

TreatmentExplorer: an Interactive Decision Aid For Medical Risk Communication and Treatment Exploration

Lyndsey Franklin
Catherine Plaisant
Kazi Minhazur Rahman
Ben Shneiderman

Human-Computer Interaction Lab and Computer Science Department
University of Maryland

ABSTRACT

Medical treatments carry unique benefits and risks which patients must understand in order to decide which option is best for them. Prior research has demonstrated that patients are ill-equipped to understand the statistical information presented to them through standard decision aids. We describe a prototype decision aid, TreatmentExplorer, which supports patients' needs by presenting treatment outcome, onset of symptoms, and treatment side effects using a novel graphic representation with staged animation and text-only narration. Our prototype also illustrates the use of a data driven personalization approach by using electronic health record data. We report on expert reviews, a pilot study (n=24) and a main study (n=42), which characterize the benefits of TreatmentExplorer over a text-only decision aid as well as a version without staged animation, and conclude with guidelines for designers.

Research Highlights

Design insights and evaluation results for the TreatmentExplorer suggest that patients may experience better knowledge gains if designers:

- Show outcomes and side effects graphically
- Use staged-animation with guided narration
- Begin each narration from the same state
- Follow a consistent narration order
- Provide both replay and skip-through options
- Explain one data point per stage
- Give users control over the flow of narration

Author Keywords

Risk Communication; Animations; Guided Narration; Evaluation; Interactive Interfaces; Information Visualization

ACM Classification Keywords

H.5.1. Information interfaces and presentation (e.g., HCI): Multimedia Information Systems.

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

Revised version to appear in *Interacting with Computers*

1. INTRODUCTION

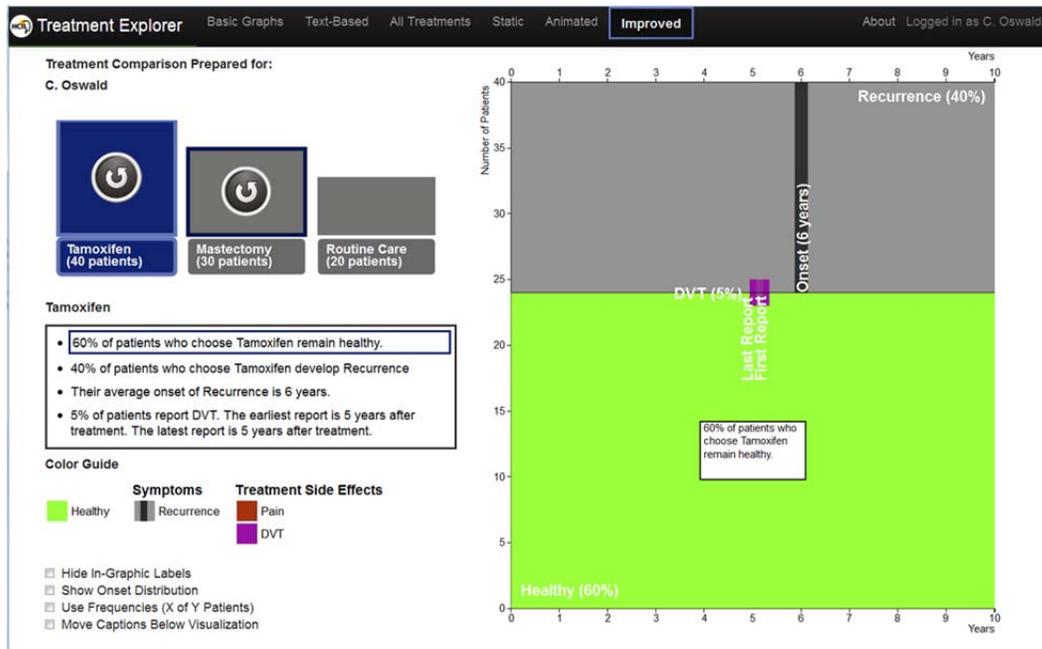
Consider the following scenario: *Donna is 60 years old and has a family history of breast cancer. She recently tested positive for the BRCA gene mutations putting her at high risk of developing cancer herself. She has to decide whether or not she wants harsh preventative treatments like surgery or only medication to reduce her risk.*

Increases in online information have started to increase patient knowledge and responsibility for their health (Sillence et al, 2007). Evidence-based medicine (Frankovich, Longhurst and Sutherland, 2011) promotes the ideal that patients like Donna should be given personalized data about their treatment options so that they can come to a logical decision based on treatments that have the greatest efficacy – usually with their physicians. They must become knowledgeable, often in a short period of time, about their condition and treatment options. For each treatment option they need to understand the expected benefits (e.g. a reduction of the probability of getting cancer) and risks (e.g. side effects or possible secondary illnesses brought by the treatment). All those benefit and risk measures also have a temporal element: when is this likely to happen? How long will this side effect last?

Unfortunately, some patients are unable to use the information presented to them about treatments and make poor choices (Brown et al, 2011; Sudore and Schillinger, 2009). Many patients are ill-equipped to understand both medical terminology and the statistical data presented to them in decision aids. Even educated patients have difficulties reading and understanding textual presentations of treatment information on occasion (Brown et al, 2011), making misinterpretation of both numeric and textual information a real hazard (Timmermans, 2005). Patients also have strong opinions on their medical care in terms of which lifestyle-impacting side effects they are willing to cope with (Timmermans, 2005). It has been reported that even with the data to reach a logical decision, patients often base treatment decisions on their emotional reactions rather than rational decision-making (Dillar et al, 2013; Edwards et al, 2008). Thus, the design of a useful decision aid faces challenges of risk communication, health literacy, and placement within a conveying expected medical workflow.

These medical challenges are a new domain for HCI researchers who are in the position to apply existing usability guidelines and heuristics to the problem of supporting patient decision making. This paper describes a prototype decision aid, TreatmentExplorer which provides patients with a novel representation of temporal treatment information, presented in an interactive interface using staged animation and text-only narration (see Figure 1-3, and the video available in the supplementary materials). We report on the evaluation of the prototype with four domain experts and the results of a pilot study (n=24) and a main user study (n=42). Finally, we present a set of guidelines based on the results of our design and evaluation experiences.

TreatmentExplorer is in early stages of design and further refinements based on the healthcare delivery needs and practices will have to be made, as well as further evaluation with different populations.



1: Treatment Explorer (See VIDEO demonstration in the supplementary materials). On the left we see the 3 buttons corresponding to the 3 treatments and their prevalence. The 1st treatment is selected. On the right we see the visual representation of the data for this treatment.

2. RELATED WORK

TreatmentExplorer draws its inspiration from the work of risk communication, health literacy, and medical decision aid research.

Risk Communication: Research investigating the communication of risk in healthcare has consistently demonstrated one great challenge: patients are generally poor with statistics, even when educated. In their commentary, Fagerlin, Zikmund-Fisher, and Ubel provide 10 recommendations for improving risk communication which include use of language, text, graphics, order of information, use of comparison, and presentation of time (Fagerlin, Zikmund-Fisher and Ubel, 2011). They note that frequencies (i.e. 1 of 4) are preferred for providing information about absolute risks and/or highlighting changes between levels of risk. They also recommend repeatedly drawing attention to the time interval over which a risk occurs and the inclusion of graphs and summary tables.

Han et al. have examined the topic of uncertainty in risk communication and formed a taxonomy of uncertainty types within health care in the attempt to help clarify the problem of its expression (Han, Klein and Arora, 2011). Follow up work produced novel visualizations capable of representing randomness and its effect on uncertainty (Han et al, 2012). Recently, a website entitled Visualizing Health started provided recommended designs for standard graphs (Visualizing Health, 2014).

Health Literacy: Health literacy research focuses on producing health-related messages that non-experts can understand and use in making informed treatment choices. The goal is to produce health materials that patients with deficient skills can comprehend as well as patients without such deficiencies. For comparing high-level information such as the quality of treatment-supporting evidence or a summarized rating of a treatment's success, work has shown that icons such as star ratings or symbols are preferred over other representations (Hildon et al, 2012; Bunge et al, 2010). Other research suggests that graphics increase the believability of

information (Hollands and Marteau, 2012). Vague verbal expressions such as “low”, “medium”, and “high” have been shown to lead to misunderstandings between physicians and patients (Timmermans, 2005).

Medical Decision Aids: Decision aids are often evaluated as part of larger clinic systems where the focus is on the feasibility of their deployment and their role in patient-physician communication. Decision aids used for physician-patient communication often produce anxiety, which interferes with patients’ ability to reason about their healthcare, leading to poorer decisions (Bekker, Hewison and Thornton, 2003). However, patients have also credited decision aids with more productive and efficient physician consultations (Belkora et al, 2009). Visual representations provide memory prompts, which reduce the cognitive load of patients during decision-making (Bekker, Hewison and Thornton, 2003). Study results suggest that decision-aided patients are able to express more emotions, use more cognitive terms, and verbalize decision-relevant feelings (Bekker, Hewison and Thornton, 2003).

A framework (Franklin, 2013) was proposed for describing the content of medical decision aids, which includes six dimensions: decision type, time scale, measurement types, data source, personalization level, and information format. It identified an understudied design space: interactive and animated decision aids. This is the gap we address with TreatmentExplorer.

Interactive information visualization and visual analytics methods will bring profound changes to personal health programs (Shneiderman, Plaisant and Hesse, 2013). Storytelling is regarded as the next step for visualization (Kosara and Mackinlay, 2013) and researchers have focused on telling stories about the analysis process or the data itself. TreatmentExplorer uses visualization and aspects of storytelling by using narration to help users learn how to read an unfamiliar visual representation. This is in contrast with the use of visual representation to describe the decision process (Karacapilidis, 2006).

We use animation which has been shown to be helpful in some cases e.g. to reveal trends over time or transformations of graphical representations during exploration (Chevalier et al, 2010; Heer and Robertson, 2007), but new examples and further user studies are needed to sharpen our understanding of which situations benefit from animation and which ones do not (Tversky, 2011; Amershi et al, 2008). Finally the design of TreatmentExplorer was inspired by the Lifeflow visualization, which summarizes temporal patterns of electronic health records (Wongsuphasawat et al, 2011). EventFlow was designed primarily for epidemiologists and clinical researchers, so it was not clear that this approach would be useful even after a complete redesign to simplify the visual representation and user interface.

3. TREATMENTEXPLORER

While this paper focuses on the information presentation aspects of the decision aids it is useful to understand the source of the data used by the decision. We first introduce our data-driven approach that draws on data from local hospitals or clinics, and then describe the main characteristics of TreatmentExplorer using the framework dimensions (Franklin, 2013).

Data-driven decision aids and personalization

Unlike the vast majority of decision aids (which present results gleaned from the medical literature) or the more recent model-based decision aids (which compute somewhat personalized information based on answers to short questionnaires, e.g. National Cancer Institute, 2013) our prototype summarizes data from Electronic Health Records (EHRs) of patients with similar profile and history, to produce personalized decision aids. This data-driven approach (Perer and Gotz, 2012) allows TreatmentExplorer to reflect the most up-to-date data relevant to a condition and patient. It also enables TreatmentExplorer to be used for a great variety of conditions. Deployed in a clinic environment, a fully-operational system would be useful to many patients if access to EHR data for enough similar patients was available from local hospitals or health organizations. Coming back to our opening scenario, Donna’s physician would first select relevant elements from her patient record to guide a search for similar patients (Perer and Gotz, 2012), then review the usefulness of the data aggregated from those similar patients, then ask Donna to use TreatmentExplorer - either on her own or together

with the physician (Wilcox et al, 2013) - to compare the effects of the different treatments for patients like her. While this vision may still be futuristic, TreatmentExplorer's interface could already be useful today using data generated by models. We acknowledge that today very few organizations have access to sufficient data to allow such personalization. Data quality is also an issue in part because historical data is only starting to be used for retrospective analysis. We believe that the demonstrated value of TreatmentExplorer and other similar tools will provide further incentive for increased EHR adoption and for data quality improvements.

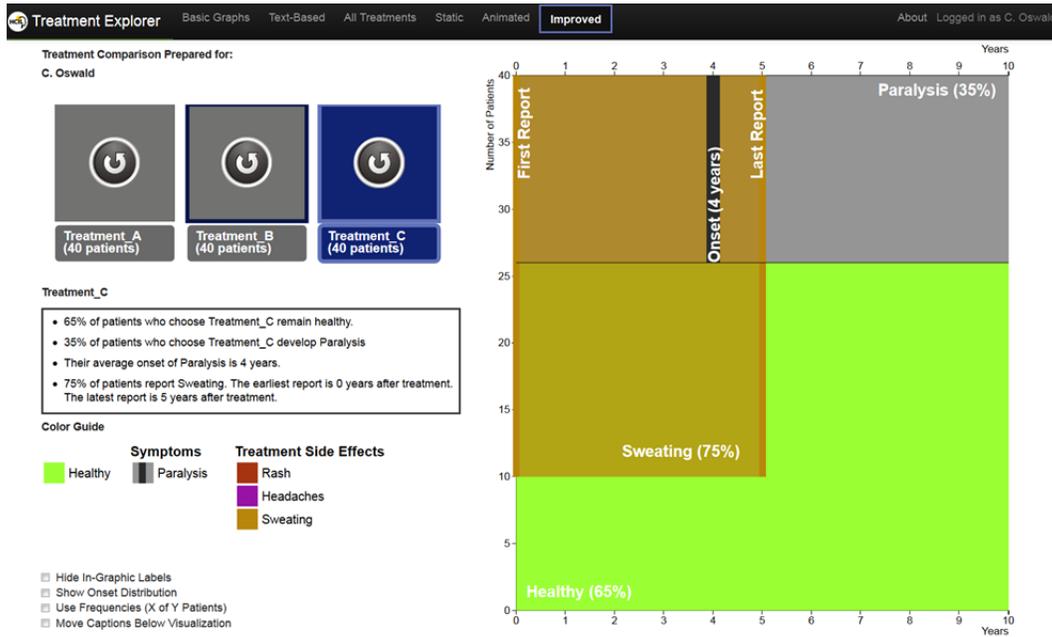


Figure 2: This TreatmentExplorer's interactive decision aid (as used in the experiment) was prepared for patient C. Oswald who has a (imaginary) condition leading to paralysis. She can see that there are 3 treatments (A, B, C) and review each treatment. All three prevalence (the 3 bars for A, B and C are of same height). C is currently selected. On the right a visualization summarizes treatment success rates over 10 years i.e. the proportion of healthy patients (green) versus patients who develop paralysis (gray), the average onset of paralysis (4 years), the proportion of patients experiencing side effects (75% of them) and when those side effects were reported (between 0 and 5 years). This complex display is introduced step by step using staged animation and text-only narration

We now review the main characteristics of TreatmentExplorer and how it differs from existing decision aids.

Multi-Option Decision Support

TreatmentExplorer supports the visualization and comparison of multiple treatment options. By supporting comparison, patients using TreatmentExplorer are relieved of some of the cognitive burden of assimilating data from multiple decision aids using different formats. Patients are able to explore the differences between treatments as well as learn the details of individual treatments. Animated transitions facilitate comparison between treatments, first slowly guided by text-only narration, then more rapidly for back and forth comparison.

Measurement Types

The majority of decision aids focus on comparing a single measure of effectiveness for each treatment, for example the proportion of patients who get cancer (Franklin, 2013). Only text decision aids have attempted to provide the rich picture of the likely future of patients selecting a treatment over another one. Decision aids using graphics become quickly crowded with the many charts and graphs needed for each measure so limit themselves to one or two measures. In

contrast TreatmentExplorer attempts to combine multiple measures in a single interactive visual display: i.e. the proportions of patients 1) remaining healthy after treatment, 2) experiencing a specific outcome (e.g. breast cancer), and 3) afflicted by side effects and other secondary outcomes linked with the treatment (e.g. another type of cancer). Vertical sizing of each visual element is controlled by the proportion of patients represented. The average onset time of a condition is shown (an exact cumulative distribution of onsets is available as an advanced option). Information about side effects associated with treatments include the number of patients reporting each side effect as well as the date of the earliest reporting and the date of the last reporting of the side effect. As warehouses of electronic health records are expanding rapidly we believe that patients and physicians will soon be able to learn from the experience of the other patients from their hospitals, clinics or health organization. This information will complement the lessons learned from small clinical trials that may be decades old or involve only white men. Treatment Explorer could be used with clinical trial data as well.

Multiple Projected Points of Time

While most decision aids provide information about one or two time points, e.g. chance of having cancer at the 5 year or 10 year mark (Franklin, 2013), TreatmentExplorer provides the average time of the outcome (e.g. on average women following this treatment who develop cancer will develop cancer after 8 years on average) or the time range of side effects (e.g. during the 1st six months). To provide that information the horizontal axis of the TreatmentExplorer visualization represents time. Each treatment is represented using a consistent horizontal scale, allowing patients to compare the average time of condition onset as well as the duration of side effects.

Staged Animation and Guided Narration

The decision aid framework (Franklin, 2013) review of existing decision aids highlights that the majority of decision aids use text or traditional simple graphics (e.g. bar charts) as information format. Interaction is mostly nonexistent, with few exceptions (see Jones et al, 2011 for best example). In contrast TreatmentExplorer proposes a novel visual display combining all the information needed to compare treatments. Staged animation and synchronized text-only guided narration explain elements of the visualization (see Figure 3) and progressively increase the richness of the display. This animation draws from the lessons of (Plaisant, Shneiderman, 2005).

4. DESCRIPTION OF THE INTERFACE

A typical user experience with TreatmentExplorer, illustrates its expected use (drawn from the controlled experiment, Section 6.3). First, the user sees the title information i.e. “Prepared for patient C. Oswald” (who has been told that she has a condition leading to paralysis.) She can see that there are 3 treatments (A, B, C) with a bar chart showing the relative prevalence of the three treatments in her health organization. In the experiment, the 3 treatments had data from the same number of patients i.e. the 3 bars for A, B and C were the same height (left side of Figure 2 and 3). A text box invites her to click on one of the treatments.

First treatment: When the user selects Treatment C the axes of the visualization appear on the right, and she is told that she will be guided step by step by pressing the Right Arrow button. When she does, the staged animation begins:

- 1) A green rectangle fills the appropriate section of the visualization (vertically from the bottom) to match the relative portion of patients who remained healthy after receiving the selected treatment (step 1 of Figure 3). Pressing the arrow again animates to the next step:
- 2) A gray box fills the remaining portion of the screen to indicate the proportion of patients that became paralyzed (as explained in the dialog box – see step 2 of Figure 3).
- 3) A black bar is animated from the left (treatment time) to the right (here the 4 year mark) to indicate the average onset of the paralysis, which is also described in the text narration.
- 4) A brown rectangle is added to indicate treatment side effects, i.e. the proportion of patient getting them and average range of reports (i.e. likely start and stop of the effect). In this example, side effects partly overlay the healthy patient groups as well as the paralysis

patients to show that patients in both sets may experience this treatment-associated side effect. While this example includes a single side effect, multiple such treatment risks can be displayed when appropriate.

Each step of the narration is cued by the patient by clicking in the visualization with the mouse or by a key press so that patients can take as much time to read captions and study the visualization as they want. Each step begins with the display of the text and captions, so users have time to read them, then after a brief delay the animation of the graphic elements begins. When the animation has completed for a step, the prompt appears to alert the patients that they may continue when they are ready.

Once all steps have been completed, the patients can either review the same data on their own (i.e. replay the staged animation, or mouse over elements of the display to re-read the corresponding narration text) or select a different treatment. When a different treatment is selected the screen is cleared and the information for the new treatment is presented using the same method. When all treatments have been reviewed in detail at least once, the speed of animation is increased so that the final visualization can be accessed quickly for back and forth comparison. In the faster animation, elements animate from the old position to the new position (e.g. when switching from treatment B to C, the time of onset would animate from the position it had in the B treatment to the position for the C treatment, allowing patients to see how the time of onset changes to a point earlier or later).

To the best of our knowledge no other decision aid attempts to present such rich information in a personalized interactive environment. It would take seven traditional graphs to represent this information (with timelines and bar charts, see video). While the final visualization is novel and may be perceived as very complex we believe that the use of staged animation and guided narration enable patients to learn the meaning of the visual representation, which they could then use to remember and compare the benefits and risk associated with each treatment.

To evaluate TreatmentExplorer and its suitability for use as a decision aid, our evaluation process involved expert reviews, a pilot study, and two controlled experiments comparing TreatmentExplorer. We discuss these evaluations and the feedback we received in the following sections.

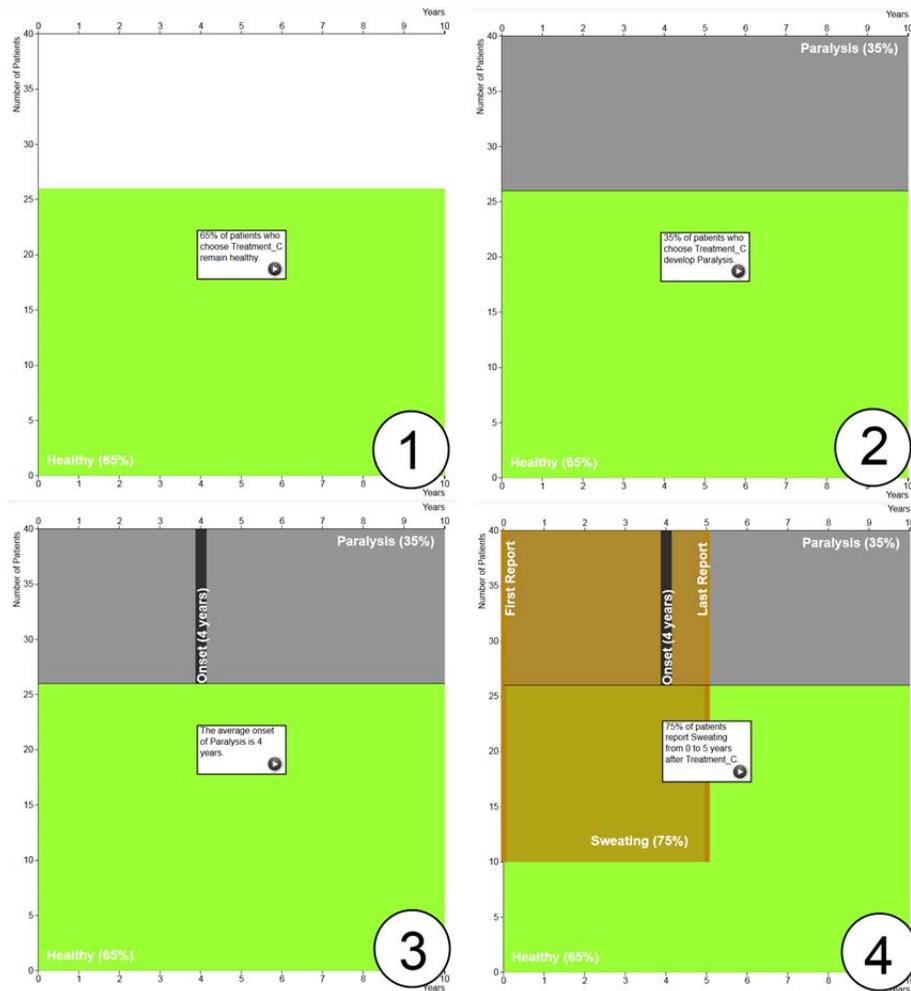


Figure 3 Animation steps in TreatmentExplorer. 1) first the proportion of patients who remained healthy after receiving the selected treatment appears. i.e the green color fills the appropriate section of the visualization. 2) next, the remaining patients i.e those afflicted with the paralysis condition appear (i.e. gray fills the remaining portion of the screen). 3) the average onset of the paralysis is added. 4) Treatment side effects appear. Each step is described with a text-only narration. A right arrow button cues users to step through at their own pace

5. EXPERT REVIEWS

Four experts in the fields of medicine, public health, and risk communication were given a demo of the TreatmentExplorer prototype and interviewed for design improvements in sessions lasting about 1 hour. Suggestions and feedback were then incorporated into the next version (sometime as options) before interviewing the next expert. Feedback on our evaluation plans and an adequate baseline for comparison were also gathered

The demonstration used a synthetic dataset simulating 90 patient records with rates of cancer and side effects mimicking the breast cancer fact sheets available from (National Cancer Institute, 2013b) and (Susan G. Komen Breast Cancer Foundation, 2013). Three treatment options were available: routine care (no treatment), medication (based on Tamoxifen), and surgery (based on double prophylactic mastectomy). Both the medication and surgery treatment options carried a risk of some side effect based on side effects reported in the fact sheets.

TreatmentExplorer was received very well and experts said they hoped to get a version of TreatmentExplorer to investigate with their own data. Desired improvements centered on

improving the visualization and making it more accessible to patients. Some comments were straight forward (e.g. reminding us to make sure the colors were suitable for patients with color deficiencies), others were more specific:

- Further simplify the animations so that patients only need to follow one moving object at a time
- Include a meta-description of the represented health records (i.e. How many patients are represented and how long their records provide data for)
- Use a single consistent time scale across treatments
- Animate visual elements from the bottom of the chart (where the zero is located)
- Overlay side effects on the area for both healthy and condition-developing patients to reflect the real distributions
- Add an option to show the distribution of onsets to show skew or outliers (instead of only the average, but experts agreed that by default showing the average was preferable)

We discussed the possible role of TreatmentExplorer in clinic environments and refined a use scenario where both patients and physicians select treatment options together during an initial consultation. The patients could be introduced to TreatmentExplorer and have any initial questions answered. The patients would then be given access to TreatmentExplorer to use on their own time after the consultation, so that they could continue to review the information and explore treatment options. We discussed the importance for patients of including a “no treatment” as one of the possibilities.

While five alternative designs were including in our prototype we focus on the three versions that were evaluated. Other versions are mentioned in the discussion, and demonstrated briefly in the video demonstration (see supplementary materials).

6. USER STUDIES

Our user studies compared a text-only interface, a static version of the TreatmentExplorer prototype, and the full-featured prototype with animation. A 10-item questionnaire was used to determine knowledge gained by participants.

6.1 Research Questions

The study was designed to answer the following research questions:

1. Are there statistically significant differences in the number of times participants must consult the decision aid in order to answer questions accurately?
2. Are there statistically significant differences in the number of incorrect responses to questions participants provide between the three different decision aids?
3. Are there statistically significant differences in the subjective ratings of usefulness?

6.2 Pilot Study

A pilot study with 24 participants allowed us to refine our testing procedure and identified usability issues which were corrected before the main study. For example: rapid animation of graphic elements to show the differences between two treatments after both have been reviewed slowly at least once. While we had only recorded the total number of incorrect answers before reaching the correct answer, our observations suggested that participants could recall more information from the first trial with the full interface so we decided to record more precisely at what time participants were providing the correct answer, i.e. either from recall, or after looking at the interface once more, twice, etc. Participants’ comments were promising and some limited differences identified. The first five questions of the questionnaire focused on the prevalence and risk of scenario side effects. A one-way independent-measures ANOVA revealed significant differences in the accuracy of participant responses for the five questionnaire questions dealing with scenario side effects ($p < 0.05$, $F(2,21) = 4.97$). A post-hoc Tukey’s HSD test revealed significant differences between the text only interface and the full-featured interface with regards to accuracy. Participants using the full-featured interface were able to answer questions about side effects with significantly greater accuracy than participants with a

text-only interface. Nevertheless results also suggested that participants had difficulties understanding the difference between condition onset and risk of the condition itself so additional text labels were added on the visualization and the color legend was refined. More details about the pilot study is available in the Master's thesis of the first author (Franklin, 2013)

6.3 Main Study

Participants

Our main study recruited 42 participants on campus through emails to mailing lists, paper fliers, and verbal advertisements. These participants came from the undergraduate and graduate population and were compensated \$10 for their participation.

Data

We created a synthetic dataset for a fictitious condition so that no participant would have prior experience or knowledge of the condition, its treatments, or its side effects (see Figure 2 and 3). Participants would thus need to use the decision aids provided to them to complete the questionnaire. This synthetic data and fictitious condition also eliminated the risk of participants ever developing the condition themselves in the future and drawing on information and experiences from this study as part of their actual personal health decision process. The synthetic dataset consisted of fictitious patient records for 120 patients dealing with the medical condition and three possible treatments, each with a single adverse side effect.

Experimental Design

The study followed a 1x3 between subjects design with participants using only one of three possible decision aids to complete a short questionnaire about a fictitious condition. The three decision aids were presented in a website displayed in a maximized Firefox web browser on a 36x18 inch monitor set to a resolution of 1920 x 1080 p. Participants could use a mouse to scroll through the decision aid and point and click as desired.

All three versions are demonstrated in the video available in the supplementary materials.

Decision Aid 1: Text-Only

A text-only decision aid was created based on the layout and contents of a 2-page summary of Type 2 Diabetes oral medications produced by Consumer Reports Health: Best Buy Drugs (Consumer Reports Health, 2013). This decision aid (Figure 4) was thus a realistic analog of other text-only decision aids that patients would likely consult when trying to choose between multiple treatment options available for a single condition.

Decision Aid 2: Static-graphic TreatmentExplorer

A functionally limited version of TreatmentExplorer was used as a second interface to provide a limited experience. This version used the same layout and visualization as the full-featured TreatmentExplorer, however the guided narration, staged animation, progressive disclosure, and reinforcing captions were removed. This eliminated all interactive features of TreatmentExplorer so that participants would have access only to the static visualization for each treatment. This interface was intended to isolate the effects of interaction and animation on the usability and effectiveness of TreatmentExplorer.

Decision Aid 3: Full-Featured TreatmentExplorer

The final interface made use of the fully-featured TreatmentExplorer prototype, including guided narration, staged animation, progressive disclosure, and reinforcing captions.

Treating Crimson Blot Syndrome

Crimson Blot Syndrome is a paralysis-inducing condition associated with dry, reddened skin on the hands, arms, legs, and feet. Patients often describe a tingling sensation as the condition progresses. Early treatment is crucial for preventing the partial and complete paralysis associated with later stages of Crimson Blot Syndrome.

Treatment A

Treatment A is a recently approved treatment for Crimson Blot Syndrome. It has few known side effects but is associated with a 1 in 4 (25%) chance of severe rash. Patients typically experience this rash after 4 years of treatment with Treatment A but report rashes as late as 7 years after treatment. The risk of Paralysis when taking Treatment A has been shown to be as low as 25% (or 1 in 4 patients). Patients taking Treatment A also tend to develop Paralysis at much later stages of the condition with an 8 year average onset.

Treatment B

Treatment B has been used for several years to treat Crimson Blot Syndrome. It has a known history of increased risk of headaches with 1 out of 2 (50%) patients reporting increased headaches starting 3 years after beginning Treatment B. These headaches continue while the patients continue to take Treatment B. Treatment B has been shown to successfully delay the onset of Paralysis to an average of 6 years. Treatment B is associated with a higher risk of Paralysis (3 of 4 patients or 75%)

Treatment C

Treatment C is a once per day medication for Crimson Blot Syndrome associated with a lower risk of paralysis (35%). 3 of 4 (75%) Patients choosing Treatment C report increased or excessive sweating when first starting to take Treatment C. These side effects are reported only for the first few years with the latest reports 5 years after treatment with Treatment C begins. Most patients taking Treatment C remain healthy but the minority of patients develop Paralysis after an average of 4 years.

Treatment	Advantages	Disadvantages
Treatment A	<ul style="list-style-type: none"> • Low risk of Paralysis • Later Onset of Paralysis • Take Once per Day 	<ul style="list-style-type: none"> • Small risk of severe rash
Treatment B	<ul style="list-style-type: none"> • Take Once per Day 	<ul style="list-style-type: none"> • Risk of moderate to severe headache while taking medication • Lower reduction in risk of Paralysis
Treatment C	<ul style="list-style-type: none"> • Low risk of Paralysis • Take Once per Day 	<ul style="list-style-type: none"> • Risk of Increased to Excessive Sweating

Figure 4: the text only version

Dependent Variables

A 10-item questionnaire was filled out by participants while they used one of the three decision aid interfaces. This questionnaire was adapted from the “Questions You May Want to Ask Your Doctor” sections of the National Cancer Institute’s guide for breast cancer treatments (NCI, 2013). The questionnaire reflected questions that patients in a realistic healthcare situation would likely need to answer. Time was recorded as the cumulative sum of times participants needed to consult their decision aid. Researchers noted whether or not participants made full use of their time exploring their decision aid or requested the questionnaire early. The cumulative sum of incorrect questionnaire responses provided by participants was recorded. After each chance to complete the questionnaire, researchers marked all responses as correct, incorrect, or unanswered. Responses were either completely correct or incorrect, no partial credit was given. Participants were allowed to provide new answers after exploring treatment options with the decision aid. These new answers were also marked and added to the cumulative sum of responses.

While the classic uninterrupted time-to-completion metric was considered we decided not to use it because of concerns that interaction effects would weaken the results (e.g. Are participants taking longer because it is harder to answer or because the interaction style is more engaging and they are playing longer before answering? Are they faster because they are simply overwhelmed and guessing which leads to more errors?). The uninterrupted time to correct answer metric was also thought to be problematic, as participants would simply fish for the right answer by submitting many erroneous ones until they got it right. In contrast our selected metric lets users focus on learning about the treatment, not merely answering specific questions, and gives some insight on to when in the process of using the decision aid participants began to understand the information.

Procedure

The following procedure was used:

1. Participants read the imaginary scenario (in short “you have just been diagnosed with Crimson Blot Syndrome and need to learn about the three possible treatments”). They received no training about the interface.
2. Participants were taken to the website of their decision aid and given 3 minutes to investigate their treatment options. Participants did not have access to the questions during this period.
3. The decision aid was taken away and participants were given their questionnaire and a maximum of two minutes to answer as many questions as they could. Participants did not have access to their decision aid during the questionnaire period so answers were based on recall.
4. After the 2 minutes researchers marked all answers on the questionnaire as correct, incorrect, or unanswered.
5. Participants were asked for their subjective ratings (from 1 to 10) of how well prepared they felt after their first interaction with their decision aid.

Steps 2-4 were repeated until the questionnaire was completely and correctly filled out.

6. Participants were debriefed on the nature of the study and given the opportunity to ask questions. Two subjective debriefing questions were also asked:
 - i. What feature of this decision aid did you find the most or least helpful?
 - ii. What other additional information would you want to see as a patient making a treatment choice with this decision aid?

During the session, participants were not allowed to use calculators, consult any other materials, or take notes. Participants did not have access to both their decision aids and questionnaires at the same time. They were given the same questionnaire with their previously marked answers to continue working from.

Results

One participant using the full-featured interface lost focus on the task. He continuously commented on the interface rather than paying attention to the content. His performance with 13 errors meets Chauvenet’s criterion (Chauvenet 1960) as spurious and was removed as an outlier. This resulted in 14 participants in the text interface group, 14 in the static interface group, and 13 in the full featured interface group.

A one-way ANOVA (three treatments) assessed the total number of incorrect responses for all ten questions. The means were 7.21 for text-only, 6.00 for static-graphic, and 3.77 for the full-featured interface ($p < 0.05$, $F(2,38) = 3.74$) (Figure 5).

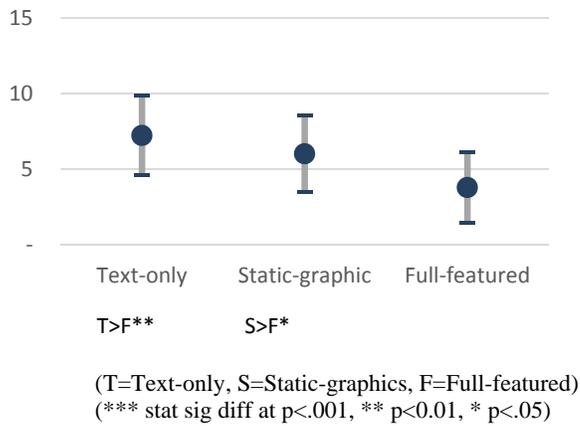


Figure 5. Total number of incorrect responses

We split the ten questions in two groups: five questions about the main symptom of the condition (i.e. paralysis), and five questions about the side effects associated with the treatments (i.e. sweating, headache and rash). For the first group (questions about the main symptom), there were significant differences among the number of incorrect responses across the interfaces. The means were 2.21 for text-only, 2.43 for static-graphic, and 0.69 for full-featured ($p < 0.01$, $F(2, 38)=8.32$) (Figure 6).

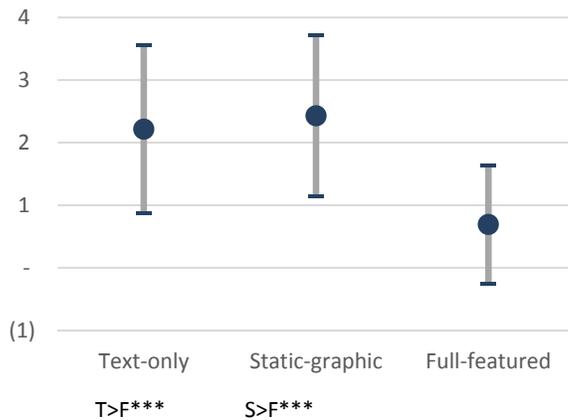


Figure 6. Number of incorrect responses for symptom-related questions

For the second group (questions about side effects) there was no significant difference. The means were 5.00 for text-only, 5.37 for static-graphic, and 3.08 for full-featured ($p > 0.05$, $F(2,38)=2.10$) (Figure 7).

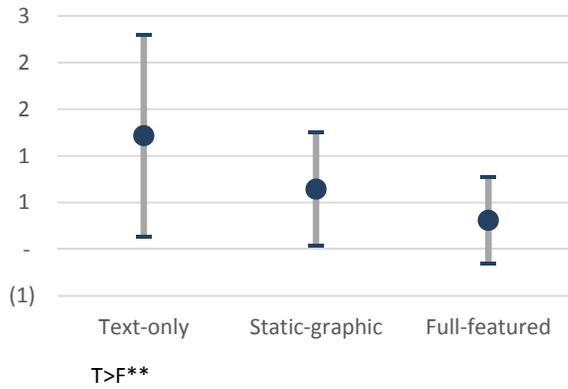


Figure 7. Number of incorrect responses for listing all the possible side effects

Looking in more detail at the individual questions related to side effects revealed significant differences in participant's total number of incorrect responses for the question which asked to list all the possible side effects of all treatments (1.21 for text-only, 0.64 static-graphic, 0.31 full-featured, $p < 0.05$, $F(2,38)=4.46$) and for the question asking to name the most common of the three side effects (1.93 text-only, 0.79 static-graphic, 0.85 full-featured, $p < 0.01$, $F(2,38)=5.31$) (Figure 8).

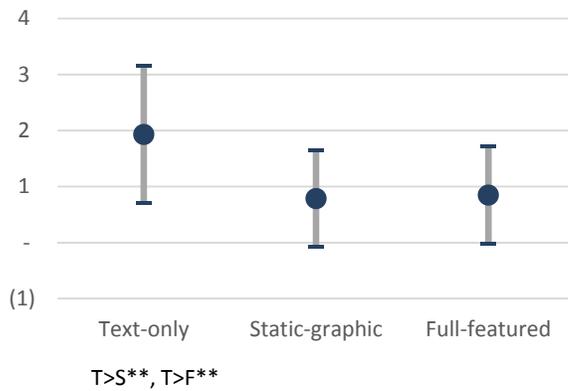


Figure 8. Number of incorrect responses for the most common of the three side effects

The mean number of correct responses after the first trial (i.e. based entirely on recall) was not significantly different with means of 4.50 for the text-only interface, 4.29 for the static-graphic interface and 5.3 for the animated one ($p > 0.05$, $F(2,38)=2.92$) (Figure 9).

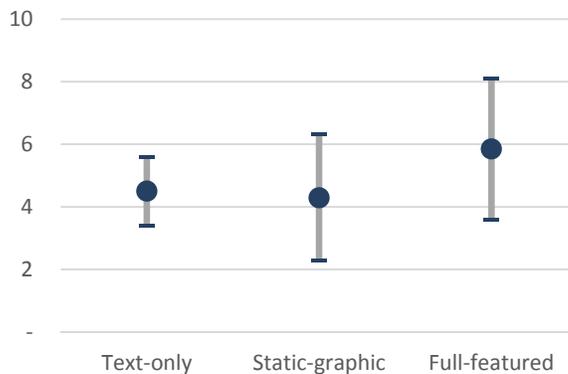


Figure 9. Number of correct responses after the first trial

Similarly one-way (three treatments) ANOVA analysis found no significant differences between interfaces for the number of times participants had to consult the decision aid to answer all questions correctly. Participants' subjective ratings were also not significantly different (note that participants only used one of the three versions).

Discussion

Results suggest that participants using the full-featured TreatmentExplorer interface demonstrated better knowledge gains about the treatments' risks and benefits than participants using the static or text-only interfaces. Since there was no significant difference in terms of time or preference rating it can only be said that using the fully-featured version did not negatively impact participant completion times or experience. That is, walking through a series of guided narrations did not slowdown participants when compared with participants using familiar formats such as text. This early evaluation was focused on determining approximately when participants began to learn from their decision aids. Some variation of a true time-to-completion may be needed to establish performance time of decision aids provided it can separate time spent completing a questionnaire from time spent interacting with the decision aid.

During the study, participants seemed more engaged when using the full-featured and static interface compared to the text-only interface. For example, participants using the text-only interface seemed to lose interest and get bored after reviewing the information once and waited for the trial time to end, especially after the first one or two sessions. On the other hand, participants who used the full-featured or static interfaces went back and made full use of the trial time. Seven participants who used the text-only interface commented that they would like to see the data in tables or charts. Text-only interface users who had time to see the full-featured interface after the experiment concluded all (100%) reacted positively to it – mentioning that it would be a more effective solution. Patient satisfaction with a decision aid would be a useful metric to consider in future evaluations. Only anecdotes and user comments suggest that TreatmentExplorer may be more satisfactory than text-only decision aids. Our between subjects design for these initial studies precludes us from yet establishing statistical claims of greater user satisfaction. Interfaces such as TreatmentExplorer are likely to augment traditional text only decision aids (which remain accessible to users with visual impairments).

Additional observations and feedback from participants suggest possible differences between versions. For example, three of the static and full-featured interface participants commented that they could easily remember how the proportion of healthy and paralysis-afflicted patients resized between treatments. One participant even drew the final visualization for a treatment on the questionnaire when he struggled to answer a question. During the debriefing six participants spontaneously made comments about the usefulness of color, e.g. “the color coding helped to distinguish elements”, or “the green and grey were always balancing each other but the yellow-ish (i.e side effect) colors changed from one treatment to the next”.

A number of comments referred to the complex graphics used in the static-graphic and full-featured interfaces. Six out of thirteen participants using the static-graphic version said that the graphic was either “confusing” or “not-friendly”; one commented that while the proportion of healthy and paralysis-afflicted patients was obvious, other features of the (static) visualization were more confusing. However, there were fewer negative comments about graphics from participants who used the full-featured interface and these were more specific. One participant said that there were “too many numbers (and that) she panicked”. Another said that it was “tricky to understand” and a third one said that it was “confusing at first but made sense later”. Other comments focused on smaller aspects such as the axis labels. These comments indicate that further refinements and testing are needed to push TreatmentExplorer beyond the stages of early prototyping.

There were positive comments about the staged animation of the full-featured interface being helpful. For example one participant commented that terminology like “onset” was confusing at first but that it became clear once she watched the bar representing onset animate between treatments. Two participants commented that they appreciated the text description of the visualization as it animated and felt they learned it quickly with the text narration to guide them.

We saw participants who used the full-featured version gesture with their hands and mimic the animations while they were trying to remember what they had seen.

While the full-featured version with visualization and narration seemed useful there are still many aspects of the design that could be improved. Showing a view of all three treatments side-by-side may help patients compare among alternative treatments, but it may also be found overwhelming by others (see Figure 10 and the video). Other possible improvements include: changing the color of the graph from green to gray after the average onset of paralysis rather than at the time zero; moving the notification text-box to the side so that it does not overlap any part of the graph; animating the graph elements from left to right to give a sense of time continuity. While the prototype included the option to show the distribution of the time of onset instead of the average we asked participants not to use it (to reduce the variability). Several participants who saw it after the end of the experiment commented that it was useful as an option.

We observed participants clicking rapidly at the end of the first session to speed the animation in order to finish in time. Further studies may vary the speed of animation to determine the best possible timing so that users do not need to wait but can easily link graph items with the annotations.

Another area of improvement could be the terminology used. Several of the participants had problems understanding the definition of 'first onset', 'average onset' and 'last onset'. Furthermore, when participants were asked to identify the side effects, nineteen among the forty-one participants initially included the main effect of the disease (i.e. paralysis) as a side effect. A brief description of these terms may help patients to make better comparison among treatments. Many comments suggest that treatment cost would be an important issue for comparing treatments.

TreatmentExplorer is in an early stage of design and further refinements based on the healthcare delivery needs and practices will have to be made. How might such an interface be used by a physician alone before the visit or in collaboration with the patient? How do personal differences impact the performance of complex decision aids such as TreatmentExplorer? Further evaluation is needed to confirm that our results are generalizable to older populations instead of students. Results are likely to vary between participants with varying levels of statistical training, health literacy, reading ability, or computer literacy.

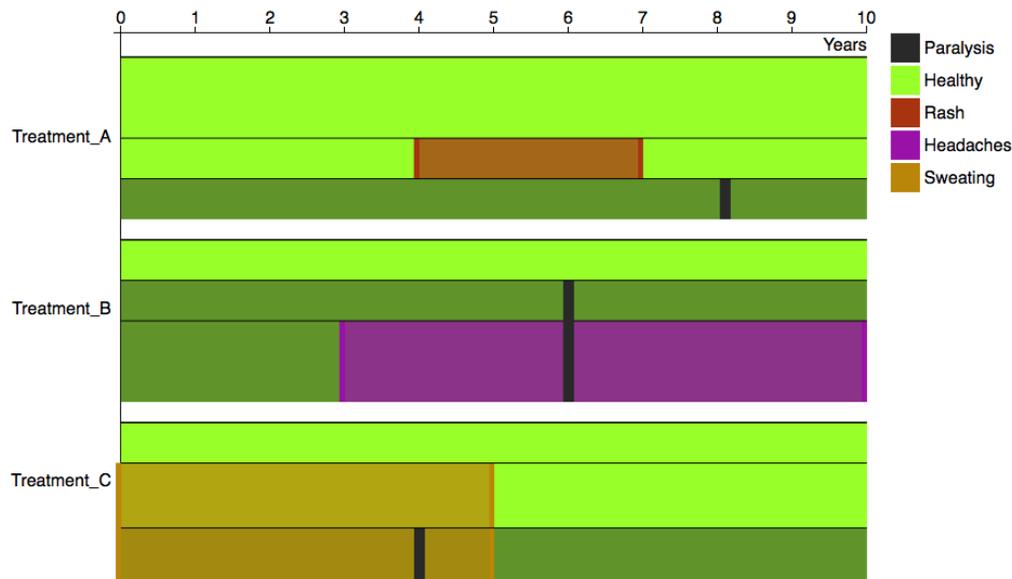


Figure 10: Showing multiple treatments at once as an option for advanced users

7. GUIDELINES FOR ANIMATION AND INTERACTION

Feedback from experts and observations during the two user studies encourage us to propose guidelines for designers. They complement existing guidelines for use of animation and text-narration (e.g. Heer and Robertson, 2007).

- **Begin Each Narration from the Same State:** Patients are perceptive of subtle unintended differences and will question the meaning of inconsistencies when they appear. To avoid this accidental confusion, all narrations should begin from the same clear starting state.
- **Follow a Consistent Narration Order:** Patients who watch a series of narrations will learn the order of information and begin to expect data points to follow the prescribed sequence. This not only helps them to structure how they think about the data but also provides a way to help patients compare data points across narrations. Patients who are learning a narrative interface will also be less intimidated if their expectations about what will happen next are not defied.
- **Provide Both Replay and Skip-Through Options:** Some patients will find value in re-watching a narration if they feel they missed some information in prior viewings. Other patients will be quick to absorb information and may only need a quick reminder of the final state of a narration later. Providing options for patients to quickly skip to the end of a narration will allow them to refresh their memory while full replay options will support both sets of patients.
- **Explain One Data Point per Stage:** Stages of narration should be simple to allow patients to focus their attention to one data point at a time. Related data points should be grouped into a sequence of stages.
- **Give Users Control Over the Flow of Narration:** Patients interacting with a guided narration will be trying to understand the information presented to them and learn from the narration. Some patients may need to re-read text or re-think accompanying animation several times to feel comfortable. Patients should then have the control to decide when the narration advances to the next stage. They may also want to turn-off all the animation after they become familiar with novel displays.

8. CONCLUSIONS

TreatmentExplorer is a prototype decision aid allowing patients to explore their treatment options and educate themselves alone or with their physicians on the benefits and risks of each treatment. To the best of our knowledge no other decision aid attempts to present such rich information in a personalized interactive environment (they typically present only one or two measures, at a single point in time, and simply list possible side effects). While the combined visualization of all the needed information may appear complex we believe that the use of staged-animation and text-only narration enables patients to learn the meaning of the visual representation, which they can then use to remember and compare the benefits and risk associated with each treatment, to ultimately make more informed decisions.

Our empirical evaluation demonstrated that with the use of staged animation and guided narration TreatmentExplorer users with no prior experience were able to learn about treatment options, and they did so more easily than with a text based or static decision aid. Further refinements and in-depth studies with real patients learning about real treatment options will be needed to fully understand the role of novel interactive decision aids such as TreatmentExplorer, and how they might be best integrated into the clinical workflow. We believe the use of data from the patient's health organization, local hospital or from their own doctors will be seen by patients as more relevant, thereby engendering greater trust in the information they receive. We hope that this early work will inspire further research on medical decision aids, and ultimately help patients get a greater understanding of their options in order to take greater responsibility for their care and to confidently assert their preferences.

Our initial evaluation was conducted with participants recruited among campus students. They may not be representative of the global patient population, but probably closer to the growing population of savvy computer users seeking advice from online interactive decision aids, or to physicians willing to use computers with their patients while discussing treatment options. Further studies will need to evaluate refined designs with different types of participants in order to refine our understanding of the benefits and limitations of the approach, and appropriate situations of use. Finally, the data centric approach to decision aid requires data standardization, adequate privacy protection, and general validation before becoming a reality. Nevertheless, the TreatmentExplorer approach could be used today to display aggregate data from clinical trials and models. We hope that decision aids such as TreatmentExplorer can provide incentive to address the problems of EHR data quality. While today very few organizations have access to data warehouses with enough patients and sufficient data quality to allow adequate personalization of decision aids, the momentum is growing quickly.

We hope that our prototype, video and proposed guidelines will also inspire new designs in other human-computer interaction application domains. For example we believe that tools like Treatment Explorer could be used as decision aids to help consumers choose products (by comparing the experience of other consumers with multiple products over several years), or to help students compare training programs or degrees to pursue (e.g. BS, Master or PhD).

ACKNOWLEDGMENTS

We would like to thank Beth St. Jean, Linda Aldoory, Min Qi Wang for their expertise and advice during the development of the TreatmentExplorer prototype. This work is supported in part by Grant No. 105592 for Patient-Centered Cognitive Support under the Strategic Health IT Advanced Research Projects Program (SHARP) from the Office of the National Coordinator for Health Information Technology.

REFERENCES

- Amershi, S., Carenini, G., Conati, C., Mackworth, A. K., Poole, D. Pedagogy and usability in interactive algorithm visualizations: Designing and evaluating CIspace, *Interact. Comput.* 20, 1 (2008) 64-96
- Bekker, H. L., Hewison, J. and Thornton, J. G. ,Understanding why decision aids work: linking process with outcome, *Patient Education and Counseling*, vol. 50, no. 3, pp. 323-329, 2003.
- Belkora, J. K., Loth, M. K. , Volz, S. and Rugo, H. S., Implementing decision and communication aids to facilitate patient-centered care in breast cancer: A case study, *Patient Education and Counseling*, vol. 77, no. 3, pp. 360-368, 2009.
- Brown, S. M., Culver, J. O. , Osann, K. E., MacDonald, D. J., Sand, S., Thornton, A. A. , Grant, M. , Bowen, D. J. , Metcalfe, K. A., Burke H. B. and others, Health literacy, numeracy, and interpretation of graphical breast cancer risk estimates, *Patient education and counseling*, vol. 83, no. 1, pp. 92-98, 2011.
- Bunge, M., Muhlhauser, I. , Steckelberg, A. and others, What Constitutes Evidence-Based Patient Information? Overview of Discussed Criteria, *Patient Education and Counseling*, vol. 23, no. 3, pp. 229-235, 2010.
- Chauvenet 2014: Chauvenet's criterion, In Wikipedia. Retrieved May 15, 2014, from http://en.wikipedia.org/wiki/Chauvenet's_criterion
- Chauvenet, William. *A Manual of Spherical and Practical Astronomy* V. II. 1863. Reprint of 1891. 5th ed. Dover, N.Y.: 1960. pp. 474–566.
- Chevalier, F., Dragicevic, P., Bezerianos, A. and Fekete, J.-D., Using text animated transitions to support navigation in document histories, *Proceedings of ACM CHI '10*, pp. 683-692, 2010.

- Civan-Hartzler, A., McDonald, D. W., Powell, C., Skeels, M. M., Mukai, M., and Pratt, W., Bringing the field into focus: user-centered design of a patient expertise locator. Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI '10). ACM, New York, NY, USA, 1675-1684, 2010
- Consumer Reports Health, Treating Type 2 Diabetes: The Oral Diabetes Drugs, Comparing Effectiveness, Safety, and Price, December 2012. [Online]. Available: <http://www.consumerreports.org/health/best-buy-drugs/type2diabetes.htm>. [Accessed 30 June 2013].
- Dillar, A. J. , Scherer, L. , Ubel, P. A. , Smith, D. M. , Zikmund-Fisher, B. J. , McClure, J. B., Greene, S. Stark, A. and Fegerlin, A., Breast Cancer Anxiety's Assisatons with Responses to a Chemoprevention Decision Aid, *Social Science & Medicine*, vol. 77, pp. 13-19, 2013.
- Edwards, A., Gray, J. , Clarke, A. , Dundon, J. , Elwyn, G. , Gaff, C. , Hood, K., R. Iredale, S. Sivell, C. Shaw and H. Thornton, Interventions to improve risk communication in clinical genetics: Systematic review, *Patient Education and Counseling*, vol. 71, no. 1, pp. 2-25, 2008.
- Fagerlin, A. , Zikmund-Fisher, B. J. and Ubel, P. A. , Helping Patients Decide: Ten Steps to Better Risk CCommunication, *Journal o the National Cancer Institute*, vol. 103, no. 19, pp. 1436-1443, 2011.
- Feldmant-Stewart, D. and Brundage, M. D., Challenges for Designing and Implementing Decision Aids, *Patient Education and Counseling*, vol. 54, no. 3, pp. 265-273, 2004.
- Franklin, L., Plaisant, C., Shneiderman, B, An Information-Centric Framework for Designing Patient-Centered Medical Decision Aids and Risk Communication, in *Proceedings of AMIA Annual Symposium*, pp. 456-465, 2013.
- Franklin, L., Designing Interactive Decision Aids for Medical Risk Communication and Exploration of Treatment Options, Masters Dissertation from the Department of Computer Science, 2013 - <http://www.cs.umd.edu/localphp/hcil/tech-reports-search.php?number=2013-20>
- Frankovich, J., Longhurst, C. A. and Sutherland, S. M. , Evidence-Based Medicine in the EMR Era, *New England Journal of Medicine*, vol. 365, no. 19, pp. 17-58-1759, 2011.
- Han, P. K. , Klein W. M. and Arora, N. K., Varieties of Uncertainty in Health Care: A Conceptual Taxonomy, *Medical Decision Making*, vol. 31, no. 6, pp. 828-838, 2011.
- Han, P. K. , Klein, W. M., Killam, B., Lehman, T., Massett, H. and Freedman, A. N., Representing Randomness in the Communication of Individualized Cancer Risk Estimates: Effects on Cancer Risk Perceptions, Worry, and Subjective Uncertainty About Risk, *Patient Education and Counseling*, vol. 86, no. 1, 2012.
- Heer J., and Robertson, G. , Animated transitions in statistical data graphics, *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1240-1247, 2007.
- Hildon, Z., Allwood, D. and Black, N. , Making Data More Meaningful: Patients' Views of the Format and Content of Quality Indicators Comparing Health Care Providers., *Patient Education and COUNseling.*, vol. 88, no. 2, pp. 298-394, 2012.
- Hollands, G. J. and Marteau, T. M., The Impact of Using Visual Images of the Body Within a Personalized Health Risk Assessment: An Experimental Study, *British Journal of Health Psychology*, 2012.
- IOM - Institute of Medicine of the National Academies, Social Networking Sites and the Continuously Learning Health System: A Survey, 2014 [Online] <http://www.iom.edu/Global/Perspectives/2014/Sharinghealthdata.aspx> [Accessed February 21, 2014]

- Jones, J. , Shah, N. R. , Bruce, C. A. and Stewart, W. F., Meaningful Use in Practice: Using Patient-Specific Risk in an Electronic Health Record for Shared Decision Making, *American Journal of Preventative Medicine*, vol. 40, no. 5, pp. s179-s186, 2011.
- Karacapilidis, N., Koukouras, D., A Web-based System for Supporting Collaboration towards Resolving Oncology Issues. *Oncology Reports*, Vol. 15, No 4, 2006, pp. 1101-1107
- Kosara, R. , and Mackinlay, J., Storytelling: the Next Step for Visualization, *Computer*, vol. 46, no. 5, pp. 44-50, 2013.
- National Cancer Institute, Breast Cancer Risk Assessment Tool, National Cancer Institute, 16 May 2011. [Online]. Available: <http://www.cancer.gov/bcrisktool/>. [Accessed 16 September 2013].
- National Cancer Institute, National Cancer Institute, National Institutes of Health, 2013. [Online]. Available: www.cancer.gov. [Accessed 1 May 2013].
- National Cancer Institute, What You Need to Know About Breast Cancer, 26 September 2012. [Online]. Available: <http://www.cancer.gov/cancertopics/wyntk/breast/page7>. [Accessed 30 June 2013].
- PatientsLikeMe®, www.patientslikeme.com [Accessed February 21, 2014]
- Perer, A., and D. Gotz, Data-Driven Exploration of Care Plans for Patients, in *ACM SIGCHI CHI 2013 Extended Abstracts*, Paris, France, 2012.
- Plaisant, C., Shneiderman, B., , Show Me! Guidelines for Producing Recorded Demonstrations, *IEEE Symposium on Visual Languages and Human-Centric Computing*, pp. 171-178, 2005.
- Shneiderman, B., Plaisant, C., and Hesse, B. W., Improving health and healthcare with interactive visualization methods, *IEEE Computer - Special Issue on Challenges in Information Visualization* 46, 5, pp. 58 – 66, 2013.
- Sillence, E, Briggs, P., Harris, P. and Fishwick, L., Going online for health advice: Changes in usage and trust practices over the last five years, *Interact. Comput.* 19, 3, pp 397-406, 2007.
- Sudore, R. L. and Schillinger, D. , Interventions to Improve Care for Patients with Limited Health, *Journal of Clinical Outcomes Management*, vol. 16, no. 1, pp. 20-29, 2009.
- Susan G. Komen Breast Cancer Foundation [Online] Available: <http://ww5.komen.org/Default.aspx>. [Accessed 8 July 2013].
- Timmermans, D. R., Prenatal screening and the communication and perception of risks, in *Gynaecology, Obstetrics, and Reproductive Medicine in Daily Practice: Proceedings of the 15th Congress of Gynaecology, Obstetrics and Reproductive Medicine*, 2005.
- Tversky, B., Visualizing Thought, *Topics in Cognitive Science*, vol. 3, pp. 499-535, 2011.
- Visualizing Health [online] <http://www.vizhealth.org> [Access February 19, 2014]
- Wilcox, L., R. Patel, A. Back, M. Czerwinski, P. Gorman, E. Horvitz and W. Pratt, Patient-Clinician Communication: the Roadmap for HCI, *CHI EA '13*, pp. 3291-3294, 2013.
- Wongsuphasawat, K., J. Guerra Gomez, C. Plaisant, T. Wang, M. M. Taieb and B. Shneiderman, LifeFlow: Visualizing and Overview of Event Sequences, *Proceedings of ACM CHI '11*, pp. 1747-1756, 2011.