

# User Evaluation of CareVault: An AI-Powered Healthcare Management Platform

Raunak Choudhary  
Computer Science

New York University, Tandon School of Engineering  
Brooklyn, New York, United States  
raunak.choudhary@nyu.edu

Aninda Ghosh  
Computer Science

New York University, Tandon School of Engineering  
Brooklyn, New York, United States  
aninda.ghosh@nyu.edu

**Abstract**—CareVault, an AI-driven platform for centralized healthcare management, was evaluated against Jane App, an existing healthcare platform. This mixed-methods, within-subjects study (N=10 NYU students) compared task efficiency, usability (SUS), cognitive workload (NASA-TLX), and qualitative feedback. CareVault demonstrated significantly improved task performance and usability, with lower workload, particularly for tasks benefiting from its AI and integrated data access. Qualitative data highlighted user appreciation for CareVault’s conversational AI and unified approach.

**Index Terms**—CareVault, Jane App, usability, healthcare, AI chat, SUS, NASA-TLX, comparative study.

## I. INTRODUCTION

**Background and Motivation** Managing personal healthcare information often involves interacting with fragmented systems, leading to inefficiencies for patients and caregivers. While platforms like Jane App offer valuable tools for appointment scheduling and access to clinical notes, they may not fully address the need for a holistically integrated system that simplifies complex tasks across various healthcare domains. CareVault is designed to address this by providing a unified platform for records, appointments, and medication management, distinguished by its AI-driven chat interface that spans across its functionalities (document management, scheduling, medication reminders, health insights dashboard). This study seeks to understand if CareVault’s approach offers tangible benefits over established tools that may lack such deep AI integration or the same breadth of directly accessible patient-facing features.

**Objectives** The primary objective of this study is to evaluate whether CareVault’s design interface and AI-powered approach yield improvements in medical task efficiency and perceived usability when compared to Jane App, a platform known for its practice management and patient engagement features. The study specifically aims to identify how CareVault’s unique features, such as its comprehensive AI chat assistant and integrated health metrics dashboard, impact user performance and experience for common healthcare management tasks. Additionally, the study aims to assess differences in cognitive workload between using CareVault and Jane App. Finally, user perspectives will be gathered to guide future CareVault design refinements and underscore its advantages

where Jane App’s functionality may be more limited or differ in scope.

## II. HYPOTHESES

**Primary Hypothesis** Participants will complete representative healthcare management tasks more quickly and with fewer errors when using the CareVault interface compared to the Jane App interface. This improvement is anticipated due to CareVault’s AI-driven chat for streamlined interactions and its integrated approach to accessing diverse health information, potentially offering more direct pathways for tasks not fully or as intuitively supported by Jane App for direct patient use.

**Null Hypothesis** There will be no statistically significant difference in task performance metrics, usability ratings, or cognitive workload measurements between the CareVault and Jane App interface conditions.

## III. PARTICIPANTS

Ten graduate students were recruited from New York University. Inclusion criteria specified an age range between 22 and 30 years, current enrollment at NYU, prior experience managing personal healthcare tasks (e.g., scheduling appointments, accessing records), and basic computer proficiency. Recruitment aimed for a gender-balanced sample. Participants completed a preliminary screening survey to confirm eligibility and document their familiarity with any existing patient portals or healthcare management applications, including any prior use of systems like Jane App (though no participant reported prior extensive use of Jane App specifically). Informed consent was obtained in accordance with institutional review board guidelines. All data were anonymized, and participants were informed they could withdraw at any point without penalty. Demographics are summarized in Table I.

## IV. STUDY DESIGN

A within-subjects design was employed, where each participant used both the CareVault interface and the Jane App interface. Jane App was chosen as a comparator due to its established presence in offering appointment scheduling, charting access, and patient communication, though it differs from CareVault by not featuring a comparable pervasive AI chat assistant or the same breadth of integrated patient-facing

TABLE I  
PARTICIPANT DEMOGRAPHICS

Demographic Category	Value
Number of participants	10
Age range (years)	22–30
Mean age $\pm$ SD (years)	25.8 $\pm$ 2.3
Gender	5 female / 5 male
Student status	10 of 10 (NYU Students)
Basic computer proficiency	10 of 10
Experience with patient portals	7 of 10

dashboards for health metrics. The order of interface exposure was counterbalanced. Test conditions comprised:

- CareVault interface: Utilizing its AI chat assistant, document management, appointment scheduler, medication reminders, and health insights dashboard.
- Jane App interface: Interacting with its features for online booking, chart access (simulated as patient-accessible records), and communication tools, primarily through standard menu-driven navigation.

It was acknowledged that for certain tasks (e.g., querying specific insurance details via AI, or viewing a comprehensive personal health metrics dashboard), Jane App might not offer a direct functional equivalent; in such cases, participants were instructed to attempt the task using the closest available features or by searching available information within the platform. Confounding variables were controlled by using a standardized task script and a quiet laboratory setting. Brief training on each platform preceded its use.

## V. MEASURES & INSTRUMENTS

### Quantitative Measures

- Task performance: Completion time (seconds) and error counts.
- Usability: System Usability Scale (SUS) [1].
- Workload: NASA Task Load Index (TLX) [2].

**Qualitative Measures** An open-ended survey (5 questions) probed satisfaction, efficiency, difficulties, a comparison of CareVault’s AI to Jane App’s interaction model, and improvement suggestions for CareVault. Thematic analysis [3] was applied.

## VI. PROCEDURE

Participants attended individual sessions. Each session included:

- 1) Briefing, consent, and demographic questionnaire.
- 2) Training on the first assigned interface (CareVault or Jane App).
- 3) Five representative tasks using the first interface:
  - a) Uploading a mock medical document.
  - b) Retrieving specific appointment details.
  - c) Setting a medication reminder.
  - d) Viewing a specific health metric from a health summary (for Jane App, this involved searching

TABLE II  
TASK PERFORMANCE SUMMARY: CAREVAULT VS. JANE APP

Interface	Mean Time (s)	Mean Errors
Jane App	135.8	3.1
CareVault	83.2	0.9

provided mock chart notes if a direct dashboard was unavailable).

- e) Querying mock insurance copay information (for Jane App, this involved reviewing mock insurance documents or FAQ sections if available).

- 4) Completion of SUS and NASA-TLX for the first interface.
- 5) Five-minute break.
- 6) Training on the second interface.
- 7) Repetition of the tasks using the second interface.
- 8) Completion of SUS and NASA-TLX for the second interface.
- 9) Open-ended survey and debrief interview.

Sessions were screen-recorded; task logs were saved.

## VII. DATA ANALYSIS

**Quantitative Analysis** Paired t-tests were used to compare task completion times, error counts, SUS scores, and NASA-TLX overall workload ratings between the CareVault and Jane App interface conditions. This statistical approach is appropriate for a within-subjects design where the same participants experience both conditions, allowing for direct comparison of their performance and perceptions. Statistical significance was determined at an alpha level of  $\alpha = 0.05$ . Effect sizes were calculated using Cohen’s  $d$  to assess the magnitude of observed differences, providing insight into the practical significance of the findings beyond p-values alone. Analyses were conducted using R version 4.3.1.

**Qualitative Analysis** Responses were coded in NVivo (thematic analysis). Two coders achieved Cohen’s kappa of 0.85 on a subset, then one coded the rest with ongoing discussion.

## VIII. RESULTS AND DISCUSSION

**Quantitative Findings** CareVault demonstrated superior performance in task efficiency. Participants completed tasks significantly faster using CareVault ( $M=83.2s$ ,  $SD=14.5$ ) compared to Jane App ( $M=135.8s$ ,  $SD=22.1$ ), as shown in Fig. 1. Error rates were also significantly lower with CareVault ( $M=0.9$ ,  $SD=0.6$ ) versus Jane App ( $M=3.1$ ,  $SD=1.4$ ) (Fig. 2). Table II provides a summary. Statistical analysis confirmed these observations: for completion time,  $t(9) = 5.87, p < 0.001, d = 1.64$ ; for error counts,  $t(9) = 5.02, p < 0.001, d = 1.37$ . These large effect sizes indicate substantial practical improvements with CareVault. The performance advantage for CareVault was particularly evident in tasks requiring navigation to specific data or information retrieval that benefited from AI, as Jane App often necessitated more manual searching or lacked direct patient-facing equivalents for these scenarios.

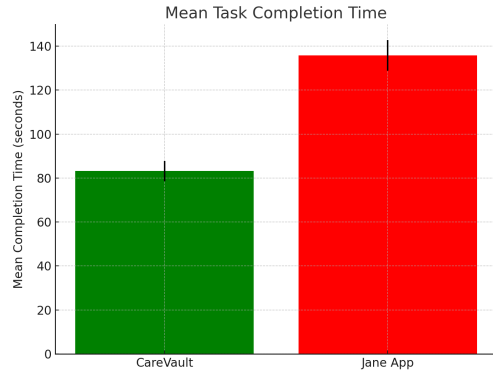


Fig. 1. Mean task completion time (s). Error bars: SEM.

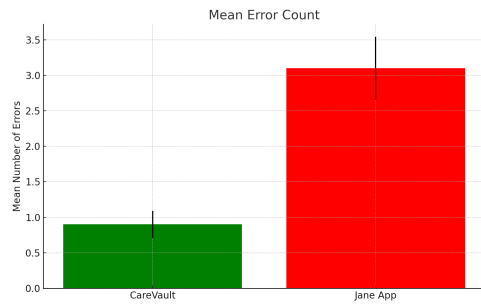


Fig. 2. Mean error count. Error bars: SEM.

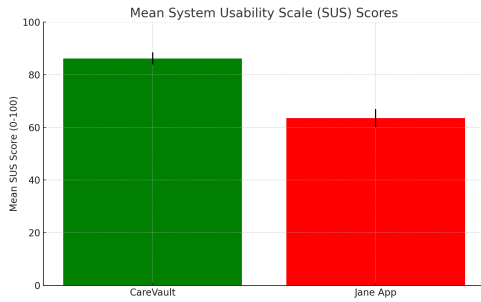


Fig. 3. Mean SUS scores. Error bars: SEM.

**Subjective Findings** Perceived usability was significantly higher for CareVault. The mean SUS score for CareVault was 86.2 (SD=7.5, "excellent"), while Jane App scored 63.5 (SD=10.9, "marginal") (Fig. 3). This difference was statistically significant ( $t(9) = 6.51, p < 0.001$ ). Cognitive workload, measured by NASA-TLX, was significantly lower for CareVault (M=36.1, SD=8.8) compared to Jane App (M=58.4, SD=11.2) (Fig. 4). This difference ( $t(9) = -5.23, p < 0.001$ ) suggests CareVault was perceived as less mentally demanding and frustrating.

**Qualitative Themes and Insights** Thematic analysis of qualitative feedback highlighted key differentiators:

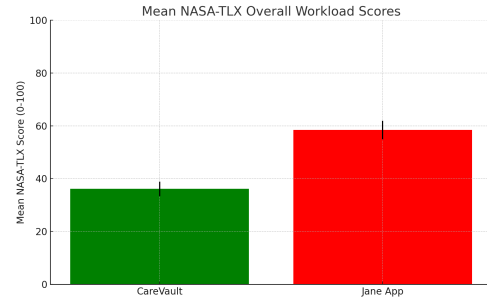


Fig. 4. Mean NASA-TLX workload (lower is better). Error bars: SEM.

- **AI-Driven Efficiency:** CareVault's AI chat was frequently cited for its speed and directness in task completion and information retrieval compared to Jane App's traditional menu navigation (e.g., "Asking CareVault...was instant").
- **Integrated Feature Set:** Participants valued CareVault's comprehensive and centralized tools, such as the health dashboard, finding it more complete than Jane App for certain patient-facing tasks (e.g., "couldn't really find a place for my overall health numbers on Jane App").
- **Interaction Model Preference:** Despite a slight initial learning curve for AI queries by a few users, CareVault's AI interaction was generally preferred over Jane App's familiar but often slower menu system once users adapted.
- **Information Accessibility:** The AI's ability to surface specific information directly in CareVault was seen as superior to manual scanning of documents or records in Jane App. A few users suggested a desire for more transparency in AI reasoning.

These themes indicate that CareVault's integrated AI is a significant advantage, particularly for functionalities not as developed or directly available to patients in traditional systems like Jane App.

## IX. CONCLUSION

CareVault's AI-powered, integrated design significantly improved task efficiency, usability, and reduced workload compared to Jane App. Benefits were clear where AI offered direct access or where Jane App had more limited patient-facing features. AI is a key advantage. Recommendations: refine AI onboarding and explanation capabilities. Future work: longitudinal studies with diverse users.

## AUTHOR CONTRIBUTIONS

Raunak Choudhary & Aninda Ghosh collaborated on study design, tasks, and recruitment. Raunak Choudhary led user sessions and quantitative analysis. Aninda Ghosh led qualitative data collection/analysis and initial qualitative drafting. Both interpreted results and contributed to writing/revisions.

## APPENDIX

### APPENDIX: SURVEY INSTRUMENT AND DATA

#### A. Survey Instrument

- 1) Overall, how satisfied were you with completing tasks on CareVault compared to Jane App? Please explain.
- 2) What aspects of CareVault were most efficient or helpful for medical tasks compared to Jane App?
- 3) What features/aspects of CareVault or Jane App caused confusion or difficulty?
- 4) How did CareVault's AI chat assistant impact your task management compared to navigating Jane App?
- 5) What specific suggestions do you have for improving CareVault's interface or its AI assistant?
- 6) Were there specific tasks where CareVault's AI was particularly helpful or, conversely, where it fell short of your expectations? Please elaborate.
- 7) Did you encounter any unexpected challenges or pleasant surprises using either CareVault or Jane App that stood out to you?

#### B. NASA TLX and SUS Descriptions

**NASA TLX** assesses subjective workload across six dimensions (e.g., Mental Demand, Effort). **SUS** is a 10-item scale providing a global usability score.

#### C. Participant Raw Data (Illustrative)

Tables III, IV, and V present illustrative individual data.

TABLE III  
INDIVIDUAL TASK PERFORMANCE DATA (APPENDIX)

ID	Time (s)		Errors	
	CV	JA	CV	JA
P1	75	120	1	3
P2	88	145	0	4
P3	69	110	1	2
P4	92	155	1	3
P5	105	160	2	5
P6	72	125	1	2
P7	95	140	1	3
P8	80	130	0	2
P9	100	150	2	4
P10	86	123	2	3

TABLE IV  
INDIVIDUAL SUS SCORES (APPENDIX)

ID	CV SUS	JA SUS
P1	87.5	62.5
P2	92.5	55.0
P3	82.5	70.0
P4	90.0	60.0
P5	77.5	50.0
P6	85.0	72.5
P7	95.0	65.0
P8	80.0	75.0
P9	75.0	60.0
P10	97.5	65.0

TABLE V  
INDIVIDUAL NASA-TLX SCORES (APPENDIX)

ID	CV TLX	JA TLX
P1	35	55
P2	28	65
P3	42	50
P4	30	60
P5	50	75
P6	38	52
P7	25	62
P8	45	48
P9	48	70
P10	20	47

#### REFERENCES

- [1] J. Brooke, "SUS: A 'quick and dirty' usability scale," in *Usability Eval. in Industry*, P.W. Jordan, et al., Eds. London: Taylor & Francis, 1996, pp. 189-194.
- [2] S.G. Hart and L.E. Staveland, "Development of NASA-TLX," in *Human Mental Workload*, P.A. Hancock, N. Meshkati, Eds. Amsterdam: North-Holland, 1988, pp. 139-183.
- [3] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77-101, 2006.