# Development of an Electronic Healthcare Tool to Elicit Patient Preferences in Older Adults Diagnosed with Hematologic Malignancies

Amy Cole[1]([✉]) [iD], Amro Khasawneh[2] [iD], Karthik Adapa[1] [iD], Lukasz Mazur[1,2], and Daniel R. Richardson[3] [iD]

[1] Carolina Health Informatics Program, University of North Carolina, Chapel Hill, NC, USA
amy_cole@med.unc.edu
[2] Department of Radiation Oncology, University of North Carolina, Chapel Hill, NC, USA
[3] UNC Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

**Abstract.** The objective of this study was to develop and evaluate iterative prototypes for an electronic healthcare tool (EHT) using three versions of a discrete choice experiment (DCE) designed to elicit the treatment preferences of older adults with hematologic malignancies. We used a mixed-methods approach including qualitative assessments (think-aloud sessions and semi-structured interviews) to develop an affinity diagram for thematic analysis, and questionnaires (Post-Study System Usability and the National Aeronautical and Space administration's Task Load Index [NASA-TLX]) to evaluate human-computer interaction, human factors and ergonomics standards on the perceived usability of, and the cognitive workload (CWL) required to perform tasks within the prototypes. DCEs included object case, profile case and multi-profile case. Iterative changes to the prototype were planned after each 5 participants. Overall, 15 healthy volunteers completed all assessments with 3 prototypes. Participants reported the prototypes were easy to complete and straightforward but usability issues around definitions, instructions, information overload, and navigation were revealed. Participants also reported feeling overwhelmed at the information presented in the DCEs and having difficulty understanding definitions. Usability and CWL levels were acceptable for all prototypes. The profile case DCE had higher frustration scores than the other versions (NASA-TLX subscale, p = 0.04). Iterative improvements were guided by usability principles and included easier access to definitions, the addition of instructive videos and the inclusion of a more straightforward DCE (object case). This process should improve the validity of results from the DCE and the feasibility of clinical implementation of the EHT.

**Keywords:** Cognitive workload · Electronic healthcare tools · Shared decision-making · Oncology

# 1 Introduction

## 1.1 Background

Hematologic malignancies, including leukemia, lymphoma, and myeloma, account for approximately 10% of cancer cases each year and nearly 57,000 deaths [1]. These cancers predominantly affect older adults over the age of 60. Treatment decision-making is complex for many patients. Chemotherapy causes substantial side effects and offers uncertain benefits. Oncologists and patients can engage in shared decision-making to clarify patient values and preferences to guide chemotherapy decisions. Multiple stakeholders have called for using validated methods to elicit patients' values and preferences to inform shared decision-making [2–7].

Quantitative methods using conjoint analysis to elicit patient preferences and values are increasingly used in healthcare to elicit patient preferences for treatment outcomes [8]. A recent systematic review and meta-analysis demonstrated that the use of these methods can improve values-congruent care [9]. We previously developed two distinct versions of a discrete choice experiment (DCE) (a multi-profile case and an object case) to elicit the preferences of patients with acute myeloid leukemia [10–12]. An electronic health tool (EHT) may facilitate the use of these instruments to improve shared decision-making and allow for real-time reporting of results to patients and clinicians. To the best of our knowledge, DCEs have not been designed and developed as electronic health tools for older patients with hematologic malignancies.

The utilization of EHTs has been shown to increase knowledge of treatment options, improve risk perception, and improve communication between providers and patients, including those diagnosed with hematologic malignancies [13, 14]. Some older adults lack confidence in using technology or have diminished physical abilities, which may present challenges when they try to utilize EHTs in which their specific needs are not considered in the design [15]. A recent systematic review found that older adults recommend that EHTs are designed with detailed instructions, intuitive user interfaces, and attention to the accessibility of the text and layout [15]. Further, for an EHT to be beneficial for newly-diagnosed older patients with hematologic malignancies, it must be understandable, usable, functional and optimize the cognitive workload (CWL) associated with treatment decisions.

# 2 Methods

## 2.1 Participants

We recruited healthy volunteers, aged 21 years of age and older, between August and September 2021. This development study enrolled 15 healthy volunteers, chosen both for ease of recruitment, and for the potential to identify approximately 90% of usability issues before testing with our target patient population (older adults with hematologic malignancies) [16]. Before enrollment, each participant was asked to review and sign an

online consent form. Participants received a \$25 gift card upon completion of the study. We obtained approval from the Institutional Review Board at the University of North Carolina for this study.

## 2.2  Baseline Characteristics of Participants

After being enrolled in the study, participants self-reported demographics including gender, race, ethnicity, education level, employment status, and comfort level with technology through REDCap, a secure online database.

## 2.3  Study Design

We designed a mixed-methods study using an iterative co-design approach [17, 18] including 3 cohorts of participants (total n = 15) in the evaluation of medium-fidelity prototypes (developed in AdobeXD-v.40.0.22). Participants were scheduled to attend a usability session, which was held in the Human Factors Laboratory housed within the Department of Radiation Oncology at the University of North Carolina at Chapel Hill. Recruitment efforts stopped after we reached a point of thematic saturation, in which our final testing sessions were not producing new data to evaluate [19].

**Discrete Choice Experiments.** Patients were asked to complete choice tasks within DCEs to elicit their preferences for treatment outcomes. Prototypes of an EHT containing three alternative DCEs were used: object case, profile case, and multi-profile case. Each EHT prototype contained two DCE versions. Cohort 1 completed the profile case and multi-profile case and cohorts 2–3 completed the object case and multi-profile case. The order of cases presented to each participant was randomly assigned to account for the effect of order and to ensure information from one case type was not transferred across cases [20].

Object Case: The choice tasks within this version of the DCE, also referred to as a best-worst scaling instrument, were adapted from our prior study to identify which outcomes are most important and least important to participants (see Fig. 1). Our initial study used an object case DCE to identify which outcomes patients with leukemia were most worried about. We altered the object case in order to allow for direct comparison between DCEs and to better inform shared decision-making. The following seven attributes were included: maintain day-to-day activities, avoid long-term side effects, avoid short-term side effects, living longer, avoid hospitalizations, avoid becoming dependent on others, and avoid high financial costs. These attributes are condensed from the prior study that included 10 attributes. Each choice task included 4 of the 7 attributes. Participants completed 7 choice tasks.

**Fig. 1.** Example object case discrete choice experiment



**Fig. 2.** Example profile case discrete choice experiment

Profile Case: The choice tasks within this version of the DCE followed the same logic as the object case where participants chose the most important and least important items among a series of options (see Fig. 2). In this version, however, participants were presented with an attribute at a specific level that varied throughout the choice tasks (e.g., avoid mild short-term side effects, avoid moderate short-term side effects, or avoid severe short-term side effects). Each choice task in this DCE included 5 attributes, with each attribute represented at 3 distinct levels. Participants completed 10 choice tasks. Attributes and levels were identical to those in the multi-profile case.

Multi-Profile Case: The choice tasks in this DCE differ from the object case and the profile case. Each choice task required participants to choose which of two profiles they preferred each with 5 attributes at varying levels (see Fig. 3).

This DCE is unchanged from our prior work [11]. Each choice task included 5 attributes (event-free survival, complete remission, time in hospital, short-term side effects, and long-term side effects) each with three levels. Participants completed 10 choice tasks.



**Fig. 3.** Example multi-profile case discrete choice experiment

The choice tasks within each DCE were designed in a linear fashion that provided a "back" and "continue" button to progress through the pages. The initial pages of the prototype contained a set of definitions for the terms that would be used as well as a static example on how to complete the choice tasks.

**Think-Aloud Sessions and Semi-structured Interviews.** Participants were prompted to think aloud as they evaluated the prototype. Each session was audio-recorded and transcribed. Semi-structured interviews were conducted after completing each version of the DCE to elicit feedback on participants' (1) understanding of the provided definitions of the attributes, (2) ability to distinguish between the levels of attributes as presented, (3) preference for DCE version, and (4) perspective on whether patients and family caregivers would utilize and/or trust this tool. Our interview guide was developed in consultation with a qualitative research expert (see Appendix A).

**Subjective Usability and Cognitive Workload Assessments.** After completing each assigned version of the DCE, participants evaluated the usability by completing the post-study system usability questionnaire (PSSUQ), consisting of 16 questions divided into three sub-constructs: system usefulness, information quality, and interface quality [21]. Each question within the PSSUQ is rated on a 7-point Likert Scale, with overall scores calculated by averaging all sub-constructs. Benchmark scores derived from Sauro and Lewis provide the means depicted in Table 2 to interpret overall and subconstruct PSSUQ scores, with better performance and satisfaction reflected in a lower PSSUQ score [22].

Following completion of each assigned DCE, participants also assessed the CWL, quantified subjectively using the National Aeronautical and Space administration's Task Load Index (NASA-TLX) questionnaire. The NASA-TLX measures six dimensions of CWL (mental, physical, and temporal demands, frustration, effort, and performance) with scores $\geq 55$ associated with reduced performance in numerous settings including oncology [23].

### 2.4   Data Analysis

Descriptive statistics were reported for the demographic information.

**Think-Aloud Sessions and Semi-structured Interviews.** Qualitative data from both think-aloud sessions and semi-structured interviews were coded by four members of our study team (AC, AK, KA, DRR) in brainstorming sessions to create an affinity diagram. An affinity diagram uses inductive reasoning to gain insight into the key requirements of a system, including reliability and performance [24]. A recent analysis described the use of modified affinity diagramming techniques that were originally developed from from Holtzblatt's methodology and have since been used for prototype evaluation in Human Computer Interaction and interaction design [25]. Holtzblatt's classic approach to affinity diagramming is intended to gain insight during the discovery stages of design and focuses on contextual inquiries, work modeling, consolidation and affinity diagram building, storyboarding and paper prototyping. Lucero's analysis suggests that once interactive prototypes are established, contextual inquiries, work modeling and paper prototyping are no longer relevant [25]. As we had created medium-fidelity prototypes based on paper

versions from previous studies, we chose to use a modified approach that uses affinity diagramming for prototype evaluations. This approach includes four stages, including creating notes, clustering notes, walking the wall, and documentation. Data gathered during usability testing was consolidated, and team members were invited to group the user feedback into themes. Team members began placing their suggested themes on the wall, inviting others to "walk the wall", by creating additional subgroups and relocating cards, until all final themes had emerged. Upon completion, the diagramming was converted into an electronic format. We evaluated these findings by selecting relevant quotes within the themes and translated them into specific design recommendations.

**Subjective Usability and Workload Assessments.** Descriptive statistics for subjective assessments of usability and CWL were calculated using Microsoft Excel. Descriptive statistics established for the subjective assessments (PSSUQ and NASA-TLX) were broken down by case type and cohort.

We compared the obtained usability to existing standards to assess if each case type met acceptable usability levels. PSSUQ scores recorded from each cohort, by assigned case type, were compared to existing standards to assess overall usability, system usefulness, information quality, and interface quality. Paired t-tests were performed, using JMP 15 Pro, to examine differences in overall usability scores between assigned case types within cohorts, and between case types across cohorts. We considered P values of less than .05 to have statistical significance.

We compared the obtained CWL scores to existing CWL standards to assess if the cases met acceptable levels of user interactions with EHTs. The six dimensions of CWL (mental, physical, and temporal demands, frustration, effort, and performance) were compared for each case type and broken down by cohort. Paired t-tests were performed, using JMP 15 Pro, to examine differences between versions of the DCE.

## 3 Results

### 3.1 Participant Characteristics

15 participants divided into 3 cohorts were enrolled in the study and completed all assessments. Table 1 summarizes demographic information and comfort level with technology assessments. The median age of all participants was 29 years (range: 21–50). Most participants were female (78.6%) (Table 1). Participants identified as Caucasian (53.3%), Asian (26.7%), African American (13.3%), and Black/White (6.7%). Most participants identified as not Hispanic/Latino (86.7%). All participants were college-educated, with most having completed a bachelor's degree or higher (93.3%). The median household income was $50,000 to $74,999. Most participants indicated they are very comfortable with technology (93.3%).

### 3.2 Participant Preference for Version of the DCE

When asked to specify the preferred version of the DCE, 80% (n = 4) of participants in cohort 1 preferred the multi-profile case, with 20% (n = 1) equally preferring the profile

**Table 1.** Baseline characteristics of participants

| Characteristic | Cohort 1(n = 5) | Cohort 2 (n = 5) | Cohort 3 (n = 5) |
|---|---|---|---|
| **Age, y** | | | |
| Mean (SD) | 29.8 (11.4) | 25.4 (2.9) | 30.8 (11.6) |
| Range | 23–50 | 23–30 | 21–50 |
| **Sex** | | | |
| Male | 2 | 1 | 0 |
| Female | 3 | 4 | 4 |
| Chose not to specify | | | 1 |
| **Race** | | | |
| Caucasian | 2 | 3 | 3 |
| African American | 1 | 1 | |
| Asian | 1 | 1 | 2 |
| Other | 1 (Black/white) | | |
| **Ethnicity** | | | |
| Hispanic/Latino | | | 2 |
| Not Hispanic/Latino | 5 | 5 | 3 |
| **Education** | | | |
| Some college, no degree | 1 | | |
| Bachelor's degree | 3 | 4 | 2 |
| Graduate or professional degree | 1 | 1 | 3 |
| **Household income** | | | |
| Less than $25,000 | 1 | 2 | 2 |
| $25,000 to $34,999 | 1 | | |
| $35,000 to $49,999 | 2 | | 1 |
| $50,000 to $74,999 | 2 | 2 | 2 |
| Prefer to not answer | | 1 | |
| **Employment status** | | | |
| 35 h a week or more | 3 | 3 | 1 |
| Less than 35 h a week | | | 3 |
| Unemployed | 2 | | |
| Other (please specify) | | 2 (Student) | 1 (Student) |
| **Comfort level with technology** | | | |
| I am very comfortable using technology with little or no help from others | 5 | 5 | 4 |

**Table 1.** (*continued*)

| Characteristic | Cohort 1(n = 5) | Cohort 2 (n = 5) | Cohort 3 (n = 5) |
|---|---|---|---|
| I am somewhat comfortable using technology, but need help getting started | | | 1 |

case and multi-profile case. In cohort 2, 80% (n = 4) preferred the multi-profile case, with 20% (n = 1) equally preferring the object case and multi-profile case. In cohort 3, 60% (n = 3) preferred the multi-profile case, with 40% (n = 2) preferring the object case.

### 3.3 Thematic Analysis

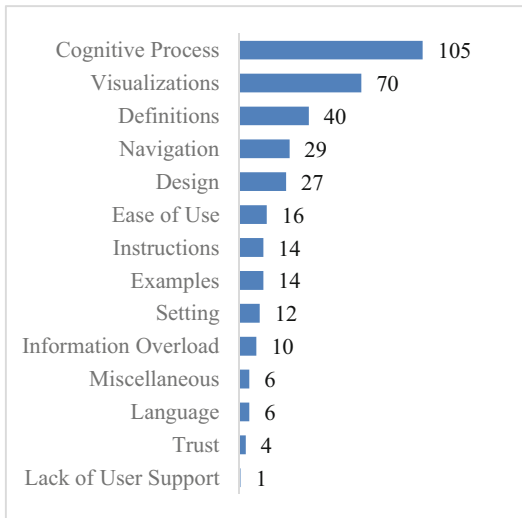Overall, 14 themes emerged during our affinity diagramming sessions (Fig. 4). The themes and subgroupings associated with usability and CWL, and that led to specific design improvements are presented in (Fig. 5). Additional themes emerged that focused on trust for using the tool and appropriate settings (e.g., clinical setting with assistance from a provider) for completing the prototypes. Data from these additional themes will be analyzed when evaluating this tool for clinical implementation.



**Fig. 4.** Affinity diagramming themes

All participants indicated that the prototypes were not challenging to complete. Some participants indicated that while the user interface made it easy to select their answers, cognitive challenges arose when asked to complete the DCEs. These challenges were most often reported with the profile cases, especially the single profile case.

Overall, 4 of the 14 themes that emerged, led to most of the key design change decisions, and are focused on definitions, instructions, information overload, and navigation.

**Theme 1: Definitions.** All participants (n = 15) reported understanding the attribute definitions well enough to complete the choice tasks, however, 33% needed to reference the definitions. Many participants commented on the need for clarity and having easy access to definitions of the attributes and levels. Study participants mentioned that having the definitions presented before answering the questions was helpful, however, by
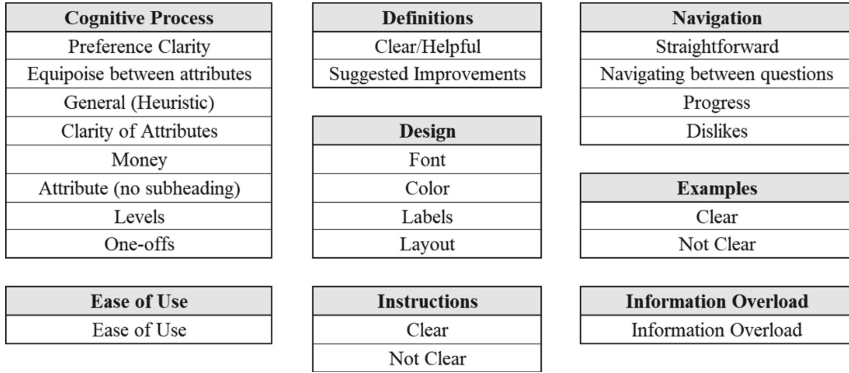
| Cognitive Process |
|---|
| Preference Clarity |
| Equipoise between attributes |
| General (Heuristic) |
| Clarity of Attributes |
| Money |
| Attribute (no subheading) |
| Levels |
| One-offs |

| Definitions |
|---|
| Clear/Helpful |
| Suggested Improvements |

| Design |
|---|
| Font |
| Color |
| Labels |
| Layout |

| Navigation |
|---|
| Straightforward |
| Navigating between questions |
| Progress |
| Dislikes |

| Examples |
|---|
| Clear |
| Not Clear |

| Ease of Use |
|---|
| Ease of Use |

| Instructions |
|---|
| Clear |
| Not Clear |

| Information Overload |
|---|
| Information Overload |

**Fig. 5.** Affinity diagram

the time they were asked to complete the questions, many were uncertain they could accurately recall how a term was defined. Participants suggested adding the ability to review the definitions while answering the questions.

*"If there is a way to add a feature where if you could click on like complete remission or something, it could show that definition again" (Female, Age 31)*

*"Maybe when they start actually going through like the quiz part, having like you know how you can like hover over something and the definition pops up, a reminder of maybe what it means." (Female, Age 26)*

Participants also emphasized the need to clarify the definitions, as terms such as event-free survival and complete remission were unfamiliar to them. Participants indicated that emphasizing the importance of understanding the definitions would improve the prototype. Many suggested that the terms used with side effects, such as mild, moderate, and severe, can have different meanings for different people, and providing examples would make it easier to distinguish between the varying levels of side effects.

*"What is event free survival?" (Male, Age 23).*

*"People without medical background may not know what this meant." (Female, Age 25)*

*"It's like after reading the top paragraph, I think I understand event free survival, but yeah, I don't know, I've never heard that term before. At first, I was like oh, remission, and then it says it does not mean you are in remission. And I feel like understanding what event free survival is important for the next questions I'm going to have to respond to, and like I understand the definition, but it's just not really intuitive what event free survival means." (Female, Age 31)*

*"What is moderate, what is severe, that was the lingering question." (Female, Age 30)*

*A disease related event, could that also be any other health thing such as acute respiratory disease or some other complication from the drug? Does that count as an event? I need clarification. (Female, Age 50)*

*"What does a moderate long term side effect mean right? Does that mean? Oh, I may not be able to have kids." (Female, age 27)*

Design Changes: After completing user testing sessions with cohort 1, we added an "information icon" to each choice task page. Within the object case, the icon was located above the attributes and provided definitions of all attributes when selected. For the multi-profile case, the information icon ("i") was located on the left side of each attribute. Cohort 2 noted the icon resembled a bullet point, and that its function was unclear. In response, it was changed to a question mark, relocated to the right side of the attribute, and enlarged by 50% to make it more visible. Text was added to the multi-profile case to state users can "click on any word to display the definition."

With the support of a health literacy expert, we modified the terminology used to define each attribute by removing medical jargon to not exceed a sixth-grade reading level as recommended by the American Medical Association and the Agency for Healthcare Research and Quality [26, 27].

We also simplified the language for attribute levels and incorporated graphs for visual representation. For example, we utilized anthropomorphic representations to indicate the chance of achieving complete remission, and we utilized bullet charts to indicate differences between mild, moderate, and severe (as shown in Figs. 6 and 7 respectively).

We also removed the static "text-based" definition pages and replaced these with audio/video overviews of the simplified definition. Participants can pause or replay the videos if desired.

**Theme 2: Instructions.** One common preference among participants was the need to have clearer instructions at the onset of the prototype. Participants indicated that providing an overview page or table of contents for the surveys would improve the prototype.

*"An overview to see everything that is available on the prototype on one page first might be good because it's when you start going through those continued questions it is not immediately clear what is going to come next. So, it's not clear how much information you are missing. For example, the page that said benefits and risks, it would have been nice to know this would be elaborated on a different page." (Female, Age 25)*

*"It would be nice if there were a Table of Contents. Let them know the number of questions at the beginning." (Male, Age 23)*

*"Maybe in the beginning, more of a title to the slides because the title was kind of small up there and I had to sometimes say, like, what am I doing again?" (Female, Age 50)*
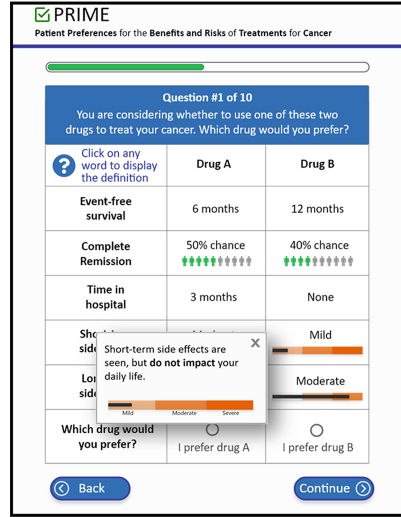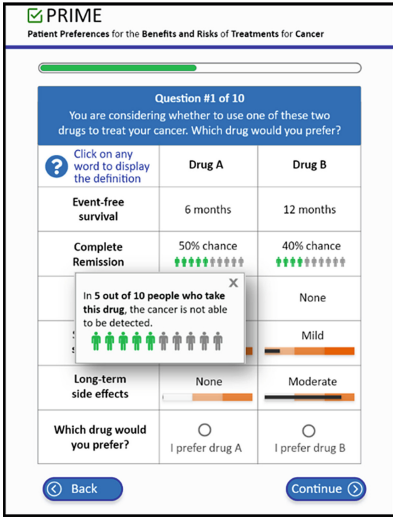
**Fig. 6.** Multi-profile case with 50% chance of complete remission selected to show pop-up definition.

**Fig. 7.** Multi-profile case with mild short term side effects select to show pop-up definition.

Many felt that the instructions did not provide a clear direction on what they were expected to do, nor how many questions were included. Participants commented that the instructions did not make them aware of the importance of thoroughly understanding the definitions.

> *"Maybe the survey should reference earlier on, I didn't know this is like a survey. It might have been a little confusing." (Female, Age 27)*

> *"It's kind of the way that the mind works if I know this is something that I have to retain until the very end of all of this stuff, then my mind is going to process it differently." (Female, Age 50)*

> *"I don't really understand those directions. I couldn't imagine like older person doing that like that was kind of confusing." (Female, Age 26)*

Participants felt examples of the DCE choice tasks were very helpful. However, for many participants, the static examples contained too much text and caused confusion rather than clarification.

> *"The first example is very confusing. I didn't know what it was asking me to do." (Male, Age 23)*

> *"Oh, this is his choice, not my choice in this example. I did not read that properly." (Male, Age 23)*

Design Changes: We added a welcome video that provided an overview of each prototype for participants to view before interacting with the prototypes. A title page was added to each prototype, as well as an introductory page that provides information about the assigned prototype. The welcome video includes a statement about the importance of understanding the definitions for completing the prototypes.

We also replaced the "text-based" examples with short example videos that walk the participant through the choice tasks and provide them information about how to access definitions, if needed.

**Theme 3: Information Overload.** Participants stated they felt a sense of information overload and noted that there was too much content on each page. Recommendations were made to simplify the information presented. Although these statements referred to the appearance of the prototype generally, participants noted that the information overload made understanding definitions and answering choice tasks within the DCE difficult. Many participants found it challenging to comprehend and retain the information due to the amount of text and the terminology used.

*"A lot of information, it was an overload. That made it hard especially when it is something difficult to think about. It was so much. People have different ways of taking in information. There was a lot of text." (Female, Age 27)*

*"A couple of the screens were a bit too wordy. There was just too much information on the screen and they just needed to be broken up so I can process. I think fit into multiple pages." (Female, Age 50)*

*"It felt long and time consuming. It felt like it was a lot for one sitting." (Male, Age 24)*

*"OK, now I got to read all this. I thought it was just going to be pictures." (Female, Age 30)*

*"I think this could be streamlined into more like one or two sentences. This page looks complicated on first glance." (Female, Age 32)*

Design Changes: We simplified the terminology, focusing primarily on improving the attribute definitions. This resulted in a reduced word count on each page, thereby reducing the clutter and improving the overall appearance.

**Theme 4: Navigation.** Some parts of the initial prototype contained a navigation bar at the top of the screen. Many commented it was not necessary and oftentimes caused confusion as to its purpose.

*"I like that there's not the thing on the top there because I was wondering if I need to do something up there." (Reference to the navigation bar) (Female, Age 50)*

*"I noticed this survey doesn't have the buttons on top, which I think makes it look cleaner because I didn't really need it for them in the last one." (Reference to the navigation bar) (Male, Age 24)*

Each prototype was initially designed with an overall progress bar, located at the top of the screen, and an additional numerical bar located at the bottom of each choice task page to indicate which choice task they were on, as seen displayed in (Figs. 1, 2 and 3). Participants indicated that having two progress bars was confusing.

> *"So the progress bar thing or the advancement thing has changed and that's maybe a little bit jarring." (Referencing the addition of the numerical choice task progress bar) (Female, Age 31)*

> *"When I see this, I feel like I can like choose which question." (Referencing the numerical choice task progress bar) (Age 23)*

Design Changes: We removed the top navigation bar. We also removed the choice task progress bar located at bottom of all DCEs. We added text to the top of each choice task page indicating which question participants were on (e.g. Question 1 of 10).

### 3.4  Usability

Usability was reported as acceptable for all DCEs, with mean PSSUQ scores below the validated mean scores presented by Sauro and Lewis. [22] Table 2 includes detailed PSSUQ results broken down by DCE type and cohort. When examining the subscales by case type, the best subscale scores for all case types were system usefulness. There was no difference in overall PSSUQ scores between assigned case types within cohorts. When evaluating between case types across cohorts, cohort 1 reported lower mean usability scores for the multi-profile case compared to cohort 2 ($-2.8$, p $= 0.02$). The multi-profile case was the only case presented to each cohort, of which we saw an improved overall mean score from 2.13 (SD 0.59) to 1.66 (SD 0.30) as well as improved scores in each subscale.

### 3.5  Cognitive Workload

Global NASA-TLX scores for each DCE case type were all below 55, indicating acceptable cognitive workload. Table 3 includes detailed NASA-TLX results broken down by case type and cohort. There was no difference in global NASA-TLX scores between assigned case types within cohorts. There was a significant difference in the frustration rating between the profile case and object case (cohort 2) ($-2.23$, p $= 0.04$) and the profile case and object case (cohort 3) ($-2.33$, p $= 0.04$).

**Table 2.** Usability scores

| Type of DCE (Cohort) | System Usefulness (SYSUSE) | Information Quality (INFOQUAL) | Interface Quality (INTERQUAL) | Overall PSSUQ |
|---|---|---|---|---|
| **Multi-Profile** | Mean (SD) | | | |
| Cohort 1 | 1.57 (0.40) | 2.35 (0.81) | 2.05 (0.51) | 2.13 (0.59) |
| Cohort 2 | 1.10 (0.08) | 1.28 (0.47) | 1.25 (0.32) | 1.25 (0.39) |
| Cohort 3 | 1.49 (0.44) | 1.46 (0.43) | 1.85 (0.30) | 1.66 (0.30) |
| **Profile** | | | | |
| Cohort 1 | 1.63 (0.53) | 2.43 (0.98) | 2.25 (0.47) | 2.26 (0.60) |
| **Object** | | | | |
| Cohort 2 | 1.27 (0.53) | 1.32 (0.64) | 1.20 (0.40) | 1.30 (0.60) |
| Cohort 3 | 1.24 (0.31) | 1.45 (0.39) | 1.50 (0.57) | 1.45 (0.44) |
| *Sauro & Lewis [22] | 2.80 | 3.02 | 2.49 | 2.82 |

*Recommended means provided by Sauro and Lewis to interpret PSSUQ scores. Better performance and satisfaction are reflected in lower PSSUQ scores [22].

**Table 3.** Cognitive workload scores

| Type of DCE | Global | Mental | Physical | Temporal | Performance | Effort | Frustration |
|---|---|---|---|---|---|---|---|
| **Multi-Profile** | Mean (SD) | | | | | | |
| Cohort 1 | 12.32 (3.93) | 55.00 (18.71) | 0.00 (0.00) | 9.00 (13.57) | 48.00 (40.82) | 40.00 (17.61) | 16.00 (11.58) |
| Cohort 2 | 6.03 (3.93) | 22.00 (25.81) | 2.00 (4.00) | 2.00 (4.00) | 42.00 (47.39) | 5.00 (7.75) | 2.00 (4.00) |
| Cohort 3 | 14.73 (4.83) | 69.00 (22.23) | 0.00 (0.00) | 7.00 (7.48) | 85.00 14.14 | 35.00 (32.71) | 10.00 (20.00) |
| **Profile** | 16.13 (6.21) | 58.0 (17.94) | 2.0 (2.24) | 31.0 (21.29) | 50.0 (32.53) | 37.0 (25.43) | 49.0 (29.50) |
| Cohort 1 | | | | | | | |
| **Object** | | | | | | | |
| Cohort 2 | 7.27 (2.91) | 26.00 (10.20) | 2.00 (2.45) | 1.00 (2.00) | 41.00 (48.21) | 13.00 (12.08) | 11.00 (22.00) |
| Cohort 3 | 14.67 (5.82) | 67.00 (26.94) | 0.00 (0.00) | 14.00 (15.94) | 78.00 (34.29) | 36.00 (32.77) | 10.00 (12.65) |

* Global Scores $\geq$ 55 are associated with reduced performance [23].

# 4   Discussion

## 4.1   Principal Findings

The objective of this study was to evaluate and compare standards on the perceived usability and cognitive workload of healthy volunteers interacting with a prototype EHT containing three distinct versions of a DCE: multi-profile case, profile case, and object

case. We found similarities between case types, mainly from the results of the PSSUQ and NASA-TLX assessments that suggest usability and the mental workload levels were acceptable. Overall, the qualitative feedback received was positive, with most participants indicating the EHT prototypes containing the DCEs were straightforward, user-friendly, and easy to use.

Utilizing a modified affinity diagramming process for evaluating our prototypes was an effective approach to gain insight from usability testing. Consolidating the qualitative data and creating groupings led to the discovery of usability issues as common themes began to emerge during the diagramming process, including definitions, instructions, information overload, and navigation. Specific design improvements were made to the prototypes based on the discovery of these usability issues. This is consistent with previous studies that indicated this modified approach helped the study team prioritize the issues that needed to be addressed and ultimately led to improved prototypes [25, 28, 29].

Data from DCEs are derived from individual choice tasks where participants indicate their preference for one attribute over another. Poor understanding of attributes fundamentally undermines the validity of results from DCEs. Participants in our study had difficulty recalling the definitions of the attribute terms used in the DCEs. We found that participants desired the option to reference definitions in real-time. Similar to other studies evaluating EHTs, we also found that the terminology used to define each attribute included some medical jargon that was challenging for participants to comprehend [30–32]. We simplified definitions, eliminated medical jargon, and added an information icon to allow for easy reference to definitions. These changes were essential to maintain the validity of the choice tasks by ensuring that participants clearly understood each attribute.

We saw high levels of frustration (NASA-TLX subscale) with the profile case DCE. Participants also reported that they disregarded several attributes in this DCE in order to simplify the process. Previous studies have demonstrated that this strategy, termed attribute non-attendance, is frequently employed by participants in complex DCE choice tasks and may undermine the validity of results [33]. Therefore, we decided to remove the profile case from the prototype and replace it with the object case. This change resulted in decreased levels of frustration and should result in higher levels of confidence about the validity of future results.

Although prototypes were not directly compared by individual participants, usability improved for the multi-profile case in each prototype suggesting that prototypes were progressively more usable. These findings are supported by interviews with participants who indicated that improvements, such as adding the ability to review definitions made the tool more usable.

Published recommendations for developing DCEs for use in healthcare do not currently include a recommendation to evaluate CWL, nor is it a standard practice to evaluate the usability of DCEs embedded in EHTs [34, 35]. This study demonstrates that these methods are useful at identifying potential challenges to the validity of DCE results. Participants universally reported on questionnaires that the DCEs were easy to understand; however, it was clear that many participants did not understand the definitions of

attributes and felt frustrated completing choice tasks. These findings suggest that subjective reporting may not be a reliable metric of participant understanding. As we intend to use this EHT to inform shared decision-making regarding chemotherapy, ensuring participant understanding of attributes is critical.

## 4.2 Limitations

Participants in this study were all healthy volunteers. These participants were younger and likely have different life-experience to our target population of older adults with hematologic malignancies. Our target population will likely experience the EHT differently than these volunteers. Although we attempted to make versions of the DCE similar by containing similar attributes, our study was not designed to directly compare versions of the DCE. For example, the number of choice tasks varied for each DCE. Also, the profile case was only presented to one cohort of patients and was replaced by the object case due to high levels of frustration and subjective report. This reduced the number of participants that completed the profile case. In addition to our small sample size, these considerations make definitive conclusions regarding preference for DCE versions impossible. While our sample size was small, this was adequate for informing the design recommendation of the prototypes, as we were able to identify common themes and achieve thematic saturation. Previous studies using healthy volunteers, and similar sample sizes have shown to be an effective method to resolving usability problems prior to definitive work in our target population [32, 36].

## 4.3 Conclusion

Shared decision-making in oncology requires a clear understanding of what matters most to each patient. Methods to reliably capture patient preferences at the point-of-care have not been fully developed. We used an innovative approach with healthy volunteers to design efforts that would leverage insights on completing choice tasks within iteratively refined prototypes and improve the processes involved in capturing patient preferences. When implemented, an EHT to elicit patient preferences has the potential to improve shared decision-making and patient-centered care across oncology.

# Appendix A

**Semi-structured Interview Guide (asked after completing each prototype):**

1.  What is your overall impression of the prototype?

    a.   What did you like about the prototype?

   b.  What did you not like about it?
   c.  What additional features would you like?

2. Did you find the prototype challenging to complete?

   a.  Can you tell me more about that?

3. [Multi-Profile Case] You were presented with 10 questions that asked you to choose between drug a and drug b. Can you walk me through how you made a decision regarding treatment preferences in the prototype?
4. [Profile Case/Object Case] You were asked to think about what was most important and least important. Can you walk me through how you made a decision regarding treatment preferences in this prototype.

**Semi-structured Interview Guide (asked after completing both prototypes):**

5.  Did the definitions of the attributes make sense to you as they were presented?
6.  Did you feel that you understood the definitions of the attributes well enough to complete the prototype?
7.  Could you distinguish between the levels of each attribute as they were presented in the prototype? For example, the levels presented for remission were 40%, 50%, or 60%.

    a.  Can you tell me more about that?

8.  When you read the words mild, moderate, and severe, what kinds of things came to mind?
9.  You were asked 10 questions about drug a and drug b. You were also asked 7 questions about what was most important and least important.

    a.  Which of these question series, if either, would help you have a more informed discussion with your providers?

10. Do you think patients with newly diagnosed cancer and their family caregivers will use this? If they do use this, do you think they will trust it?

    a.  Can you tell me more about that? (What are the barriers and facilitators?)

# References

1. Lymphoma Survival Rate | Blood Cancer Survival Rates | LLS. https://www.lls.org/facts-and-statistics/facts-and-statistics-overview. Accessed 09 Oct 2021
2. Smith, M.Y., et al.: Patient engagement at a tipping point-the need for cultural change across patient, sponsor, and regulator stakeholders: insights from the DIA conference, 'patient engagement in benefit risk assessment throughout the life cycle of medical products.' Ther. Innov. Regul. Sci. **50**(5), 546–553 (2016)
3. Sekeres, M.A., et al.: American society of hematology 2020 guidelines for treating newly diagnosed acute myeloid Leukemia in older adults. Blood Adv. **4**(15), 3528–3549 (2020)

4. Hunter, N.L., O'Callaghan, K.M., Califf, R.M.: Engaging patients across the spectrum of medical product development: view from the US food and drug administration. JAMA **314**(23), 2499–2500 (2015)
5. Rocque, G., et al.: Engaging multidisciplinary stakeholders to drive shared decision-making in oncology. J. Palliat. Care **34**(1), 29–31 (2019)
6. Text - S.1597 - 114th Congress (2015–2016): Patient-Focused Impact Assessment Act of 2016 | Congress.gov | Library of Congress. https://www.congress.gov/bill/114th-congress/senate-bill/1597/text. Accessed 21 Oct 2021
7. Text - H.R.34 - 114th Congress (2015–2016): 21st Century Cures Act | Congress.gov | Library of Congress. https://www.congress.gov/bill/114th-congress/house-bill/34/text. Accessed 21 Oct 2021
8. Cheung, K.L., et al.: Using best-worst scaling to investigate preferences in health care. Pharmacoeconomics **34**(12), 1195–1209 (2016)
9. Witteman, H.O., et al.: Clarifying values: an updated and expanded systematic review and meta-analysis. Med. Decis. Making. **41**, 801–820 (2021)
10. Seo, J., Smith, B.D., Estey, E., Voyard, E., O'Donoghue, B., Bridges, J.F.P.: Developing an instrument to assess patient preferences for benefits and risks of treating acute myeloid leukemia to promote patient-focused drug development. Curr. Med. Res. Opin. **34**(12), 2031–2039 (2018)
11. Richardson, D.R., et al.: Age at diagnosis and patient preferences for treatment outcomes in AML: a discrete choice experiment to explore meaningful benefits. Cancer Epidemiol. Biomark. Prev. **29**(5), 942–948 (2020)
12. Bridges, J.F., Oakes, A.H., Reinhart, C.A., Voyard, E., O'Donoghue, B.: Developing and piloting an instrument to prioritize the worries of patients with acute myeloid leukemia. Patient Prefer. Adher. **12**, 647–655 (2018)
13. Stacey, D., et al.: Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst. Rev. **4**, CD001431 (2017)
14. Walker, J.G., Licqurish, S., Chiang, P.P.C., Pirotta, M., Emery, J.D.: Cancer risk assessment tools in primary care: a systematic review of randomized controlled trials. Ann. Fam. Med. **13**(5), 480–489 (2015)
15. McLean, B., et al.: Providing medical information to older adults in a web-based environment: systematic review. JMIR Aging **4**(1), e24092 (2021)
16. Faulkner, L.: Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav. Res. Methods. Instrum. Comput. **35**(3), 379–383 (2003)
17. Sumner, J., Chong, L.S., Bundele, A., Lim, Y.W.: Co-designing technology for ageing in place: a systematic review. Gerontologist. **61**, e395-409 (2020)
18. Vandekerckhove, P., de Mul, M., Bramer, W.M., de Bont, A.A.: Generative participatory design methodology to develop electronic health interventions: systematic literature review. J. Med. Internet Res. **22**(4), e13780 (2020)
19. Watling, C.J., Lingard, L.: Grounded theory in medical education research: AMEE guide No. 70. Med. Teach. **34**(10), 850–861 (2012)
20. Between-Subjects vs. Within-Subjects Study Design. https://www.nngroup.com/articles/between-within-subjects/. Accessed 09 Oct 2021
21. PSSUQ (Post-Study System Usability Questionnaire) - UIUX Trend. https://uiuxtrend.com/pssuq-post-study-system-usability-questionnaire/. Accessed 31 May 2021
22. Sauro, J., Lewis, J.R.: Standardized usability questionnaires. In: Quantifying the User Experience: Practical Statistics for User Research, Waltham, MA 02451, pp. 185–240. Elsevier, USA (2012)
23. Mazur, L.M., et al.: Positive effects of neurofeedback intervention on radiation oncology physicians' workload during go/no-go performance test. Trav. Hum. **80**(1), 113 (2017)

24. Holtzblatt, K., Beyer, H.: The affinity diagram. In: Contextual Design, pp. 127–146. Elsevier (2017)
25. Lucero, A.: Using affinity diagrams to evaluate interactive prototypes. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9297, pp. 231–248. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22668-2_19
26. Weiss, B.D.: Health Literacy - A Manual for Clinicians. American Medical Association Foundation and American Medical Association Chicago, IL (2006)
27. Health Literacy | PSNet. https://psnet.ahrq.gov/primer/health-literacy. Accessed 29 Jan 2022
28. Buur, J., Soendergaard, A.: Video card game: an augmented environment for user centred design discussions. In: Proceedings of DARE 2000 on Designing Augmented Reality Environments - DARE 2000, New York, New York, USA, pp. 63–69 (2000)
29. Cox, D., Greenberg, S.: Supporting collaborative interpretation in distributed Groupware. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work - CSCW 2000, New York, New York, USA, pp. 289–298 (2000)
30. De Clercq, P.A., Hasman, A., Wolffenbuttel, B.H.R.: A consumer health record for supporting the patient-centered management of chronic diseases. Med. Inform. Internet Med. **28**(2), 117–127 (2003)
31. Pyper, C., Amery, J., Watson, M., Crook, C.: Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract **54**(498), 38–43 (2004)
32. Britto, M.T., Jimison, H.B., Munafo, J.K., Wissman, J., Rogers, M.L., Hersh, W.: Usability testing finds problems for novice users of pediatric portals. J. Am. Med. Inform. Assoc. **16**(5), 660–669 (2009)
33. Lew, D.K., Whitehead, J.C.: Attribute non-attendance as an information processing strategy in stated preference choice experiments: origins, current practices, and future directions. Mar. Resour. Econ. **35**(3), 285–317 (2020)
34. Bridges, J.F.P., et al.: Conjoint analysis applications in health–a checklist: a report of the ISPOR good research practices for conjoint analysis task force. Value Health **14**(4), 403–413 (2011)
35. Reed Johnson, F., et al.: Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health. **16**(1), 3–13 (2013)
36. Bansback, N., Li, L.C., Lynd, L., Bryan, S.: Development and preliminary user testing of the DCIDA (dynamic computer interactive decision application) for 'nudging' patients towards high quality decisions. BMC Med. Inform. Decis. Making **14**, 62 (2014)