Usability testing of radiotherapy systems as a medical device evaluation tool to inform hospital procurement decision-making

Mingyin Jiang, Xuancheng Tu, Wanchao Xiao, Jinhui Tang, Qiang Li, Dongjie Sun and Daoxiong Wang
Department of Medical Equipment, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Abstract
Purpose: Poor usability designs of radiotherapy systems can contribute to use errors and adverse events. Therefore, we evaluated the usability of two radiotherapy systems through radiation therapists' performance, workload, and experience that can inform hospital procurement decision-making about the selection of appropriate radiotherapy system for radiation therapist use.

Methods: We performed a comparative usability study for two radiotherapy systems through user testing. Thirty radiation therapists participated in our study, in which four typical operational tasks were performed in two tested radiotherapy systems. User performance was measured by task completion time and completion difficulty level. User workloads were measured by perceived and physiological workload using NASA-TLX questionnaires and eye motion data. User experience was measured by the USE questionnaire.

Results: Significantly less task completion time and an easier task completion difficulty level were shown with the Varian Trilogy than with the XHA600E. The study results suggest that higher perceived and physiological workloads were experienced with the XHA600E than with the Varian.

Corresponding authors:
Dongjie Sun, Department of Medical Equipment, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1095 JieFang Avenue, Wuhan 430030, China.
Email: 2217644701@qq.com

Daoxiong Wang, Department of Medical Equipment, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1095 JieFang Avenue, Wuhan 430030, China.
Email: dxwang988@yeah.net

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Trilogy. Radiation therapists reported better user experience with the Varian Trilogy than with the XHA600E. Five paired t-tests regarding user performance, user workload, and user experience between the Varian Trilogy and the XHA600E were performed, showing that the Varian Trilogy radiotherapy system has a better usability design than the XHA600E radiotherapy system.

**Conclusions:** Based on study results, we confirmed that the Varian Trilogy radiotherapy system has a better usability design than the XHA600E radiotherapy system. Furthermore, the study results provide valuable evidence for hospital procurement decision-making regarding the selection of a suitable radiotherapy system for radiation therapists to use.

**Keywords**
Radiotherapy system, usability testing, decision-making, usability design, evaluation tool

**Introduction**
It has been proposed that approximately 50% of cancer patients should receive radiation therapy. However, even though recent technological advances have significantly improved the quality of radiation treatment, the complexity of radiotherapy administration has been accompanied with numerous human use errors, resulting in injury to patients. For instance, a U.S. Nuclear Regulatory Commission (NRC) report revealed that more than 60% of radiation therapy related accidents resulted from human user errors. A separate report by Task Group 100 on failures and effects of intensity-modulated radiation therapy (IMRT) treatment confirmed NRC findings. The traditional reaction of the medical institution to adverse events was blame the clinicians for being careless. On the flipside, other reports show that human use errors occurring during radiotherapy treatment are occasioned by poorly designed medical devices, which mostly overlook the characteristics of end users.

The importance of medical device designs in ensuring safe use has been echoed by several government regulatory bodies and international medical associations. For example, the U.S. Food and Drug Administration (FDA) guidelines requires that a medical device should be well designed and must be safe and effective for the intended users, uses, and use environments. Moreover, manufacturers should provide usability data of medical devices before floating them on the market. The FDA is charged with suitability assessments of medical equipment before making recommendations for their use. In addition, the International Electrotechnical Commission (IEC) has stipulated several quality standards required for design and use of medical devices. However, most manufacturers do not disclose technical functionality aspects of their medical equipment, which are not deliberately to get a statistical significance in validating the medical device usability design. One common method of evaluating suitability of a system or medical device is performing usability tests. Indeed numerous usability tests have led to modification and improvement of design for several medical devices such as ventilators, infusion pumps, defibrillators, mobile medical ultrasound devices, and laparoscopic devices. Several usability studies on radiation therapy have revealed that poor designs results in human use errors and reduces user performance. Flaws in
the user interface of radiotherapy systems results in usability errors during clinical practice. Currently, most medical devices in clinical practice such as radiotherapy systems do not undergo usability tests prior procurement, thus safety and the suitability of the selected systems remain largely unclear.

The intent of this study is to use the results of user testing to compare the usability of two different radiotherapy systems. This study evaluated the usability design of different radiotherapy systems in three research areas, including user performance, user workload, and user experience. The methodology used in this study can be applied by end users to select the best radiotherapy systems currently available on the market prior procurement.

Materials and methods

Radiotherapy system

This study was performed in a university affiliated hospital in Wuhan, China. The tested radiotherapy systems in our usability study were the Varian Trilogy® linear accelerator system (Varian Medical Systems, California, USA) and the Shinva XHA600E system (Shinva Medical Instrument, Shandong, China).

Participants

A total of 30 experienced radiation therapists participated in our usability study. The radiation therapists have the basic knowledge of radiotherapy and experience with radiotherapy systems (16 radiation therapists with 1–5 years of experience, 7 with 5–10 years, and 7 with more than 10 years of experience) participated in this study. Before formal user testing, the researcher provided a radiotherapy system (the Varian Trilogy® linear accelerator system and the Shinva XHA600E system) operation training course for all radiation therapists who participated in our usability study. A researcher would be available to answer any questions that the radiation therapists. Informed consent for this usability study was collected from the radiation therapists before formal user testing.

Scenarios and tasks

A total of four test scenarios were designed by our research team for participants to complete to evaluate the usability of the radiotherapy system. The specific tasks for participants to complete for the radiotherapy system are shown in Table 1.

User performance

User performance was based on task completion time and task completion difficulty level. Participants’ completion time for each task was recorded by the researcher. Moreover, participants rated the difficulty level of task on a Likert scale ranging from 1 (very easy) to 9 (very difficult). The lesser the time spend to
complete a task and the lesser the completion difficulty, and the better user performance.

**User workload**

The workload of the participants was evaluated with physiological and perceived workload.

| Scenario                                                                 | Test tasks                                                                 | Task content                                                                 |
|--------------------------------------------------------------------------|---------------------------------------------------------------------------|----------------------------------------------------------------------------|
| 1. The radiotherapy treatment plan of specific patient is in the         | Searching for the patient's treatment plan and uploading it                | • Search patient radiotherapy treatment plan with given ID number |
| radiotherapy system. The participant should find and review it and then   |                                                                           | • Check patient name and gender                                           |
| upload it to system.                                                      |                                                                           | • Upload the radiotherapy treatment plan                                   |
|                                                                          | Reviewing the patient's treatment plan information                        | • Check patient radiotherapy treatment plan type                           |
|                                                                          |                                                                           | • Check the total number of times that the patient needs to receive       |
|                                                                          |                                                                           | radiotherapy treatment                                                     |
|                                                                          |                                                                           | • Check the number of times the patient has received radiotherapy treatment|
|                                                                          |                                                                           | • Check the number of times fields in this radiotherapy treatment plan     |
|                                                                          |                                                                           | • Check the number of treatment fields in this radiotherapy treatment plan |
|                                                                          |                                                                           | • Check the number of treatment fields in this radiotherapy treatment plan |
|                                                                          |                                                                           | • Check whether every treatment field has treatment accessory             |
| 2. The radiotherapy treatment plan of patient has been uploaded to       |                                                                           | Report the name of the treatment field undergoing beam delivery           |
| system. The participant needs to review the treatment plan information.  |                                                                           | Report the number of treatment fields that have been treated              |
|                                                                          |                                                                           | Report the value of plant treatment time and actual dose rate for the      |
|                                                                          |                                                                           | current treatment field                                                    |
|                                                                          |                                                                           | Report the value of Couch Vrt, Couch Lng, and Couch Lat for the current    |
|                                                                          |                                                                           | treatment field                                                           |
|                                                                          |                                                                           | Identify the interlock alarm type                                          |
|                                                                          |                                                                           | Report the interlock alarm type                                           |
| 3. The participant should report the parameter value of the current      | Reading and reporting the treatment parameter values during beam delivery  | • Report the name of the treatment field undergoing beam delivery           |
| treatment field while the patient is receiving radiation therapy         |                                                                           | • Report the number of treatment fields that have been treated            |
|                                                                          |                                                                           | • Report the value of plant treatment time and actual dose rate for the    |
|                                                                          |                                                                           | current treatment field                                                    |
|                                                                          |                                                                           | • Report the value of Couch Vrt, Couch Lng, and Couch Lat for the current   |
|                                                                          |                                                                           | treatment field                                                           |
|                                                                          |                                                                           | Identify the interlock alarm type                                          |
|                                                                          |                                                                           | Report the interlock alarm type                                           |
| 4. The participant should report the alarm content to the researcher     | Reading and reporting alarm content                                        | • Identify the interlock alarm type                                        |
| when the radiotherapy system alarm sounds                                |                                                                           | • Report the interlock alarm type                                          |
Physiological workload. Physiological workload was assessed using blink rate and blink duration. The eye motion data were sampled through Tobii Glasses 2 Eye Tracker (Tobii Technology, Danderyd, Sweden). Several published studies have applied the blink rate and blink duration to evaluate participants’ mental fatigue. These studies have been conducted according to data from ventilators, surgeons, and radiation therapists. Overall, blink rate and blink duration are inversely proportional to the degree of physiological workload.

Perceived workload. Perceived workload was evaluated based on the NASA Task Load Index (NASA-TLX) questionnaires. The subjective workload perception was based on six parameters; mental demand, temporal demand, physical demand, frustration, performance, and effort. The NASA-TLX questionnaires have been extensively used in evaluating workload experienced by radiation therapists. The output of the NASA-TLX questionnaires is scored from 0 to 100, with lower scores indicating the lower perceived workload.

User experience

User experience was evaluated using the USE questionnaire, which relies on four parameters; usefulness, ease of use, ease of learning, and satisfaction. The USE questionnaire comprises 30 questions, each rated on a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The USE questionnaire has previously been used in the healthcare settings to evaluate user experience of ventilators, radiologists, and M-health care providers. Higher USE scores correspond with better user experience.

Study procedure

The user testing was conducted at a university affiliated hospital in Wuhan, Hubei Province, central China. The protocol for this study was approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (IORG No.: IORG0003571).

Two radiotherapy systems were assigned for participants to perform user testing tasks administered by study researchers in random order. Before the tests, the eye tracker was calibrated according to the manufacturer’s protocol. After completing the calibration process, a researcher would stand nearby to inform participants of the task to complete. Each participant only had one attempt to perform each task. Once a task was completed, the participants’ task completion time and the eye motion data were recorded. Then, the task completion difficulty level questionnaire was completed by participants for the task. After participant completed four tasks in one radiotherapy system, the NASA-TLX questionnaires and the USE questionnaire were completed by the participant. Afterward, the participant could continue to the rest of the radiotherapy system to perform the same tasks. Participants could have rest whenever they wanted.
**Data collection**

When participants performed task in radiotherapy system, their eye motion data were sampled through the eye tracker. Moreover, the task completion time was recorded, and the task completion difficulty level was rated for each task. After the participant completed all four tasks in one radiotherapy system, the NASA-TLX questionnaires and the USE questionnaire were completed by the participant before he or she performed tasks in another radiotherapy system.

**Statistical analysis**

Continuous variables were expressed as mean ± SD. Differences in user performance between the radiotherapy systems were compared using paired student’s t tests. The relationships among user performance, user workload, and user experience of participants when using a particular radiotherapy system were analyzed using Spearman correlation coefficient. p < 0.05 was considered statically significant. Data was analyzed using SPSS software, V. 20 (IBM Corporation, Armonk, New York).

**Results**

**User performance**

The time taken to complete each task and the total time taken for all tasks for each radiotherapy system are shown in Figure 1. Significant differences were found for task completion times. Participants needed more time to complete their review of patients’ treatment plan information with the Shinva XHA600E (63.90 ± 22.34) than with the Varian Trilogy (42.50 ± 14.32). For the task of reading and reporting alarm content, participants need to spend more time to obtain alarm information with the Shinva XHA600E (17.97 ± 7.57) than with Varian Trilogy (12.10 ± 5.86). Overall, participants used less time to complete all tasks using the Varian Trilogy (129.07 ± 26.26) than with the Shinva XHA600E system (159.73 ± 32.74), implying that the Varian Trilogy displays better performance than the Shinva XHA600E system.

Regarding difficulty in completing a task (Table 2), overall, Varian system offered the best platform for generating and organizing a patient’s treatment plan (Varian: 1.23 ± 0.50; Shinva: 3.20 ± 2.02), reviewing patients treatment plan (Varian: 1.60 ± 1.00; Shinva: 3.23 ± 1.83), interpreting treatment outcome (Varian: 1.53 ± 0.94, Shinva: 3.20 ± 1.73), and interpreting patient report (Varian: 1.63 ± 0.93, Shinva: 2.83 ± 1.80). Moreover, participants also thought it was easier to perform all tasks with the Varian Trilogy (1.77 ± 0.94) than with the Shinva XHA600E (3.30 ± 1.66), which indicates that participants performing tasks (tasks list in Table 1) with the Varian Trilogy can complete their work more easily.
**Task 3.** Task completion difficulty level between radiotherapy systems.

| Task                                                                 | Varian Trilogy | Shinva XHA600E | t     | df  | p     |
|----------------------------------------------------------------------|----------------|----------------|-------|-----|-------|
| Task 1: searching for the patient’s treatment plan and uploading it | 1.23 0.50      | 3.20 2.02      | -5.03 | 29  | <0.001|
| Task 2: reviewing the patient’s treatment plan information          | 1.60 1.00      | 3.23 1.83      | -4.28 | 29  | <0.001|
| Task 3: reading and reporting the treatment parameter values during beam delivery | 1.53 0.94      | 3.20 1.73      | -4.95 | 29  | <0.001|
| Task 4: reading and reporting alarm content                         | 1.63 0.93      | 2.83 1.80      | -3.46 | 29  | 0.002 |
| Task completion difficulty level for all tasks                      | 1.77 0.94      | 3.30 1.66      | -4.49 | 29  | <0.001|

df: degree of freedom; Mean: the difficulty level of task on a Likert scale ranging from 1 (very easy) to 9 (very difficult). Lower the Mean scores, easy for participant complete the tasks on radiotherapy system; p: p-value. The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. Statistically significant at the 0.05 level. Bold indicates that the significance level is less than 0.05; SD: standard deviation.

**User workload**

**Physiological workload.** Table 3 shows the physiological workload results for the two radiotherapy systems. For blink duration, significant differences were found for the task of searching for patient’s treatment plan and uploading it (Varian: 0.14 ± 0.04, Shinva: 0.11 ± 0.04) and reading and reporting alarm content (Varian: 0.14 ± 0.04, Shinva: 0.11 ± 0.04). The participants showed a shorter blink duration with the Shinva XHA600E than with the Varian Trilogy. Moreover, the average blink duration for all tasks showed that the participants showed a shorter blink duration with the Shinva XHA600E (0.12 ± 0.03) than with the Varian Trilogy (0.14 ± 0.02).

For blink rate, significant differences were also found in the task of searching for patient’s treatment plan and uploading it (Varian: 0.40 ± 0.39, Shinva: 0.21 ± 0.20) and reading and reporting alarm content (Varian: 0.55 ± 0.55, Shinva: 0.24 ± 0.26); the participants showed a lower blink rate with the Shinva XHA600E than with the Varian Trilogy. Furthermore, the average blink rate for all tasks showed that the participants showed a lower blink rate with the Shinva XHA600E (0.25 ± 0.21) than with the Varian Trilogy (0.45 ± 0.37).

According to a published study, a shorter blink duration and a lower blink rate indicates an increasing physiological workload. As shown in Table 3, participants
performing tasks with the Shinva XHA600E experienced a larger physiological workload than those performing tasks with the Varian Trilogy.

**Perceived workload.** Table 4 shows the perceived workload assessment for each workload parameter for the two radiotherapy systems. The TLX of Shinva XHA600E (26.08 ± 14.30) was substantially higher than that of Varian Trilogy (26.08 ± 14.30), indicating the participants felt it was comparatively harder to perform tasks using the Shinva XHA600E. Moreover, the physical demand rating was higher for the Shinva system (5.09 ± 6.89) than the Varian system (2.06 ± 3.62). A comparable trend was observed for temporal demand (8.42 ± 5.55 for Shinva and 5.72 ± 4.50 for Varian) and performance (9.54 ± 7.08 for Shinva and 4.68 ± 4.49 for Varian).

**User experience**

User experience was significantly different between the two radiotherapy systems. Varian Trilogy registered High USE (5.86 ± 0.96) than Shinva XHA600E (5.17 ± 0.96), implying participants had better experience with the Varian Trilogy than with the Shinva XHA600E system. Also, Varian registered higher scores for usefulness (5.89 ± 0.97 vs 4.98 ± 1.00) and ease of use (5.73 ± 1.07 vs 5.20 ± 0.96)
Table 3. Comparison of physiological workload for radiotherapy systems.

| Indicator     | Task                                                                 | Varian Trilogy | Shinva XHA600E | t     | df | p     |
|---------------|----------------------------------------------------------------------|----------------|----------------|-------|----|-------|
| Blink duration| Task 1: searching for the patient’s treatment plan and uploading it (Units: S, Second) | 0.14 0.04      | 0.11 0.04      | 2.81  | 29 | 0.009 |
|               | Task 2: reviewing the patient’s treatment plan information (Units: S, Second) | 0.13 0.04      | 0.12 0.03      | 1.60  | 29 | 0.121 |
|               | Task 3: reading and reporting the treatment parameter values during beam delivery (Units: S, Second) | 0.14 0.03      | 0.13 0.04      | 0.75  | 29 | 0.459 |
|               | Task 4: reading and reporting alarm content (Units: S, Second) | 0.14 0.04      | 0.11 0.04      | 2.77  | 29 | 0.010 |
|               | Average blink duration of all tasks (Units: S, Second) | 0.14 0.02      | 0.12 0.03      | 2.98  | 29 | 0.006 |
| Blink rate    | Task 1: searching for the patient’s treatment plan and uploading it (Units: Count/Second) | 0.40 0.39      | 0.21 0.20      | 2.21  | 29 | 0.035 |
|               | Task 2: reviewing the patient’s treatment plan information (Units: Count/Second) | 0.42 0.47      | 0.28 0.31      | 1.20  | 29 | 0.242 |
|               | Task 3: reading and reporting the treatment parameter values during beam delivery (Units: Count/Second) | 0.41 0.32      | 0.26 0.24      | 1.94  | 29 | 0.062 |
|               | Task 4: reading and reporting alarm content (Units: Count/Second) | 0.55 0.55      | 0.24 0.26      | 2.66  | 29 | 0.013 |
|               | Average blink rate of all tasks (Units: Count/Second) | 0.45 0.37      | 0.25 0.21      | 2.30  | 29 | 0.029 |

df: degree of freedom; Mean: a lower blink rate and blink duration indicate an increasing physiological workload; p: p-value. The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. Statistically significant at the 0.05 level. Bold indicates that the significance level is less than 0.05; SD: standard deviation.
Table 4. Comparison of perceived workload between radiotherapy systems.

| NASA-TLX workload | Varian Trilogy | Shinva XHA600E | t    | df  | p   |
|-------------------|----------------|----------------|------|-----|-----|
|                   | Mean | SD  | The range of scores | Mean | SD  | The range of scores |      |     |     |
| Mental demand     | 6.00 | 5.34| 0.00–20.00             | 7.02 | 4.79| 0.67–18.67             | −0.946| 29  | 0.352|
| Physical demand   | 2.06 | 3.62| 0.00–16.67             | 5.09 | 6.89| 0.00–33.33             | −2.30 | 29  | 0.029|
| Temporal demand   | 5.72 | 4.50| 0.00–20.00             | 8.42 | 5.55| 1.33–20.00             | −2.40 | 29  | 0.023|
| Performance       | 4.68 | 4.49| 0.00–16.67             | 9.54 | 7.08| 0.00–21.67             | −3.61 | 29  | 0.001|
| Effort            | 5.33 | 3.79| 0.00–16.67             | 6.49 | 5.54| 0.00–26.67             | −0.96 | 29  | 0.345|
| Frustration       | 2.29 | 3.25| 0.00–13.33             | 3.78 | 4.23| 0.00–16.00             | −1.56 | 29  | 0.129|
| TLX (task load index) | 26.08 | 14.30| 4.00–54.33           | 40.34 | 16.29| 9.67–74.00             | −3.62 | 29  | 0.001|

df: degree of freedom; Mean: this subjective workload evaluation questionnaire evaluates the participant’s perceived workload based on six evaluation dimensions: mental demand (lower scores are better), temporal demand (lower scores are better), physical demand (lower scores are better), frustration (lower scores are better), performance (lower scores are better), and effort (lower scores are better). The output of the NASA-TLX questionnaires is scored from 0 to 100, with lower scores indicating the lower perceived workload; p: p-value. The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. Statistically significant at the 0.05 level. Bold indicates that the significance level is less than 0.05; SD: standard deviation.
and satisfaction (5.77 ± 1.12 vs 4.80 ± 1.16), implying that the Varian Trilogy system displayed better performance. This data is shown in Table 5.

**Table 5.** Comparison of USE questionnaire scores between systems.

|                      | Varian Trilogy | Shinva XHA600E | t    | df  | p   |
|----------------------|----------------|----------------|------|-----|-----|
|                      | Mean  | SD   | The range of scores | Mean  | SD   | The range of scores |
| Usefulness           | 5.89  | 0.97 | 3.63–7.00            | 4.98  | 1.00 | 1.75–6.88            | 4.73  | 29   | <0.001 |
| Ease of use          | 5.73  | 1.07 | 3.00–7.00            | 5.20  | 0.96 | 2.45–7.00            | 2.98  | 29   | 0.006  |
| Ease of learning     | 6.04  | 1.04 | 3.25–7.00            | 5.68  | 1.05 | 3.50–7.00            | 1.66  | 29   | 0.108  |
| Satisfaction         | 5.77  | 1.12 | 2.14–7.00            | 4.80  | 1.16 | 1.86–7.00            | 3.49  | 29   | 0.002  |
| USE questionnaire scores | 5.86  | 0.96 | 3.62–7.00            | 5.17  | 0.96 | 2.39–6.97            | 3.67  | 29   | 0.001  |

df: degree of freedom; Mean: the USE questionnaire includes 30 questions, and every question was rated on Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Higher scores on the USE questionnaire correspond to better user experience; p: p-value. The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. Statistically significant at the 0.05 level. Bold indicates that the significance level is less than 0.05; SD: standard deviation.

Correlation among user performance, user workload, and user experience

**Correlation between user performance and user workload.** The association between user performance and workload is shown in Table 6. For the Varian Trilogy, we found a positive correlation between user performance and perceived workload ($r = 0.387, p = 0.035$). For the Shinva XHA600E system however, there was a strong negative correlation between user performance and physiological workload ($r = -0.487, p = 0.008$). However, there was a strong positive correlation between user performance and perceived workload for the Shinva XHA600E system.

According to the correlation results shown in Table 6 between user performance and user workload, a longer task completion time and a higher task completion difficulty level are associated with a shorter blink duration, a lower blink rate, and a higher NASA-TLX score.

**Correlation between user performance and user experience.** The correlation results between user performance and user experience are shown in Table 6. For the Varian Trilogy, user performance was negatively correlated with user experience ($r = -0.586, p = 0.001$). For the Shinva XHA600E, user performance was negatively correlated with user experience ($r = -0.520, p = 0.003$). Furthermore, for the Shinva XHA600E, significant negative correlations ($r = -0.482, p = 0.007$) were also found between user performance and user experience.
Based on our study results in Table 6, we found that a longer task completion time and a higher task completion difficulty level are associated with a poorer user experience.

**Correlation between user workload and user experience.** The results of the correlation between user workload and user experience are shown in Table 6. For the Varian Trilogy, physiological workload was positive correlated with user experience ($r = 0.407, p = 0.026$). Moreover, perceived workload (NASA-TLX) was negatively correlated with user experience ($r = -0.535, p = 0.002$). For the Shinva XHA600E, significant negative correlations were found between perceived workload and user experience.

According to the results of the correlation between user workload and user experience in Table 6, we find that a better user experience is associated with a longer blink duration, a higher blink rate, and lower NASA-TLX scores.
Table 7. Comparison of two radiotherapy systems usability.

| Evaluation dimensions | Comparison | Varian Trilogy vs Shinva XHA600E | The better |
|-----------------------|------------|----------------------------------|------------|
| User performance      | Less task completion time | 129.07 ± 26.26 vs 159.73 ± 32.74 | Varian Trilogy |
|                       | Lower task completion difficulty level | 1.77 ± 0.94 vs 3.30 ± 1.66 | Varian Trilogy |
| User workload         | Lower physiological workload (blink duration) | 0.14 ± 0.02 vs 0.12 ± 0.03 | Varian Trilogy |
|                       | Lower physiological workload (blink rate) | 0.45 ± 0.37 vs 0.25 ± 0.21 | Varian Trilogy |
|                       | Lower perceived workload | 26.08 ± 14.30 vs 40.34 ± 16.29 | Varian Trilogy |
| User experience       | Better USE questionnaire scores | 5.86 ± 0.96 vs 5.17 ± 0.96 | Varian Trilogy |

Summary of the usability study results

Table 7 presents the usability study results of two radiotherapy systems. According to the results for the three usability dimensions evaluated for the two radiotherapy systems tested, we found that the Varian Trilogy shows better performance, a lower task completion difficulty level, lower physiological workload, lower perceived workload, and better user experience than the Shinva XHA600E ($p < 0.05$), which indicates that the Varian Trilogy radiotherapy system is more suitable for end users.

Discussion

This study conducted user testing to evaluate the usability design of two radiotherapy systems that are used in a university affiliated hospital in China. Through user testing, the study results revealed that the different usability designs of radiotherapy systems could have different influences on end user performance, user workload, and user experience. Moreover, the results of this study can be further used for hospital radiotherapy system procurement decisions in the selection of a suitable radiotherapy system for end users.

End users have been blamed for the adverse events occurring during radiotherapy. However, experts in patient safety have shown that accidents related to radiotherapy result from errors in user interface and poor designs of medical devices.\textsuperscript{8,9} Overwhelming evidence has also implicated faulty software systems and poor designs,\textsuperscript{31} for accidents during radiation therapies treatment. They include incorrect reading of R&V data\textsuperscript{32} excessive requirements for radiation therapist’s knowledge, skill and personal capabilities that lead to incidents,\textsuperscript{33} treatment set-up parameters
are not consistent with written instructions in the treatment chart, and incorrect data entry into the system or improper overriding of patient data that making radiation therapist led to a use error cause incident. This underscores how designs for radiotherapy systems influence the usability of medical equipment. Herein, we also found the user interface substantially influences usability of radiotherapy systems. Therefore, manufacturers should pay attention to the usability design of radiotherapy system. In our study, several usability problems of radiotherapy system user interface were observed by our researchers. For example, the Shinva XHA600E radiotherapy system provided limited information (only the patient’s name and ID), which made it challenging for radiation therapists to confirm that the patient’s treatment plan. This increases the likelihood of putting a patient on a wrong treatment plan. However, this usability problem has been noted by several published studies. Usability technicalities have been reported in several studies. For instance, Chan et al. found that radiotherapy systems that provide limited patient information increase the risk of wrong diagnosis and treatment. Jiang et al. further found that radiation therapy that give very little patient information increases the risk of radiation treatment accidents. For alarm, the color of the alarm in the Shinva XHA600E radiotherapy system was not consistent with the end user’s common sense, as it used yellow as the alarm color, which sometimes confused the end user in the confirmation of radiotherapy system status. Furthermore, the alarm content for the Shinva XHA600E radiotherapy system had no independent area to display information, which requires further action beyond clicking a button to view. Additionally, all alarms for the radiotherapy system were shown together, which can make the display of information difficult to view. For the Varian Trilogy radiotherapy system, several usability problems have also been observed. For instance, on the page of patient radiation treatment plan selection, this system does not inform radiation therapists on the patient’s identifying information, which may lead to the treatment of the wrong patients. This problem has been reported in the studies of Chan et al. and Jiang et al. Moreover, the sex of the patient was represented by “♀” or “♂” instead of other common manifestations for men or women, which can sometimes confuse the end user. Overall, the current radiotherapy systems need tremendous improvement.

Even so, the Varian Trilogy system is more efficient than the Varian Trilogy system. This conclusion was further supported by the results of task completion difficulty level (Varian: 1.77 ± 0.94, Shinva: 3.30 ± 1.66). The participants considered that performing patient radiation treatment tasks with the Varian Trilogy radiotherapy system would be easier than with the Shinva XHA600E. Compared to published radiotherapy system usability studies, our study also used user workload as the usability evaluation indicator to evaluate radiotherapy system usability. In our study, we comprehensively evaluated user workload through perceived workload and physiological workload. For perceived workload, we used the NASA-TLX questionnaires to perform the evaluation. The results suggest that end users perform radiation treatment with the Shinva XHA600E with more perceived workload than with the Varian Trilogy (Varian: 26.08 ± 14.30, Shinva: 40.34 ± 16.29).
Furthermore, the results also revealed that conducting radiation treatment tasks with the Varian Trilogy required lower physical demand and temporal demand than with the Shinva XHA600E and had better performance than with the Shinva XHA600E. Using only perceived workload results to evaluate the workload may not comprehensively assess the workload of end users. Quantitative physiological workload data for end users were collected in this study. We used blink duration and blink rate as evaluation indicators to assess physiological workload. Published studies have confirmed that a shorter blink duration and a lower blink rate indicate an increasing physiological workload.\textsuperscript{15,23,35} Significant differences were also found for blink duration and blink rate between the two radiotherapy systems when participants performed tasks with these systems. According to the results in Table 3, a longer blink duration (Varian: $0.14 \pm 0.02$, Shinva: $0.12 \pm 0.03$) and a higher blink rate (Varian: $0.45 \pm 0.37$, Shinva: $0.25 \pm 0.21$) were observed for participants with the Varian Trilogy, which means a lower physiological workload was observed for the Varian Trilogy. Our study also proved that the eye motion data blink duration and blink rate can be useful indication tools to evaluate the physiological workload of radiation therapists. Moreover, our usability of radiotherapy systems also takes user experience into consideration. Through the USE questionnaires, a better user experience was found for the Varian Trilogy. Furthermore, better usefulness, ease of use, and satisfaction were also observed for the Varian Trilogy to further support that the Varian Trilogy has better user experience.

Furthermore, the correlations among user performance, user workload and user experience were also analyzed in our study. The results in Table 6 indicate that a better user performance is associated with lower user workload and better user experience, and these conclusions can be supported by the study results in Table 6. However, published studies\textsuperscript{16,36–39} have confirmed that selecting a suitable medical device for end users should take the medical device usability design into consideration when hospital medical device procurement decision-making is conducted. In Table 7, the study results show that the Varian Trilogy has better user performance, lower user workload, and better user experience than the Shinva XHA600E, which means the Varian Trilogy radiotherapy system may be more suitable for end users. This study further emphasizes the importance of usability design for medical equipment, a factor which should be considered prior to procurement of medical equipment.

**Limitations**

There are several limitations in this usability study of radiotherapy systems. First, the tested radiotherapy system can perform more functions than tested in our study, we only tested some of the functions that radiation therapists perform daily. Second, in our study, the participants were radiation therapists that the study results cannot be applied to other types user of radiotherapy system (such as radiation physicists, radiotherapy system maintenance engineers). Finally, our study was conducted in a controlled simulated clinical usage environment where some usability problems of radiotherapy system may not be found.
Conclusions
This study evaluates the usability design of two radiotherapy systems based on radiation therapists’ user performance, user workload and user experience. According to the study results, we can infer that the Varian Trilogy radiotherapy system has a better usability design than the Shinva XHA600E system, with better user performance, lower user workload, and better user experience. Currently, government regulatory agency and hospital organization emphasize medical device usage safety to ensure patient treatment safety. Therefore, it is important to select a better usability design of medical device for intended end users. This study provides experimental evidence to inform those involved in hospital procurement decision-making in the selection of a suitable radiotherapy system for radiation therapists to perform radiation treatment for patients.

Acknowledgements
We acknowledge the support of the radiation therapists from Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, who made this study possible.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Health Commission of Hubei Province scientific research project (No.: WJ2021M134)

ORCID iD
Daoxiong Wang https://orcid.org/0000-0002-6682-5705

References
1. Delaney G, Jacob S, Featherstone C, et al. The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. Cancer 2005; 104(6): 1129–1137.
2. Yorke E, Gelblum D and Ford E. Patient safety in external beam radiation therapy. Am J Roentgenol 2011; 196(4): 768–772.
3. Shafiq J, Barton M, Noble D, et al. An international review of patient safety measures in radiotherapy practice. Radiother Oncol 2009; 92(1): 15–21.
4. Huang G, Medlam G, Lee J, et al. Error in the delivery of radiation therapy: results of a quality assurance review. Int J Radiat Oncol Biol Phys 2005; 61(5): 1590–1595.
5. Duffey RB, Duffey RB, Duffey R, et al. Know the risk: learning from errors and accidents: safety and risk in today’s technology. Amsterdam: Butterworth-Heinemann, 2003.
6. Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: application of risk analysis methods to radiation therapy quality management. *Med Phys* 2016; 43(7): 4209–4262.

7. Israelski EW and Muto WH. Human factors risk management as a way to improve medical device safety: a case study of the therac 25 radiation therapy system. *Jt Comm J Qual Saf* 2004; 30(12): 689–695.

8. Crowley JJ and Kaye RD. Identifying and understanding medical device use errors. *J Clin Eng* 2002; 27(3): 188–193.

9. Perry SJ. An overlooked alliance: using human factors engineering to reduce patient harm. *Jt Comm J Qual Saf* 2004; 30(8): 455–459.

10. Lum TE, Fairbanks RJ, Pennington EC, et al. Profiles in patient safety: misplaced femoral line guidewire and multiple failures to detect the foreign body on chest radiography. *Acad Emerg Med* 2005; 12(7): 658–662.

11. Food and Drug Administration. *Applying human factors and usability engineering to medical devices: guidance for industry and Food and Drug Administration staff*. Washington, DC: FDA, 2016.

12. International Electrotechnical Commission. *62366-1: medical devices part 1: application of usability engineering to medical devices*. Geneva: IEC, 2015.

13. International Electrotechnical Commission. *IEC/TR 62366-2: medical devices part 2: guidance on the application of usability engineering to medical devices*. Geneva: IEC, 2016.

14. Kendler J and Strochlic AY. *Usability testing of medical devices*. Boca Raton, FL: CRC Press, 2015.

15. Jiang M, Liu S, Gao J, et al. Comprehensive evaluation of user interface for ventilators based on respiratory therapists’ performance, workload, and user experience. *Med Sci Monit* 2018; 24: 9090–9101.

16. Liu K, Chan FY, Or CK, et al. Heuristic evaluation and simulated use testing of infusion pumps to inform pump selection. *Int J Med Inform* 2019; 131: 103932.

17. Fairbanks RJ, Caplan SH, Bishop PA, et al. Usability study of two common defibrillators reveals hazards. *Ann Emerg Med* 2007; 50(4): 424–432.

18. Wichrowski M. Usability engineering in the prototyping process of software user interfaces for mobile medical ultrasound devices. *Comput Sci* 2015; 16(3): 219–236.

19. McCrory B, Lowndes BR, LaGrange CA, et al. Comparative usability testing of conventional and single incision laparoscopic surgery devices. *Hum Factors* 2013; 55(3): 619–631.

20. Jiang M, Liu S, Gao J, et al. A usability study of 3 radiotherapy systems: a comparative evaluation based on expert evaluation and user experience. *Med Sci Monit* 2019; 25: 578–589.

21. Chan AJ, Islam MK, Rosewall T, et al. The use of human factors methods to identify and mitigate safety issues in radiation therapy. *Radiother Oncol* 2010; 97(3): 596–600.

22. Chan AJ, Islam MK, Rosewall T, et al. Applying usability heuristics to radiotherapy systems. *Radiother Oncol* 2012; 102(1): 142–147.

23. Zheng B, Jiang X, Tien G, et al. Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surg Endosc* 2012; 26(10): 2746–2750.

24. Mazur LM, Mosaly PR, Hoyle LM, et al. Subjective and objective quantification of physician’s workload and performance during radiation therapy planning tasks. *Pract Rad Oncol* 2013; 3(4): e171–e177.
25. Rubio S, Diaz E, Martin J, et al. Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Appl Psychol* 2004; 53(1): 61–86.

26. Mazur LM, Mosaly PR, Tracton G, et al. Improving radiation oncology providers’ workload and performance: can simulation-based training help? *Pract Rad Oncol* 2017; 7(5): e309–e316.

27. Mazur LM, Mosaly PR, Jackson M, et al. Quantitative assessment of workload and stressors in clinical radiation oncology. *Int J Radiat Oncol Biol Phys* 2012; 83(5): e571–e576.

28. Lund AM. Measuring usability with the use questionnaire. *Usability Interface* 2001; 8(2): 3–6.

29. Martynov P, Mitropolskii N, Kuikkola K, et al. Testing of the assisting software for radiologists analysing head CT images: lessons learned. *BMC Med Imaging* 2017; 17(1): 59.

30. Lestantri ID, Sabiq A and Suherlan E. Developing and pilot testing M-health care application for pregnant and toddlers based on user experience. *J Phys Conf* 2018; 978: 012067.

31. Bissonnette JP and Medlam G. Trend analysis of radiation therapy incidents over seven years. *Radiother Oncol* 2010; 96(1): 139–144.

32. Patton GA, Gaffney DK and Moeller JH. Facilitation of radiotherapeutic error by computerized record and verify systems. *Int J Radiat Oncol Biol Phys* 2003; 56(1): 50–57.

33. Clark BG, Brown RJ, Ploquin JL, et al. The management of radiation treatment error through incident learning. *Radiother Oncol* 2010; 95(3): 344–349.

34. Yeung TK, Bortolotto K, Cosby S, et al. Quality assurance in radiotherapy: evaluation of errors and incidents recorded over a 10 year period. *Radiother Oncol* 2005; 74(3): 283–291.

35. May JG, Kennedy RS, Williams MC, et al. Eye movement indices of mental workload. *Acta Psychol* 1990; 75(1): 75–89.

36. Ginsburg G. Human factors engineering: a tool for medical device evaluation in hospital procurement decision-making. *J Biomed Inform* 2005; 38(3): 213–219.

37. Vincent CJ and Blandford A. How do health service professionals consider human factors when purchasing interactive medical devices? A qualitative interview study. *Appl Ergon* 2017; 59: 114–122.

38. Keselman A, Tang X, Patel VL, et al. Institutional decision-making for medical device purchasing: evaluating patient safety. *Stud Health Technol Inform* 2004; 107(2): 1357–1361.

39. Keselman A, Patel VL, Johnson TR, et al. Institutional decision-making to select patient care devices: identifying venues to promote patient safety. *J Biomed Inform* 2003; 36(1–2): 31–44.

**Author biographies**

Mingyin Jiang received the BS degree in Biomedical Engineering from Hubei University of Science and Technology, China, in 2014, and his PhD degrees in Biomedical Engineering from Huazhong University of Science and Technology, China, in 2019. His research interests include medical device management, ergonomics, and usability engineering of medical device.
Xuancheng Tu received his BS degree in Auditing from Zhongnan University of Economics and Law, China, in 1995. His research interests include medical device management and training.

Wanchao Xiao received the BS degree in Economics from Hubei University of Economics, China, in 1993, and the master's degree in Economics from Zhongnan University of Economics and Law, China, in 2008. His research interests include medical device management and quality control of medical device.

Jinhui Tang received the BS degree in Paediatrics from Tongji Medical University, China, in 1990, and the master's degree in Paediatrics from Tongji Medical University, China, in 1997. His research interest is medical device management.

Qiang Li received his BS degree in Biomedical Engineering from Huazhong University of Science and Technology, China, in 2006. His research interests include medical device management and training.

Dongjie Sun received the BS degree in engineering from Wuhan University of Technology, China, in 1989. His current research interests include medical device management and design.

Daoxiong Wang received the BS degree in Clinical Medicine from Jianghan University, China, in 2001, and the master's degree in Engineering from Central China Normal University, China, in 2020. His research interests include medical device management, training and quality control of medical device.