Interactive Visual Analysis of Patient-Reported Outcomes for Improved Cancer Aftercare

- Detailed Results of an Expert Evaluation Study According to our Questionnaire

Domain experts were asked to fill out a questionnaire about the medical relevance of and the visualization techniques used in our Interactive Visual Analysis tool of Patient-Reported Outcome (PRO) measures for Improved Cancer Aftercare. In total, 20 domain experts (6 otolaryngologists from the Ear-, Nose- and Throat (ENT) department, 9 radiotherapists / radio-oncologists, 4 medical researchers, and 1 nurse) evaluated the visual approach by answering 26 and 5 closed-ended and open-ended questions, respectively. The questionnaire was inspired by *Ergonomics of Human System Interaction* defining "dialogue principles" (ISO 9241-110:2006) and the questions were ranked using a six-point Likert scale (1, 2, 3, 4, 5, 6). For the interpretation of the predefined Likert scores, we provide the number *S* of experts who chose the individual categories.

Procedure. Before the study, each participant was given an introductory tutorial. The experimenters explained the aim of the presentation and research, the structure of the study, the PRO itself and the approximately needed time. Furthermore, they declared the functionalities of the original PRO visualization for physicians (Fig. 1 in the paper), followed by a training session, which comprised one trial. The training session was designed to enable participants to familiarize themselves with the user interface and typical visual patterns. To minimize interruption of the tight clinical schedule and attract more attendees, we decided to have just one training trial. Afterward, the participants independently performed the following four closed-ended evaluation tasks investigating one additional patient case:

1. Mark the current value in the functional aspect of "health and quality of life".
2. Mark the patient’s development from the last to the current value in the functional aspect "pain".
3. Mark the answer to the penultimate question "dysphagia weight loss" in the functional aspect "swallowing".
4. Mark the suspicious answers.

In performing the individual tasks, the participants could choose one out of three, three, and four possible answer options for tasks 1, 2, and 3 and multiple out of 10 answer options for task 4. During the autonomously performed tasks, the time was measured and the participants had to fill the answers of each task in a paper-based form. The same procedure was repeated for the new presentation using two different patient cases alternatingly. Additionally, the tasks for the new presentation included five further assignments to assess the functionality of the newly implemented features:

1. Open the development visualization over time for the functional aspect "fatigue".
2. Set the number of tolerated missing values.
3. Write down the distribution in the first year after therapy of the functional aspect "fatigue".

4. Select all patients with a similar development in the functional aspect "fatigue" and write down the number.
5. How many employed patients had a bad condition in the second year after therapy in the functional aspect "anxiety"?

In this context, task 1 and 2 comprises assignments, whereas in tasks 3 ~ 5 the numbers needed to be filled in. After completing the tasks, the participants were asked to answer the survey.

Time and Error Comparison Between the Original and the New Visual Presentation. To contrast the original and new visual presentation of functional aspects in head and neck cancer aftercare, the timings for performing the same tasks in the two presentations were compared. The mean time for answering the tasks using the original presentation was $t_{original}$ = 1.38 minutes, whereas performing the same tasks using the new presentation took on average only t_{new} = 1 minute (Figure 1). The paired two-tailed t-test of the time savings resulted in a p-value of p = 0.0038, which is considered to be statistically significant by conventional criteria. Even when considering only participants familiar with the original presentation, the time needed to perform the same tasks was improved ($t_{original}$ =1.75 minutes vs. t_{new} = 1.25 minutes). Furthermore, the answer accuracy was higher using the new presentation and the increase was statistically significant (p = 0.0101) (Figure 2). Probable reasons therefore were the simplified presentation of a patient's current condition, the less needed space, and a good usability. Additionally, the visual encoding using icons instead of traffic lights was stated to be more intuitive. Especially in healthcare, time-saving is very important. The saved time can be used to have further conversations and investigations with the patient.

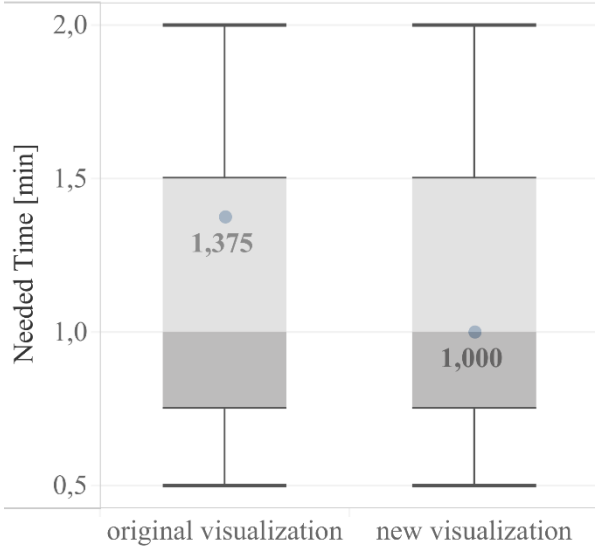


Figure 1: Box plot presentation of needed time [min] to perform the same evaluation tasks in original and new visualization. The blue dots indicate the average needed time.

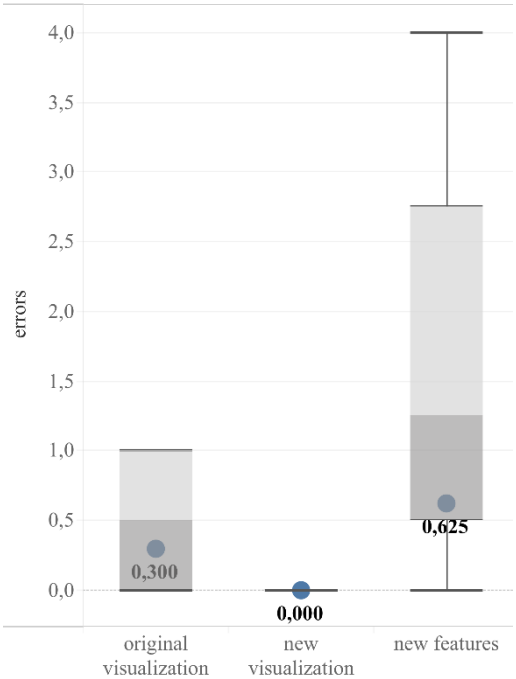


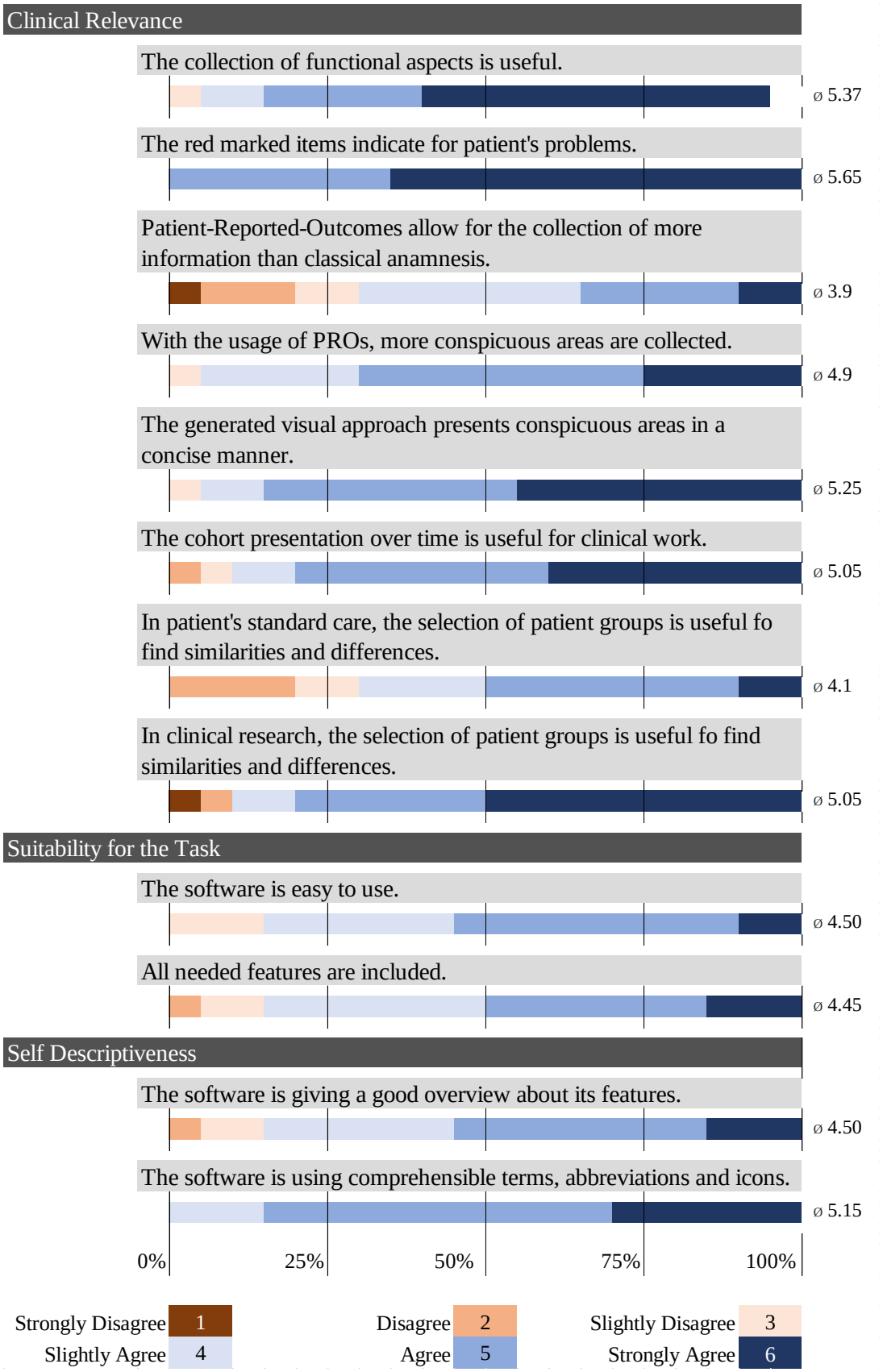
Figure 2: Box plot presentation of erroneously performed tasks using original and new visualization as well as using the new features, e.g. investigating the cohort development. The blue dots indicate the average number of performed errors.

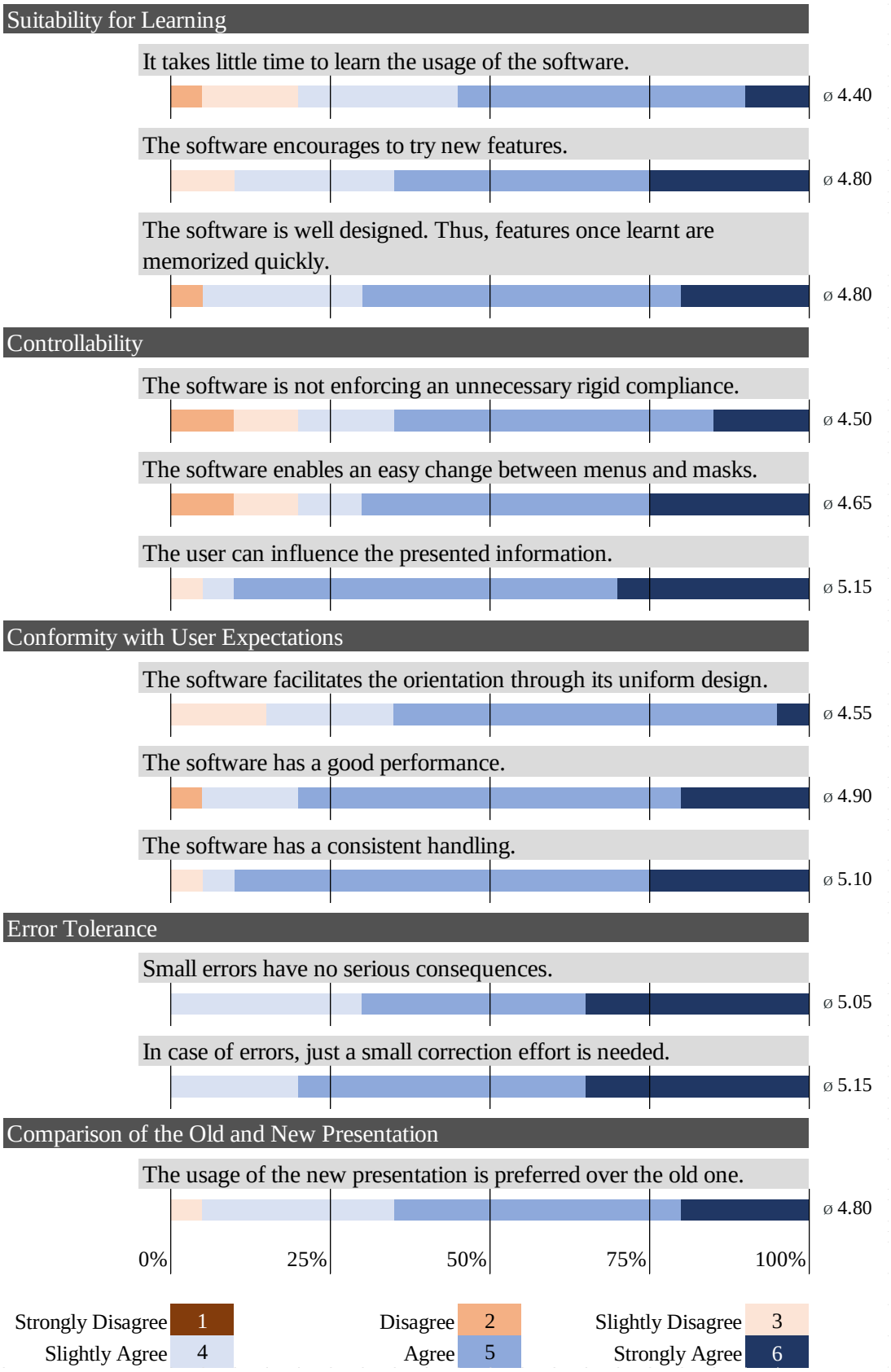
Clinical Relevance. Almost all experts emphasized the importance of PROs in general and our new visual approach. They especially highlighted the Patient Status view and the Patient Development view to be very useful in identification of a patient's conspicuous functional aspects.

Evaluation of New Functionalities. Investigating the performance of the new features, such as investigating the cohort development, it can be seen that on average 0.625 errors were performed (Figure 2). The main causes laid in the understanding and usage of the comparative visualization. Two participants claimed especially the selection of patient groups as not intuitive. However, this could be due to the fact that each participant had just one trial for practice and that this visual presentation is not as familiar as e.g. bar charts. By having more practice, a better comprehensibility of the comparative visualization could be achieved and thus, fewer errors are likely. Additionally, the comprehensibility could be improved by adding more textual explanations. Overall, the comparative visualization was stated to be useful especially for clinical research.

Evaluation of the Survey. To evaluate the survey, the averages of scores for each question and topic were calculated and investigated (**Responses of our Domain Expert Evaluation**). Additionally, the open-ended questions related to violations in each topic were examined. Overall, the averages were located in the range [4.4,5.15]. In particular, the topic error tolerance as well as several questions related to the topics self-descriptiveness, controllability, and conformity with user expectations received good ratings. Reasons for this were the very familiar icon presentation and easily comprehensible functionality of patient status and patient development view (Fig. 3). However, the cohort development view downgraded the rating. Therefore, e.g., the topic suitability for the task as well as parts of self-descriptiveness, suitability for learning, controllability, and conformity with user expectations received lower ratings. Furthermore, participants stated the wish for more explanations and fewer clicks on icons. During the evaluation tasks, it was observed that the click positions could be improved. Participants mostly intuitionally clicked on the functional aspects' labels and not on the icons. Furthermore, the double click on the icons to visualize the comparative visualization was pointed out to be not easily memorizable. A better way would be to present a highlighted button to open the lower part. The hiding of the lower visualization in the beginning, however, was appreciated. Otherwise, the initial view would be too overloaded. Furthermore, during the patient consultations, the comparative visualization was stated to be rarely needed. Overall, the new presentation was preferred to the original one (average score of 4.8). For that reason, the integration of the new presentation into the PRO and thus, its clinical use is scheduled for April this year.

Responses of our Domain Expert Evaluation




Suitability for Learning
It takes little time to learn the usage of the software.
ø 4.40
The software encourages to try new features.
ø 4.80
The software is well designed. Thus, features once learnt are memorized quickly.
ø 4.80
Controllability
The software is not enforcing an unnecessary rigid compliance.
ø 4.50
The software enables an easy change between menus and masks.
ø 4.65
The user can influence the presented information.
ø 5.15
Conformity with User Expectations
The software facilitates the orientation through its uniform design.
ø 4.55
The software has a good performance.
ø 4.90
The software has a consistent handling.
ø 5.10
Error Tolerance
Small errors have no serious consequences.
ø 5.05
In case of errors, just a small correction effort is needed.
ø 5.15
Comparison of the Old and New Presentation
The usage of the new presentation is preferred over the old one.
ø 4.80
0%
25%
50%
75%
100%
Strongly Disagree 1
Disagree 2
Slightly Disagree 3
Slightly Agree 4
Agree 5
Strongly Agree 6