

Burnout and the Quantified Workplace: Tensions around Personal Sensing Interventions for Stress in Resident Physicians

DANIEL A. ADLER, MS* and EMILY TSENG, MS*, Cornell Tech, USA

KHATIYA C. MOON, MD, Zucker Hillside Hospital, USA

JOHN Q. YOUNG, MD, Zucker School of Medicine at Hofstra/Northwell Health and Zucker Hillside Hospital, USA

JOHN M. KANE, MD, Zucker Hillside Hospital, USA

EMANUEL MOSS, PHD, Cornell Tech and Data & Society Research Institute, USA

DAVID C. MOHR, PHD, Northwestern University, USA

TANZEEM CHOUDHURY, PHD, Cornell Tech, USA

Recent research has explored computational tools to manage workplace stress via personal sensing, a measurement paradigm in which behavioral data streams are collected from technologies including smartphones, wearables, and personal computers. As these tools develop, they invite inquiry into how they can be appropriately implemented towards improving workers' well-being. In this study, we explored this proposition through formative interviews followed by a design provocation centered around measuring burnout in a U.S. resident physician program. Residents and their supervising attending physicians were presented with medium-fidelity mockups of a dashboard providing behavioral data on residents' sleep, activity and time working; self-reported data on residents' levels of burnout; and a free text box where residents could further contextualize their well-being. Our findings uncover tensions around how best to measure workplace well-being, who within a workplace is accountable for worker stress, and how the introduction of such tools remakes the boundaries of appropriate information flows between worker and workplace. We conclude by charting future work confronting these tensions, to ensure personal sensing is leveraged to truly improve worker well-being.

CCS Concepts: • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: mental health, burnout, privacy, human-centered design

ACM Reference Format:

Daniel A. Adler, MS, Emily Tseng, MS, Khatiya C. Moon, MD, John Q. Young, MD, John M. Kane, MD, Emanuel Moss, PhD, David C. Mohr, PhD, and Tanzeem Choudhury, PhD. 2022. Burnout and the Quantified Workplace: Tensions around Personal Sensing Interventions for Stress in Resident Physicians. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 430 (November 2022), 48 pages. <https://doi.org/10.1145/3555531>

*equal contribution, listed in alphabetical order by last name.

Authors' addresses: Daniel A. Adler, MS; Emily Tseng, MS, Cornell Tech, New York, NY, USA; Khatiya C. Moon, MD, Zucker Hillside Hospital, Queens, NY, USA; John Q. Young, MD, Zucker School of Medicine at Hofstra/Northwell Health and Zucker Hillside Hospital, New York, NY, USA; John M. Kane, MD, Zucker Hillside Hospital, New York, NY, USA; Emanuel Moss, PhD, Cornell Tech and Data & Society Research Institute, New York, NY, USA; David C. Mohr, PhD, Northwestern University, Chicago, IL, USA; Tanzeem Choudhury, PhD, Cornell Tech, New York, NY, USA.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2022 Copyright held by the owner/author(s).

2573-0142/2022/11-ART430

<https://doi.org/10.1145/3555531>

1 INTRODUCTION

Over the past decade, research in HCI and CSCW has explored using computational tools to manage workplace stress and well-being. This research agenda, framed as the Quantified Workplace [88], is timely—burnout is on the rise, especially among healthcare workers, who as a labor force have been particularly taxed by the demands of the COVID-19 pandemic [19, 30, 87]. Burnout is an attractive target for computer-supported measurement and intervention: since work-related stress is by definition the root cause of burnout, managers equipped with knowledge of when their workers are nearing burnout can act to ameliorate it by creating more supportive workplace structures. Such knowledge is also made newly possible by the growing fields of ubiquitous computing and mobile health sensing, coined as personal sensing, in which researchers have developed techniques for measuring behaviors associated with workplace stress via novel data streams from technologies like smartphones, wearables, and personal computers [15, 47, 73]. While research is still exploring the validity and technical feasibility of these measurement techniques, their proliferation raises a range of sociotechnical questions around how these systems might be used to improve worker well-being in practice. For example, what data best represents workers' well-being and burnout, and how can that data be visualized and communicated? Who should be held accountable for the management of this data and consequently, workers' mental health? How can these programs be built to ensure workers' best interests are served?

In this study, we examine these questions through a design provocation set in one particularly stressful workplace: a U.S. physician residency program. Developed historically as a form of apprenticeship, a residency can be a grueling experience, requiring doctors-in-training to work long, unpredictable, and taxing shifts during which they can face many life-or-death decisions. Resident physicians ("residents") are a vital labor force within the hospital, rotating quickly through different clinical settings and working with many different supervising attending physicians ("attendings"). At the same time, residents are students within a postgraduate educational program, in which senior doctors simultaneously train them and evaluate their suitability for a career in medicine. Stressors during a residency often seriously affect residents' own health and well-being: in some programs, residents may work on shifts requiring them to stay awake seeing patients for upwards of 24 hours at a time [8]. The cumulative stress and physical toll of residency may contribute to residents experiencing depression at twice the rate (25-33%) [49] of other young adults (8-15%) [29].

Faced with long-brewing challenges around residents' stress and rising burnout throughout the medical profession, the Accreditation Council for Graduate Medical Education (ACGME), which governs American residency programs, issued in 2017 a new mandate adding attention to well-being as a core program requirement. In this, ACGME places responsibility on programs to not only train residents on well-being self-management, but also to create structural supports, e.g. managing work hours and intensity, attending to workplace safety, and identifying and ameliorating burnout [4]. The mandate represents a major shift in the relationship between residents and their residencies, which was previously understood to be one of teaching and supervision. Residents' direct supervisors trained them day-to-day and ensured their patients received high-quality care, and residency program directors ensured residents had a quality educational experience. But neither party was responsible for residents' health, as this was viewed as a conflict of interest—e.g., residents experiencing mental health difficulties were to seek treatment elsewhere.

With this mandate, ACGME created a new type of relation, one that strains traditional boundaries in workplace supervision and educational mentorship—but remains distinct from the treatment relationship between a patient and their doctor. This new relation begs the question: What are the responsibilities of a supervisor to their workers' well-being, particularly in mixed educational and supervisory contexts like residencies? ACGME's mandate stops short of offering programs

specific implementation guidance, opening the door for residency programs to experiment with novel methods for enshrining this new relation in curriculum, practice, and evaluation.

We therefore studied how personal sensing technologies might be integrated into a workplace well-being program for resident physicians, as an exploratory step towards how Quantified Workplace ideas might be used to realize these new relations. This paper reports a qualitative design study with a psychiatry residency program at an urban hospital system in New York City. Through formative interviews with 11 residents, followed by design provocation sessions with 5 residents and 5 supervising attendings, we sought to examine how well-being management driven by personal data might be enacted in this setting. Drawing inspiration from literature in HCI and CSCW on design probes and speculative data work [10, 31], we presented participants with a medium-fidelity mockup of a dashboard depicting (a) real-time behavioral data on residents' sleep, activity and productivity; (b) self-reported data on residents' levels of burnout; and (c) free text boxes where residents could contextualize or comment on their own well-being. We used semi-structured interviewing to elicit participants' reactions to the provocation and two potential scenarios for its use: (1) a *self-reflection* tool for residents, and (2) a *data sharing* tool giving residents' supervisors and program directors a data-driven perspective on their well-being.

Our findings illuminate the design space around the use of personal sensing for workplace well-being (summarized in Table 2). Residents were less interested in using the tool for self-reflection, but saw potential in the data sharing use case as a way to encourage their programs to manage their well-being. Across both use cases, participants described tensions around how resident well-being can and should be measured (Section 4.1), the need for supervisors to be held accountable to act on well-being data (Section 4.2), and how the introduction of such tools might remake the boundaries of appropriate information flows between residents and their programs (Section 4.3). We conclude by charting future work confronting these tensions (Section 5), to ensure personal sensing is leveraged to truly improve well-being in high-stress workplaces.

2 RELATED WORK

Our work lies at the intersection of three lines of inquiry: worker health monitoring and the "Quantified Workplace" (Section 2.1), theories of privacy in personal sensing (Section 2.2), and empirical work to confront the problem of burnout in medical residencies (Section 2.3).

2.1 Worker Well-being and the Quantified Workplace

A growing line of research has collected and processed data from ubiquitous, everyday technologies, including sensors embedded in a smartphones, wearables, and other networked devices, to create near-continuous measures of behaviors associated with mental health and well-being [62, 72, 77, 90]. These novel data streams use sensor data including accelerometer and GPS for approximating behaviors such as sleep duration, physical activity, and sociality, or interactions such as computer scrolling or typing speed, to identify novel digital biomarkers associated with mental health and well-being from the fabric of individuals' lives. This area of research has taken many names, including *digital phenotyping* [33] and *behavioral sensing* [61]. Following Mohr et al. [60], we will use the term *personal sensing*, which highlights the "*sensitive form of surveillance*" we engage in when collecting personal data from devices. Personal sensing data can be used in conjunction with existing self-reported or diagnostic measures of mental health and well-being to provide a more complete picture of individuals' health [17, 41].

As researchers have developed the technical methods to enable personal sensing, they have in parallel explored how these techniques can be used to improve workplace well-being by identifying sources of stress, or by creating real-time tools to assist workers in stressful situations and enhance productivity. A CSCW workshop held in 2016 coined the term *Quantified Workplace* to describe

these efforts, defining the field as a paradigm “*in which the dynamics and health of organisations can be quantified through ubicomp technologies and visualised in order to offer collective intelligence to the managerial stakeholders and act as individual triggers for self-reflection and behavioural change*” [48]. This work has only accelerated as more and more of work becomes digitally mediated. For example, the Tesseract project has collected wearable sensor and social media data from over 700 information workers throughout the United States to investigate how sensor data and online social interactions can measure characteristics including job performance and psychological traits [51]. Kaur et al. used facial features collected from a worker’s webcam, heart rate, eye movement, software logs, and task progress to detect affect, which was then integrated into a productivity stress score that could be used to recommend work breaks [38]. Zakaria et al. used in-office WiFi-sensed location data to measure movement and group interaction patterns associated with worker stress and depression [101]. Lastly, a body of work has explored how passively sensed stress and affect detection can be integrated into virtual collaboration tools, for example Zoom or Microsoft Teams, to give presenters feedback on audience reactions, or to gauge meeting productivity [63, 78].

As research explores the technical feasibility of the Quantified Workplace, scholars have also noted the limited frame of mental health assumed by the underlying technologies which power personal sensing. As Semel notes, personal sensing is anchored to a line of inquiry which posits there are “biological mechanisms driving the observable symptoms of mental illness and are therefore more objective indicators than a person’s self-report” [79]. This frame, of mental health as a biological phenomena and of personal sensing data as a “more objective” indicator for lived experience, has motivated further debate within psychology and psychiatry, as well as in computer and information science. Not all clinicians and scholars agree that observable symptoms are a reflection of underlying mental health, questioning the validity of personal sensing to track symptoms compared to self-reports and patient interviews [85]. Our study extends this debate by prompting reflection on a personal sensing dashboard to explore whether and how well-being and burnout might be reflected in passively sensed behaviors.

2.2 Privacy in Personal Sensing for Well-being

Amidst debate around the extent to which passively measured behavioral signals map to mental health and well-being, scholars across the spectrum agree: collecting and analyzing personal sensing data poses real challenges to extant norms around privacy, especially in the relationship between workers and their supervisors. As Mohr et al. [60] highlight in their work arguing for adoption of the term *personal sensing* for this field, such data are highly intimate, especially when captured in real time. The inferences possible from raw personal sensing can also have material consequences for individuals: consider, for example, the use of activity data to set life insurance premiums [32].

Quantified Workplace researchers argue that informed consent is one tool to remediate concerns surrounding data sharing and privacy [101]. While the motivation for consent is well-intentioned, there are practical limitations to using informed consent as an agreement for long-term data sharing. First, from a historical perspective, the purpose of informed consent was to authorize a specific doctor to perform a specific treatment, on a specific patient [65]. Consent is not an ongoing data use agreement between two parties, particularly where the individuals involved in the agreement may change over time, and the understanding of what could be uncovered within the data, or how the data could be used and re-used, is ambiguous [59, 96].

Researchers in machine learning and computer security and privacy have investigated methods like *differential privacy* and *federated learning* [3, 100] to enable ongoing data sharing and re-use without exposing sensitive attributes at the individual level. These techniques are built to enable population-level summary statistics or model development, for example understanding commuting patterns or conducting the U.S. Census, without exposing sensitive information about an individual

who was part of the population. It is tempting to imagine that Quantified Workplaces might similarly allow supervisors to manage workplace well-being at a structural level without compromising individual workers' privacy—however, individual-level data may hold the most utility for stress management, as it creates an opportunity to target interventions for those most in need.

If personal sensing is most useful to well-being intervention at the individual level, then we must turn to understanding how data sharing can occur without harming a person by compromising their privacy. Such sharing requires clear enforcement of norms around appropriate data collection and use, to ensure it is used for well-being improvement only. Nicholas et al. [66] explored these sensitivities in their survey of attitudes towards sharing personal sensing data in and around health contexts. This work used Nissenbaum's theory of privacy as *contextual integrity* [67] as an analytical framework for the contextual factors that might influence a person's willingness to share intimate data, delineated as the *data subject*, the *data type* (including its sensitivity, in the well-being case), the *sender*, the *recipient*, and the *transmission principle*, or the set of constraints on information flow set by the norms of the context. Ultimately, their work found that people are most comfortable sharing personal sensing data with clinicians who have a formal role in health management, but less comfortable sharing data with others, including family and friends [66].

This finding points to the idea that sharing personal sensing data may be appropriate given established contextual norms ensuring the recipient uses it only to directly improve the data subject's health and well-being. However, this definition of appropriate information flow is insufficient for the challenges of worker well-being in the Quantified Workplace. What are the norms to be enacted around the role of workers' supervisors in managing their work-related stress? Of course, workers could choose to independently collect personal sensing data, and review it themselves or with a clinician to make decisions about well-being—and indeed, the Quantified Self [88] movement and associated research into self-tracking has developed numerous tools for personal health management. But when it comes to measuring and intervening in burnout and workplace-related stress, self-tracking places the burden of stress management directly on the worker, instead of on managers who have more power to change systemic factors that affect worker well-being.

These tensions around the appropriate flow of personal sensing information motivated the two use cases presented in our provocation: *self-reflection* and *data sharing* (see Section 3.1 for elaboration). Our study contributes to this growing literature by confronting how data flows in either case might exacerbate the power asymmetries between worker and supervisor, or resident and attending. In this, we provide an empirical case amplifying the need for what McDonald and Forte articulate as a theory of privacy as attention to *vulnerability*, rather than privacy as a matter of aligning information flow to dominant social norms [54].

2.3 Resident Physicians and Burnout

The stressors of the ongoing COVID-19 pandemic have exacerbated concerns about burnout within the global healthcare workforce. Burnout is a syndrome characterized by emotional depletion and maladaptive detachment that develops after prolonged occupational stress [91]. Burnout is occurring among healthcare workers in the U.S. *en masse*: a survey of healthcare workers by the American Medical Association found that 49% of physicians, 57% of nurses, 50% of non-physician clinical staff (e.g. social workers), 52% of clinical support staff (e.g. lab technicians), and 46% of non-clinical support staff (e.g. housekeepers) reported experiencing some level of burnout between April 2020 and April 2021 [7]. This alarming trend has the potential to devastate society, as the very workers tasked with caring for us leave their profession in droves. In response, researchers in medicine and public health have sought to characterize and intervene in burnout, in not only physicians but also the many roles within U.S. healthcare where workers are afforded less compensation and power (e.g. nurses, non-physician clinical staff, and support staff) [21, 26, 56, 92].

We contribute a study towards one segment of the healthcare workforce with unique burnout concerns: resident physicians. In medicine, a *residency* is the postgraduate training for individuals who have finished medical school, and is a requirement to obtain board certification to practice medicine in the U.S. Physicians enrolled in a residency program are called *resident physicians*, or informally, *residents*. Day-to-day, residents treat patients under the supervision of board-certified physicians, called *attending physicians*, or informally, *attendings*. Residents are additionally supervised by *program directors*, attending physicians who lead residency programs. Residency programs are designed as apprenticeships: Though attendings and program directors supervise residents' work, the educational nature of residency means they simultaneously hold a teacher-student relationship with residents. Good supervisors are role models for their residents, balancing workloads with opportunities for teaching and learning, and supporting them as not just workers or doctors-in-training, but also as people [83]. This constant fluctuation between teacher and supervisor makes the resident-attending relationship complex, particularly in high-stakes medical environments where established hierarchies and power dynamics dictate accountability and communication norms amongst clinical team members [11, 24].

Organizationally, residency programs are specific to a *specialty* (e.g., cardiology, internal medicine, or psychiatry), which determines the length and structure of post-graduate training required to become board-certified. Within a given year, residents will participate in multiple *rotations*, which expose them to different areas within their specialty. For example, in a general surgery residency, first-year residents might spend four weeks in a pediatric surgery rotation, followed by four weeks in thoracic surgery rotation [55]. Residency programs are accredited by the Accreditation Council for Graduate Medical Education (ACGME), which sets requirements for training programs [1].

Even prior to the COVID-19 pandemic, residents were known to work long hours and experience many emotional stressors, often resulting in burnout. Among physicians and physicians-in-training, burnout occurs at higher rates among residents (40-80%) compared to medical students (43-45%) and practicing specialists and general practitioners (22-60%) [52]. Burnout is dangerous not only for residents but also for their patients: for example, internal medicine residents experiencing burnout are >2x as likely to self-report suboptimal patient care, and are at increased risk for depression and substance abuse during their residency [80].

Interventions to reduce burnout and improve well-being among residents have taken many forms, including program-level structural interventions, the introduction of workplace well-being programs, and the promotion of individual self-care [19]. One of the more notable structural interventions occurred in 2004, when the ACGME capped resident work hours at 80 hours per week [44]. Prior to this change, residents often self-reported working up to 140 hours per week [8]. The reduction in work hours improved sleep and reduced attentional failures [44], but burnout still remains prevalent, and work hours may remain under-reported. Other attempted structural interventions include reducing rotation or shift lengths [46, 70], and hiring medical assistants to complete data entry tasks that previously took residents' time away from patient care [42].

Programs may also institute workplace well-being programs, e.g. stress management and spirituality workshops [50, 69], as well as protected time for small-group discussion sessions [98]. These workshops are associated with short-term improvements in some components of burnout within-subjects [53], as well as potential long-term improvements [98], but the scalability of these sessions has not been assessed [14], and effects are mediated by continued adherence [69]. Lastly, self-interventions for burnout have focused on meditation and targeted behavior change [94, 97].

Alongside these workplace interventions, research shows only a small percentage of residents choose to seek mental health support [2]. Residents' perceived barriers to mental healthcare include not only society-wide concerns like a lack of time and an aversion to the cost of care, but also concerns specific to the challenges around seeking care within the medical profession [2]. Physicians

in training perceive lowered career prospects for those who seek support for mental health concerns [99], and physicians encounter discrimination by medical license boards and malpractice insurers for disclosing mental health diagnoses and/or treatment on re-licensing questionnaires [16, 58].

Motivated by these challenges, recent research has explored how residents' stress might be measured through personal sensing data, with the goal of enabling more convenient and lower-cost tools for handling mental health. Kalmbach et al. found associations between sleep and activity data collected from a wearable device and daily mood [37]. Adler et al. predicted wearable sensing behavior changes associated with increased depression symptoms during the first year of a residency [5]. While this literature has shown these measurements might be feasible, there remain open questions around how personal sensing can be implemented for resident well-being management, e.g., how the data can or should be interpreted, who within the program should be accountable to act on it, and how the data can be used appropriately. Our study takes a formative first step towards addressing these questions.

3 METHODS

We conducted a qualitative study of how residents and their supervising attendings might react to a *design provocation* suggesting a possible future in which supervisors are equipped with intimate knowledge of their residents' well-being. Methodologically, we draw on interpretive inquiry methods like design probes [10], speculative dashboards, and data work [31], which have been used in HCI and CSCW to study potential futures for our era of data-driven sociotechnical systems. The goal of this study, in line with the design provocation tradition, was to explore and interpret participants' reactions to a suggestion of this future, to understand tensions that need to be addressed should technologists seek to realize it. In this section, we detail our methodology, including the design of our provocation (Section 3.1). We then describe the study procedures we employed to ensure our research activities protected our participants to the best of our ability (Section 3.2). All study procedures were approved by the coauthors' institutional review boards (IRBs).

3.1 Provocation Design

We began with a set of formative interviews and focus groups with residents to better understand the complex challenges they faced in managing their workplace well-being (See Appendix A for the interview guide). From these initial interviews and from the literature on burnout and personal sensing [61], we derived three possible affordances to examine: (a) the provision of quantified data on residents' sleep, activity, time working, and level of burnout; (b) the use of anonymization as an attempt to preserve individual residents' privacy; and (c) the ability for residents to provide contextualizing information on their wellness, in the form of written text. We then created medium-fidelity mockups of a dashboard that appears to provide these data streams to residents, their attendings, and directors in a unified program that includes two intersecting use cases: (1) *self-reflection*, in which residents view the data themselves as part of a personal well-being management strategy, and (2) *data sharing*, in which the same well-being data are provided to their supervising attendings and program directors for the purpose of intervention. We also added highlighting on the dashboard that suggests the use of "machine intelligence" to select worrying values, as an invitation for users to fill in what they would want from such a predictive or explanatory model if they were to see one in practice.

Our provocation centered on a dashboard depicting an individual resident's behavioral data, self-reported burnout scores, and written reflections over the last 14 days (Figure 1). Personal sensing data streams were provided as bar charts of sleep, activity, and time working. These are examples of behavioral indicators considered to be constitutive of worker well-being and burnout in formative interviews, and in particular ones for which personal sensing research has sought to develop reliable

measurements from pervasive data [5, 37]. The presented burnout scores were based upon the commonly used Maslach Burnout Inventory (MBI), which considers burnout along three dimensions: emotional exhaustion (EE), depersonalization (D), and sense of personal accomplishment (PA) [68]. MBI scores are shown as single measurements to emphasize they are cross-sectional and likely to be assessed less frequently. The scale of all measurements was intentionally unspecified to draw focus on relative comparisons over numerical absolutes, both between-subject (across residents and rotations, as shown in Figure 2) and within-subject (individual variation, as shown in Figure 1).

We presented these affordances to participants in variant scenarios specifying both the self-reflection and data sharing use cases. We added variant scenarios depicting different anonymization schemes, namely the use of “Resident A” versus the use of a real name (written as “John Smith” or “Jane Doe”). We also added variant scenarios including different messages residents might write in the provided text box adding context for anything in their lives that might have affected their wellbeing (e.g. “My attending has given me more work than I can possibly complete.” and “I have been moving my family across town.”).

Additional “screens” depicted aggregated views, including one where the behavioral measures for each resident in a rotation were placed side-by-side and one where statistics across rotations were compared (Figure 2). These views also included a small envelope icon similar to an unread notification on an email inbox, intended to indicate to the viewer that a new written reflection has been submitted within a specific rotation or on a specific resident’s dashboard.

Resident and attending participants were shown provocations sequenced per their participant group. The resident provocation session began by depicting the individual dashboard, prompting participants to reflect on the utility of the tools for their individual well-being first before then

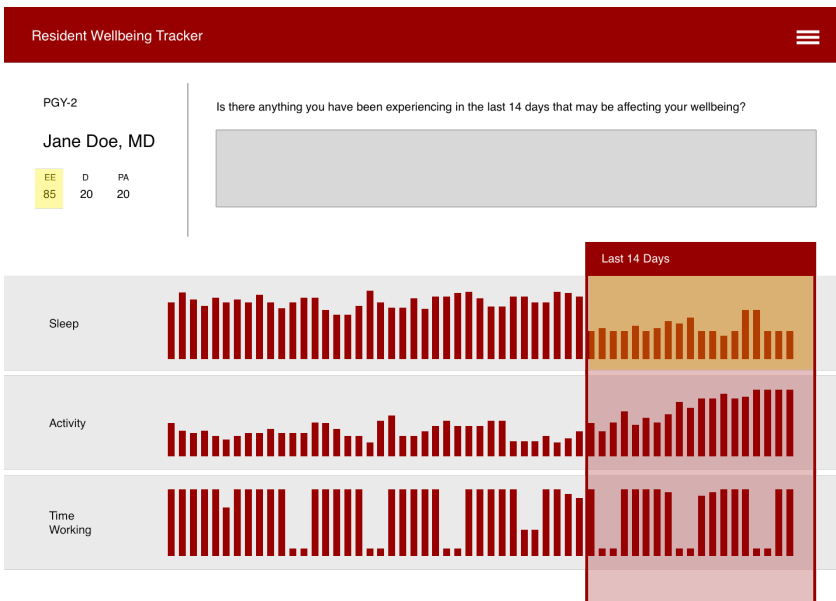


Fig. 1. Medium-fidelity individual tracking dashboard presented in our design provocation sessions. Behavioral data streams on sleep, activity and time working (bottom) are shown alongside burnout measures—EE (emotional exhaustion), D (depersonalization), and PA (personal accomplishment), following the Maslach Burnout Inventory (top left). The resident’s name could be anonymized (top left), and she is given a text box where she may self-report contextual information that may be affecting her well-being (top right).

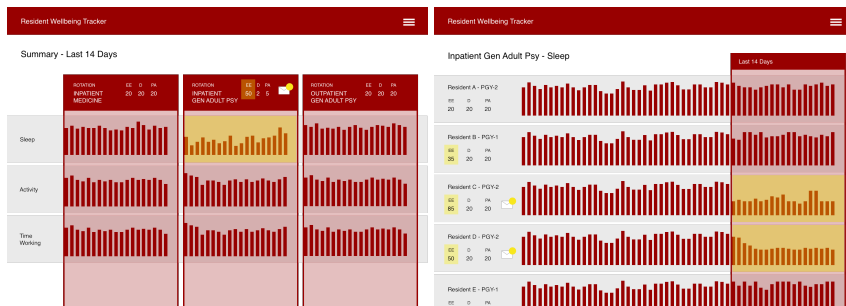


Fig. 2. Between-rotation and within-rotation dashboards presented in the provocation.

depicting the aggregated views. The attending provocation began with the aggregated views, accompanied with interview prompts on how they imagined leading their teams differently given the aggregated information, before moving into individual dashboard views to discuss their views on residents' privacy. The interview guides used to scaffold resident and attending sessions are in Appendices C and D respectively.

3.2 Study Procedures

Recruitment. We enrolled as participants resident physicians and attendings who work in a psychiatry residency program in an urban hospital in New York City, NY, USA. We chose a psychiatry program to take advantage of these mental healthcare providers' increased familiarity with stress, burnout, and well-being. Participants were recruited via a combination of convenience and purposive sampling [25], and were assured their participation would be anonymized, and would not impact their relationships with their residency programs, or with our research institution.

We sought to conduct an in-depth and interpretive qualitative design study, creating interviewing environments to allow for on-the-spot question adaptation and probing that would reveal insights from participants not possible from standardized survey methodologies [9]. Doing so also required us to balance the burdens that participating in such research would place on our participants. Physicians work long, intensive hours, and have limited bandwidth to engage in research. This bandwidth was further reduced during our study period: we recruited in late 2020 and early 2021, during which hospital systems throughout the U.S. were strained by the ongoing COVID-19 pandemic, and many residents and attendings were actively treating COVID-19 patients. Given the in-depth qualitative nature of our work, and the fact that our study population faced so many pressures at the time of our study, we spoke with 11 residents in the formative interviewing stage, followed by provocation sessions with 5 residents and 5 attendings.

Table 1 summarizes the demographic information self-reported by study participants. Because our study was conducted in the context of a single residency program, we did not take the additional step of asking participants to report more detailed information about their identities, or to reflect on how their identities might uniquely influence their responses to the provocation. We elaborate on this further in the **Ethics and Researcher Positionality** section below.

Data Collection and Analysis. Interviews and provocation sessions were held via 60-minute Zoom meetings. In the formative interviews, the two co-first-authors discussed with residents their experiences with burnout, their opinions on workplace well-being within their residency program, and their attitudes towards passive health sensing (see Appendix A for the formative interview guide). For the provocation sessions, the two co-first-authors presented participants with

Formative interviews (n=11 residents)	
Mean (SD) age	32 (2)
Gender	7 women, 2 men, 2 unreported
Resident Provocations (n=5)	
Mean (SD) age	31 years (2)
Gender	4 women, 1 man
Attending Provocations (n=5)	
Mean (SD) age	43 (14)
Gender	2 women, 3 men

Table 1. Self-reported demographic information of study participants. 2 residents did not complete the study intake survey, and an additional resident did not self-report their age. SD = standard deviation

the provocations via screen sharing and elicited their reactions via semi-structured interviews (see Appendices C and D for the provocation interview guides). All study sessions were audio-recorded with participants' permission and transcribed using a professional service. The transcriptions were then manually anonymized by the two co-first-authors.

These transcripts, alongside notes from each session, were analyzed by the two co-first-authors using a reflexive and constructionist thematic analysis approach adapted from Braun & Clarke [12]. Guided by the reviewed literature and research questions, the two co-first-authors first open-coded the data to develop both semantic and latent sets of codes. These codes were clustered and further combined into the topics reported in the codebooks in Appendices B and E. The topics in these codebooks were then further developed into the themes reported as our findings in Section 4 and summarized in Table 2. Examples of codes from the formative interviews (Appendix B) included *residents using fitness trackers*, *utility of program interventions for coping with stress*, *privacy implications of program reviewing data*. Codes from the resident and attending provocations (Appendix E) included *sleep as a well-being proxy*, *program shifts well-being responsibility back on resident*, *dashboard only useful if supervisors or directors act*, *whether anonymity is possible*.

Ethics and Researcher Positionality. Given the stigmas associated with mental health in the medical workforce (reviewed in Section 2.3), it was paramount to protect the anonymity of our participants from their present and future employers.

The two co-first-authors are both graduate students in computer and information science, and are not employed by the hospital system in our study, though one has an unpaid research affiliation which does not involve interacting with the participants in our study. These authors conducted all participant-facing study activities, including recruitment and elicitation sessions, and were responsible for anonymizing the session transcripts before they were analyzed. The second author, who is a resident in the same program as our participants, assisted with data analysis, but did not partake in participant-facing activities or see transcript data before it was anonymized. All other authors contributed only to the final manuscript and did not see the data. The third and fourth authors are supervisors of the residency program at the heart of our study, and helped to facilitate recruitment by sending a mass email, but were not a part of participant-facing sessions and do not know the identities of our participants. They further contributed only to the presentation of results in this paper, and did not see data before it was anonymized.

Our efforts to maintain participants' confidentiality and anonymity also mean we do not report participants' identities in this publication beyond the basic age and gender demographics presented in Table 1. We additionally did not ask them about personal identity in our interviews. Identity is an incredibly important element of the power relationships we study in this work, and more

broadly an incredibly important element of privacy [54]. But after careful consideration of diversity and equity issues in U.S. medical residencies, we avoided specifically studying identity in order to protect participants who may be minoritized within the small residency program we have studied, and therefore more identifiable based only on their responses: e.g., there may be just 1-2 Black or Latinx residents within a given program. After analysis, it was clear that in our data, participants had not volunteered detailed disclosure of their intersectional identities or reflected on how they may influence their responses. We leave an identity-focused study to future work that can protect minoritized participants by, for example, enrolling across multiple sites, and focus the present study on the specific power asymmetries between a resident and their employer.

Lastly, we took additional care to ensure our research activities did not interfere with participants' existing pathways for mental health and social support. Discussion of burnout and labor conditions can be emotionally stirring, and our participants took time out of intensive work schedules to speak with us. Throughout our sessions, we ensured participants knew they could always pause or stop the session, decline to answer a specific question, or withdraw from the study completely.

4 FINDINGS

Our analysis finds a rich set of tensions that must be addressed before personal sensing technologies can be implemented in a Quantified Workplace intervention for resident physician well-being. We describe these tensions in three overarching themes, which are summarized in Table 2.

At a low level, participants saw the dashboard's affordances could, with some refinement, provide the right behavioral and self-reported data to uncover both individual and program-wide resident well-being changes (Section 4.1). But when asked about how these measurements could be implemented in their workplaces to drive well-being improvement, participants highlighted challenges in both the *self-reflection* and *data sharing* use cases implied in the provocation. Participants were generally less interested in using the provocation for *self-reflection* and more interested in the *data sharing* use case presented. Many residents felt an extra visual representation of their own well-being data would not offer new information or help alleviate stress, and instead confirm their own understanding of their own negative behavioral and well-being trends (Section 4.2). This sentiment on self-reflection echoes the broader personal sensing literature, in which studies have shown people are ambivalent towards affordances for self-reflection alone [6, 71, 95].

Participants were more intrigued by the tool's potential to better enable supervising attendings and program directors to implement well-being interventions—the ultimate intention of the *data sharing* use case implied in the provocation. To realize this potential, however, participants raised that such a program would need clear and well-enforced accountability measures ensuring supervisors would act to help residents when given the information (Section 4.2). Residents and attendings also raised concerns around the intrusion into residents' personal lives that the personal sensing data would constitute, and voiced that implementing such tools would require remaking notions of privacy and appropriate information flows (Section 4.3).

We describe each theme in detail in this section per the outline in Table 2, and in Section 5 further explicate the implications of each theme for research into and implementation of workplace well-being programs. Participants are quoted using a participant- and interview-specific identification number, and a letter specifying whether the quoted participant was an attending (A) or resident (R).

4.1 Refining personal sensing towards legible measures of workplace well-being

The concept of personal sensing for workplace well-being rests on the ability of behavioral data streams to represent some notion of worker well-being and its inverse, burnout. Participants' reactions to the provocations yielded valuable insights on how personal sensing measures—independent of use case—might be refined to better align with their experiences of these constructs.

Theme	Finding	Implications
Passive behavior measures must be refined and accompanied by self-reports and reference ranges to better capture lived experiences of well-being and burnout.	Sleep duration is most useful as a low-level indicator for sleep quality, which better informs well-being. (Section 4.1.1)	Further research is needed to develop indicators that focus on sleep quality, available leisure time, and productivity. Further research might also consider better ways to combine self-reported context with passive measurements and co-constructed notions of wellness, in ways that provide reference ranges for individuals sensitive to diversity in what is considered well-being. (Section 5.1)
	Physical activity might be less useful as a well-being indicator than a measure of leisure time. (Section 4.1.1)	
	Time working is most useful as a low-level indicator for productivity or overwork, which would inform well-being. (Section 4.1.1)	
	Passive tracking is perceived as a less arduous method to measure behavior, but loses the contextual information self-reports provide. (Section 4.1.2)	
	Passively tracked data requires both within-subject and between-subject reference ranges to be useful. (Section 4.1.3)	
Workplace well-being measurement is only useful if the data is used to ameliorate stress-causing working conditions, with clear accountability within the program.	Residents do not want their workplace well-being to be measured for self-reflection, but instead to create accountability for those in power (e.g. supervisors) to improve their working conditions (Section 4.2)	Prior to implementation of personal sensing for workplace well-being, workplaces should develop policy mechanisms that hold supervisors accountable for resolving employees' stress in response to data shared about their well-being. Frameworks for responsible data usage must be integrated with worker protections to prevent retaliation and ensure these programs truly steward worker well-being. (Section 5.2)
	However, it is not clear who within the supervisory structure of a residency program should be held accountable to institute these improvements. (Section 4.2.1)	
	Supervisors are much more comfortable using well-being data to approach residents when the data show sources of stress considered exclusively work-related. (Section 4.2.2)	
	Residents see opportunities for retaliation if data sharing shifts accountability for maintaining workplace mental health from the program to the resident. (Section 4.2.3)	
Under current norms, sharing workplace well-being data with supervisors poses major risks to residents' privacy and autonomy.	Sharing personal sensing data with supervisors blurs the boundaries between residents' work and personal lives, thereby violating residents' privacy. (Section 4.3.1)	Existing approaches to privacy preservation, e.g. anonymization, are insufficient to address the challenges to autonomy created by personal sensing in the Quantified Workplace. New relational norms that prevent abuse of the power differential between data subjects and recipients must be created and enforced if this technology is to be widely used. (Section 5.3)
	Passive data sharing violates data subjects' control over disclosure, and may cast residents as patients rather than as trainees or junior colleagues of their supervisors. (Section 4.3.2)	
	Within tightly knit residency programs, anonymity is perceived as impossible, hampering the group participation required for guarantees like <i>k</i> -anonymity. (Section 4.3.3)	

Table 2. Summary of major themes, supporting findings, and implications of our study for the research and implementation of Quantified Workplace technologies.

At a high level, participants appreciated the idea of having visualizations of behavioral data for well-being management purposes—even if some data streams were less useful to some than others. As one resident described:

“I’m someone who loves visuals. I don’t know if time working is useful for me because I have an acute awareness of how much I’m working, but I think putting activity and sleep into a quantifiable visual is [helpful towards understanding well-being].” (R13)

Other participants suggested many improvements that might make a tool like the provocation better reflect their intrinsic notions of well-being. Drawing on their lived experience with their own well-being and burnout, as well as their clinical experience as psychiatrists, our participants considered how best to measure well-being and burnout along three overarching themes. They first described that these data streams must be refined with context to encapsulate their intrinsic notions of well-being. Still, they reported that the gold-standard ways to provide adequate context might create additional burdens in their lives and workflows. Lastly, they grappled with the need for reference ranges to help them further contextualize individual well-being data, and any potential relationship with burnout.

4.1.1 Refining well-being measurements with necessary context. At each level of the provocation, participants described refinements that might help align the provocation’s implied data streams more closely with their experiences of well-being and burnout. Some of this arose from their own training as psychiatrists: most participants agreed that **sleep** may be a “good proxy for well-being” because in psychiatry “the first place you look for concerns are sleep disturbances” (A01). In accordance with existing scholarship on sleep sensing for mental wellness [75], several participants stated that sleep numbers would be more useful as an indicator of well-being if they reflected on not simply the duration of sleep but instead a more subjective notion of *sleep quality*: a daily feeling of energy or sense of renewal, which may be a product of consistent or otherwise high-quality sleep.

Similarly, participants felt **activity** measurement might be refined as specifically leisure activity, to reflect whether a resident had “time to decompress, reflect, and do things that aren’t work related” (R13). Such a refined measure would avoid the pitfalls of more raw behavioral interpretations of activity like step counts, which participants also described would be easily biased by whether a resident was on a more sedentary rotation, e.g., on outpatient versus inpatient. It would also align with a more positive definition of well-being, as not only the lack of overwork but also the ability to best make use of time off of work.

Participants had the most refinements for the **time working** measure implied in the provocation. First, following concerns around working hours described in Section 2.3, many asked if this measure was intended to reflect actual hours on the job or reported “duty hours”, the latter of which were a reporting requirement set by the ACGME [44]. For well-being purposes, residents described that recorded duty hours would be an imperfect proxy as they typically underreported total hours worked, not accounting for time spent on administrative tasks or scholarly activities, and it was this totality of working hours that might be most indicative of (a lack of) well-being. As one alternative, several floated the potential for a “productivity” measure, and offered that their program already tracked this for each resident—however, they were unsure of how, precisely, these metrics were calculated. Participants’ lack of clarity around how productivity can be measured echoes literature grappling with how best to define resident productivity—e.g., a metric tracking patients seen per hour is potentially inappropriate for psychiatry residents, because it focuses on volume and not service quality, and ignores the longer-term nature of psychiatric care [13, 34, 35]. Beyond refinements for what precisely might be measured in this construct, residents also worried about “the belief that time working is synonymous with productivity” allowing “programs to milk us for as much work as we can possibly do in four years” (R14). (We unpack this tension in Section 4.2.)

Refinements to better incorporate context extended beyond the behavioral data streams into the self-reported **burnout scores**. Most participants were unfamiliar with the Maslach Burnout Inventory, the validated scale reported on the dashboard [68] and summarized in Section 3.1. One participant found the polarity differences between the components confusing, specifically, that higher emotional exhaustion (EE) and depersonalization (D) indicated higher burnout, but higher personal accomplishment (PA) indicated lower burnout. While they were unfamiliar with the scale, most agreed at face value that it did seem to capture constructs that could be construed as burnout. Interestingly, several attendings stated that in their work, they preferred simpler methods for assessing burnout, such as off-the-cuff Likert scales. Attendings and residents alike also shared that work satisfaction was another potentially useful subjective measure of burnout, potentially as a subcomponent of PA.

4.1.2 For well-being measurement, self-reported and passively collected data must be considered complementary. In contrast to the numerical data streams described above, residents agreed that the free text field, offering a **written well-being report** in response to the prompt “*Is there anything you have been experiencing in the last 14 days that may be affecting your wellbeing?*”, did have the potential to offer sufficiently contextualized information on well-being. Residents described that this field better aligned with how they would assess patients for burnout, in which they would “*not use scales when talking to patients, and be conversational*” (R03) to understand “*what it is specifically that’s causing the burnout*” and “*find out exactly what it is that patient is experiencing*” (R11). Along these lines, residents described the written well-being report provided an opportunity to add contextual information supporting the personal nature of burnout, its potential causes, and effects.

However, residents also immediately latched onto the burdens this form of measurement would create. One resident highlighted that despite this data stream capturing “*the most nuanced information, it also requires the most mentally*” (R14). Several offered that they were already bombarded with survey instruments as-is, and would be unlikely to engage with an even more time-consuming wellness measure despite its ability to provide sufficient context.

Issues of burden in self-report extended beyond the written well-being comments to behavioral data streams as well. Participants compared and contrasted two potential methods of behavioral data collection: gathering data passively using devices (e.g. a wearable or smartphone), vs. self-report. Generally, residents believed that behavioral self-reports would be arduous: as one resident stated, “*I think people wouldn’t want to [self-report sleep]*” (R15).

However, participants in both the resident and attending subgroups stated that passive measurement should be used as a well-being indicator not in place of but in concert with some form of self-report. Part of this concern was due to the limited accessibility of wearables: one resident suggested that a smartphone may be a more reliable device for passive measurement because the phone was “*more universal*” and “*not everyone has a wearable*” (R02). In addition, several supervising attendings also affirmed that self-reports hold fundamentally different meaning than passive behavioral data, and that both types of information would complement each other to provide a more complete picture of well-being. As an example, one offered:

“I think both [self-report and passive measurement] would be great because you’d be able to see if they confirm or don’t confirm your experience. Let’s say according to your tracker, you slept plenty, but according to how you feel, you haven’t slept enough. Is it because you were preoccupied with work, or is it because you really did not sleep well?” (A01)

Providing this complete picture, with two reinforcing data streams, was seen as potentially useful to the overall goal of understanding and championing residents’ well-being—still, one resident cautioned that in using both sensed behavioral data and self-reported well-being, the program must take care to not imply the supremacy of one data source over the other. As they described:

“If I were really feeling overwhelmed by a person giving me more work than I could possibly do, I would hope that in a program who cares a lot about me, if I went to the trouble of making a meeting with my program director to be like, ‘I cannot do this. This person is making my life too hard’—I would hope they would take my self-report as seriously as they would weight seeing the actual numbers of me not sleeping.” (R14)

4.1.3 The need for comparison and reference ranges to capture well-being. Many of the attendings in our study noted they would want to understand where residents fell with respect to a *reference range*—a set of values for a given measure understood to indicate normal functioning. As one attending described, they needed “*some standard to compare to*” (A01). As shown in the highlighting feature of the provocation, participants imagined an alarm that triggered upon deviation outside this range, akin to vitals monitoring used in hospitals.

Participants offered many different ideas for where they felt these reference ranges ought to be drawn from—what behaviors or well-being measures might be seen as deviating from the norm. Some residents jumped straight to the need to identify individual baselines, perhaps via a screening survey taken when a resident could be verified as not yet burnt out, to account for within-resident changes. But while the utility of individual fluctuations in well-being was clear, participants described it was not clear when an individual’s baseline well-being might be assessed. As one resident described:

“I think you could have underlying chronic burnout in a way, so maybe then [deviation would be] harder to pick up on. At the same time, if you’re living really, really good, what state do you assess against? I mean, I just came back from vacation, so I feel good.” (R10)

Others identified that reference ranges might be drawn from between-resident comparisons (implying use case 2 - data sharing), setting the expectation that well-being and burnout might be defined as deviations from your cohort’s norm. One attending posited:

“Different rotations are going to have very different activities. That makes me feel like there’s too much confounding for behavior to be useful as a well-being indicator, unless it’s comparing your value to the average value [within a rotation], and I see that you’re an outlier within your cohort.” (A05)

Participants also saw utility for well-being to be managed as deviations from the norm at a program level, looking across rotations. One attending identified that between-resident well-being data could be used to identify problems with a rotation, or even with one attending supervising multiple burnt out residents.

“If I see [multiple residents with alarming values], and I can tell if they’re [residents] on the same rotation and the same unit, I can tell that maybe there was a problem with the unit. Maybe there’s something that’s beyond just the individual. Maybe there’s a problem between the residents or with the attending supervising those residents.” (A03)

4.2 Well-being measurement must be accompanied by accountability for supervisors to intervene

Our findings in Section 4.1 show that, with refinement, the dashboard could offer residents and their supervisors program new behavioral and contextual information to measure resident well-being. Given refined measures, residents and supervisors speculated on the utility within the two intersecting use cases: as a tool for (1) *self-reflection*, for residents to self-manage their well-being, and as a (2) *data sharing* platform, in which residents could share their data with supervising attendings and program directors to better enable program-level well-being interventions.

Many residents were ultimately disinterested in the dashboard as a tool for well-being self-reflection and self-management (use case 1). Those who did have some interest were already motivated by health tracking tools like FitBits in their daily lives, but others were resistant to the behavior change that self-reflection implied. As one resident described, *“I know I’m very sedentary so I don’t want it to confirm things I already know I should do more of.”* (R11). Part of this reluctance was historical: the residency program had given out pedometers in a previous year, and one resident observed, *“I haven’t actually used mine since the challenge ended”* (R01).

On use case 2, the data sharing program, participants in our study agreed on one central idea: that sharing well-being data might give supervising attendings and program directors substantial power to ameliorate residents’ burnout and suffering. But, residents were skeptical that the information, if centralized in a tool like the dashboard, would lead to any meaningful changes in their program.

In past experiences, residents described, when they raised concerns about well-being, their program had responded by providing *“emotional support in the moment, [...] but then they don’t really change anything”*. This made residents question, *“what’s the point of sharing these things?”* (R15). One resident stated that residents already *“get a lot of surveys [on well-being]”*, and that *“if the program is assessing burnout and they’re not going to listen to our suggestions, [personal sensing] would actually be more annoying.”* (R10). Thus, if the program were to institute a new data collection program, there would need to be some amount of accountability on the program to make changes that improve well-being. Otherwise, as one resident stated, tracking this information could be viewed as an additional stressor:

“I think if I were truly burnt out and feeling burnt out, having a number to put on it may just make me feel worse, and think, ‘Well, what do you want me to do to fix this? Are you going to give me a vacation or something like that?’ ” (R16)

In this section, we further unpack tensions around the use of passively sensed behavioral data in the *data sharing* use case. Specifically, we describe findings surrounding (1) who within the program should be held accountable to residents’ well-being, (2) how perceptions of accountability differed by the source of resident stress, and (3) residents’ fear of shifting accountability and retaliation.

4.2.1 Who in the program should be held accountable? Given residents’ hesitancy to share data, we probed residents on who within the program was accountable for well-being—e.g. attendings, program directors, or some other third-party auditor. Residents stated that since program directors supervised residents throughout the entirety of their programs, and not in a single rotation, having well-being information across residents available in a centralized location could help program directors understand resident well-being year-by-year, and potentially instigate systemic changes. Supervising attendings, on the other hand, only worked with individual residents during short-term (e.g. 2-week) rotations, and did not have the same level of awareness or connection with residents. As one resident expressed:

“You’re working with attendings for a brief period of time. And attendings’ goals are ‘have the resident work and learn.’ I am having trouble thinking of many attendings who would really care about things outside of those two spheres in the resident’s life.” (R12)

Many residents in our study agreed that the degree to which individual attendings might want to attend to their residents’ well-being was highly variable, even within their program. For those attendings who did care, however, residents observed that the introduction of well-being data into their relationship might actually subvert their bond:

“The supervisor that I value the most, she really only supervises me. So, if she got this data, I don’t know if it would be valuable. We would both be like, ‘Isn’t it weird that I’m giving you all of my sleep numbers instead of us talking about how I’m doing?’ ” (R14)

Though we expected that supervising attendings had an implied expectation to maintain day-to-day resident well-being, our attending participants were sometimes hesitant to take on this role. Several observed that being provided with well-being data seemed to fundamentally alter their relationships with their residents in unwanted ways. (We further elaborate on the tensions surrounding new relational norms in Section 4.3, and unpack the issues raised by these alterations in Section 5.3). As one described:

“I don’t know how I would feel. I would be happy to do this maybe more on an ad hoc basis with someone who may be struggling or may need the help, but then you’re taking the role of a caregiver, which is different from the role that we [attendings] have with residents.” (A03)

At the same time, another supervisor stated it was often difficult to assess resident well-being because program directors often did not relay information to attendings about personal issues that may impact residents’ work. This supervisor stated that if they had a tool like the dashboard, they may feel more comfortable approaching residents to have a conversation about their mental health.

4.2.2 Perceptions of accountability change with source of stress. Supervisors’ hesitancy to engage residents on their well-being became particularly apparent when they were asked to reflect on varying sources of residents’ stress. As described in Section 3.1, we presented a free text box on the dashboard, and in variant scenarios supplied the box with two different messages: (1) a message from a resident who wrote that they were stressed because they were *“moving their family across the city”*, and (2) a message a resident who stated *“my attending has been giving me more work than I can possibly complete”*.

Supervisors appeared much more comfortable approaching residents regarding scenario 2 than scenario 1. One supervisor directly stated, *“maybe in [scenario 2], it’s something that will resolve on its own.”* and *“the intervention [for scenario 2] doesn’t focus on the resident, which allows me to address the concern that resident has”* (A03). In other words, for this attending, they felt more comfortable intervening when the source of stress was work-related, and not personal. We pressed supervisors on this further, stating if the dashboard was de-anonymized, and they had the capability to directly approach the resident from scenario 1, would they take action?

Interviewer: *“The resident is de-anonymized. Does that de-anonymization change your approach?”*

A02: *“Not really, because in this case I wasn’t really going to intervene much anyway. Give her boxes maybe, I don’t know.”*

That supervisors’ willingness to intervene differed by scenario comes as no surprise. While personal detail is critical to measuring well-being (cf. Section 4.1) following up with the resident implies de-anonymization of well-being data, the privacy implications of which we will unpack in Section 4.3. Indeed, in conversations with residents, one anticipated an approach to sharing the type of life-stress information described in scenario 1 that may remediate supervisors’ hesitancy to act: potentially, the dashboard could have a button *“to trigger some kind of alert where [the attending] doesn’t have details but it’s just like, ‘Hey, you should check in with your resident.’”* (R13).

4.2.3 Datafication of well-being and burnout might lead to retaliation against workers. Residents were fearful that quantified behavioral data on their well-being could lead their program to retaliate, by shifting accountability for workplace well-being from the program back to the resident. First, by sharing data on well-being, residents highlighted the potential for *datafication*, or the quantification of subjective aspects of well-being and performance [74]. Most notably in the context of productivity, residents recalled several past instances in which datafied behavior enabled their programs to create additional unwanted burdens on residents:

“Throughout this year, we have had random times where the program told us, ‘Hey, here’s some news. Your numbers are kind of low, so here’s five more patients for you.’ We don’t know why these came down, or how these numbers were calculated, but we all fear it could happen again.” (R03)

This fear was heightened by the complicated power structures that exist within a residency. Residency programs are technically educational programs, and indeed hospital systems receive reimbursements for training residents [57]. Yet residents are held to a similar standard of care as attendings, despite being paid significantly less [36, 64]. Residents in our study felt these work conditions meant they were uniquely vulnerable to exploitation, and were concerned that data-driven perspectives on their well-being would only further their vulnerability. As one described:

“If residents had more protection, in terms of being employees, then intervening to help residents out would be great. But I’m so skeptical of programs solely having that intent at the heart of what they’re doing or how they would use this data.” (R13)

In summary, residents were hesitant to share their behavioral and mental health data with the program without assurances of accountability. That being said, who should receive this information and the context surrounding which type of information should be shared remained unclear. This resulted in speculation around negative repercussions surrounding data sharing.

4.3 Contending with privacy

Across all interviews and sessions in our study, both resident and attending participants expressed that various features of the provocation, and the proposition of a data sharing program (use case 2), violated some notion of residents’ privacy, broadly defined. But residents in our study varied in their expressions of how much such violations offended their innate sensibilities, or might drive them to action. As one stated plainly:

Interviewer: *“Let’s say there’s something on in the background that collects your heart rate. How long would you be comfortable with this kind of data being collected and shared?”*
 R02: *“Forever. I don’t care.”*

For the residents in our study who shared this sentiment, the privacy violations implied by sharing their health and activity data with their supervisors were outweighed by the potential benefits. Some described they were actually already engaging in passive self-tracking for personal or clinical purposes, for example by wearing a Fitbit, or, in one case, sharing a daily mood log with their psychotherapist. Extending this personal tracking to their residency program would be fine, these residents said, if it would help program directors improve the resident experience.

However, not all resident participants shared this permissiveness. For some, the privacy violations created by allowing the residency program to track residents’ activities and well-being were simply a non-starter. Such a program would require placing enormous trust in the hands of the program, which not all participants shared. Even for those participants who did think their specific residency program could be a trusted receiver of this information, the provocation raised concerns for the standard it would set in other programs. As one attending reflected:

“Honestly, my first reaction is, this is big brother. This is something that can be used for good or not. And if you trust that the residency program is looking after their residents, they can do good with it, but there might be programs that don’t [look after their residents], that use this to monitor when residents leave or how many patients they see.” (A03)

In addition to expressing their broad preferences around personal data sharing and their levels of trust (or lack thereof) in residency programs, our data show participants considered the nature of the implied privacy violation along three interlocking dimensions. Participants were concerned

that the data would (1) blur boundaries between work and life contexts and (2) inadvertently cast residents not as attendings' coworkers, trainees or mentees, but as their patients. In addition, participants were concerned that (3) anonymity, however enacted, might be functionally impossible within their workplace.

4.3.1 Blurring work and life. In both the attending and resident groups, participants voiced discomfort with supervisors providing feedback on residents' personal lives, even under the assumption that such feedback would be provided to help them manage the demands of their work. Many felt this discomfort extended across both the behavioral and contextual data streams implied in the provocation. As one resident described:

"I can only imagine coming in in the morning and having the supervisor be like, 'You got two hours of sleep last night? What's going on?' That's the panopticon, your supervisor watching you at that level. Is there a big difference between the wearable data, versus [a resident's answer to] 'do you feel like you're losing interest in your work?' I'd say they're actually pretty similar." (R07)

Attendings, for their part, agreed that sleep data in particular felt "intrusive", but were able to imagine using such detailed information to help train residents on how to cope with the demands of their jobs. As one speculated:

"It might be a good conversation starter. 'I notice you're sleeping less.' I mean, as soon as I say that, it sounds a little bit intrusive, to put it mildly, for me to be talking to a resident about their sleep. But in the context of looking at this data together, it could be a useful prompt. 'I noticed you're spending a lot more time working, can we talk about how much work you have and your ability to organize to get it done properly?' " (A05)

Residents also expressed that the provision of detailed wearable data might give their supervisors license to attribute workplace difficulties to their personal situations, rather than to work-related challenges over which they might have some control. As one expressed:

"There's sort of an intrusion there on my personal life, and making my own choices about what I do with my time, in a way that I fear sort of gives them numbers to blame me for my own lack of well-being. They'll be like, 'well, you slept two hours last night, of course you hate this patient this morning.' " (R14)

4.3.2 Treating residents like patients. For several attendings, the provocation evoked the metaphor of the resident as a patient. Detailed personal sensing data might be useful in a doctor-patient relationship, these participants said, for the social contract implied therein, in which patients allow doctors to track highly personal information for the explicit purpose of diagnosing and treating medical conditions. But this type of relation was different from the resident-attending relationship, in which such disclosure would be considered far more invasive. As one attending said, *"This is a lot like monitoring patients, right? Which is part of the intrusiveness of using it with residents"* (A05).

Beyond the perceived violation of exposing this information within resident-attending relationships, participants also pointed out that the passive nature of the real-time sensing specifically violated additional sensibilities around privacy. At issue was the tendency of passively sensed data to violate the data subjects' ability to control which data was shared, and for what purpose—in any type of relationship, under any type of distributional norm. As one attending participant described:

"To me, the concern is about the patient's—I mean, the resident's autonomy. If they're willing to disclose this to me explicitly as a tool for me to help them in a particular scenario, then yes, it may be helpful. But just to have this information out of that context makes me uncomfortable." (A03)

That this participant accidentally refers to the hypothetical resident as a patient highlights the slippage between these roles created by the provocation. What's more, this participant's framing of the dashboard as a violation of autonomy echoed some residents' observations of how passive tracking might also subvert distributional norms within doctor-patient relationships. Asked for their impressions of the use of this type of highly personal and passively tracked data in mental health and well-being generally, one participant said it violated a core tenet of mental healthcare: the patient's right to "self-presentation", or the ability to talk about oneself in the way that one wants. Elaborating on the necessity of self-presentation in psychotherapy, this participant said:

"My patient has the right to lie to me. It's part of the process for them to feel comfortable, to get to a point where they feel they can tell me the truth. Sometimes people don't tell me something for months, but then they get to a point that they can tell me. And that's important for the therapeutic relationship. But if you already know what's going on, it interferes with that relationship." (R10)

The right to self-presentation, this participant described, was key to the therapeutic relationship in a doctor-patient interaction—and the use of passively tracked data constituted nothing less than "giving up your right to talk about yourself in the way that you want" (R10). We further unpack this finding in Section 5.

4.3.3 Anonymity is not possible, realistically. Lastly, participants contended with whether anonymity measures might offer a way to retain the benefits of the provocation for improving residents' well-being, while still affording them some measure of privacy. Averaging the data for attendings' and program directors' views might achieve anonymity, but participants were quick to point out that reporting only average values for a rotation or service would defeat the purpose of helping residents' supervisors identify that individuals were struggling relative to their peers. Turning to the problem of reporting individual values, participants considered whether anonymization might be possible by removing identifying information (*blinding*), or by aggregating and sanitizing the data such that a single resident in a table of residents was no longer identifiable. Pressing participants further on how these measures might be actualized, however, revealed that most participants had significant qualms over whether anonymization was possible at all within the constraints of their program.

For individual-level reports, participants considered anonymization to be impossible given that their program was small and closely knit. Names could be easily removed, but the personal nature of both the behavioral data and the textual self-reports was such that participants felt they could easily figure out which data belonged to whom. One attending pointed out that on some services, they might only supervise one or two residents at a time, thus making it very simple to map the shift schedule implied in the sleep numbers for "Resident A" to the resident's actual identity. Another pointed out that the written contextual self-reports were by definition difficult to sanitize, especially when presented in conjunction with an individual's behavioral data. As one attending described:

"To me, there's not a good way of blinding this. It is going to be difficult to keep it anonymized in such a closed system where everybody knows each other, more or less. It may actually be a misleading assumption to think that you're going to type something there and no one's going to recognize you." (A03)

If individual-level reports could not be realistically blinded, could aggregation be used to retain the benefits of the dashboard for program-level intervention while affording residents anonymity? Following the principle of k -anonymity set out by Sweeney [89], anonymity-via-aggregation could be achieved by reporting only values that had been abstracted to less specific categories (e.g. reporting an age bucket rather than a specific age) and removing very sensitive attributes altogether — sanitization techniques called *generalization* and *suppression* in the literature. Such measures

could be used to ensure that a single resident could not be uniquely identifiable within a set of k residents. But regardless of the blinding or sanitizing techniques used, one resident was skeptical that a sufficiently large pool of residents would consent to the program to create such guarantees. As one resident described:

“There is still, unfortunately, this mindset of, ‘I don’t want to appear somehow weak, or incapable of holding up my portion of the work.’ Asking for accommodations is somehow adding to that, and admitting that you have outside obligations is still contributing to that. So if only some residents are participating in this, then it becomes even weirder to do it or not do it. Then it just becomes very skewed. You almost need everyone’s buy-in, and I really think you would struggle to get that to broadcast this data to everyone.” (R13)

As this resident highlights, the provocation implied a well-being tracking program with such far-reaching consequences for the residency program overall that there could be no partial participation. The *broadcasting* required to retain the benefits of the program—specifically, using transparent and (hopefully) anonymized reporting to fight stigmas around well-being and burnout—would, in this resident’s view, be too large of a hurdle for some to overcome. Assuming that some residents would participate and others would not, mathematical guarantees of privacy would be degraded for the entire pool—and what’s more, the data visible to program directors and supervisors would only represent those residents most comfortable with sharing or broadcasting their discomfort, creating bias or skew that might ultimately doom any downstream attempts at data-driven intervention.

5 DISCUSSION

Our study has highlighted a range of sociotechnical challenges in the design space around using personal sensing technologies for Quantified Workplace well-being interventions. While data-driven “solutions” for well-being and burnout may be appealing to well-intentioned researchers and managers, our findings demonstrate that their implementation must be accompanied by guardrails to prevent them from instead reifying existing power asymmetries, or exacerbating potential for abuse. We see urgent need for further work establishing new norms around data-driven workplace well-being management solutions that better center workers’ needs, and provide protections for the workers they intend to support. Such work must be a precondition of introducing Quantified Workplace technologies like well-being dashboards into workers’ social arrangements. We would like to note a benefit of conducting a study to highlight these tensions in the context of a psychiatry residency program, specifically that experts on treating mental health and well-being themselves were not sure how to leverage well-being data for meaningful workplace intervention, and concerned about the privacy implications of doing so. If experts cannot resolve these tensions, continued research to establish appropriate data sharing norms is of utmost necessity prior to workplaces more broadly implementing similar systems. We join the strands of work in Quantified Workplace research that have acknowledged its tensions with privacy [38, 51, 101], and also highlighted how these monitoring systems can feed into the at-times-coercive dynamics of scientifically managed workplaces, which can optimize for the productivity of the worker over the well-being of the worker themselves [38, 43]. In this section, we build further on the themes in Table 2 to unpack additional considerations for researchers and workplaces who, like residency programs, may be exploring the use of Quantified Workplace technologies to improve worker well-being.

5.1 Triangulating behavior, biology, and perception in ground truths for well-being measurement

Our study has highlighted not only problems in how Quantified Workplace ideas can be implemented (which we unpack in subsequent sections 5.2 and 5.3), but also in how personal sensing itself can

be refined to better measure well-being and burnout. Section 4.1 highlights the trade-offs our participants described for including low-level behavioral (e.g. sleep duration), high-level behavioral (e.g. sleep quality), and self-reported contextual information (e.g. written well-being reports) within a resident well-being management dashboard. Residents generally appreciated that passive tracking via personal devices provided a less burdensome way to understand behavior. However, residents and attendings also acknowledged that passively tracked data may offer conflicting information about resident mental health when compared to self-reports, and that the gap between these two types of data may itself be cause for concern. As one attending summarized, a wearable tracker may show a resident is sleeping well even as a residents' perceived sleep quality remains poor—and that perception may lend itself to the resident's felt experience of burnout. This attending continued that a follow-up conversation with the resident might provide the necessary contextual information to interpret these differences, as it might give the attending the chance to help the resident make sense of their experience. This additional step we see as a form of co-construction and collaborative sensemaking, and this anecdote highlights how passively sensed, self-reported, and co-constructed information complement each other to provide a fuller and more actionable approach to measuring resident well-being.

This finding on the role of co-construction also calls attention to epistemological problems in the measurement of well-being and burnout— problems that chart ample ground for future research. In the ubiquitous computing and digital phenotyping communities, researchers often calculate low-level behavioral features from passive sensing data, and then evaluate these features against more traditional mental health rating scales [5, 76, 93]. However, these rating scales often also contain questions on higher-level behavioral and well-being constructs [39, 84]: e.g. the PHQ-9 depression questionnaire asks respondents to rate how often within the last two weeks they experienced “*feeling bad about yourself*” or “*little interest or pleasure in doing things*”. Thus these rating scales may be an inappropriate ground truth for personal sensing, and as demonstrated by Das Swain et al. [20], we should not expect this higher-level information to always align with the lower-level, machine-observable signals on which personal sensing is based. This approach, which has evolved in line with recent trends in machine learning, reifies the quantification-focused and neurobiologically essentialist perspective on mental health, which scholars including Stark [85, 86], Semel [79], and Docherty & Biega [22] highlight is contested within psychological science as the singular approach to understanding the nuanced human experience of well-being.

As an alternative to developing measurements against these ground truths, we urge the research communities around measurement of mental health to expand methodologically, towards a broader conception of well-being. A full understanding of the complex phenomena that is a person's mental health and well-being requires triangulating behavioral signals alongside self-reports, biological indicators, and their cultural, relational and sociotechnical environment. What's more, as scholars including Stark [85] and Ruckenstein and Schull [74] observe, the datafication of health has matured to the point where it has become important to consider how algorithmic categorization itself affects individuals' self-regard. If we understand that the phenomena we are attempting to understand—a person's mental health and well-being—is grounded not only in behavior but also in perception and self-regard, biology and genetics, and their relationships to social and technical infrastructures, we open avenues for collaborative inquiry between researchers trained in computational and interpretive methods. An expanded view of what it means to be well or unwell might also help account for the diversity of human experiences, and ensure we do not force one normative conception of well-being onto society, but rather appreciate well-being in its many forms. Such inquiry might help us all drive towards better measurement of mental health.

5.2 The catch-22 between engagement, utility, and accountability

In addition to the problems in measurement discussed in Section 5.1, we see that our study highlights fundamental tensions in the implementation of measurements in well-being improvement programs. We observe a feedback loop between residents' engagement in data sharing, and the utility they perceive in such data being used for well-being management. As shown in Section 4.2, residents were hesitant to share their personal sensing data without assurances that their participation would result in supervisors acting to improve their well-being. But these possibilities were contingent on mass participation in data sharing. Maintaining any illusion of privacy also depended on mass participation: as Section 4.3 highlights, anonymity was more likely if a sufficiently large pool of residents participated in the program. Yet, at the same time, anonymity was seen to decrease the potential utility of the program for individual-level well-being improvement, since supervisors would have no way to identify which individuals, precisely, were struggling relative to others.

This conundrum is tightly related to a common issue in technology-mediated health programs more broadly: people must engage in these programs to experience their benefits, yet they are often reluctant to engage before they perceive they would benefit. To address this, remote measurement programs often use social incentives to increase motivation for engagement, including social sharing, creating competition among program participants, and gamification techniques that tie participation to monetary rewards [71, 82]. These engagement techniques are integral to many workplace well-being programs [18], but their use is often dependent on the sensitivity of the information shared—e.g. workers may be more willing to share step counts versus detailed data on their sleep. Thus these types of incentives may not be ideal within a residency, where workers have limited bandwidth for well-being management [2], or in workplaces where there is pervasive stigma towards seeking mental health support [16, 58, 99].

Sharing health data with supervisors who can contextualize the information and encourage sustained use may be a more promising incentive for engagement that simultaneously demonstrates utility [6, 95]. But our study shows this source of incentive to engage is absent in our context: as shown in Section 4.2, residents were skeptical that personal sensing data would incentivize supervisory attendings and program directors to act to improve their well-being. Supervising attendings, for their part, were reluctant to take accountability for residents' burnout, stating they were not a "caregiver", and that a caregiving relationship is "*different from the role that we have with residents*" (A03). Residents in our study also voiced that ample prior experience had convinced them program-level change would be hard to come by.

Taken together, if residents' engagement is contingent upon their program's accountability, how do we create incentives for programs to rise to the occasion? For residency programs, one would think that the ACGME well-being competency could create such an incentive [4], but this assumption may be flawed. For example, despite the ACGME imposing a work limit of 80 hours per week [27], many residents continue to work beyond this limit [23, 40], suggesting that the duty hours reported to the ACGME are inaccurate. There is a possible future where the ACGME acts as a third-party auditor and uses personal sensing as an "objective" measurement tool to directly track residents' hours worked and well-being. This direct line might have the effect of forcing transparency and accountability in a medical culture that has encouraged overwork. Of course, resident protections would need to be in place if the ACGME penalizes programs as a result of data shared, and sharing data with the ACGME could be viewed as an additional form of unwanted surveillance into residents' lives. Any data sharing program also risks inviting misinterpretation, reducing mental health and well-being to data streams which may or may not capture the intended constructs of interest [74]. Finally, the ACGME as a third-party auditor still infringes upon residents' right to "self-presentation", which we will further describe in the following section. Despite these

obstacles, a personal sensing equipped ACGME could support worker well-being and incentivize workplace change. Broadening this idea to personal sensing as a whole calls for continued research into closed-loop systems where workers experience direct benefits for engaging in well-being programs driven by personal sensing.

5.3 Privacy, the Quantified Workplace, and “the right to self-presentation”

Beneath the problems in measurement discussed in Section 5.1 and the challenges of accountability in Section 5.2 sits the persistent problem of privacy. As Section 4.3 demonstrates, participants were consistently sensitive to how the implementation of personal sensing in well-being management might rearrange current boundaries in the relationship between residents and their supervisors, and therefore their notions of appropriate information flows, or privacy. Participants’ uncertainties naturally reflect the ambiguity created by the ACGME’s guidance, which mandates that programs attend to resident well-being, but gives little specific guidance on how residencies should go about it. Notably, these problems are not localized to the use of well-being data in workplace contexts: as one resident physician observed, the introduction of such tools into a doctor-patient relationship also constitutes the removal of the patient’s “right to self-presentation” (R10). Following Mohr et al. [60], we consider that personal sensing may be so disruptive a paradigm that to reap its purported benefits—the ability to drive structural improvements for well-being—requires a fundamental reworking of the relation between the data subject, the receiver, and those with the authority to drive structural change. What, then, should we strive for in these new relations? In the absence of established norms for the insertion of well-being management into supervisory or teaching relations, what extant norms should we look to—and how do we ensure the new norms are built to protect against abuses of power?

Our findings hint that extant relational norms in mental healthcare may not be the right analogues for the new norms for workplace well-being management—in part because at present, they do not sufficiently account for the new affordances brought by personal sensing. As our participant describes, monitoring well-being via personal sensing challenges our current conception of a clinician-patient relationship by altering the process of forming and maintaining the therapeutic alliance: “if you already know what’s going on, it interferes with that relationship” (R10). We see this finding as an expression of what the psychotherapy literature describes as the *therapeutic frame*: the set of expectations and boundaries that guides psychotherapeutic practice [28, 45]. Key to the frame is the notion that what happens in the time-bounded interaction of a session is the primary material for the therapy. Our finding opens the possibility that the introduction of external tools, like personal sensing data tracked in between sessions, expands the boundaries of the frame to the detriment of the intense interpersonal exchange necessary for a successful therapeutic alliance: in R10’s words, the patient’s “right to lie”, and the work it takes for a clinician to develop their trust.

Notably, the therapeutic frame in psychotherapy is not the only kind of norm through which clinicians and patients might interact. Approaches to mental healthcare that draw less on the therapy tradition and more on medical treatment, for example remote patient monitoring, may be the extant relations that are more instructive for personal sensing. Yet here, too, we find discomfort. As we have described in Section 5.1, personal sensing can, at worst, organize the care of fundamentally subjective human experiences like well-being into statistical rearranging of observable behaviors, removing the human interpretation critical to care in favor of what is machine-legible. Unlike illnesses solidly grounded in biological mechanisms, mental health and well-being is so fluid, psychosocial, and personal a construct that its indicators are necessarily up to individual interpretation—to monitor sleep and sociality in a potentially depressed person is not to monitor blood glucose in a person with diabetes. Removing a patient’s ability to provide

interpretation—their “*right to self-presentation*”—risks endorsing a fully data-centric perspective on well-being, one that might flatten necessary subjectivity.

If the use of personal sensing in psychodynamic therapy and in medical treatment requires such careful guardrails, the use of personal sensing in workplace well-being requires even further scrutiny. We contend that our findings complicate the idea that personal sensing data might be appropriately shared for well-being management if it is simply provided to a third-party auditor operating under the norms of clinical care, e.g. an independent well-being officer. Such audits have the potential to provide a check on programs or supervisors that are harming residents, if the accountability structure posited in the previous section came to fruition. Still, to enable such changemaking while protecting data subjects from downstream abuse, we need the novel relation between the well-being manager and the data subject to ensure their data are collected and processed in ways that allow for “*self-presentation*” and intervention—e.g. through mechanisms for data subjects to voluntarily disclose specific slices of sensed data, while foreclosing the ability for continuous monitoring, and giving data subjects voice in how their data are interpreted. Such novel relational norms must also emphasize the guardrails of current clinical relationships, e.g. clinician-patient confidentiality. In apprenticeship contexts like the residency program in our study, we emphasize protections for data subjects are even more necessary to create proper balance in the teaching or supervisory relationships that can determine the course of a resident’s life and career.

These discomforts are fruitful and necessary ground for future research not only in the Quantified Workplace, but also broadly in personal sensing for mental health. We urge the ubiquitous computing community to work not only with clinical experts from both medical and allied health disciplines, but also with ethicists, privacy scholars, and researchers in information science, to establish these new relational norms in ways that center the data subject’s agency first and foremost. These efforts might draw lessons from broader work to make data science more trustworthy by incorporating reflexive analyses of power, and better connecting with affected communities [81]. Doing so can help us ensure that the allied fields of psychiatry, psychotherapy, counseling, and workplace well-being can truly use this technology to improve conditions for those who are suffering.

5.4 Limitations & Future Work

Our study used a design provocation to elicit formative insights towards the development of workplace well-being interventions for resident physicians; as such, it faces the inherent limitations of all such methodologies. Our results should not be understood as broadly generalizable to all workers or all healthcare providers. We studied these tensions in the context of one small residency, and concerns around data anonymization may be amplified relative to larger programs. We worked with participants during the fall and winter of 2020 and the spring of 2021, during which the pressures of being a healthcare worker in New York City during the COVID-19 pandemic may have amplified participants’ sensitivities to stress and burnout and, in turn, their reactions to the provocation. Future research should examine well-being and burnout in temporal and geographic contexts impacted by the pandemic in different ways.

There is considerable future work to be done examining these tensions in other contexts entirely, e.g. in surgery residencies or other physician training programs with significant burnout challenges. Across the healthcare workforce, we also see considerable future research addressing burnout in roles that draw less attention than physicians and physicians-in-training, e.g. nursing staff, social workers, home health workers, therapists, orderlies and janitorial staff, and much more. Additionally, while our study examined problems in vulnerability and privacy for residents in their roles as workers and trainees, we did not take the additional step of studying what other identity-related factors might contribute to participants’ reactions to personal sensing. We leave a fully intersectional approach to future work. Future research might also consider triangulating

our results against results from alternate methodologies, e.g. randomized trials of intervention programs, for a fuller picture of the tensions at play.

6 CONCLUSION

We report a qualitative study of how personal sensing might be used in a psychiatric residency program in the United States for the management of medical residents' well-being. Our findings highlight that the use of the data for this purpose raised three fundamental tensions: (1) how best to measure well-being and burnout; (2) how best to hold residency programs accountable for using this knowledge to actually improve resident well-being; (3) navigating privacy when sharing behavioral and well-being data between worker and workplace. We discuss paths forward for future work establishing the new caring relations that might ensure personal sensing is leveraged to truly improve well-being, in the workplace context and beyond.

ACKNOWLEDGMENTS

First, we thank all of our anonymous participants for lending their time, perspectives, and expertise to our study. Funding for this work was provided by National Institute of Mental Health (NIMH) Grant No. P50 MH-113838-01. DA is supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under Grant No. DGE-2139899. DCM was supported by National Institute of Mental Health (NIMH) Grant No. R01-MH111610. ET was supported by a Digital Life Initiative (DLI) Doctoral Fellowship, and EM was additionally supported by a DLI Postdoctoral Fellowship. We thank the DLI for vibrant intellectual community. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, NIMH, or DLI.

REFERENCES

- [1] 2021. ACGME Home. <https://acgme.org/>
- [2] Alexandra L. Aaronson, Katherine Backes, Gaurava Agarwal, Joshua L. Goldstein, and Joan Anzia. 2018. Mental Health During Residency Training: Assessing the Barriers to Seeking Care. *Academic Psychiatry* 42, 4 (Aug. 2018), 469–472. <https://doi.org/10.1007/s40596-017-0881-3>
- [3] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep Learning with Differential Privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Oct. 2016), 308–318. <https://doi.org/10.1145/2976749.2978318> arXiv: 1607.00133.
- [4] ACGME Common Program and Requirements Phase 1 Task Force c. 2017. ACGME Common Program Requirements Section VI with Background and Intent. (2017), 20.
- [5] Daniel A. Adler, Vincent W.-S. Tseng, Gengmo Qi, Joseph Scarpa, Srijan Sen, and Tanzeem Choudhury. 2021. Identifying Mobile Sensing Indicators of Stress-Resilience. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (June 2021), 51:1–51:32. <https://doi.org/10.1145/3463528>
- [6] Kevin Anderson, Oksana Burford, and Lynne Emmerton. 2016. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. *PLOS ONE* 11, 5 (May 2016), e0156164. <https://doi.org/10.1371/journal.pone.0156164> Publisher: Public Library of Science.
- [7] American Medical Association. 2021. American Medical Association Coping with COVID-19 for Caregivers Survey. (2021), 47.
- [8] DeWitt C. Jr Baldwin, Steven R. Daugherty, Ray Tsai, and Michael J. Jr Scotti. 2003. A National Survey of Residents' Self-Reported Work Hours: Thinking Beyond Specialty. *Academic Medicine* 78, 11 (Nov. 2003), 1154–1163. https://journals.lww.com/academicmedicine/Fulltext/2003/11000/A_National_Survey_of_Residents_Self_Reported_Work.18.aspx
- [9] K. L. Barriball and A. While. 1994. Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing* 19, 2 (Feb. 1994), 328–335. <https://doi.org/10.1111/j.1365-2648.1994.tb01088.x>
- [10] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1077–1086. <https://doi.org/10.1145/1240624.1240789>
- [11] Philip G. Boysen. 2013. Just Culture: A Foundation for Balanced Accountability and Patient Safety. *The Ochsner Journal* 13, 3 (2013), 400–406. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3776518/>

- [12] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 18, 3 (July 2021), 328–352. <https://doi.org/10.1080/14780887.2020.1769238> Publisher: Routledge _eprint: <https://doi.org/10.1080/14780887.2020.1769238>.
- [13] Daniel F. Brennan, Salvatore Silvestri, Joanne Y. Sun, and Linda Papa. 2007. Progression of Emergency Medicine Resident Productivity. *Academic Emergency Medicine* 14, 9 (2007), 790–794. <https://doi.org/10.1111/j.1553-2712.2007.tb02353.x> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1553-2712.2007.tb02353.x>.
- [14] Kiran R. Busireddy, Jonathan A. Miller, Kathleen Ellison, Vicky Ren, Rehan Qayyum, and Mukta Panda. 2017. Efficacy of Interventions to Reduce Resident Physician Burnout: A Systematic Review. *Journal of Graduate Medical Education* 9, 3 (June 2017), 294–301. <https://doi.org/10.4300/JGME-D-16-00372.1>
- [15] Davide Carneiro, Paulo Novais, Juan Carlos Augusto, and Nicola Payne. 2019. New Methods for Stress Assessment and Monitoring at the Workplace. *IEEE Transactions on Affective Computing* 10, 2 (April 2019), 237–254. <https://doi.org/10.1109/TAFFC.2017.2699633> Conference Name: IEEE Transactions on Affective Computing.
- [16] Claudia Center, Miriam Davis, Thomas Detre, Daniel E. Ford, Wendy Hansbrough, Herbert Hendin, John Laszlo, David A. Litts, John Mann, Peter A. Mansky, Robert Michels, Steven H. Miles, Roy Proujansky, Charles F. Reynolds III, and Morton M. Silverman. 2003. Confronting Depression and Suicide in PhysiciansA Consensus Statement. *JAMA* 289, 23 (June 2003), 3161–3166. <https://doi.org/10.1001/jama.289.23.3161>
- [17] Emil Chiauzzi and Paul Wicks. 2021. Beyond the Therapist’s Office: Merging Measurement-Based Care and Digital Medicine in the Real World. *Digital Biomarkers* 5, 2 (2021), 176–182. <https://doi.org/10.1159/000517748> Publisher: Karger Publishers.
- [18] Chia-Fang Chung, Nanna Gorm, Irina A. Shklovski, and Sean Munson. 2017. Finding the Right Fit: Understanding Health Tracking in Workplace Wellness Programs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*. Association for Computing Machinery, New York, NY, USA, 4875–4886. <https://doi.org/10.1145/3025453.3025510>
- [19] Sondra Corgan, Abigail F. Winkel, Rebekah Sugarman, and John Q. Young. 2021. From Burnout to Wholehearted Engagement: A Qualitative Exploration of Psychiatry Residents’ Experience of Stress. *Academic Medicine* Publish Ahead of Print (Feb. 2021). <https://doi.org/10.1097/ACM.0000000000003912>
- [20] Vedant Das Swain, Victor Chen, Shrija Mishra, Stephen M. Mattingly, Gregory D. Abowd, and Munmun De Choudhury. 2022. Semantic Gap in Predicting Mental Wellbeing through Passive Sensing. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–16. <https://doi.org/10.1145/3491102.3502037>
- [21] Sidney Medeiros de Oliveira, Luiz Vinicius de Alcantara Sousa, Maria do Socorro Vieira Gadelha, and Vânia Barbosa do Nascimento. 2019. Prevention Actions of Burnout Syndrome in Nurses: An Integrating Literature Review. *Clinical Practice and Epidemiology in Mental Health : CP & EMH* 15 (March 2019), 64–73. <https://doi.org/10.2174/1745017901915010064>
- [22] Niall Docherty and Asia J. Biega. 2022. (Re)Politicizing Digital Well-Being: Beyond User Engagements. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–13. <https://doi.org/10.1145/3491102.3501857>
- [23] Brian C. Drolet, Matthew Schwede, Kenneth D. Bishop, and Staci A. Fischer. 2013. Compliance and Falsification of Duty Hours: Reports From Residents and Program Directors. *Journal of Graduate Medical Education* 5, 3 (Sept. 2013), 368–373. <https://doi.org/10.4300/JGME-D-12-00375.1>
- [24] Amy C. Edmondson. 2012. *Teaming: How Organizations Learn, Innovate, and Compete in the Knowledge Economy*. John Wiley & Sons. Google-Books-ID: wlnsR9ib5cC.
- [25] Ilker Etikan, Sulaiman Abubakar Musa, and Rukayya Sunusi Alkassim. 2015. Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics* 5, 1 (Dec. 2015), 1. <https://doi.org/10.11648/j.ajtas.20160501.11> Number: 1 Publisher: Science Publishing Group.
- [26] Eva Garrosa, Bernardo Moreno-Jiménez, Youxin Liang, and José Luis González. 2008. The relationship between socio-demographic variables, job stressors, burnout, and hardy personality in nurses: An exploratory study. *International Journal of Nursing Studies* 45, 3 (March 2008), 418–427. <https://doi.org/10.1016/j.ijnurstu.2006.09.003>
- [27] Dmitri V. Gelfand, Yale D. Podnos, Joseph C. Carmichael, Darin J. Saltzman, Samuel E. Wilson, and Russell A. Williams. 2004. Effect of the 80-Hour Workweek on Resident Burnout. *Archives of Surgery* 139, 9 (Sept. 2004), 933–940. <https://doi.org/10.1001/archsurg.139.9.933> Publisher: American Medical Association.
- [28] Anne Gray. 2013. *An Introduction to the Therapeutic Frame*. Routledge. Google-Books-ID: ee3eAQAQBAJ.
- [29] Constance Guille, Heather Speller, Rachel Laff, C. Neill Epperson, and Srijan Sen. 2010. Utilization and Barriers to Mental Health Services Among Depressed Medical Interns: A Prospective Multisite Study. *Journal of Graduate Medical Education* 2, 2 (June 2010), 210–214. <https://doi.org/10.4300/JGME-D-09-00086.1>
- [30] Clare Hammonds, Jasmine Kerrissey, and Donald Tomaskovic-Devey. 2020. Stressed, Unsafe, and Insecure: Essential Workers Need A New, New Deal. (2020), 38.

- [31] Michael Hockenhull and Marisa Leavitt Cohn. 2021. Speculative Data Work & Dashboards: Designing Alternative Data Visions. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 1–31. <https://doi.org/10.1145/3434173>
- [32] Christopher Ingraham. 2018. Analysis | An insurance company wants you to hand over your Fitbit data so it can make more money. Should you? *Washington Post* (Sept. 2018). <https://www.washingtonpost.com/business/2018/09/25/an-insurance-company-wants-you-hand-over-your-fitbit-data-so-they-can-make-more-money-should-you/>
- [33] Thomas R. Insel. 2017. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA* 318, 13 (Oct. 2017), 1215–1216. <https://doi.org/10.1001/jama.2017.11295> Publisher: American Medical Association.
- [34] Rebecca Jeanmonod, Donald Jeanmonod, and Ryan Ngiam. 2008. Resident productivity: does shift length matter? *The American Journal of Emergency Medicine* 26, 7 (Sept. 2008), 789–791. <https://doi.org/10.1016/j.ajem.2007.10.037>
- [35] Joshua W. Joseph, Daniel J. Henning, Connie S. Strouse, David T. Chiu, Larry A. Nathanson, and Leon D. Sanchez. 2017. Modeling Hourly Resident Productivity in the Emergency Department. *Annals of Emergency Medicine* 70, 2 (Aug. 2017), 185–190.e6. <https://doi.org/10.1016/j.annemergmed.2016.11.020>
- [36] Allen Kachalia and David M. Studdert. 2004. Professional Liability Issues in Graduate Medical Education. *JAMA* 292, 9 (Sept. 2004), 1051–1056. <https://doi.org/10.1001/jama.292.9.1051>
- [37] David A. Kalmbach, Yu Fang, J. Todd Arnedt, Amy L. Cochran, Patricia J. Deldin, Adam I. Kaplin, and Srijan Sen. 2018. Effects of Sleep, Physical Activity, and Shift Work on Daily Mood: a Prospective Mobile Monitoring Study of Medical Interns. *Journal of General Internal Medicine* 33, 6 (June 2018), 914–920. <https://doi.org/10.1007/s11606-018-4373-2>
- [38] Harmanpreet Kaur, Alex C. Williams, Daniel McDuff, Mary Czerwinski, Jaime Teevan, and Shamsi T. Iqbal. 2020. Optimizing for Happiness and Productivity: Modeling Opportune Moments for Transitions and Breaks at Work. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–15. <https://doi.org/10.1145/3313831.3376817>
- [39] Kurt Kroenke, Robert L. Spitzer, and Janet B W Williams. 2001. The PHQ-9. *Journal of General Internal Medicine* 16, 9 (Sept. 2001), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- [40] Christopher P. Landrigan, Laura K. Barger, Brian E. Cade, Najib T. Ayas, and Charles A. Czeisler. 2006. Interns' Compliance With Accreditation Council for Graduate Medical Education Work-Hour Limits. *JAMA* 296, 9 (Sept. 2006), 1063–1070. <https://doi.org/10.1001/jama.296.9.1063>
- [41] Stephen M. Lawrie, Sue Fletcher-Watson, Heather C. Whalley, and Andrew M. McIntosh. 2019. Predicting major mental illness: ethical and practical considerations. *BJPsych Open* 5, 2 (March 2019), e30. <https://doi.org/10.1192/bjo.2019.11>
- [42] Mark Linzer, Sara Poplau, Ellie Grossman, Anita Varkey, Steven Yale, Eric Williams, Lanis Hicks, Roger L. Brown, Jill Wallock, Diane Kohnhorst, and Michael Barbouche. 2015. A Cluster Randomized Trial of Interventions to Improve Work Conditions and Clinician Burnout in Primary Care: Results from the Healthy Work Place (HWP) Study. *Journal of General Internal Medicine* 30, 8 (Aug. 2015), 1105–1111. <https://doi.org/10.1007/s11606-015-3235-4>
- [43] Craig R. Littler. 1978. Understanding Taylorism. *The British Journal of Sociology* 29, 2 (1978), 185–202. <https://doi.org/10.2307/589888> Publisher: [Wiley, London School of Economics and Political Science, London School of Economics].
- [44] Steven W. Lockley, John W. Cronin, Erin E. Evans, Brian E. Cade, Clark J. Lee, Christopher P. Landrigan, Jeffrey M. Rothschild, Joel T. Katz, Craig M. Lilly, Peter H. Stone, Daniel Aeschbach, and Charles A. Czeisler. 2004. Effect of Reducing Interns' Weekly Work Hours on Sleep and Attentional Failures. *New England Journal of Medicine* 351, 18 (Oct. 2004), 1829–1837. <https://doi.org/10.1056/NEJMoa041404> Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMoa041404>
- [45] Maria Luca. 2004. *The Therapeutic Frame in the Clinical Context: Integrative Perspectives*. Psychology Press. Google-Books-ID: 35s7QSacerYC.
- [46] Brian P. Lucas, William E. Trick, Arthur T. Evans, Benjamin Mba, Jennifer Smith, Krishna Das, Peter Clarke, Anita Varkey, Suja Mathew, and Robert A. Weinstein. 2012. Effects of 2- vs 4-Week Attending Physician Inpatient Rotations on Unplanned Patient Revisits, Evaluations by Trainees, and Attending Physician Burnout: A Randomized Trial. *JAMA* 308, 21 (Dec. 2012), 2199–2207. <https://doi.org/10.1001/jama.2012.36522>
- [47] Gloria Mark, Shamsi T. Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. 2016. Email Duration, Batching and Self-interruption: Patterns of Email Use on Productivity and Stress. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1717–1728. <https://doi.org/10.1145/2858036.2858262>
- [48] Afra Mashhadi, Fahim Kawsar, Akhil Mathur, Casey Dugan, and N. Shami. 2016. Let's Talk About the Quantified Workplace. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion - CSCW '16 Companion*. ACM Press, San Francisco, California, USA, 522–528. <https://doi.org/10.1145/2818052.2855514>
- [49] Douglas A. Mata, Marco A. Ramos, Narinder Bansal, Rida Khan, Constance Guille, Emanuele Di Angelantonio, and Srijan Sen. 2015. Prevalence of Depression and Depressive Symptoms Among Resident Physicians: A Systematic

- Review and Meta-analysis. *JAMA* 314, 22 (Dec. 2015), 2373–2383. <https://doi.org/10.1001/jama.2015.15845> Publisher: American Medical Association.
- [50] D. A. Matthews, D. C. Classen, J. L. Willms, and J. P. Cotton. 1988. A program to help interns cope with stresses in an internal medicine residency. *Academic Medicine* 63, 7 (July 1988), 539–47. https://journals.lww.com/academicmedicine/Abstract/1988/07000/A_program_to_help_interns_cope_with_stresses_in_an.5.aspx
- [51] Stephen M. Mattingly, Julie M. Gregg, Pino Audia, Ayse Elvan Bayraktaroglu, Andrew T. Campbell, Nitesh V. Chawla, Vedant Das Swain, Munmun De Choudhury, Sidney K. D’Mello, Anind K. Dey, Ge Gao, Krithika Jagannath, Kaifeng Jiang, Suwen Lin, Qiang Liu, Gloria Mark, Gonzalo J. Martinez, Kizito Masaba, Shayan Mirjafari, Edward Moskal, Raghu Mulukutla, Kari Nies, Manikanta D. Reddy, Pablo Robles-Granda, Koustuv Saha, Anusha Sirigiri, and Aaron Striegel. 2019. The Tesseract Project: Large-Scale, Longitudinal, *In Situ*, Multimodal Sensing of Information Workers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–8. <https://doi.org/10.1145/3290607.3299041>
- [52] Laura W. McCray, Peter F. Cronholm, Hillary R. Bogner, Joseph J. Gallo, and Richard A. Neill. 2008. Resident Physician Burnout: Is There Hope? *Family medicine* 40, 9 (Oct. 2008), 626–632. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2903755/>
- [53] Jack D. McCue and Carol L. Sachs. 1991. A Stress Management Workshop Improves Residents’ Coping Skills. *Archives of Internal Medicine* 151, 11 (Nov. 1991), 2273–2277. <https://doi.org/10.1001/archinte.1991.00400110117023>
- [54] Nora McDonald and Andrea Forte. 2020. The Politics of Privacy Theories: Moving from Norms to Vulnerabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. <https://doi.org/10.1145/3313831.3376167>
- [55] Stanford Medicine. 2022. Rotations. <https://medicine.stanford.edu/residency/program-overview/rotations.html>
- [56] Katherine A. Meese, Alejandra Colón-López, Jasvinder A. Singh, Greer A. Burkholder, and David A. Rogers. 2021. Healthcare is a Team Sport: Stress, Resilience, and Correlates of Well-Being Among Health System Employees in a Crisis. *Journal of Healthcare Management* 66, 4 (July 2021), 304–322. <https://doi.org/10.1097/JHM-D-20-00288>
- [57] Lori Mihalich-Levin. 2012. Residents Should Know About Medicare Funding for Their Training Programs. *Academic Medicine* 87, 7 (July 2012), 834. <https://doi.org/10.1097/ACM.0b013e318257d09a>
- [58] Steven H. Miles. 1998. A Challenge to Licensing Boards: The Stigma of Mental Illness. *JAMA* 280, 10 (Sept. 1998), 865. <https://doi.org/10.1001/jama.280.10.865>
- [59] Geoffrey Miller. 2012. The Smartphone Psychology Manifesto. *Perspectives on Psychological Science* 7, 3 (May 2012), 221–237. <https://doi.org/10.1177/1745691612441215> Publisher: SAGE Publications Inc.
- [60] David C. Mohr, Katie Shilton, and Matthew Hotopf. 2020. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. *npj Digital Medicine* 3, 1 (March 2020), 1–2. <https://doi.org/10.1038/s41746-020-0251-5> Number: 1 Publisher: Nature Publishing Group.
- [61] David C. Mohr, Mi Zhang, and Stephen M. Schueller. 2017. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual review of clinical psychology* 13 (May 2017), 23–47. <https://doi.org/10.1146/annurev-clinpsy-032816-044949>
- [62] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K. D’Mello, Munmun De Choudhury, Gregory D. Abowd, and Thomas Plötz. 2019. Prediction of Mood Instability with Passive Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 1–21. <https://doi.org/10.1145/3351233>
- [63] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. AffectiveSpotlight: Facilitating the Communication of Affective Responses from Audience Members during Online Presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. <http://doi.org/10.1145/3411764.3445235>
- [64] Brendan Murphy. 2022. 6 things medical students should know about physician compensation. <https://www.ama-assn.org/residents-students/resident-student-finance/6-things-medical-students-should-know-about-physician>
- [65] Peter M. Murray. 1990. The History of Informed Consent. *The Iowa Orthopaedic Journal* 10 (1990), 104–109. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2328798/>
- [66] Jennifer Nicholas, Katie Shilton, Stephen M. Schueller, Elizabeth L. Gray, Mary J. Kwasny, and David C. Mohr. 2019. The Role of Data Type and Recipient in Individuals’ Perspectives on Sharing Passively Collected Smartphone Data for Mental Health: Cross-Sectional Questionnaire Study. *JMIR mHealth and uHealth* 7, 4 (April 2019), e12578. <https://doi.org/10.2196/12578> Company: JMIR mHealth and uHealth Distributor: JMIR mHealth and uHealth Institution: JMIR mHealth and uHealth Label: JMIR mHealth and uHealth Publisher: JMIR Publications Inc., Toronto, Canada.
- [67] Helen Nissenbaum. 2004. PRIVACY AS CONTEXTUAL INTEGRITY. *Washington Law Review* 79 (2004), 39.
- [68] National Academy of Medicine. 2018. Valid and Reliable Survey Instruments to Measure Burnout, Well-Being, and Other Work-Related Dimensions. <https://nam.edu/valid-reliable-survey-instruments-measure-burnout-well-work-related-dimensions/>

- [69] Doug Oman, John Hedberg, and Carl E. Thoresen. 2006. Passage meditation reduces perceived stress in health professionals: a randomized, controlled trial. *Journal of Consulting and Clinical Psychology* 74, 4 (Aug. 2006), 714–719. <https://doi.org/10.1037/0022-006X.74.4.714>
- [70] Christopher S. Parshuram, Andre C.K.B. Amaral, Niall D. Ferguson, G. Ross Baker, Edward E. Etchells, Virginia Flintoft, John Granton, Lorelei Lingard, Haresh Kirpalani, Sangeeta Mehta, Harvey Moldofsky, Damon C. Scales, Thomas E. Stewart, Andrew R. Willan, and Jan O. Friedrich. 2015. Patient safety, resident well-being and continuity of care with different resident duty schedules in the intensive care unit: a randomized trial. *CMAJ: Canadian Medical Association Journal* 187, 5 (March 2015), 321–329. <https://doi.org/10.1503/cmaj.140752>
- [71] Wei Peng, Shupeí Yuan, and Bree E. Holtz. 2016. Exploring the Challenges and Opportunities of Health Mobile Apps for Individuals with Type 2 Diabetes Living in Rural Communities. *Telemedicine and e-Health* 22, 9 (Sept. 2016), 733–738. <https://doi.org/10.1089/tmj.2015.0180> Publisher: Mary Ann Liebert, Inc., publishers.
- [72] Haroon Rashid, Sanjana Mendu, Katharine E. Daniel, Miranda L. Beltzer, Bethany A. Teachman, Mehdi Boukhechba, and Laura E. Barnes. 2020. Predicting Subjective Measures of Social Anxiety from Sparsely Collected Mobile Sensor Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (Sept. 2020), 1–24. <https://doi.org/10.1145/3411823>
- [73] Pablo Robles-Granda, Suwen Lin, Xian Wu, Gonzalo J. Martinez, Stephen M. Mattingly, Edward Moskal, Aaron Striegel, Nitesh V. Chawla, Sidney D’Mello, Julie Gregg, Kari Nies, Gloria Mark, Ted Grover, Andrew T. Campbell, Shayan Mirjafari, Koustuv Saha, Munmun De Choudhury, and Anind K. Dey. 2021. Jointly Predicting Job Performance, Personality, Cognitive Ability, Affect, and Well-Being. *IEEE Computational Intelligence Magazine* 16, 2 (May 2021), 46–61. <https://doi.org/10.1109/MCI.2021.3061877> Conference Name: IEEE Computational Intelligence Magazine.
- [74] Minna Ruckenstein and Natasha Dow Schüll. 2017. The Datafication of Health. *Annual Review of Anthropology* 46, 1 (2017), 261–278. <https://doi.org/10.1146/annurev-anthro-102116-041244> _eprint: <https://doi.org/10.1146/annurev-anthro-102116-041244>.
- [75] Sohrab Saeb, Thaddeus R. Cybulski, Stephen M. Schueller, Konrad P. Kording, and David C. Mohr. 2017. Scalable Passive Sleep Monitoring Using Mobile Phones: Opportunities and Obstacles. *Journal of Medical Internet Research* 19, 4 (2017), e118. <https://doi.org/10.2196/jmir.6821>
- [76] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of Medical Internet Research* 17, 7 (July 2015). <https://doi.org/10.2196/jmir.4273>
- [77] Koustuv Saha, John Torous, Emre Kiciman, and Munmun De Choudhury. 2021. Understanding Side Effects of Antidepressants: Large-scale Longitudinal Study on Social Media Data. *JMIR Mental Health* 8, 3 (March 2021), e26589. <https://doi.org/10.2196/26589>
- [78] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3411764.3445615>
- [79] Beth M. Semel. 2022. Listening Like a Computer: Attentional Tensions and Mechanized Care in Psychiatric Digital Phenotyping. *Science, Technology, & Human Values* 47, 2 (March 2022), 266–290. <https://doi.org/10.1177/01622439211026371> Publisher: SAGE Publications Inc.
- [80] Tait D. Shanafelt. 2002. Burnout and self-reported patient care in an internal medicine residency program. *Annals of Internal Medicine* (2002), 358–367.
- [81] Katie Shilton, Emanuel Moss, Sarah A. Gilbert, Matthew J. Bietz, Casey Fiesler, Jacob Metcalf, Jessica Vitak, and Michael Zimmer. 2021. Excavating awareness and power in data science: A manifesto for trustworthy pervasive data research. *Big Data & Society* 8, 2 (July 2021), 20539517211040759. <https://doi.org/10.1177/20539517211040759> Publisher: SAGE Publications Ltd.
- [82] Sara Simblett, Ben Greer, Faith Matcham, Hannah Curtis, Ashley Polhemus, José Ferrão, Peter Gamble, and Til Wykes. 2018. Barriers to and Facilitators of Engagement With Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings. *Journal of Medical Internet Research* 20, 7 (July 2018), e10480. <https://doi.org/10.2196/10480> Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.
- [83] Joanne Sinai, Richard G. Tiberius, Janet de Groot, Aileen Brunet, and Peter Voore. 2001. Developing a Training Program to Improve Supervisor-Resident Relationships, Step 1: Defining the Types of Issues. *Teaching and Learning in Medicine* 13, 2 (April 2001), 80–85. https://doi.org/10.1207/S15328015TLM1302_2 Publisher: Routledge _eprint: https://doi.org/10.1207/S15328015TLM1302_2.
- [84] Robert L. Spitzer, Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe. 2006. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine* 166, 10 (May 2006), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>

- [//doi.org/10.1001/archinte.166.10.1092](https://doi.org/10.1001/archinte.166.10.1092) Publisher: American Medical Association.
- [85] Luke Stark. 2018. Algorithmic psychometrics and the scalable subject. *Social Studies of Science* 48, 2 (April 2018), 204–231. <https://doi.org/10.1177/0306312718772094> Publisher: SAGE Publications Ltd.
 - [86] Luke Stark and Jesse Hoey. 2021. The Ethics of Emotion in Artificial Intelligence Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 782–793. <https://doi.org/10.1145/3442188.3445939>
 - [87] Madeline R. Sterling, Emily Tseng, Anthony Poon, Jacklyn Cho, Ariel C. Avgar, Lisa M. Kern, Claire K. Ankuda, and Nicola Dell. 2020. Experiences of Home Health Care Workers in New York City During the Coronavirus Disease 2019 Pandemic: A Qualitative Analysis. *JAMA Internal Medicine* (Aug. 2020). <https://doi.org/10.1001/jamainternmed.2020.3930>
 - [88] Melanie Swan. 2013. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data* 1, 2 (June 2013), 85–99. <https://doi.org/10.1089/big.2012.0002> Publisher: Mary Ann Liebert, Inc., publishers.
 - [89] Latanya Sweeney. 2002. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 5 (2002). <https://doi.org/10.1142/S021848850200165X>
 - [90] Sara Ann Taylor, Natasha Jaques, Ehimwenma Nosakhare, Akane Sano, and Rosalind Picard. 2017. Personalized Multitask Learning for Predicting Tomorrow’s Mood, Stress, and Health. *IEEE Transactions on Affective Computing* (2017), 1–1. <https://doi.org/10.1109/TAFFC.2017.2784832>
 - [91] Larissa R. Thomas, Rita Nguyen, Arianne Teherani, Catherine R. Lucey, and Elizabeth Harleman. 2020. Designing Well-Being: Using Design Thinking to Engage Residents in Developing Well-Being Interventions. *Academic Medicine* 95, 7 (July 2020), 1038–1042. <https://doi.org/10.1097/ACM.00000000000003243>
 - [92] Dnika J. Travis, Erica Leeanne Lizano, and Michàlle E. Mor Barak. 2016. ‘I’m So Stressed’: A Longitudinal Model of Stress, Burnout and Engagement among Social Workers in Child Welfare Settings. *The British Journal of Social Work* 46, 4 (June 2016), 1076–1095. <https://doi.org/10.1093/bjsw/bct205>
 - [93] Vincent W.-S. Tseng, Akane Sano, Dror Ben-Zeev, Rachel Brian, Andrew T. Campbell, Marta Hauser, John M. Kane, Emily A. Scherer, Rui Wang, Weichen Wang, Hongyi Wen, and Tanzeem Choudhury. 2020. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. *Scientific Reports* 10, 1 (Sept. 2020), 15100. <https://doi.org/10.1038/s41598-020-71689-1> Number: 1 Publisher: Nature Publishing Group.
 - [94] Ospina-Kammerer V and Figley Cr. 2003. An evaluation of the Respiratory One Method (ROM) in reducing emotional exhaustion among family physician residents. *International Journal of Emergency Mental Health* 5, 1 (Jan. 2003), 29–32. <http://europepmc.org/article/med/12722487>
 - [95] Harshvardhan Vathsangam and Gaurav S. Sukhatme. 2014. Using phone-based activity monitors to promote physical activity in older adults: A pilot study. In *2014 IEEE Healthcare Innovation Conference (HIC)*. 42–47. <https://doi.org/10.1109/HIC.2014.7038870>
 - [96] Effy Vayena and Alessandro Blasimme. 2017. Biomedical Big Data: New Models of Control Over Access, Use and Governance. *Journal of Bioethical Inquiry* 14, 4 (Dec. 2017), 501–513. <https://doi.org/10.1007/s11673-017-9809-6>
 - [97] Christopher J. Weight, Jacob L. Sellon, Collette R. Lessard-Anderson, Tait D. Shanafelt, Kerry D. Olsen, and Edward R. Laskowski. 2013. Physical Activity, Quality of Life, and Burnout Among Physician Trainees: The Effect of a Team-Based, Incentivized Exercise Program. *Mayo Clinic Proceedings* 88, 12 (Dec. 2013), 1435–1442. <https://doi.org/10.1016/j.mayocp.2013.09.010>
 - [98] Colin P. West, Liselotte N. Dyrbye, Jeff T. Rabatin, Tim G. Call, John H. Davidson, Adamarie Multari, Susan A. Romanski, Joan M. Henriksen Hellyer, Jeff A. Sloan, and Tait D. Shanafelt. 2014. Intervention to Promote Physician Well-being, Job Satisfaction, and Professionalism: A Randomized Clinical Trial. *JAMA Internal Medicine* 174, 4 (April 2014), 527–533. <https://doi.org/10.1001/jamainternmed.2013.14387>
 - [99] Leslie A Wimsatt. 2015. Predictors of Depression Stigma in Medical Students. *Am J Prev Med* (2015), 12.
 - [100] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology* 10, 2 (Jan. 2019), 12:1–12:19. <https://doi.org/10.1145/3298981>
 - [101] Camellia Zakaria, Rajesh Balan, and Youngki Lee. 2019. StressMon: Scalable Detection of Perceived Stress and Depression Using Passive Sensing of Changes in Work Routines and Group Interactions. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 37:1–37:29. <https://doi.org/10.1145/3359139>

A EXPLORATORY INTERVIEW GUIDE

We would like to start with a few questions to learn a little bit about burnout.

- (1) First, in your own words, how would you define burnout?
- (2) In broad strokes, without telling us anything specific or identifying, what's an example of a case where you have encountered burnout – in yourself, in a colleague? What was the situation?
- (3) What is the difference between burnout and demoralization or depression?
- (4) Again, without telling us anything specific or identifying, have you noticed burnout in people you work with? Co-residents, supervisors, training program staff, others in your work environment?
 - (a) Probe: How did you know?
 - (b) Probe: When you've seen someone that's burnt out, what are they likely to do? How are they likely to behave?
- (5) If you were taking care of a patient who showed signs of burnout, what kinds of things would you track to see if they were getting better or worse?
 - (a) Probe: Do you think this is adequately captured in the screens you regularly use?
 - (b) Probe: Are there any specific assessments you use for burnout?
- (6) What scales or measures would you use to assess resident well-being (e.g., PHQ-9/GAD-7)?
- (7) Are there signs of burnout that would be specific to residents?
- (8) How frequently do you believe one would need to be assessed for these measures, in order to get an accurate picture of resident well-being?
- (9) Would you respond to a smartphone survey that asked you for these measures of stress, burnout, emotional well-being, and recorded them, at this frequency? Is there a better method of delivery?

We're also interested in your thoughts and experiences around technologies you may or may not use to assess your own health and well-being.

- (1) Do you have a wearable (e.g. Fitbit, Apple Watch), or any device other than a smartphone, that you use to track daily behaviors (e.g. steps, exercise, heart rate, sleep)?
- (2) If so, what device(s)?
 - (a) Do you integrate the information together (e.g. using Apple HealthKit or Google Fit)? If so, when you look at it, what do you specifically focus on?
 - (b) In what ways do you use this information? Have you ever used it to change your behavior?
- (3) If you do not use this information, why not?
- (4) How would you feel if you had a smartphone application, or wearable, that tracked your behaviors (e.g. sleep, physical activity, phone usage) specifically to measure stress, burnout and psychological well-being?
 - (a) Probe: Which behavioral information, if any, would be helpful for you, to actively prevent or manage burnout?
- (5) How long would you be comfortable with this data being collected?
 - (a) How long do you think you'd actually comply with data collection?
 - (b) Would compliance be device dependent (e.g. smartphone vs. wearable)?
- (6) If we wanted to understand when burnout occurs throughout a residency, how long do you think we should collect this data to get an accurate representation?
 - (a) Are there specific time periods within your residency (e.g. rotations) that contribute more to burnout?
 - (b) When do these occur?

- (7) Would you want data collected on your well-being to be shared with a therapist, or other clinicians that are monitoring your health?
 - (a) Would you want it to be shared with anyone else?
 - (b) What would you want them to do with this information?
- (8) What care settings do you think would be most benefited by access to this information? (example - collaborative care setting, psychiatrist, therapist, primary care doctor, other)
How so? (example - promote communication/information sharing among providers, ease of communication with you, could react to deterioration more quickly)
- (9) Under what circumstances would you be willing to have information on your behaviors, burnout and psychological well-being shared with your supervisors, including the residency program director?
 - (a) Probe: What if it were anonymized?
 - (b) Probe: What if it were aggregated?
 - (c) Probe: What level of detail would you be comfortable sharing?
 - (d) Probe: What would you want them to do with this information?
- (10) Is there anything else you'd like to tell us today?

B EXPLORATORY INTERVIEWS CODEBOOK

Theme / Code	
Burnout definitions	Burnout measurement
Burnout as inadequacy	Frequency
Burnout as jadedness	Functional impairment
Burnout as mood fluctuations	General indicators for burnout
Burnout as persistent fatigue	Internal measurement self-report
Burnout vs depression	Interpretive assessment
Difficulty connecting with patients	Measurement to action
	Personalization
	Physical indicators
	Situational indicators
Individual stress management	Program-level stress intervention
Barriers to clinical support	Program vs individual responsibility
Clinical support	
Creating routines	
Drivers of clinical support	
Physical activity	
Protected space	
Resilience	
Self-medication	
Well-being apps	
Social stress management	Personal data privacy
Connectedness between residents	Anonymity helps preserve privacy
Family and friend support	Clinician might review data
Openness between residents	Invasiveness
	Program might review data
Utility of intervention or measurement	Contextual factors
Accuracy	Effects of COVID-19 on residents
Data perpetuates negative feelings	
Passive data for individual use	
Passive data for program use	
Too cumbersome	
Validated clinical scales	
	Program stressors
	Program provided fitness trackers

Table 3. The codebook that resulted from our thematic analysis of formative interviews with residents (see Section 3.1).

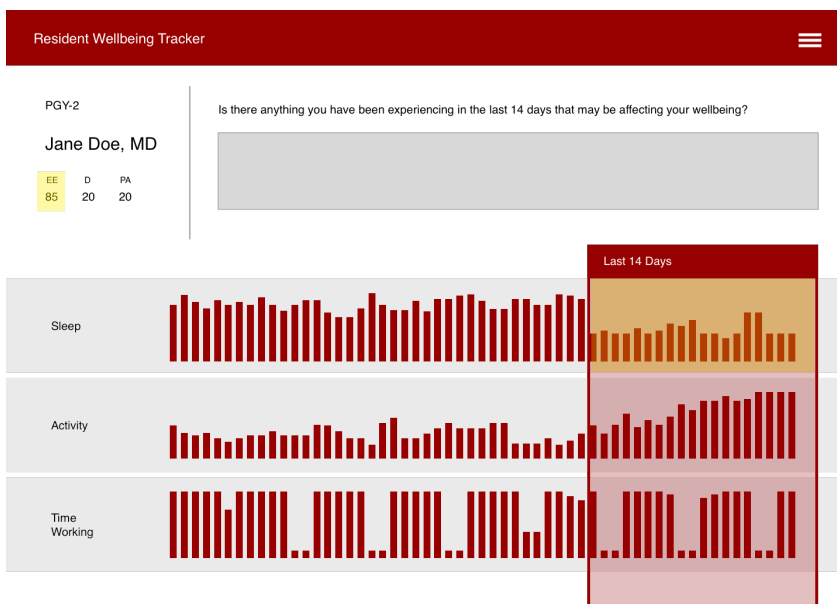
C PROVOCATION INTERVIEW GUIDE - RESIDENTS

Thank you for your participation today. We're part of a research team interested in understanding how digital technology can support resident well-being. As part of that, today we're going to describe to you a tool that could be used to understand and improve resident physician well-being. We'll then ask you specific questions regarding your perceptions and usage of this tool, and receive your feedback on the tool. This session is estimated to take a maximum of one hour. I am now going to begin the session, and describe the setting for using this tool.

You're a resident physician. The ACGME has recently given your program the mandate to incorporate well-being as a Professionalism competency for residents, but program directors don't have guidance on what, specifically, to measure.

- (1) What are the components of well-being, to you?
- (2) How do you self-assess these components, if at all?

We are working on a tool that you could use to better manage your well-being. I am now going to show you some early prototypes of what it could look like, and then ask you questions to prompt your feedback on these ideas. Please feel free to ask us questions about the prototypes – we are happy to explain.

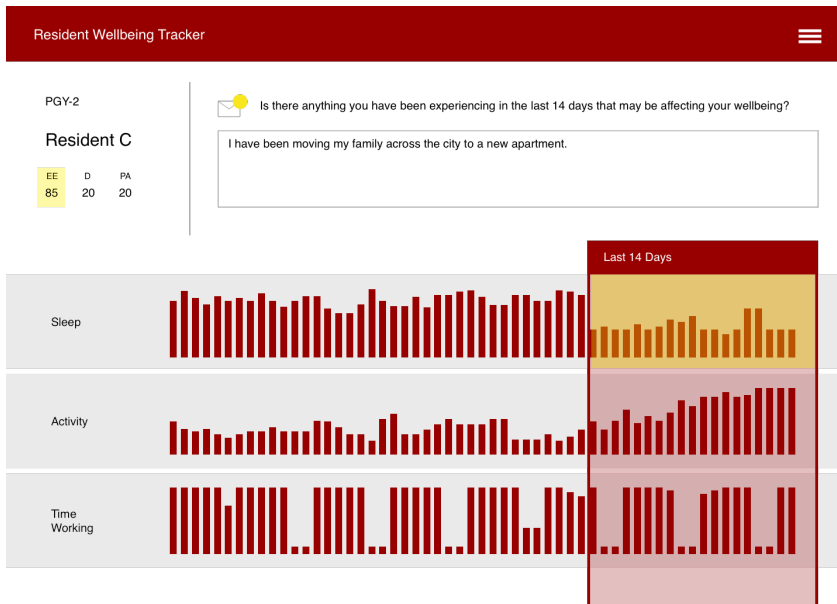


Let's say you're Jane Doe, a second-year resident. Your residency program has given you access to this dashboard for managing your own well-being.

- (1) Please take a look at this screen. What are your first impressions?
- (2) Does the information presented capture your understanding of well-being?
- (3) What do you think about the sleep and activity data?
- (4) What do you think about the productivity data (time working)?
- (5) What do you think about the free text box?
 - (a) Probe: How would you imagine using it? What would you enter, how regularly?
- (6) The three numbers under your name are measurements from a self-reported burnout screener you completed part way through your current rotation. What do you think about those numbers?

- (7) How would you rank these four types of well-being information (MBI, behavioral, work productivity, free text) in terms of usefulness to your self-management of well-being?
 - (a) Probe: What makes these more or less useful to you?
- (8) How does this compare to other methods you may currently use to assess your well-being?
- (9) Is there anything you would add or remove from this?
- (10) Do you have any other comments regarding this screen that we have not covered?

We are now going to show how a screen representing the same information could be seen by an attending physician who supervises residents.



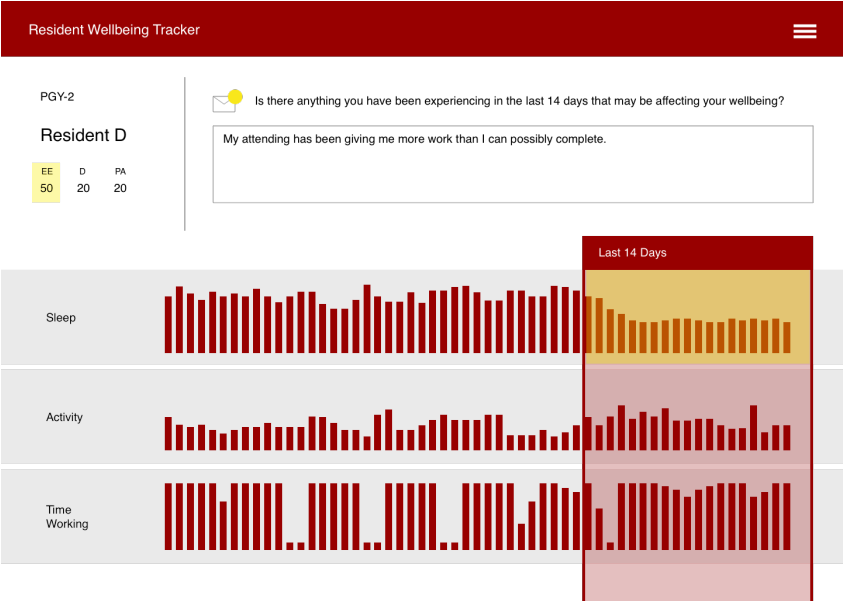
This is a screen showing how the information from the last screen could be viewed by your supervising attending. Let's say you provided context on your sleep numbers by entering into that free text box that you recently moved.

- (1) What do you think about attendings and program directors having access to this screen?
- (2) Notice that your name has been anonymized to Resident C. Do you believe Resident C can be truly anonymous to their supervisors? Does it matter?
- (3) Do you think this particular message is helpful for your supervision?
- (4) Do you think it's helpful to Resident C that their supervisor sees this information?
- (5) Is there anything you would add to or remove from this screen?
- (6) Are there other ways you would want to see your well-being data contextualized for your supervisors?

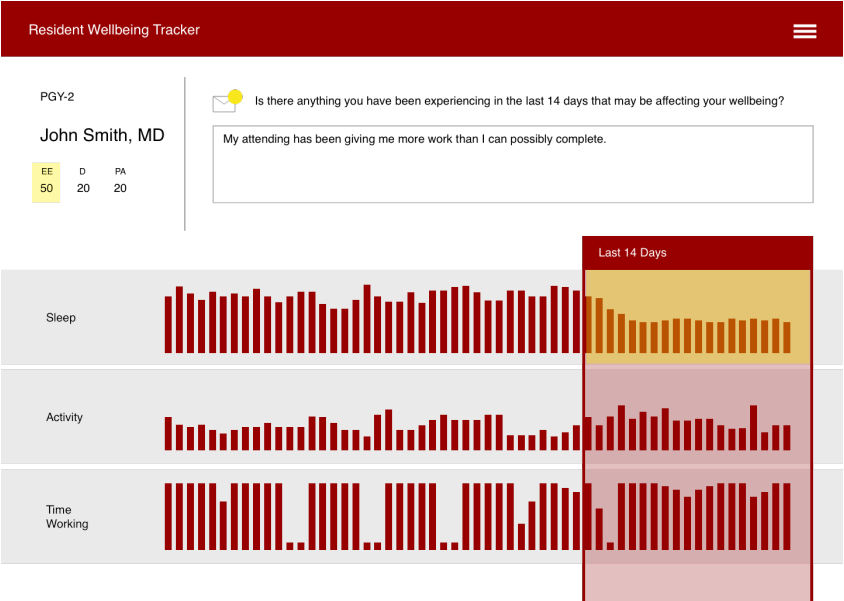
Now we are going to show a message for a different resident in the free-text box, which could be viewed by a supervising attending.

This screen shows similar information to the last screen, for a different resident, "Resident D". There is a different message in the free text box describing that this resident's attending has been giving them too much work.

- (1) What do you think about this message as compared to the last one?
- (2) Do you think this particular message is helpful to supervisors?



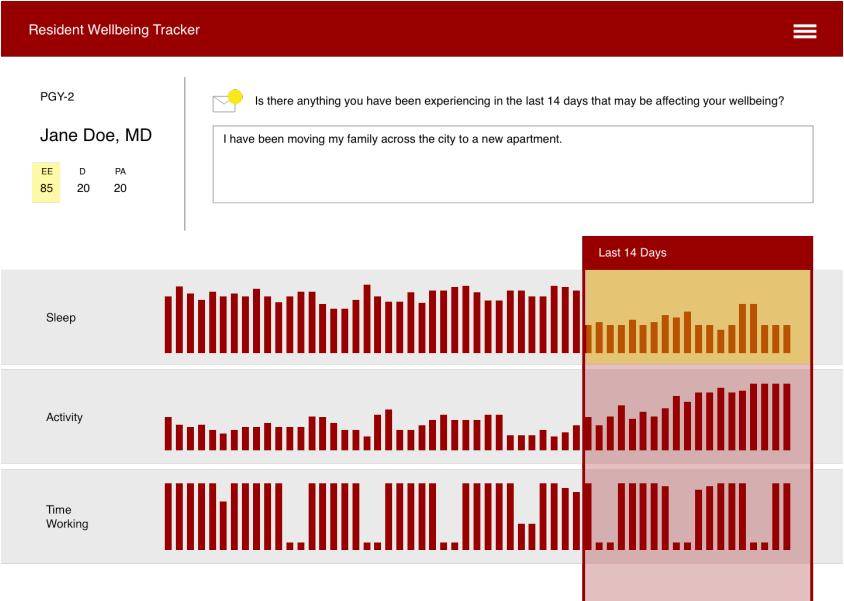
(3) Do you think it's helpful to Resident D that his or her supervisors see this particular message? Now we are going to show this same screen, but imagine a scenario where the attending physician sees de-anonymized information.



Now the supervising attending sees Resident D’s information, and knows Resident D is “John Smith”.

- (1) What do you think about supervising attending physicians and program directors having access to the de-anonymized screen?

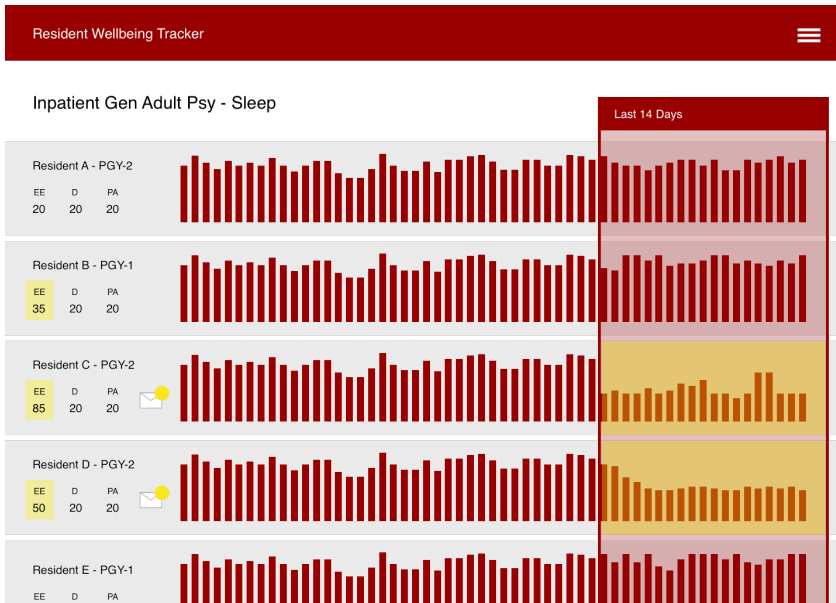
Now we are going to go back and show your screen (Resident C), with your name de-anonymized as “Jane Doe”.



We are now seeing your (Resident C's) comment again, de-anonymized to the supervising attending, showing that you are Jane Doe.

- (1) What do you think about supervising attending physicians and program directors having access to the de-anonymized screen?

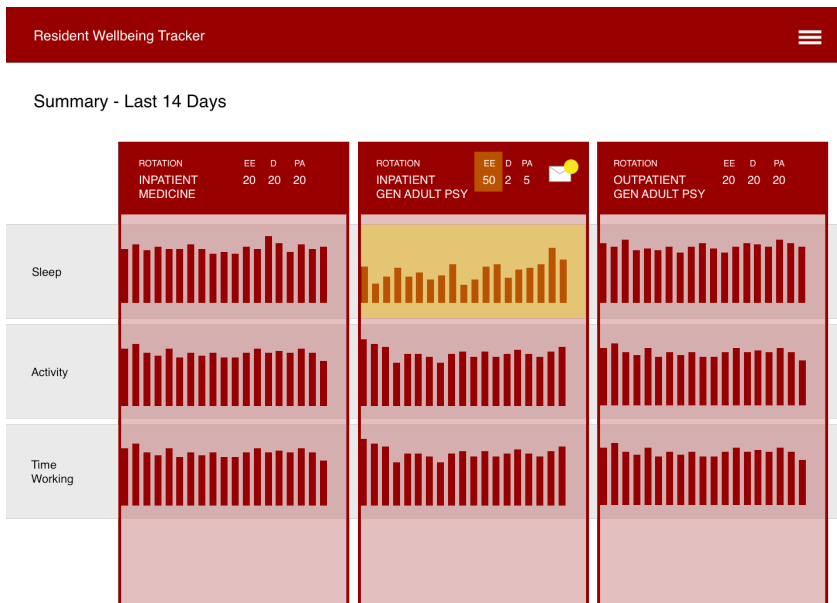
These screens so far have been your view and a supervisor's view of individual information. We'll now show you an aggregated view of your data in context with your peers'.



This screen shows well-being data from residents within your rotation, which is the Inpatient General Adult Psychiatry rotation.

- (1) Please take a look at this screen. What are your first impressions?
- (2) Notice that resident names are anonymized. Do you believe they are truly anonymous to supervisors? Does it matter?
- (3) *(If they do not notice C)* As you may have noticed, we have highlighted that Residents C and D are sleeping considerably less than the other two residents. Residents C and D also have higher emotional exhaustion (EE) scores, compared to other residents. What do you think about this presentation of within-rotation burnout?
- (4) What do you think about supervising attending physicians and program directors having access to this screen?
- (5) Is there any other information not presented on this view, that you would want supervising attending physicians and program directors to know to understand resident well-being across residents in the same rotation?
- (6) Do you have any other comments regarding this screen that we have not covered?

This is a within-rotation view of aggregated and anonymized data; we'll now switch to a between-rotation view.



This screen shows well-being data averaged across residents from three different rotations: Inpatient Medicine, Inpatient General Adult Psychiatry, and Outpatient General Adult Psychiatry.

- (1) Please take a look at this screen, what are your first impressions?
- (2) What do you think about supervising attending physicians and program directors having access to this screen?
- (3) Is there any other information not presented on this view, that you would want supervising attending physicians and program directors to know to understand resident well-being across residents in different rotations?
- (4) Do you have any other comments regarding this screen that we have not covered?

Thank you for responding to our questions, and for providing feedback on our prototypes. Are there any more comments you would like to share regarding the tool we presented?

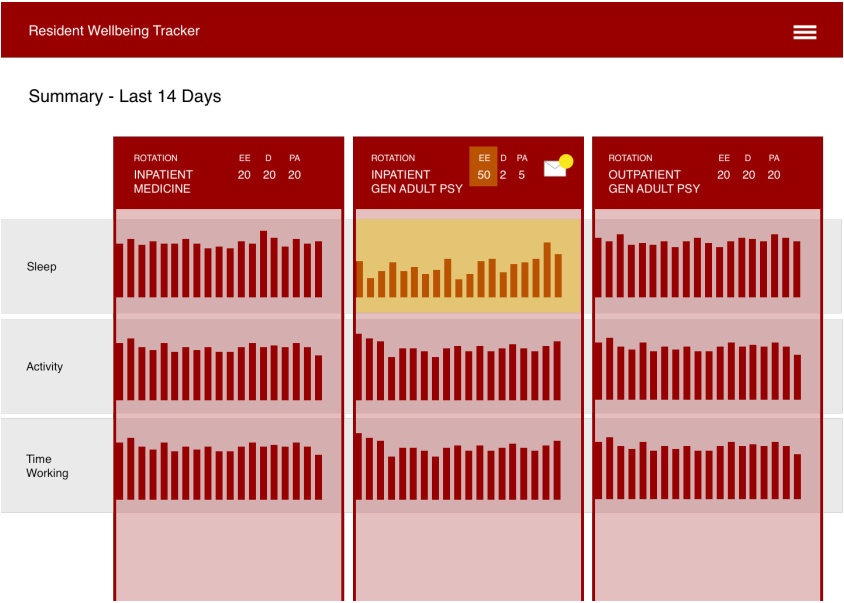
D PROVOCATION INTERVIEW GUIDE - ATTENDINGS

Thank you for your participation today. We're part of a research team interested in understanding how digital technology can support well-being. As part of that, today we're going to show you prototypes of a tool that could be used to understand and improve resident physician well-being. We'll then ask you specific questions regarding your perceptions and usage of this tool, and receive your feedback on the tool. This session is estimated to take a maximum of one hour. Do you have any questions before we begin?

I am now going to begin the session, and describe the setting for using this tool. You're an attending supervising residents, and the ACGME has recently given your program the mandate to incorporate well-being as a Professionalism competency for your residents. But you don't have guidance on what, specifically, to measure.

- (1) What are the components of well-being, to you?
- (2) How do you assess these components, if at all?

We have created a tool that could be used to understand resident well-being. I am now going to describe this tool to you, show you some early prototypes of what it could look like, and then ask you questions to understand your thoughts and feedback about the prototype. Please feel free to ask us questions about the prototypes as well.



This would be the first screen that you, as a supervisor, would see using our tool.

- (1) Taking a look at this screen, what do you notice?
- (2) Does the information presented capture your understanding of well-being?
- (3) What do you think about the sleep and activity data?
- (4) What do you think about the productivity data (time working)?
- (5) The three numbers alongside each rotation name are measurements from a burnout screener residents took partway through their current rotation. What do you think about those numbers?
- (6) How would you rank these four types of well-being information in terms of usefulness to assess resident well-being?

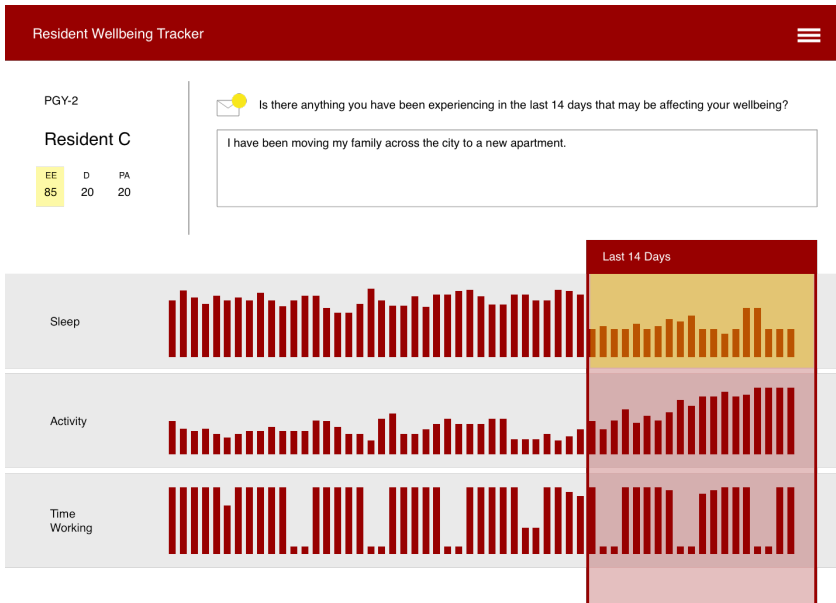
- (7) How does this compare to other methods you may currently use to assess resident well-being?
- (8) Is there anything you would add or remove from this?
- (9) What would you do with this information?
- (10) Do you have any other comments regarding this screen that we have not covered?

We are now going to move to another screen to learn more about the well-being of Inpatient General Adult Psychiatry residents over the past few months. We can move to this screen by clicking on the area of the screen that contains data on “Inpatient General Adult Psychiatry” residents, and the data we are interested in, specifically “sleep.”



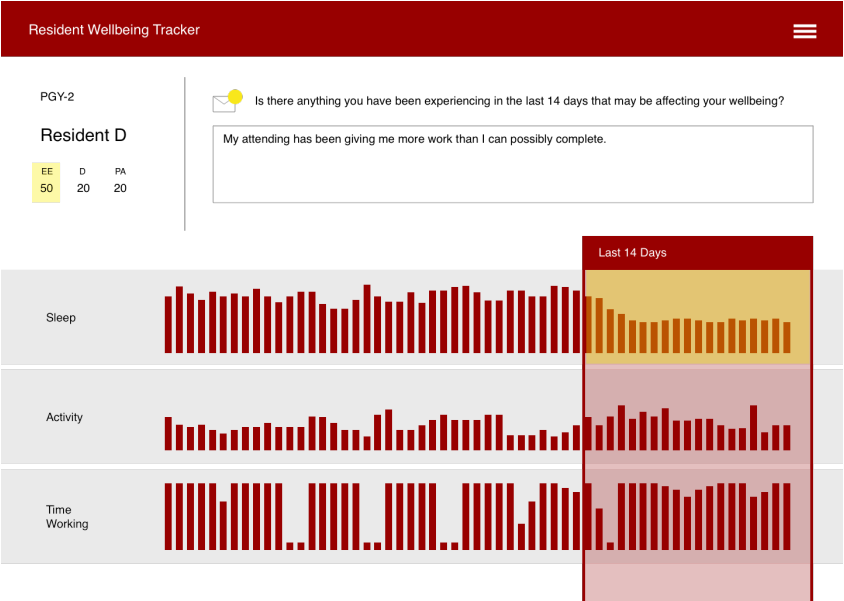
- (1) Looking at the screen, what's the first thing you notice?
- (2) Is there any other information not presented on this screen that you would want to know to understand individual-level resident well-being?
- (3) On this screen, the residents are anonymized. What do you think about this? Is it helpful to you?
- (4) What would you do with this information?
- (5) Do you have any other comments regarding this screen that we have not covered?
- (6) *(If they do not notice C or D)* As you may have noticed, the tool has highlighted that Residents C and D are sleeping considerably less than the other two residents. Residents C and D also have higher emotional exhaustion (EE) scores, compared to other residents. What would you do with this information?

We are now going to move to another screen to learn more about Resident C's well-being, specifically. We can move to this screen by clicking on the area of the screen that contains data from Resident C.



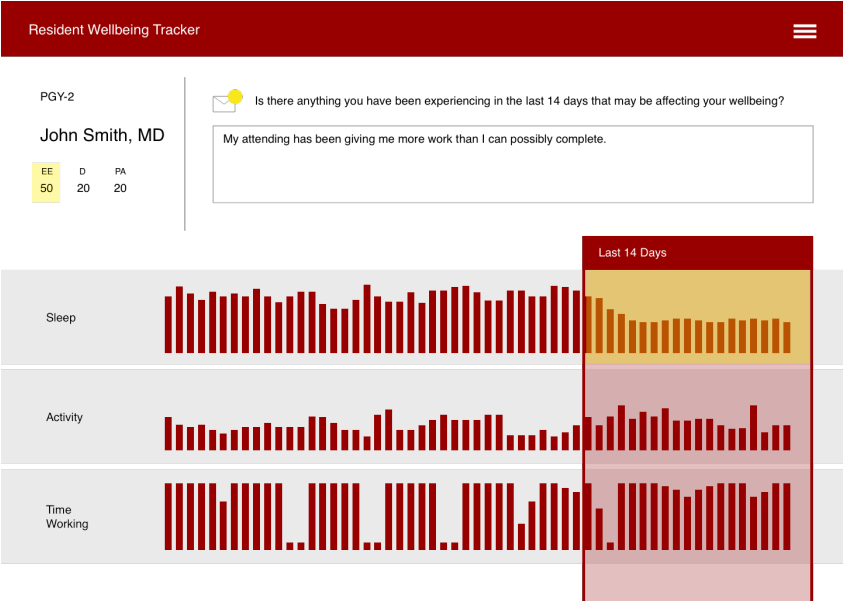
- (1) Please take a look at this screen, what do you notice?
- (2) Is there any other information not presented on this screen that you would want to know to understand Resident C's well-being?
- (3) The name "Resident C" is being used instead of the resident's real name. What do you think about this anonymity measure? Is it helpful to you?
- (4) We have four types of information on this screen: behavioral data, work productivity data, the burnout measurement, and a message. What do you think about each type of information? Is one more useful than the others?
- (5) What would you do with this information?
- (6) Do you have any other comments regarding this screen that we have not covered?

Let's say instead of Resident C, (CLICK BACK TO LAST SCREEN), we clicked on a different resident, Resident D (GO FORWARD 2 SCREENS). This is what we would see.



- (1) What do you think about Resident D’s message?
- (2) How would you act upon this information?

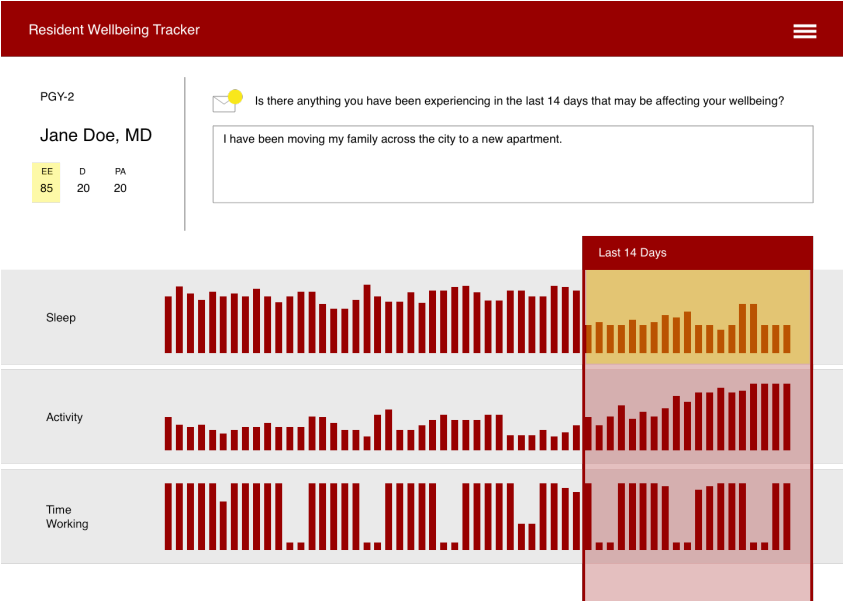
Now we’ll look at an alternate view of Resident D’s individual well-being tracker.



Resident D has been de-anonymized, and we found out that Resident D is John Smith.

- (1) What do you think about knowing that Resident D is John Smith?
- (2) Is it helpful to know that Resident D is John Smith? Why or why not?
- (3) How would you act on this message, now that you know Resident D is John Smith?

Now let's think back to Resident C, and take a look at their de-anonymized individual wellbeing tracker.



Resident C has been de-anonymized, and we found out that Resident C is Jane Doe.

(1) How would you act on this message, now that you know Resident C is Jane Doe?

Thank you for responding to our questions, and for providing feedback on our screens. Are there any more comments you would like to share regarding what we presented?

E PROVOCATION INTERVIEW CODEBOOK

Theme / Code	
Design suggestions	Interventions
Explicit requests for help Prioritization tools for attendings Reference ranges Retrospective sharing tools	Check in with resident Group vs. targeted No immediate action Role of program directors vs attendings Supporting via changes to work conditions Supporting via life or non-work assistance
Methods for measuring well-being	Refining well-being proxies
Barriers to clinical support Clinical support Creating routines Drivers of clinical support Physical activity Protected space Resilience Self-medication Well-being apps	Activity measurement as leisure time Activity measurement as physical activity Effects on patients Internal sense of well-being Productivity as duty hours vs actual hours worked Productivity as sense of work-life balance Sleep as a well-being proxy Work satisfaction as a well-being proxy
Negative repercussions	Privacy
Increased workloads Performatory solutions Responsibility back on resident Misrepresentation	Invasiveness of measurement Residents' willingness to report Residents' willingness to be monitored Utility to intervention vs. anonymization Whether anonymity is possible
Utility of dashboard	
Centralizing information Comparing and contrasting well-being between residents Conversation-starter Need all residents' buy-in Only useful if supervisors or directors act Self-reflection Validating existing inquiry	

Table 4. The codebook that resulted from our thematic analysis of design provocation sessions with residents and attendings (see Section 3.1).

Received January 2022; revised April 2022; accepted May 2022