computing and mental healthcare.

2 Lessons Learned and Practical Recommendations

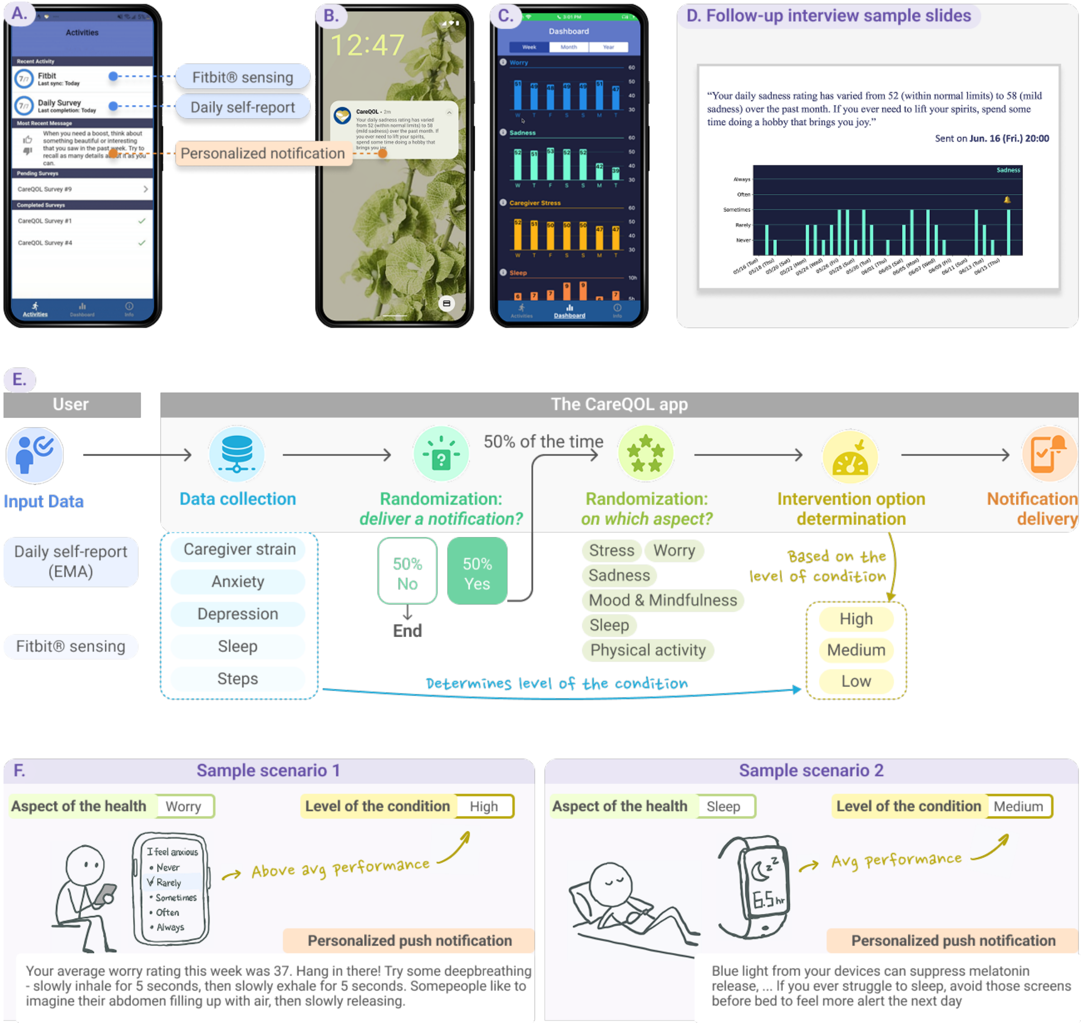
2.1 For the personalized mHealth intervention design

2.1.1 Data-driven personalization sometimes did not match users' needs and preferences. While we successfully delivered notifications based on participants' self-reported data (e.g., caregiver strain, depression, and anxiety scores) and passive sensing data (e.g., sleep, steps), we found that this approach may be insufficient to ensure meaningful engagement as it failed to account for the complex, dynamic nature of caregivers' lived experiences. During follow-up interviews, participants reported receiving suggestions at inappropriate times (e.g., while actively caregiving, at work) and noted that if they could not act on it right then its value would be lost, as the message *"is not something that kind of stayed with me all day."* Others found certain types of content (e.g., prompts to "talk to a friend" or practice mindfulness) irrelevant or unappealing due to personal circumstances or preferences.

From our conversations with participants, we observed that some more engaged users developed their own workarounds: intentionally delaying adherence until a more convenient time, taking actions when they later recalled the message, or writing down personal lists of helpful messages to revisit during moments of need. However, requiring users to independently develop these strategies places an additional burden on this already-strained population, making it challenging for other users to replicate the efficacy of the intervention.

Recommendation: Allow users to shape notification delivery toward their preferences through lightweight interactions. We recommend that personalized, adaptive interventions offer users ways to communicate their preferences when receiving notifications, giving them agency to navigate mismatched messages and shape their intervention experience. Such a user-initiated personalization approach can operate on two complementary levels. First, through direct user control. Systems can enable users to immediately adjust interventions to match their current situation. For timing issues, introduce features like "snoozing" or "micro-planning" to help users reschedule suggestions they want to act on but cannot address immediately. For content relevance, allow users to "opt out" of certain categories of suggestions or "favorite" helpful messages for future reference to help curate their experience. Second, beyond these immediate adjustments, we recommend treating preference expression as an ongoing human-in-the-loop [1, 16] mechanism that drives intervention tailoring. Users' explicit actions (e.g., marking messages as favorites or rescheduling suggestions) may inform the system's future personalization decisions. This creates a feedback loop where user preferences continuously refine the intervention's relevance and timing.

2.1.2 Lack of context in dashboard displays hindered self-reflection on emotional states. Dashboard visualizations of personal health data are common in mHealth interventions. The CareQOL app included one such dashboard that presented users' strain, sadness, and worry levels graphically as t-scores based on daily self-reports. These graphical visualizations of user data included weekly, monthly, or yearly views (Figure 1-C). Encouragingly, some participants found these visualizations helped them stay aware of their mental health. However, the dashboard included limitations for supporting deeper reflection. Participants expressed greater interest in understanding the "why" behind their past feelings rather than simply reviewing the performance. As one noted, *"I was more concerned about why I answered [the self-report questions] the way I did... I try to remember what happened during that previous time period..."* Over time, this became harder, especially when viewing t-scores rather than their original 5-point ratings. Several participants said they would have benefited from documenting the underlying causes of their feelings alongside daily self-reports, *"If you had a stressful day, you might make a note of why it was, just for future reference… if it's specifically about caregiving duties, or there are other things going on… "*



When you need a boost, think about something beautiful or interesting that you saw in the past week. Try to recall as many details about it as you can.

Your daily sadness rating has varied from 52 (within normal limits) to 58 (mild sadness) over the past month. If you ever need to lift your spirits, spend some time doing a hobby that brings you joy.

"Your daily sadness rating has varied from 52 (within normal limits) to 58 (mild sadness) over the past month. If you ever need to lift your spirits, spend some time doing a hobby that brings you joy."

Sent on **Jun. 16 (Fri.) 20:00**

Your average worry rating this week was 37. Hang in there! Try some deepbreathing - slowly inhale for 5 seconds, then slowly exhale for 5 seconds. Somepeople like to imagine their abdomen filling up with air, then slowly releasing.

Blue light from your devices can suppress melatonin release, ... If you ever struggle to sleep, avoid those screens before bed to feel more alert the next day

Figure 1: (A) The CareQOL app Activities page showing the completion status of the daily self-monitoring and Fitbit syncing tasks and the most recently received personalized self-care push notifications; (B) The delivered personalized self-care push notifications, (C) The CareQOL app Dashboard page; (D) An example of the slides used during the follow-up data-prompted interview; (E) Flow of personalized push notification delivery; (F) Sample scenarios of personalized notifications delivered based on self-monitoring data.

Recommendation: Contextualize self-reflections in the dashboard through annotations. We recommend supporting users by allowing for the addition of optional, contextual annotations that can be included alongside the dashboard visualizations. Notably, some participants proactively expressed a willingness to invest additional effort in the self-report process to document underlying causes, likely because they saw personal value in doing so. These annotations could preserve contextual details influencing their mental states, transforming isolated data points into meaningful insights. This recommendation aligns with prior work that users' ability to recall the context surrounding specific data points often degrades over time [15]. Without such support, users may struggle to recall or make sense of earlier emotional states, limiting the utility of long-term self-monitoring.

2.2 For the study methodology

2.2.1 Long study duration posed challenges for participant recall during follow-up interviews. As mentioned earlier, we conducted follow-up interviews with a subset of participants to facilitate our critical reflection process. One key challenge in these interviews, following our six-month trial, was helping participants recall their reactions to the personalized push notifications received and how specific messages had impacted them. To address this, we adopted a data-prompted interview approach, using each participant's own trial data to "stimulate discussion during the interview" [14, p. 1191]. This decision turned out to be helpful in providing context and facilitating conversations.

We presented representative messages received by each participant, alongside their delivery timestamp and the visualizations of the participant's relevant health data (Figure 1-D). When selecting messages, we considered the following criteria to look for potentially interesting incidents: (1) whether they were delivered during periods of consistently high or low level of conditions (e.g., persistently high stress); (2) whether they coincided with noticeable changes in relevant conditions (e.g., a reduction in sadness levels observed after message delivery); and (3) whether they were delivered closer in time to the interview date. During the interview, we shared screens and guided participants through the slides.

To help the visualizations facilitate accurate recall rather than leading participants to reinterpret their past experiences in ways that conformed to the data display, we were careful in two aspects: (1) we prepared the visualizations with the goal of accurately communicating participants' health data, aiming to avoid misleading design choices; (2) we avoided displaying potentially suggestive trends (e.g., immediate improvements following message delivery) before participants had shared their own perceptions and reflections, with the intention of minimizing retrospective sense-making.

Recommendation: Apply data-prompted approaches to enhance recall in post-study interviews. When conducting follow-up interviews after extended trials, researchers can consider using data-prompted techniques [14, 18] to facilitate participants' recall and elaboration. This approach not only provides memory cues but also enables researchers to direct conversations toward critical incidents of particular interest. Notably, researchers need to take care both in the design of visualizations and in how they are presented during interviews to avoid unintentionally shaping participants' recall of their experiences. First, visualizations need to be designed to accurately represent the data without introducing misleading elements, such as inappropriate axis scales, chart types, and etc. [19]. Second, researchers need to be mindful of how the visualizations are presented. As individuals tend to "[relate] recognizable features of the data to remembered experience, routines, and known facts about one's life" to "tell a cohesive story" [9, p. 540], any biased pattern in the visualizations may influence participants' recall.

Logistically, preparing personalized visualizations can be time-consuming. To reduce this burden, we developed end-to-end Python scripts to automate the data visualization process. We encourage researchers to explore approaches for streamlining or scaling data-prompted methods to enable broader use.

2.2.2 Technical breakdowns during the study introduced inconsistency in participant experiences. The trial described here began enrollment in December of 2020, completing data collection in February of 2024. During this multi-year trial, our app occasionally broke down due to practical issues like operating system updates (e.g., iOS, Android). When these breakdowns were not immediately resolved, they created inconsistent participant experiences, including missed personalized notifications, unsynced Fitbit data, duplicated EMAs, and lost data due to API integration issues. Because we recruited participants on a rolling basis, these disruptions affected participants differently depending on when they were enrolled. For example, participants who joined just before technical failures were more likely to experience early disengagement, while those enrolled after issues were resolved may have smoother experiences. These inconsistencies complicate both the intervention experience and the interpretation of study outcomes.

Recommendation: Adopt batch enrollment and prepare fallback strategies for technical breakdowns in long-term, large-scale studies. We recommend adopting a batch enrollment strategy (i.e., enrolling participants in waves rather than on a rolling basis) [21] to reduce the impact of mid-study technical issues. For example, if 25 participants begin the study around the same time, any system failure during that period would affect the entire cohort equally. This simplifies troubleshooting and allows batch identifiers to be included as covariates in the analysis. This approach would be especially valuable in RCTs, where inconsistencies in the intervention's technical functionality can complicate data analysis. Batch enrollment also provides flexibility: if a known app update is scheduled, researchers can pause recruitment or delay the next cohort. In extreme cases, "problematic" batches can be excluded without affecting others.

We also emphasize the need for fallback plans during long-term, large-scale deployments. Unlike tightly controlled lab experiments, real-world trials need to account for external dependencies, such as mobile operating system updates, hardware (e.g., trackers) failures, and API changes, that can impact system performance. Prior planning around "what to do when the system breaks" can help preserve study integrity and ensure interventions remain meaningful even when technology does not fully function as originally intended.

3 Conclusion

In this critical reflection paper, we shared actionable insights derived from our experience with the CareQOL app, a personalized mHealth intervention promoting self-care, through a large-scale RCT. Our experience reveals key considerations for designing personalized mHealth interventions and evaluating them in longitudinal studies, offering practical guidance for researchers at the intersection of ubiquitous computing and mental healthcare. Design-wise, we propose allowing users to shape intervention message delivery through lightweight interactions, complementing existing tailoring variables from self-report and passive sensing data. We also found that data annotations, despite potentially requiring additional effort from users, were desired by participants to facilitate in-depth self-reflection within data dashboards. Methodologically, we reflected on employing data-prompted interviews to facilitate recall after long-term study participation and provided practical insights for preparing visual cues. We also pointed out systems' vulnerability to technical breakdowns in longitudinal studies, recommending batch recruitment and developing technical fallback strategies. In sum, these strategies inform future work on personalized mHealth interventions for self-care, ultimately fostering improved HRQOL in care partners and other vulnerable populations.

Acknowledgments

Work on this manuscript was supported by the National Institutes of Health grant numbers R01NR013658, R01HL146354, K24HL156896, and UL1TR002240, as well as by the University of Michigan Institute for Healthcare Policy and Innovation.

References

[1] Christian Arzate Cruz and Takeo Igarashi. 2020. A Survey on Interactive Reinforcement Learning: Design Principles and Open Challenges. In *Proceedings of*

the 2020 ACM Designing Interactive Systems Conference (DIS '20). Association for Computing Machinery, New York, NY, USA, 1195–1209.

[2] Noelle E Carlozzi, Tracey A Brickell, Louis M French, Angelle Sander, Anna L Kratz, David S Tulsky, Nancy D Chiaravalloti, Elizabeth A Hahn, Michael Kallen, Amy M Austin, Jennifer A Miner, and Rael T Lange. 2016. Caring for our wounded warriors: A qualitative examination of health-related quality of life in caregivers of individuals with military-related traumatic brain injury. *J. Rehabil. Res. Dev.* 53, 6 (2016), 669–680.

[3] Noelle E Carlozzi, Sung Won Choi, Zhenke Wu, Jennifer A Miner, Angela K Lyden, Christopher Graves, Jitao Wang, and Srijan Sen. 2021. An app-based just-in-time adaptive self-management intervention for care partners (CareQOL): Protocol for a pilot trial. *JMIR Res. Protoc.* 10, 12 (Dec. 2021), e32842.

[4] Noelle E Carlozzi, Sung Won Choi, Zhenke Wu, Jonathan P Troost, Angela K Lyden, Jennifer A Miner, Christopher M Graves, Jitao Wang, Xinghui Yan, and Srijan Sen. 2022. An app-based just-in-time-adaptive self-management intervention for care partners: The CareQOL feasibility pilot study. *Rehabil. Psychol.* 67, 4 (Nov. 2022), 497–512.

[5] Noelle E Carlozzi, Michael A Kallen, Robin Hanks, Elizabeth A Hahn, Tracey A Brickell, Rael T Lange, Louis M French, Anna L Kratz, David S Tulsky, David Cella, Jennifer A Miner, Phillip A Ianni, and Angelle M Sander. 2019. The TBI-CareQOL measurement system: Development and preliminary validation of health-related quality of life measures for caregivers of civilians and service members/veterans with traumatic brain injury. *Arch. Phys. Med. Rehabil.* 100, 4S (April 2019), S1–S12.

[6] Noelle E Carlozzi, Michael A Kallen, Phillip A Ianni, Elizabeth A Hahn, Louis M French, Rael T Lange, Tracey A Brickell, Robin Hanks, and Angelle M Sander. 2019. The development of a new computer-adaptive test to evaluate strain in caregivers of individuals with TBI: TBI-CareQOL Caregiver Strain. *Arch. Phys. Med. Rehabil.* 100, 4S (April 2019), S13–S21.

[7] Noelle E Carlozzi, Jonathan P Troost, Srijan Sen, Sung Won Choi, Zhenke Wu, Jennifer A Miner, Wendy L Lombard, Christopher Graves, and Angelle M Sander. 2025. Improving outcomes for care partners of individuals with traumatic brain injury: Results for a mHealth randomized control trial of the CareQOL app. *Arch. Phys. Med. Rehabil.* (Jan. 2025).

[8] David Cella, William Riley, Arthur Stone, Nan Rothrock, Bryce Reeve, Susan Yount, Dagmar Amtmann, Rita Bode, Daniel Buysse, Seung Choi, Karon Cook, Robert Devellis, Darren DeWalt, James F Fries, Richard Gershon, Elizabeth A Hahn, Jin-Shei Lai, Paul Pilkonis, Dennis Revicki, Matthias Rose, Kevin Weinfurt, Ron Hays, and PROMIS Cooperative Group. 2010. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J. Clin. Epidemiol.* 63, 11 (Nov. 2010), 1179–1194.

[9] Chris Elsden, David S Kirk, and Abigail C Durrant. 2016. A Quantified Past: Toward Design for Remembering With Personal Informatics. *Human–Computer Interaction* 31, 6 (Nov. 2016), 518–557.

[10] Fereshteh Ghahramani and Jingguo Wang. 2021. Intention to Adopt mHealth Apps Among Informal Caregivers: Cross-Sectional Study. *JMIR Mhealth Uhealth* 9, 3 (March 2021), e24755.

[11] Marjorie M Kelley, Tia Powell, Djibril Camara, Neha Shah, Jenna M Norton, Chelsea Deitelzweig, Nivedha Vaidy, Chun-Ju Hsiao, Jing Wang, and Arlene S Bierman. 2024. Mobile health apps, family caregivers, and care planning: Scoping Review. *J. Med. Internet Res.* 26 (May 2024), e46108.

[12] J S Kreutzer, A H Gervasio, and P S Camplair. 1994. Primary caregivers' psychological status and family functioning after traumatic brain injury. *Brain Inj.* 8, 3 (April 1994), 197–210.

[13] Jeffrey S Kreutzer, Lisa J Rapport, Jennifer H Marwitz, Cynthia Harrison-Felix, Tessa Hart, Mel Glenn, and Flora Hammond. 2009. Caregivers' well-being after traumatic brain injury: a multicenter prospective investigation. *Arch. Phys. Med. Rehabil.* 90, 6 (June 2009), 939–946.

[14] Dominika Kwasnicka, Stephan U Dombrowski, Martin White, and Falko F Sniehotta. 2015. Data-prompted interviews: Using individual ecological data to stimulate narratives and explore meanings. *Health Psychol.* 34, 12 (Dec. 2015), 1191–1194.

[15] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing (UbiComp '11)*. Association for Computing Machinery, New York, NY, USA, 405–414.

[16] Adiba Orzikulova, Han Xiao, Zhipeng Li, Yukang Yan, Yuntao Wang, Yuanchun Shi, Marzyeh Ghassemi, Sung-Ju Lee, Anind K Dey, and Xuhai Xu. 2024. Time2Stop: Adaptive and Explainable Human-AI Loop for Smartphone Overuse Intervention. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 250, 20 pages.

[17] Lori A Roscoe, Elizabeth Corsentino, Shirley Watkins, Marcia McCall, and Juan Sanchez-Ramos. 2009. Well-being of family caregivers of persons with late-stage Huntington's disease: lessons in stress and coping. *Health Commun.* 24, 3 (April 2009), 239–248.

[18] Magy Seif El-Nasr, Shree Durga, Mariya Shiyko, and Carmen Sceppa. 2015. Data-Driven Retrospective Interviewing (DDRI): A proposed methodology for formative evaluation of pervasive games. *Entertain. Comput.* 11 (Nov. 2015), 1–19.

[19] Danielle Albers Szafir. 2018. The good, the bad, and the biased: five ways visualizations can mislead (and how to fix them). *Interactions* 25, 4 (June 2018), 26–33.

[20] R Wells, J Dywan, and J Dumas. 2005. Life satisfaction and distress in family caregivers as related to specific behavioural changes after traumatic brain injury. *Brain Inj.* 19, 13 (Dec. 2005), 1105–1115.

[21] Pascale Wermuth. 2022. Participant recruitment, screening, and enrollment. In *Principles and Practice of Clinical Trials*. Springer International Publishing, Cham, 257–278.

[22] Zhiwei Yu, Mingma Tendu Sherpa, Suzannah Iadarola, Kenneth Shamlian, Samantha Daley, Grace Levine, Peter Bajorski, and Zhi Zheng. 2023. Understanding stress in children with ASD and their caregivers in daily life: A feasibility study using mobile devices. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction*. ACM, New York, NY, USA.