Healthcare and Medical Devices, Vol. 130, 2024, 47–54
https://doi.org/10.54941/ahfe1004836

Exploring the Usability and User Experience of the Symptom Checker Interface

Wei An Lin[1], Meng-Cong Zheng[2], and Li-Jen Wang[3]

[1]Department of Industrial Design Master Program of Innovation and Design, Taipei Tech, Taipei,10608, Taiwan (R.O.C)
[2]Department of Industrial Design, Taipei Tech, Taipei,10608, Taiwan (R.O.C)
[3]Doctoral Program in Design, College of Design, Taipei Tech, Taipei 10608, Taiwan (R.O.C.)

ABSTRACT

The symptom checker is an application that assists patients in self-assessment and determining whether to seek medical help. Therefore, the functionality and user experience of the interface can influence users' trust in the symptom checker. This study employed task experiments, the System Usability Scale (SUS), the User Experience Questionnaire (UEQ), the Decision Attitude Scale (DAS), and semi-structured interviews to investigate the extent to which existing symptom checkers meet users' needs. This study findings indicate that, although most participants do not fully trust the results of the symptom checker, they believe that detailed descriptions of symptoms and treatment methods in medical advice can enhance trust. In addition to providing sufficient information to improve trust in the symptom checker, the assistance of a human body model also contributes to users describing their physical condition with a better experience. The quantity of information and the coordination with the aid of a human body model will be crucial aspects of design. These results can serve as references for subsequent design improvements.

Keywords: Symptom checker, Usability, User interface, User experience

INTRODUCTION

The symptom checker is an application designed to assist patients in making pre-consultation decisions, allowing patients to use these tools to supplement medical advice and as decision support tools to recommend whether they should seek immediate medical assistance(Schmieding, Mörgeli, Schmieding, Feufel, & Balzer, 2021). Many healthcare websites in the United States have begun offering symptom checkers as a service for educational purposes or to enhance the patient's consultation experience (Schmieding et al., 2021).

Symptom checkers are increasingly popular today, with over 70% of individuals aged 18 to 39 in the United Kingdom using them (Aboueid, Meyer, Wallace, Mahajan, & Chaurasia, 2021). Over 10 million Germans have also used symptom checkers (EPatient survey, 2020). Internet entrepreneur Jeff Arnold founded WebMD in 1996, attracting 75 million monthly visitors and 52 million mobile users. Currently, many healthcare websites strive to meet patients' needs (Olesch, 2019).

Most users need an understanding of the technology behind symptom checkers. Research indicates that the artificial intelligence systems employed in symptom checkers are highly opaque and challenging for users to comprehend, potentially impacting the trust users place in these systems(Tsai, You, Gui, Kou, & Carroll, 2021). For a symptom checker to contribute effectively, it must garner the trust of users, who must follow the recommendations the symptom checker provides. Lack of faith in a symptom checker can restrict its influence on the healthcare industry. Therefore, enhancing the trustworthiness of symptom checkers is crucial for better user engagement (Fan et al., 2021). According to studies, good usability enables users to find the necessary information more easily and quickly, enhancing the user experience. Conversely, poor usability in symptom checkers may lead to user confusion and erroneous judgments(Ahmad Faudzi, Che Cob, Omar, Sharudin, & Ghazali, 2023).

This study aims to research the interface of symptom checkers, encompassing aspects such as operational workflows, interface information, and the presentation of symptoms and corresponding medical advice. After testing the outcomes and incorporating them into the design, a subsequent evaluation will be performed to assess whether user experience and operations have been improved. The ultimate goal is to enhance users' trust in the symptom checker by refining its design based on empirical testing and evaluation.

METHODS

We curated a selection of medical websites currently offering symptom checkers based on criteria such as coverage, diagnostic accuracy, and appropriateness of medical advice (Stephen et al., 2020). Ultimately, two symptom checkers, Isabel and WebMD, were chosen for testing. A total of 30 participants aged 20 to 39 were invited to participate in the experiment to evaluate the usability and user experience of the symptom checker interfaces.

For this experiment, six standardized case vignettes were selected from a pool of 45 vignettes (Semigran, Linder, Gidengil, & Mehrotra, 2015) to simulate real-life scenarios of using a symptom checker for symptom inquiries, as depicted in Table 1. Each participant performed six case vignettes tasks using one of the symptom checkers, with tasks one through six being presented in random order. After completing each simulated task, participants provided answers on the symptoms and medical advice offered by the symptom checker to assess its accuracy.

Table 1. Six case vignettes tasks.

Task	Symptoms	Medical Advice	Clinical Vignettes
Task1	Kidney stones	Emergency Cases	A 45-year-old white man presents to the emergency department with a 1-hour history of sudden onset of left-sided flank pain radiating down toward his groin. The patient is writhing in pain, which is unrelieved by position. He also complains of nausea and vomiting.

(Continued)

Table 1. Continued

Task	Symptoms	Medical Advice	Clinical Vignettes
Task2	Stroke	Emergency Cases	A 70-year-old man with a history of chronic HTN and atrial fibrillation is witnessed by a family member to have nausea, vomiting, and right-sided weakness, as well as difficulty speaking and comprehending language. The symptoms started with only mild slurred speech before progressing over several minutes to severe aphasia and right arm paralysis. The patient is taking warfarin.
Task3	Influenza	Non-Emergent	A 30-year-old woman presents in January with 2-day history of fever, cough, headache, and generalized weakness. She was in her usual state of health before an abrupt onset of these symptoms. A few viral illnesses have affected her during the current winter, but not to this severity. She reports sick contacts at work and did not receive the seasonal influenza vaccine this season.
Task4	Urinary tract infection	Non-Emergent	A 26-year-old female newly wed presents complaining of painful urination, feeling of urgent need to urinate, and more frequent urination for 2 days. She denies any fever, chills, nausea, vomiting, back pain, vaginal discharge, or vaginal pruritus.
Task5	Eczema	Self-Care	A 12-year-old female presents with dry, itchy skin that involves the flexures in front of her elbows, behind her knees, and in front of her ankles. Her cheeks also have patches of dry, scaly skin. She has symptoms of hay fever and has recently been diagnosed with egg and milk allergy. She has a brother with asthma and an uncle and several cousins who have been diagnosed with eczema.
Task6	Acute bronchitis	Self-Care	Mrs. L is a 61 year-old woman who presents with 4 days of a cough productive of yellow sputum. Her symptoms started 4 days ago with rhinorrhea and productive cough. She initially had fevers as high as 101 for 2 days, but those have now resolved. In the office, she has normal vital signs and a normal physical examination. She is otherwise healthy except for high cholesterol for which she is being treated with atorvastatin. She has no drug allergies.

Following the completion of the tasks, participants were instructed to complete the System Usability Scale (SUS), the User Experience Questionnaire (UEQ), and the Decisional Attitude Scale (DAS) about their interactions with the symptom checker during the experiment. These surveys aimed to capture participants' perspectives on the usability, satisfaction, and decision confidence regarding the symptom checker interface. Subsequently, semi-structured interviews were conducted to comprehensively understand participants' overall experiences and evaluations during task operations, serving as valuable references for future improvements.

RESULTS AND DISCUSSION

According to the experimental results, under identical task conditions, the number of correct answers for symptoms was higher for WebMD (M = 3.20, SD = 0.94) compared to Isabel (M = 2.87, SD = 0.99). Isabel (M = 4.00, SD = 0.75) surpassed WebMD (M = 3.60, SD = 1.12) for the number of correct answers regarding medical advice. Nevertheless, the results of independent samples t-tests indicated no significant differences in both data sets.

We also delved deeper into the number of correct responses for symptoms and medical advice for each task (i.e., how many out of 15 participants answered correctly), as illustrated in Tables 2 and 3. Through independent samples t-tests, the statistical results showed significant differences only in Task 3 for symptoms (p = 0.005) and medical advice (p = 0.007).

Table 2. The number of participants providing correct answers for symptoms in each task.

Task	Symptom Checkers	N	Number of Correct Answers	t	p
Task1	Isabel	15	6	1.183	0.247
	WebMD	15	3		
Task2	Isabel	15	15	1.187	0.082
	WebMD	15	12		
Task3	Isabel	15	7	3.130	**0.005****
	WebMD	15	14		
Task4	Isabel	15	9	0.000	1.000
	WebMD	15	9		
Task5	Isabel	15	5	1.871	0.072
	WebMD	15	10		
Task6	Isabel	15	1	1.000	0.334
	WebMD	15	0		

* p < 0.05. ** p < 0.01. *** p < 0.001.

Table 3. The number of participants providing correct answers for medical advice in each task.

Task	Symptom Checkers	N	Number of Correct Answers	t	p
Task1	Isabel	15	14	1.058	0.301
	WebMD	15	12		
Task2	Isabel	15	15	1.468	0.164
	WebMD	15	13		
Task3	Isabel	15	13	2.928	**0.007****
	WebMD	15	6		
Task4	Isabel	15	12	0.418	0.679
	WebMD	15	11		
Task5	Isabel	15	1	1.871	0.075
	WebMD	15	5		
Task6	Isabel	15	5	0.367	0.716
	WebMD	15	6		

* p < 0.05. ** p < 0.01. *** p < 0.001.

The System Usability Scale (SUS) scores indicate that the overall usability scores for Isabel (M = 67.50, SD = 18.40) and WebMD (M = 47.50, SD = 20.59). The results of an independent samples t-test show a significant difference in the usability of the two symptom checker interfaces (p = 0.024). Combining the SUS scores to assess the usability of both systems, neither Isabel (67.50) nor WebMD (47.50) met the usability passing standard (68). They fall into the D and F rating categories, respectively, as depicted in Figure 1. This suggests that both symptom checker interfaces do not fully meet the operational needs of the participants, indicating room for improvement.

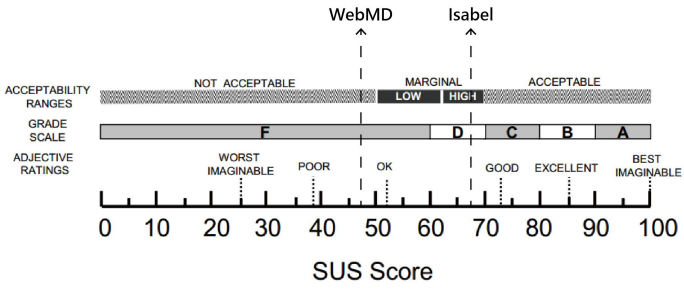

Figure 1: Average SUS score and rate of Isabel and WebMD.

According to the interpretation of "Pragmatic Quality" and "Hedonic Quality" in the −0.8 and 0.8 represent a neural evaluation of the corresponding scale, values > 0.8 represent a positive evaluation, and values < −0.8 describe a negative review. The range of the scales is between −3 (horribly bad) and +3 (extremely good). Experimental data indicate that, except for Isabel's pragmatic quality (M = 0.90, SD = 1.11) showing a positive evaluation, the remaining data in terms of "Pragmatic Quality," "Hedonic Quality," and "Overall" did not meet the criteria for a positive evaluation. This suggests areas for optimization in both Isabel and WebMD, as shown in Table 4. Independent samples t-test results indicate no significant difference in overall user experience between the two symptom checkers. However, when analyzing pragmatic quality and hedonic quality separately, a significant difference is observed in Pragmatic quality, where Isabel (M = 0.90, SD = 1.11) differs significantly from WebMD (M = −0.26, SD = 1.31) (p = 0.014).

Table 4. Statistical analysis of the UEQ scores for Isabel and WebMD.

User Experience	Symptom Checkers	N	M	SD	t	p
Pragmatic Quality	Isabel	15	0.90	1.11	2.619	**0.014***
	WebMD	15	−0.26	1.31		
Hedonic Quality	Isabel	15	0.08	1.39	0.000	1.000
	WebMD	15	0.08	1.41		
Overall	Isabel	15	0.49	1.20	1.303	0.203
	WebMD	15	−0.09	1.25		

* p < 0.05. ** p < 0.01. *** p < 0.001.

The DAS (Decision Attitude Scale) utilizes a Likert Scale with five-point scores. According to the results, the overall average score for Isabel (M = 24.80, SD = 5.71) is higher than that of WebMD (M = 22.73, SD = 3.95). However, independent samples t-test results indicate no statistically significant difference between the two. The DAS includes three components: "Satisfaction with the Choice," "Usability of Information," and "Adequacy of Information." Data reveals that WebMD only has a higher average score (M = 2.06, SD = 0.073) than Isabel (M = 1.76, SD = 0.776) in "Adequacy of Information," but without significant difference. However, in "Usability of Information," Isabel's average score (M = 2.93, SD = 0.863) is higher than WebMD (M = 2.23, SD = 0.752) with a significant difference (p = 0.025), as shown in Table 5.

Table 5. Statistical analysis of the DAS scores for Isabel and WebMD.

Decision Attitude	Symptom Checkers	N	M	SD	t	p
Satisfaction with the Choice	Isabel	15	3.08	0.867	0.947	0.352
	WebMD	15	2.82	0.565		
Usability of Information	Isabel	15	2.93	0.863	2.367	**0.025***
	WebMD	15	2.23	0.752		
Adequacy of Information	Isabel	15	1.76	0.776	0.690	0.277
	WebMD	15	2.06	0.703		

* p < 0.05. ** p < 0.01. *** p < 0.001.

We conducted interviews based on the challenges faced by participants in the operational processes of using the symptom checker. The interview content serves as supplementary information to the experiment results, and suggestions for improvement and optimization are proposed for subsequent new design proposals. Feedback from the interviews can be broadly categorized into preferences and sentiments, symptom screening processes, medical advice interface, and suggestions for additional features.

In the section on preferences and sentiments, 80% of participants indicated that they would still rely on their own experiences to interpret the results after using the symptom checker. Still, it could serve as a reference before consulting a doctor. 73% of participants expressed that WebMD is perceived as more trustworthy than Isabel and provides a better understanding and diagnosis of their conditions. Regarding the symptom screening process, 67% of participants believed that WebMD's human body model assists in identifying corresponding symptoms. Additionally, 47% of participants felt that Isabel, with the option to input users' symptoms directly, would make them think Isabel received comprehensive information. This is illustrated in Figure 2.

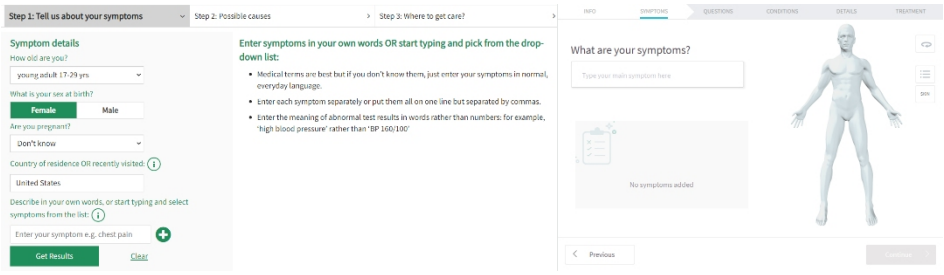

Figure 2: Isabel symptom screening process (Left), WebMD symptom screening process (Right).

In the medical advice interface section, 70% of the participants believe that WebMD provides detailed explanations and treatment methods for symptoms, allowing them to confirm better which symptom aligns more with their conditions. However, a minority of participants find the information on symptom explanations a bit too much, making them less inclined to read in detail. WebMD also presents the matching degree between the entered data and potential symptoms, with 53% of participants expressing a sense of trust in the results. Isabel indicates the urgency of symptoms, and 47% of participants find this interface and functionality clear and intuitive. However, some participants feel that the gradient colors may be visually challenging to distinguish, as shown in Figure 3.

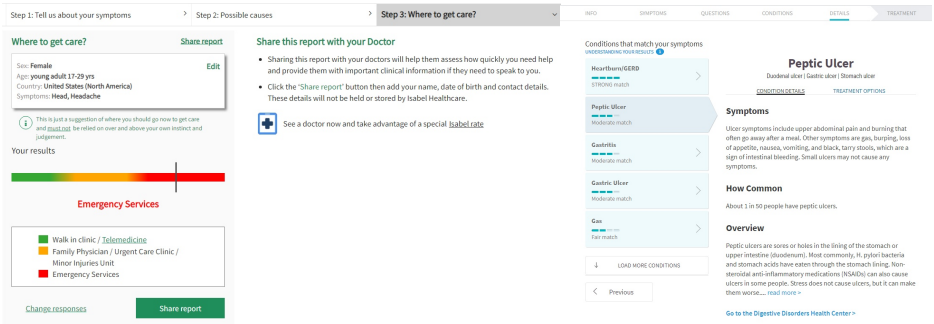

Figure 3: Isabel medical advice interface (Left), WebMD medical advice interface (Right).

In the section on suggestions for additional features, 50% of the participants believe that it would be beneficial to add the functionality of locating nearby hospitals along with recommendations for clinics or appointment scheduling. Additionally, 20% of the participants suggest including a feature allowing users to share their reports with healthcare professionals.

CONCLUSION

The experimental results reveal that the accuracy rates of the two symptom checkers for diagnosing symptom are only 48% and 53%, respectively. Moreover, based on the System Usability Scale (SUS) and the User Experience

Questionnaire (UEQ) scores, there is still significant room for improvement in the usability and user experience of the functional interfaces of both symptom checkers. Although it is currently challenging to determine whether usability and user experience impact the accuracy of the symptom checkers, enhancements and optimizations in symptom screening and the interface of medical advice could contribute to increased trust and user experience.

The findings of this study are advantageous for understanding the relationship between the interface content of symptom checkers and user experience. Future designs of symptom checkers should effectively guide users in selecting symptoms and obtaining accurate information through the presentation of interfaces, ultimately enhancing the accuracy of symptom checkers and elevating user trust levels.

REFERENCES

Aboueid, S., Meyer, S., Wallace, J. R., Mahajan, S., & Chaurasia, A. (2021). Young Adults' Perspectives on the Use of Symptom Checkers for Self-Triage and Self-Diagnosis: Qualitative Study. *JMIR Public Health Surveill, 7*(1), e22637. doi:10.2196/22637

Ahmad Faudzi, M., Che Cob, Z., Omar, R., Sharudin, S. A., & Ghazali, M. (2023). Investigating the User Interface Design Frameworks of Current Mobile Learning Applications: A Systematic Review. *13*(1), 94.

EPatient survey. (2020). Health & Care Management. Retrieved from https://www.hcm-magazin.de/epatient-survey-2020-digital-health-studie-271773/.

Fan, X., Chao, D., Zhang, Z., Wang, D., Li, X., & Tian, F. (2021). Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study. *J Med Internet Res, 23*(1), e19928. doi:10.2196/19928

Olesch, A. (2019). Symptom Checkers Will See You Now. Startups Race To Create A Perfect Health Concierge. *ICT &health*. Retrieved from https://ictandhealth.com/symptom-checkers-will-see-you-now-startups-race-to-create-a-perfect-health-concierge/news/.

Schmieding, M. L., Mörgeli, R., Schmieding, M. A. L., Feufel, M. A., & Balzer, F. (2021). Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study. *J Med Internet Res, 23*(3), e24475. doi:10.2196/24475

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ, 351*, h3480. doi:10.1136/bmj.h3480

Stephen, G., Alicia, M., Adel, B., Caoimhe, C., Jean, C., Hamish, F.,... Claire, N. (2020). How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. *BMJ Open, 10*(12), e040269. doi:10.1136/bmjopen-2020-040269

Tsai, C.-H., You, Y., Gui, X., Kou, Y., & Carroll, J. M. (2021). *Exploring and Promoting Diagnostic Transparency and Explainability in Online Symptom Checkers*. Paper presented at the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445101