Check for updates

Received: 28 November 2023 | Accepted: 31 March 2024

DOI: 10.1002/jhm.13366

Journal of Hospital Medicine | shm Society of Hospital Medicine

ORIGINAL RESEARCH

# The development and pilot of a novel mobile application to assess clinician perception of workload and work environment

Marisha Burden MD, MBA | Lauren McBeth BA | Angela Keniston PhD, MSPH

Division of Hospital Medicine, University of Colorado School of Medicine, Aurora, Colorado, USA

**Correspondence**
Marisha Burden, MD, MBA, Division of Hospital Medicine, University of Colorado School of Medicine, 12401 E. 17th Ave, Aurora, CO 80045, USA.
Email: Marisha.Burden@cuanschutz.edu;
Twitter: @marishaburden

**Funding information**
National Institute for Occupational Safety and Health, Grant/Award Number: U19OH011227


## Abstract

**Background:** Traditional measures of workload such as wRVUs may not be adequate to understand the impact of workload on key outcomes.

**Objective:** The objective of this study was to develop a mobile application to assess, in near real time, clinicians' perception of workload and work environment.

**Designs, Settings and Participants:** We developed the GrittyWork™ application (GW App) using the Chokshi and Mann process model for user-centered digital development. Study occured at a single academic medical center with hospitalist clinicians.

**Main Outcome Measures and Measures:** Measures included the System Usability Scale (SUS), use measures from GW App, electronic health record (EHR) event log data and note counts, and qualitative interviews.

**Results:** From October 28, 2022 to November 3, 2022, six hospitalist clinicians provided feedback on the early prototype of the GW App, and from February 28, 2023 to June 8, 2023, 30 hospitalist clinicians participated in the pilot while on clinical service. All 30 clinicians (100%) participated in the pilot submitting data for a total of 122 shifts. Participants reported working 10 ± 1 h per day (mean ± SD) and were responsible for an average of 11 ± 3 patients per day. The postpilot evaluation of the GW App showed a SUS score of 86 ± 11 and a participant preference toward mobile application-based surveys (73% of participants). Regarding workload measures, EHR event log data and notes data correlated with physician-reported workloads. Applying user-centered design techniques, we successfully developed a mobile application with high usability. These data can be paired with EHR event log data and outcomes to provide insights into the impact of workloads and work environments on outcomes.


This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Authors. *Journal of Hospital Medicine* published by Wiley Periodicals LLC on behalf of Society of Hospital Medicine.

# INTRODUCTION

Research suggests that high workloads in the inpatient setting may contribute to healthcare worker burnout, worse patient outcomes, and adverse organizational outcomes including financial outcomes.[1–3] To date, the mainstay of measuring clinician workloads has been through wRVUs or encounters; yet, recent research has shown the importance of incorporating measures that assess both work quantity and the impact of workload on key outcomes (i.e., workforce, patient, and organizational outcomes) as relying solely on traditional productivity measures may be insufficient.[4] Furthermore, organizations often assess workloads, employee outcomes (e.g., burnout), and patient outcomes (e.g., length of stay) independently, without connecting work processes and structure to the outcomes.

In this study, we aimed to bridge these gaps by developing a mobile application to capture clinician perception of workload and work environment. We then sought to integrate these data with electronic workload measurements from the electronic health record (EHR) to gain a more comprehensive evaluation of clinician workloads and work environments through both objective and subjective measures. This work lays the foundation for future research aimed at developing evidence-based practices for determining optimal workloads and strategies for enhancing and optimizing work environments.

# METHODS

## Study design

We utilized a multiple methods approach. We began by utilizing user-centered design methods to develop a mobile application, called the GrittyWork™ application (GW App), to assess various dimensions of clinician perception of work including measures of workload, work environment, as well as leadership support, and clinician well-being. We then integrated this approach with an ecological momentary assessment (EMA) method[5,6] to harness the benefits of understanding work experiences in near realtime as they occur in the workplace over the course of time and in different contexts. The study was approved by the Colorado Multiple Institutional Review Board. The full protocol can be found in Supporting Information S2: Supplement 1. We utilized the STROBE guidelines[7] for reporting.

## Setting, participants, and study size

The study was conducted at a 678-bed quaternary care hospital with approximately 12,000 medicine discharges per year with 58 clinical full-time equivalents (cFTE) for physicians and 42 cFTE for advanced practice providers (APPs) including physician assistants and nurse practitioners. The study was conducted from October 28, 2022 to June 8, 2023, which included prototyping and building of the mobile application, pilot testing, and qualitative work. Hospitalist physicians and APPs were invited to participate via group-wide emails with invitations being sent no more than three times. Verbal consent was obtained for those interested in participating by a nonsupervisor research team member. Participants could participate in two of the three offerings (prototype feedback, pilot test, and/or qualitative interviews following the pilot) and were offered a $10 gift card for each activity. Participants were consented on a first come, first serve basis until recruitment was complete. Clinicians that participated in the pilot were asked about their interest in participating in the qualitative interviews via the final survey sent during the pilot test.

## GW App development

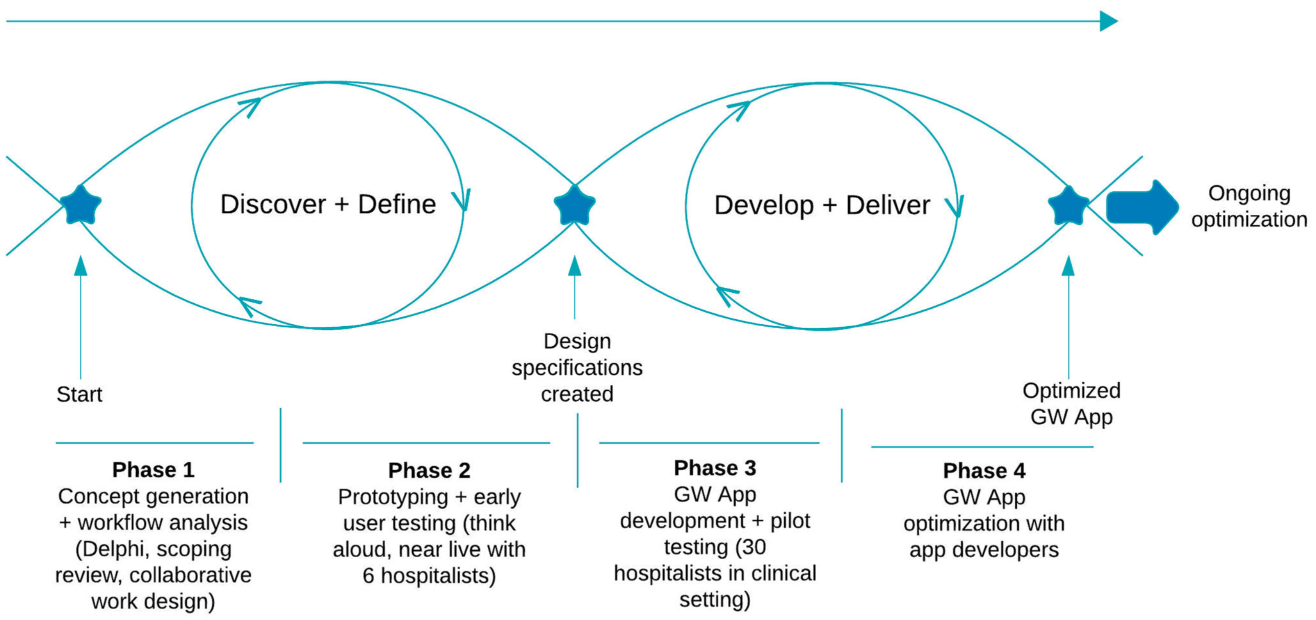
To guide the design and iterative development of the GW App, we applied the Chokshi and Mann process model for user-centered digital development (Figure 1),[8] applying all four phases with a continuous feedback loop between discover and define and develop and deliver stages.

In phase 1 of the discover and define stage, we sought to understand the concepts and processes associated with capturing workloads in near real time from hospitalists via a mobile application, utilizing both a scoping review process and Delphi panel to generate ideas to measure workloads and work environments.[4,9] This work, alongside the clinical experience and background of the research team, informed the rudimentary concept for the GW App. Research team members actively collaborated with the developers to determine the early prototype platform functionality (Android and iOS), ease of use, interface attractiveness, push notification functionality, biometric data collection, data storage, data retrieval, and the process to modify and push new questions to users.

In phase 2 of the discover and define stage, we engaged with users to understand how they might use the mobile application and observed clinicians interacting with the application using "think-aloud" and "near-live" methods.[8] Think-aloud involves users verbalizing their thoughts in real time as they navigate the mobile application. Near-live attempts to simulate a real-world situation where users interact with the mobile application as they would in their regular work lives but in a controlled setting. We recruited six hospitalist clinicians for interviews to capture insights regarding context of use, user requirements, and feedback on an early prototype of the GW App. Participants were observed downloading and completing the workforce survey using the GW App and verbalizing their thoughts as they completed the tasks. After interacting with the GW App, users were asked their opinions in a semistructured interview. See Supporting Information S2: Supplement 2 for the interview guide. Based on user feedback, we made changes considered both relevant and feasible to the GW App to optimize functionality and correct any issues identified.

For phases 3 and 4 of the develop and deliver stage, we iteratively developed, tested, and optimized the GW App in the hospital setting. The pilot testing specifically allowed us to evaluate the quality of data collection including the completeness, correctness, concordance, plausibility, and currency of data[10] by comparing

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License




Adapted from Chokshi and Mann user-centered digital development. Copyright © Sara Kuppin Chokshi, Devin M Mann. Originally published in JMIR Human Factors (http://humanfactors.jmir.org), 19.12.2018.

**FIGURE 1** Study flow.

self-reported workload measures from the application to clinician usage data from the EHR. A priori, we planned to recruit 30 hospitalists for the pilot and successfully recruited the target number including both physicians and APPs to test the GW App while on service. Before the participant's clinical service, the mobile application was installed. EMA prompts[5,6] were programmed into the mobile application (i.e., push notifications) whereby participants were asked to complete a daily survey in the GW App at the end of each shift over the course of their week on service. We targeted four consecutive weekdays for the pilot data collection to minimize the effects of variable staffing over the weekend.

## Surveys

Within the GW App, there were two surveys: a short daily survey and a longer survey. The longer survey was typically administered on the fourth day of service. See Supporting Information S2: Supplement 3 for the full list of questions. In the shorter survey, we asked each person to answer questions about their type of shift, workload, and perception of workload (e.g., shift rating with a "sad face"/"happy face" emoji). The longer version included the short survey plus additional questions about culture of safety, teamwork, leadership, work environment, and burnout. This approach was taken to minimize the time spent on surveys while striving to obtain a comprehensive understanding of participants' perceptions across various domains of work (e.g., leadership, safety, etc.). Workload questions included the number of patients a care team was responsible for, the number of patients the clinician saw, and the National Aeronautics and Space Administration Task Load Index (NASA TLX).[11,12] The NASA TLX was utilized to measure clinician task load and consists of six subscales: mental demand, physical demand, temporal demand, frustration, effort, and performance.[12] Of note, some of the NASA TLX data were not captured and thus "missing" during the pilot as the slider bar functionality malfunctioned. We utilized a two-question rapid screening for burnout ("I feel burned out from my work" and "I have become more callous toward people since I took this job"), which has been shown to correlate well with the full Maslach Burnout Inventory.[13–15] Safety, teamwork, work environment, and leadership questions were taken from the Agency for Healthcare Research and Quality (AHRQ) Hospital Survey on Patient Safety Culture™.[16–19] The maximum score for each composite item is 100% and calculated based on AHRQ methods. These questions were included in the long survey given the importance of work environment, leadership, and safety impacting organizational approaches to workloads as well as their impact on a variety of outcomes.[20–22]

Additionally, at the end of the 4-day stretch on service, participants were asked to complete a final survey in REDCap[23] including demographic questions and the validated System Usability Scale (SUS) to assess the GW App usability.[24,25] The SUS is a 10-item questionnaire with each item rated on a 5-point Likert scale from "strongly disagree" to "strongly agree" with three of the items reverse scored (see Supporting Information S2: Supplement 4 for details). The highest possible score is 100 with a higher score indicating better usability.[24,25] The survey also collected open-ended feedback and comments. Biometric data collection including number of steps and amount of sleep was enabled during the pilot using linked mobile

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

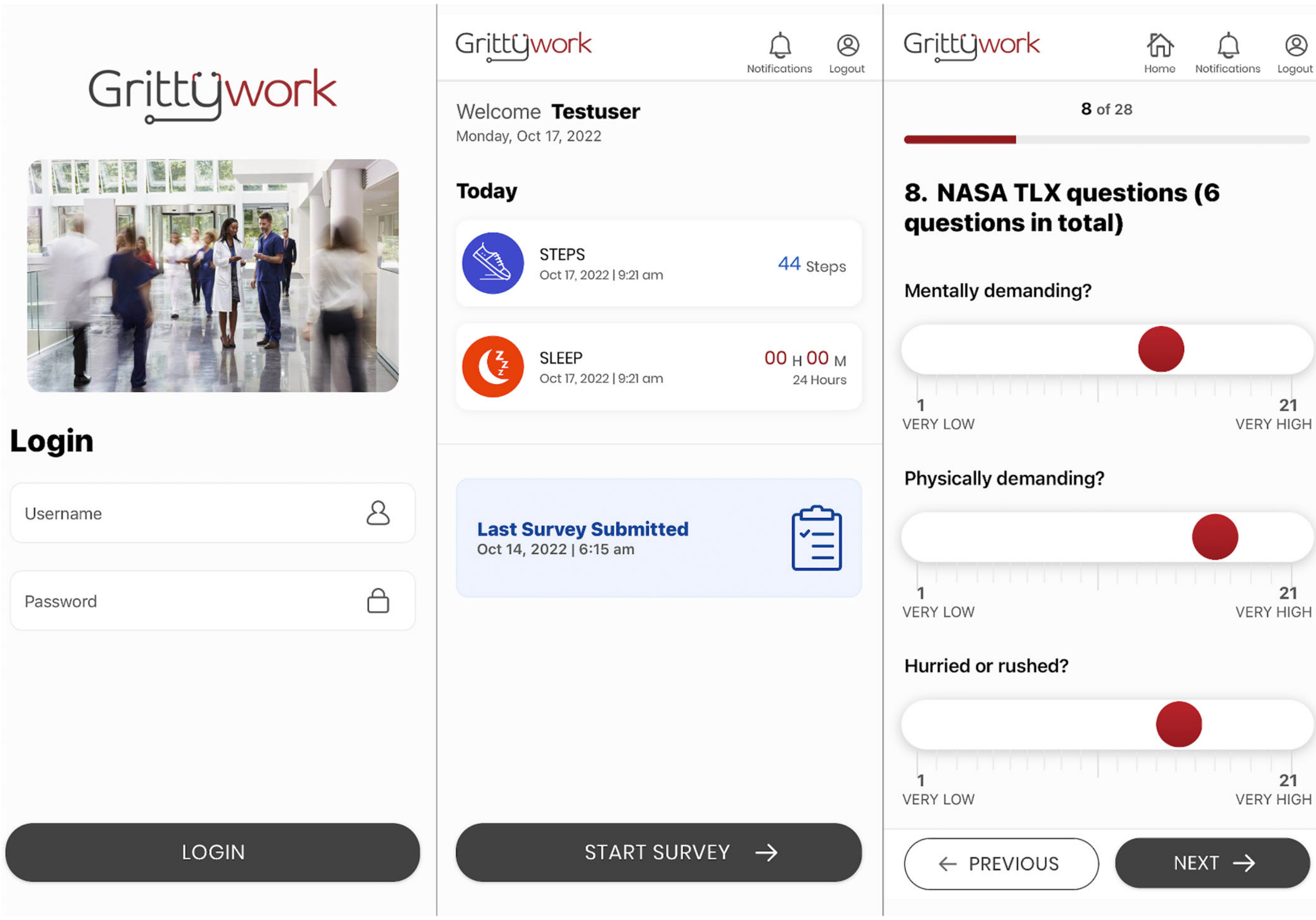



**FIGURE 2** GrittyWork™ application.

health applications; however, it was not fully functional and thus not reported.

Finally, 10 individuals who participated in the pilot were asked to participate in a semistructured qualitative interview about their experience using the GW App while on service. The interview guide is in Supporting Information S2: Supplement 5. These interviews on average lasted 20 min.

## Mobile application utilization data and EHR use data

We extracted usage metrics from the GW App, including number of active users in a given time period, variation in use patterns over time, and rates of survey abandonment/completion. If a study team member needed to access the GW App, this use data would have also been included; however, during the pilot, this was an infrequent occurrence. In addition, we used pre-existing data sets from the EHR containing a count of notes signed or cosigned by each clinician during the time periods they participated in the pilot project and worked on a day shift. Similarly, we extracted aggregated EHR event log data (vendor-derived aggregated EHR use metrics captured via clinician use of the EHR during regular daily work),[26–28] which included the average number of minutes that the clinician spent engaging in specific activities in the EHR.

## Analysis

Quantitative analysis was performed using SAS Enterprise Guide 8.3 (SAS Institute Inc.). A Student's *t*-test was used to compare the number of patients per day on teams and personally cared for by a participant, the number of hospital units, and the NASA TLX score by the shift rating reported by each participant each day. We used a Pearson correlation coefficient test to evaluate the relationship between the number of patients per day on teams and personally cared for by a participant and the NASA TLX score. Chi-square tests were used to compare the proportion of participants reporting symptoms of burnout by the shift rating reported by each participant each day. In addition, a chi-square test was used to analyze the association between teams being staffed with APPs or having learners with the shift rating reported by each participant each day. Using a Pearson correlation coefficient test, we evaluated the linear relationship between workload measures collected via the GW App and those obtained from EHR data.

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## Role of the funding source

The funding source did not have any role in study design, conduct, analysis, or reporting.

# RESULTS

A total of six clinicians participated in phase 2 of the discover and define stage from October 28, 2022 to November 3, 2022 (four physicians, two APPs). Multiple recommendations were made to improve the prototype (Supporting Information S2: Supplement 6) and the majority were then sent to the developers to modify the GW App. Figure 2 shows the GW App.

For phase 3 of the develop and deliver stage, 30 clinicians including 23 (77%) physicians and seven (23%) APPs participated in the pilot from February 28, 2023 to May 26, 2023, with the final qualitative interview conducted on June 8, 2023. Demographics are provided in Table 1. The average engagement time per session was 7 min 30 s. No surveys were abandoned (100% completion rate).

*Pilot results*: Results from the daily surveys are shown in Table 2. Shifts occurred during the day shift and predominantly on primary teams. We found that the average number of patients a team was responsible for was significantly lower during shifts where participants reported the "happy face" emoji compared to shifts where the participant selected the "sad face" emoji (mean ± SD: 11 ± 3 vs. 13 ± 2; $p = .04$) and similarly for the average number of patients that a participant personally saw (mean ± SD: 10 ± 3 vs. 13 ± 2; $p = .006$). The average NASA TLX score was significantly lower during shifts where participants selected the "happy face" shift rating compared to shifts where the participant selected the "sad face" (mean ± SD: 36 ± 13 vs. 69 ± 9; $p < .001$). Finally, we found that the average number of patients that a participant personally saw was correlated with the average NASA TLX score ($r = .42$; $p = .03$). There was no significant association between the average number of patients a team was responsible for and the average NASA TLX score ($r = .39$; $p = .05$). We did not observe a significant association between the rating of the shift ("happy face" or "sad face") and the average number of hospital units reported by participants each day, participants reporting symptoms of burnout, or teams being staffed with an APP or learner. Box plots illustrating the associations between the number of patients, the number of hospital units, and the NASA TLX score and the shift rating are provided in Supporting Information S1: Appendix Figure S1.

*Postpilot evaluation*: The postpilot evaluation of the GW App showed a SUS score of 86 ± 11 (mean ± SD) and a preference toward mobile application surveys (73% of participants) (Table 2). There was fair to moderate correlation between hours worked per the GW App and reported measures of workload for physicians; however, we did not see similar correlations for APPs (Table 3). In postpilot qualitative interviews, participants raised several important issues around successfully implementing a mobile survey application such as the GW App (Supporting Information S1: Appendix Table S1). Additional recommendations were made for future iterations (Supporting Information S2: Supplement 7).

**TABLE 1** Participant demographics, $N = 30$.

| Role | *N* (%) |
|---|---|
| Physician | 23 (77) |
| NP/PA | 7 (23) |
| Gender | *N* (%) |
| Woman | 18 (60) |
| Man | 10 (33) |
| Prefer not to answer | 0 (0) |
| Missing | 2 (7) |
| Years since training | |
| Mean ± SD | 6.5 ± 4.5 |
| Min, Max | 1, 19 |
| Missing | 2 (7) |
| Number of years working with the hospitalist group | |
| Mean ± SD | 4.1 ± 3.9 |
| Min, Max | 0.5, 14 |
| Missing | 2 (7) |
| Percent clinical effort (cFTE) | |
| Mean ± SD | 74.5 ± 24.1 |
| Min, Max | 25, 100 |
| Missing | 2 (7) |

*Note*: If missing data row not included, then no missing data.
Abbreviations: cFTE, clinical full-time equivalent; NP, nurse practitioner; PA, physician assistant; SD, standard deviation.

# DISCUSSION

Previous studies have highlighted that current measures for workload including wRVUs and encounters are likely insufficient to capture an understanding of both workloads and the impact of work design on outcomes.[4] Additionally, previous work has shown that high workloads may result in harm for patients, clinicians, and organizations.[2,29–35] Despite knowing these harms, in most medical specialties, including those of generalists, the optimal workload is still not known.

Our study offers the beginning of a paradigm shift—one of quick, strategic clinician assessments of workloads and work environments combined with electronic measures of workloads (i.e., event log data) with the goal of linking these data to outcomes to detect clinician, patient, and organizational harms early. We utilized user-centered design techniques to successfully develop a mobile application that was both simple and intuitive for capturing clinician workloads. We found that clinicians in this pilot preferred mobile application surveys compared to surveys delivered via email or text. This emphasizes the convenience and immediacy of mobile platforms which can capture

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 2** GrittyWork™ pilot evaluation (*N* = 30 participants, 122 shifts).

| Daily questions (short survey) | *N* = 122 shifts |
|---|---|
| Shift characteristics | |
| What type of shift did you work? | *N* (%) |
| Day shift | 120 (98) |
| Missing | 2 (2) |
| What type of team were you on?[a] | |
| Primary team | 105 (86) |
| Consult/comanagement team | 17 (14) |
| Admitting team | 3 (2) |
| Procedural team | 3 (2) |
| Select all of the team members that you had with you on this shift, excluding yourself.[a] | |
| NP/PA | 60 (49) |
| Attending physician | 28 (23) |
| Intern/Resident | 30 (25) |
| Medical Student | 14 (11) |
| NP/PA fellow | 34 (28) |
| Other | 25 (20) |
| Missing | 1 (1) |
| Reported workload | Mean ± SD |
| How many hours did you last work (or anticipate that you will work if filling it out for today)? | 10 ± 1 |
| How many patients was your team responsible for during this shift? | 11 ± 3 |
| How many patients did you personally see during this shift? | 11 ± 3 |
| How many units in the hospital were your patients housed on? | 4 ± 2 |
| NASA Task Load Index score | 44 ± 21 |
| Missing, *N* (%) | 96 (79) |
| Number of critical events or emergencies you had to respond to during this shift. | *N* (%) |
| 0 | 79 (65) |
| 1 | 28 (23) |
| 2 | 12 (10) |
| 3 | 2 (2) |
| Missing | 1 (1) |
| Resources, tools, and overall rating | |
| During this shift, I had the essential people, tools, supplies, and resources I needed. | *N* (%) |
| Never | 0 (0) |
| Seldom | 0 (0) |
| About half the time | 8 (7) |
| Usually | 60 (49) |
| Always | 54 (44) |
| How would you rate this shift? *N* (%) | |
| Happy face emoji | 107 (88) |

(Continued)

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 2** (Continued)

| **Daily questions (short survey)** | ***N* = 122 shifts** |
|---|---|
| Sad face emoji | 13 (11) |
| Missing | 2 (2) |
| **Once a week questions (long survey)** | ***N* = 32 shifts** |
| I feel burned out from my work or I've become more callous toward people since I took this job. | *N* (%) |
| Yes | 9 (28) |
| No | 23 (72) |
| AHRQ Hospital Survey on Patient Safety Culture composites | Mean ± SD |
| Teamwork (work together as an effective team, help each other during busy times, and are respectful) (% positive) | 89% ± 5% |
| Staffing and work pace (enough staff to handle the workload, work appropriate hours and do not feel rushed, and appropriate reliance on temporary, float, or PRN staff) (% positive) | 65% ± 16% |
| Missing, *N* (%) | 1 (3) |
| Organizational learning—continuous Improvement (work processes are regularly reviewed, changes are made to keep mistakes from happening again, and changes are evaluated) (% positive) (mean ± SD) | 74% ± 13% |
| Missing, *N* (%) | 2 (6) |
| **Final survey** | ***N* = 30 participants** |
| System usability scale, mean ± SD | 86 ± 11 |
| Missing, *N* (%) | 2 (7%) |
| What is your preferred method to submit real-time information regarding workload? | *N* (%) |
| Mobile application survey | 22 (73) |
| Email survey | 1 (3) |
| Text message survey | 5 (17) |
| Missing | 2 (7) |

*Note*: If missing data row not included, then no missing data.

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; NASA, National Aeronautics and Space Administration; NP, nurse practitioner; PA, physician assistant; SD, standard deviation.

[a]Participants could check all that applied.

near real-time experiences of clinicians. While more basic electronic measures of workload were used for this study as proof of concept, additional EHR event log data such as time to task completion or time spent in activities in the EHR could also be utilized to understand the impact of workloads on a variety of outcomes.[26,36–38]

Our primary goal of this work was to utilize user-centered design techniques to develop a usable tool that would be acceptable to clinicians. The Chokshi and Mann process allowed for a rapid cycle and iterative user-centered approach[8] to develop the application which likely led to the high usability rating. We also aimed to ensure the GW App resulted in high-quality data collection with most measures having low rates of missing data. We saw fair to high correlations for physician-reported workload measures when comparing to EHR measures; however, we did not see that with APPs. Some of these reported differences for APPs likely resulted from care models that were utilized at the study site (collaborative, shared service model) in addition to the small number of APPs included in the study. Future opportunities will include assessing which measures of workloads (reported and/or measured through the EHR) are most closely associated with outcomes.

From this study, we also gained insights into the importance of psychological safety and the necessity for clear actions by organizational leaders in response to data collected. These factors were noted to be critical for the successful dissemination and implementation of the GW App. The GW App, as it currently stands, was primarily designed for nonprocedural clinicians working in hospital settings.

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 3** Comparison of GrittyWork™ data to electronic health record event log and notes data by role.

| | GrittyWork™ | EHR event log data | r (95% confidence interval) | p-Value |
|---|---|---|---|---|
| Physicians[a] | | | | |
| Work hours, mean ± SD | 10.3 ± 1.4[b] | 2.3 ± 1.0[c] | 0.59 (0.22, 0.80) | 0.003 |
| Workload, mean ± SD | | | | |
| Number of patients responsible for | 11.0 + 3.7[d] | 9.6 + 3.1[e] | 0.78 (0.51, 0.91) | <0.001 |
| Patients personally seen | 10.4 + 3.4[f] | 9.6 + 3.1[e] | 0.98 (0.94, 0.99) | <0.001 |
| Advanced practice providers[g] | | | | |
| Work hours, mean ± SD | 9.6 ± 0.4[b] | 2.4 ± 0.6[c] | −0.56 (−0.94, 0.51) | 0.25 |
| Workload, mean ± SD | | | | |
| Number of patients responsible for | 11.4 + 1.4[d] | 5.3 + 1.1[e] | 0.47 (−0.47, 0.90) | 0.28 |
| Patients personally seen | 11.3 + 1.3[f] | 5.3 + 1.1[e] | 0.49 (−0.45, 0.90) | 0.26 |

Abbreviations: EHR, electronic health record; SD, standard deviation.

[a]Three providers missing notes data.

[b]How many hours did you last work (or anticipate that you will work if filling it out for today)?

[c]Average number of minutes the provider spent in basket, orders, notes, and clinical review per day.

[d]How many patients was your team responsible for during this shift?

[e]Number of notes signed or cosigned during shift: does not differentiate between responsible for and personally seen.

[f]How many patients did you personally see during this shift?

[g]One provider missing EHR event log data.

Additional modifications will be needed to cater to various types of specialties and settings including trainees, outpatient clinics, and those with procedurally based roles.

Finally, this work also presents an opportunity to incorporate workload monitoring into healthcare worker safety management systems to identify when workplace practices may be driving harm, rather than relying solely on traditional workforce well-being surveys, which are typically conducted infrequently and may not link directly to workloads or other aspects of work design. In future research, it will be important to determine the optimal frequency and sample size for clinician surveys. This should entail streamlining the process by minimizing the number of participants and reducing any unnecessary obstacles, potentially making it as easy as a survey capturing a "happy" or "sad" emoji response. Many of the questions on workload (i.e., number of patients seen and hours worked) can likely come entirely from EHR event log data further minimizing survey burden. Capturing biometric data could also be an important addition to future work which may also provide indicators of a stressed workforce.

## STRENGTHS AND LIMITATIONS

Our study had several strengths, including iterative development of a mobile workforce application and taking a multiple methods approach to evaluate the GW App. Limitations include that it occurred in a major urban academic setting and with clinicians that worked day shifts. Thus, the findings may not be fully applicable to other settings. However, we do believe much of the described approach and aspects of the GW App could directly pertain to other care settings recognizing that additional adaptations to the mobile application are necessary and that organizations also have differing cultures, priorities, and incentives which may impact the perceived usability of the mobile application. The question around hours worked could have been a combination of retrospectively reported hours worked and prospectively forecasted hours. For instance, if a participant received a survey push notification while still working, they could either complete the survey later, at the end of their workday, or estimate their remaining hours and fill it out accordingly. This approach aimed to streamline the survey process for participants. In future iterations, workload-related questions could primarily rely on EHR event log data. This approach would reduce survey time and ensure that only questions not extractable from other data sources are asked. There could also have been selection bias; however, we tried to mitigate this by inviting the entire group of hospitalists to participate and limited individual participation to two activities to try and capture diverse perspectives. The order of the various questions could have biased participants' responses; thus, additional assessments should be conducted on how the order of questions impact findings.

## CONCLUSION

By employing user-centered design techniques, we successfully developed a mobile application to assess clinician perception of workloads and work environments with high usability. Combining

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

clinician perception of work with EHR event log data and subsequent outcomes will offer insights into how workloads and work environments impact patient, workforce, and organizational outcomes.

## ACKNOWLEDGMENTS

This publication was supported by grant number U19OH011227 from CDC NIOSH Center for Health, Work, and Environment (CHWE), a NIOSH Center of Excellence for Total Worker Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDC, NIOSH, and CHWE. Drs. Burden and Keniston report funding from the Agency for Healthcare Research and Quality and from the American Medical Association not related to this work. Drs. Burden and Keniston have a trademark for the GrittyWork™ workforce mobile application. The authors utilized the ChatGPT language model (versions 3.5 and 4) developed by OpenAI for editing of original author content to improve readability and grammar. All information and materials in the manuscript are original.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## ORCID

*Marisha Burden* http://orcid.org/0000-0002-8262-3994

## TWITTER

*Marisha Burden* @marishaburden
*Angela Keniston* @kenistonangela

## REFERENCES

1. National Academies of Sciences, Engineering, and Medicine. *Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being*. The National Academies Press; 2019. doi:10.17226/25521
2. Burden M, Patel M, Kissler M, Harry E, Keniston A. Measuring and driving hospitalist value: expanding beyond wRVUs. *J Hosp Med*. 2022;17:760-764. doi:10.1002/jhm.12849
3. Kamalahmadi M, Bretthauer K, Helm J, et al. *Mixing It Up: Operational Impact of Hospitalist Caseload and Case-Mix*. SSRN; 2024.
4. Burden M, McBeth L, Keniston A. Salient measures of hospitalist workload. *JAMA Netw Open*. 2023;6(8):e2328165. doi:10.1001/jamanetworkopen.2023.28165
5. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1-32. doi:10.1146/annurev.clinpsy.3.022806.091415
6. Doherty K, Balaskas A, Doherty G. The design of ecological momentary assessment technologies. *Interact Comput*. 2020;32(3):257-278. doi:10.1093/iwcomp/iwaa019
7. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577. doi:10.7326/0003-4819-147-8-200710160-00010
8. Chokshi SK, Mann DM. Innovating from within: a process model for user-centered digital development in academic medical centers. *JMIR Hum Factors*. 2018;5(4):e11048. doi:10.2196/11048
9. Smith EM, Keniston A, Welles CC, et al. Inpatient clinician workload: a scoping review protocol to understand the definition, measurement and impact of non-procedural clinician workloads. *BMJ Open*. 2022;12(12):e062878. doi:10.1136/bmjopen-2022-062878
10. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013;20(1):144-151. doi:10.1136/amiajnl-2011-000681
11. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv in Psych*. 1988;52:139-183. doi:10.1016/S0166-4115(08)62386-9
12. NASA Task Load Index. National Aeronautics and Space Administration. Accessed March 5, 2024. https://humansystems.arc.nasa.gov/groups/TLX/
13. Li-Sauerwine S, Rebillot K, Melamed M, Addo N, Lin M. WestJEM 21.3 May issue. *West J Emerg Med*. 2020;21(3):610-617. doi:10.5811/westjem.2020.2.45139
14. Maslach Burnout Inventory (MBI). Accessed March 4, 2024. https://www.mindgarden.com/117-maslach-burnout-inventory-mbi
15. Maslach C, Jackson SE. The measurement of experienced burnout. *J Organ Behav*. 1981;2(2):99-113. doi:10.1002/job.4030020205
16. Adams-Pizarro I, Walker ZA, Robinson J, Kelly S, Toth M. Using the AHRQ hospital survey on patient safety culture as an intervention tool for regional clinical improvement collaboratives. In: Henriksen K, Battles JB, Keyes MA, Grady ML, eds. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign)*. Agency for Healthcare Research and Quality; 2008. https://www.ncbi.nlm.nih.gov/books/NBK43728/
17. Zebrak K, Yount N, Sorra J, et al. Development, pilot study, and psychometric analysis of the AHRQ surveys on patient safety culture™ (SOPS®) workplace safety supplemental items for hospitals. *Int J Environ Res Public Health*. 2022;19(11):6815. doi:10.3390/ijerph19116815
18. Jones KJ, Skinner A, Xu L, et al. The AHRQ hospital survey on patient safety culture: a tool to plan and evaluate patient safety programs. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign)*. Agency for Healthcare Research and Quality; 2008.
19. Agency for Healthcare Research and Quality. Surveys on patient safety culture. Accessed January 22, 2022. https://www.ahrq.gov/sops/index.html
20. Mardon RE, Khanna K, Sorra J, Dyer N, Famolaro T. Exploring relationships between hospital patient safety culture and adverse events. *J Patient Saf*. 2010;6(4):226-232. doi:10.1097/PTS.0b013e3181fd1a00
21. Tawfik DS, Adair KC, Palassof S, et al. Leadership behavior associations with domains of safety culture, engagement, and health care worker well-being. *Joint Commission J Qual Patient Saf*. 2023;49(3):156-165. doi:10.1016/j.jcjq.2022.12.006
22. Tawfik DS, Profit J, Morgenthaler TI, et al. Physician burnout, well-being, and work unit safety grades in relationship to reported medical errors. *Mayo Clin Proc*. 2018;93(11):1571-1580. doi:10.1016/j.mayocp.2018.05.014
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inf*. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010
24. Measuring U. Accessed October 8, 2022. https://measuringu.com/sus/
25. Usability.gov. Accessed October 8, 2022. https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html
26. Adler-Milstein J, Adelman JS, Tai-Seale M, Patel VL, Dymek C. EHR audit logs: a new goldmine for health services research. *J Biomed Inf*. 2020;101:103343. doi:10.1016/j.jbi.2019.103343

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

27. Cohen GR, Boi J, Johnson C, Brown L, Patel V. Measuring time clinicians spend using EHRs in the inpatient setting: a national, mixed-methods study. *J Am Med Inform Assoc*. 2021;28(8): 1676-1682. doi:10.1093/jamia/ocab042
28. Dyrbye LN, Gordon J, O'Horo J, et al. Relationships between EHR-based audit log data and physician burnout and clinical practice process measures. *Mayo Clin Proc*. 2023;98(3):398-409. doi:10.1016/j.mayocp.2022.10.027
29. Arogyaswamy S, Vukovic N, Keniston A, et al. The impact of hospital capacity strain: a qualitative analysis of experience and solutions at 13 academic medical centers. *J Gen Intern Med*. 2022;37(6): 1463-1474. doi:10.1007/s11606-021-07106-8
30. Aiken LH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*. 2002;288(16):1987-1993. doi:10.1001/jama.288.16.1987
31. Aiken LH, Cimiotti JP, Sloane DM, Smith HL, Flynn L, Neff DF. Effects of nurse staffing and nurse education on patient deaths in hospitals with different nurse work environments. *J Nurs Administr*. 2012;42(10 suppl):S10-S16. doi:10.1097/01.NNA.0000420390.87789.67
32. Aiken LH, Sermeus W, Van den Heede K, et al. Patient safety, satisfaction, and quality of hospital care: cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. *BMJ*. 2012;344:e1717. doi:10.1136/bmj.e1717
33. Cimiotti JP, Aiken LH, Sloane DM, Wu ES. Nurse staffing, burnout, and health care-associated infection. *Am J Infect Control*. 2012;40(6): 486-490. doi:10.1016/j.ajic.2012.02.029
34. Sturm H, Rieger MA, Martus P, et al. Do perceived working conditions and patient safety culture correlate with objective workload and patient outcomes: a cross-sectional explorative study from a German university hospital. *PLoS One*. 2019;14(1):e0209487. doi:10.1371/journal.pone.0209487
35. Michtalik HJ, Yeh HC, Pronovost PJ, Brotman DJ. Impact of attending physician workload on patient care: a survey of hospitalists. *JAMA Int Med*. 2013;173(5):375-377. doi:10.1001/jamainternmed.2013.1864
36. Sinsky CA, Rule A, Cohen G, et al. Metrics for assessing physician activity using electronic health record log data. *J Am Med Inform Assoc*. 2020;27(4):639-643. doi:10.1093/jamia/ocz223
37. Kannampallil T, Adler-Milstein J. Using electronic health record audit log data for research: insights from early efforts. *J Am Med Inform Assoc*. 2022;30(1):167-171. doi:10.1093/jamia/ocac173
38. Lou SS, Lew D, Harford DR, et al. Temporal associations between EHR-derived workload, burnout, and errors: a prospective cohort study. *J Gen Intern Med*. 2022;37(9):2165-2172. doi:10.1007/s11606-022-07620-3


## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Burden M, McBeth L, Keniston A. The development and pilot of a novel mobile application to assess clinician perception of workload and work environment. *J Hosp Med*. 2024;19:661-670. doi:10.1002/jhm.13366

15535606, 2024, 8, Downloaded from https://shmpublications.onlinelibrary.wiley.com/doi/10.1002/jhm.13366 by CDC NCHSTP Info Ctr, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License