Title
Evaluating a Novel Summary Visualization for Clinical Trial Reports: A Usability Study.

Authors
Tong, Maurine
Hsu, William
Taira, Ricky K

Publication Date
2016

Peer reviewed

Evaluating a Novel Summary Visualization for Clinical Trial Reports: A Usability Study

Maurine Tong, BS, William Hsu, PhD, Ricky K. Taira, PhD
University of California Los Angeles, Los Angeles, CA

Abstract

Contributions of clinical trials are captured in published reports that are unstructured and often require extensive manual review to gain a deeper understanding of the study itself. Our goal is to increase comprehension and decrease the time necessary to understand these reports through the use of visualization tools. In this paper, we specify and evaluate the visualization of a previously developed representation as well as gain insight from user input for further development. The usability experiment consisted of a two-arm study with users either having or not having access to the visualization. A user questionnaire was used to measure time spent and accuracy in comprehension; intuitiveness and reproducibility of the visualization; and preferences. We found that having the visualization required on average 28.1% less time (25.8 min vs. 35.8 min, p=0.01) while maintaining similar accuracy (73.7% vs. 67.0%). Users were then asked to create their own visualizations, with their visualizations averaging 86.1% similar to the gold standard. All participants either preferred the visualization over the status quo or preferred both equally. These results demonstrate that novel visualizations for trial reports could provide time savings and achieve similar accuracy as reviewing the paper itself. Understanding the strength and quality of clinical trials can be alleviated with a visualization that makes content explicit.

Introduction

Information within clinical trial reports can help inform clinical guidelines and support evidence-based medicine (EBM)[1]. Accurate interpretation of these results is important to characterize the quality and strength of a given clinical trial study[2]. However, this information is difficult to assess, one reason is that its representation is free-text with numerical data scattered in tables, figures, and embedded in the text. A significant amount of effort is needed to identify and interpret information scattered throughout published reports, requiring the clinicians' and researchers' to organize this information mentally[3,4]. This further complicates tasks such as assessing the strength of similar trials or comparing trials with conflicting results[2]. The overarching goal of this work is to evaluate a visualization that links scattered data of a clinical trial together to better assess and understand its contributions.

The medical community is striving towards a structured representation to facilitate the deeper understanding of contributions from clinical trials. One attempt at standardizing the type of information within clinical trial reports is the Consolidated Standards of Reporting Trials (CONSORT), which comprises a checklist of essential items that should be included in reports of clinical trials and a diagram for documenting the flow of participants through a trial[5,6]. CONSORT is a starting point to help investigators and others write or appraise trial reports, but its purpose is to promote good reporting, with no particular emphasis on appraising trial reports. Since CONSORT's founding, groups have worked on defining, structuring and standardizing information related to clinical trials. RCT Schema captures concepts related to a trial's design, basic intervention description, execution, administration, and results[7,8]. These data structures create classes for specific types of information, with a standardized way to fill in classes and attributes. While structuring of data has been done mainly for the purpose of patient recruitment, few efforts have attempted to develop a representation for capturing the context of information for further interpretation. Clinical Research Eligibility Criteria Extraction and Representation (EliXR) addresses the disconnect in ambiguous eligibility criteria and clinical data results through developing a framework for eligibility criteria text[9]. Our effort attempts to capture a wider spectrum of information and reported statistics necessary to perform quality assessment, relationships to piece information together, and a method for interactively presenting a published report to facilitate the user's understanding of the study.

We have been developing a data model, which abstracts various types of information presented in clinical trial literature and connects these pieces together, and a visualization that allows target users (e.g., translational researchers, research students) to better interpret this information. We hypothesize that this representation and its visualization can facilitate understanding and analysis of clinical trial studies. In this paper, we specify a representation to classify information presented in clinical trial reports. We evaluate the visualization and report its performance, measured on a set of non-small cell lung cancer (NSCLC) clinical trial reports.

Methods

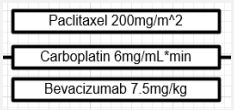
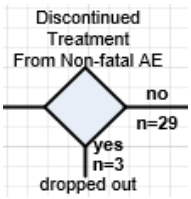
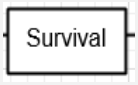
A usability study was performed to assess the benefit of using a standard visualization to understand the contents of a clinical trial report. In particular, we were interested in determining how well the visualization assists a researcher with understanding the published report's content, its ease of use, whether its presentation is intuitive to navigate and comprehend, and user preferences. Various aspects of the representation and visualization have been previously demonstrated[10], including methods for representing experimental flow[11] and statistical analysis[12]. In this paper, we briefly describe the representation model and the visualization. Results of the usability study were based on a comparison of interpreting information using the status quo published report versus the visualization.

The Representational Model

A representational model has been developed to store information presented in a published report. The first goal of this structured representation is to capture the essential elements related to recruitment, steps of the experiment, the data collection process, the data, the analyses, and the conclusions in a logical and consistent manner. The second goal is to provide context for observational data or reported statistics.

The model is functionally divided into purpose, methodology with raw data as a subgroup of methodology, statistical methods, and interpretations. These sections parallel the structure of a clinical trial report. Under the purpose section, the model captures the overall goal of the clinical trial with a free-text statement. The methodology section of the model contains a process model of nodes with a list of relevant variables, attributes, and values. Table 1 shows typical examples from a NSCLC clinical trial report[13].

Table 1. Example excerpts from clinical trial report[13].

Excerpts from report	Node	Variables within Node
"Patients received up to six cycles of carboplatin/paclitaxel. Paclitaxel (200 mg/m2) was administered over 3 hours every 3 weeks. Carboplatin dosing was based on the Calvert formula with a target area under the curve of 6 mg/mL x min and glomerular filtration rate (GFR) estimated for males as GFR = (140-age) x weight/72 x (serum creatinine)."	Treatment 	• drug name(s) • dosage(s) • duration • frequency
"Eleven patients discontinued treatment as a result of a nonfatal AE. Discontinuations occurred as a result of: hemorrhagic event (three patients) in the low-dose bevacizumab arm; a hemorrhagic event (one patient); Aspergillus lung abscess (one patient); aspiration pneumonia (one patient); thrombotic stroke (one patient); vertebral fracture (one patient); and peripheral europathy (paclitaxel-related; one patient) in the high-dose arm. In two cases, bevacizumab was discontinued following initiation of anticoagulant therapy. Bevacizumab was withheld from one patient with subclavian vein thrombosis."	Discontinued Treatment 	• Hemorrhagic event • Aspergillus lung abscess • aspiration pneumonia • thrombotic stroke • vertebral fracture • peripheral neuropathy • following initiation of anticoagulant therapy • subclavian vein thrombosis
" The overall response rate showed a trend toward improved response for patients receiving bevacizumab, with the highest response noted in the high-dose group (31.5%) and the lowest in the control group (18.8%; Table 2)....Median TTP was longer in the high-dose bevacizumab arm compared with the control arm (7.4 v 4.2 months; P = .023; Fig 1)....Survival for the high-dose bevacizumab arm was modestly longer than the control arm (17.7 v 14.9 months; P = .63; Fig 2)."	Survival 	• median survival (month) • median survival (range) • % PFS vs. time • % survival vs. time

The process model serves three purposes: (1) it documents the recruitment process using a flow chart to represent eligibility criteria, (2) demonstrates randomization by branching of nodes from one to many nodes, with each node representing a subpopulation, and (3) captures the overall study flow using nodes to represent specific types of experimental procedures. Typical types of nodes include general population, population sampling pool, decision boxes, recruitment criteria, control and intervention population, methods, and observation points. Each node is associated with a list of relevant variables, attributes, and values within a spreadsheet structure corresponding to the raw data stated in the published report. Together, the process model and linked spreadsheet allow fragmented knowledge pre-

sented in free-text, figures, and tables, to be placed in the context of the entire clinical trial "experiment." The statistical methods section of the model collects information from the text and displays it in a consistent manner with references back to the nodes of the process model and data elements from the spreadsheet. Each statistical method includes inputs from the spreadsheet, outputs as a measure of statistical significance, and a statement of significance. The interpretations section of the model captures the contributions of the clinical trial as stated by the published report.

The Visualization

The visualization assists with viewing and interacting with the contents of the representation from each clinical trial report. The visualization is divided into four panels following the structure of the model: (1) purpose of the trial, (2) process model and data grid, (3) list of statistical methods with variables compared and results, and (4) interpretations. The purpose of the trial and the endpoints are listed at the top of the visualization in free-text. The recruitment process and methodology are below. At the bottom are the interpretations summarizing the contributions of the published report listed in free-text (Figure 1).

The process of a clinical trial study is displayed as a timeline of events performed over the course of the entire study. In this example[13], the study population is recruited using three separate types of criteria: tumor status, presence of prior chemotherapy, and other clinical criteria. The study population is then randomized into three groups, control, low dose, and high dose; and each group undergoes a set of protocols. Note that the control receives no drug intervention as shown by a lack of yellow boxes in the control row (Figure 2).

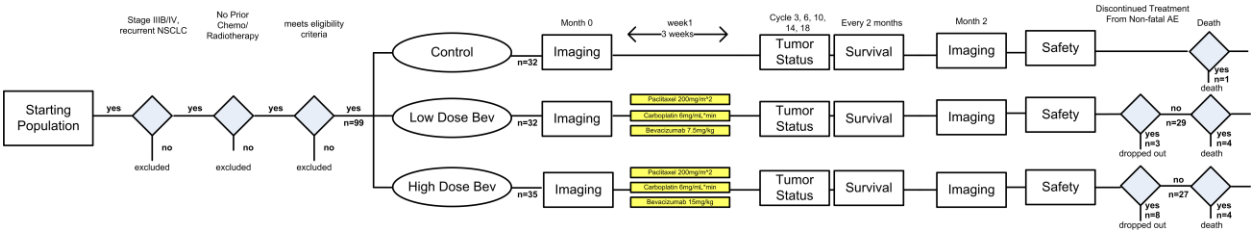


Figure 2. Flow chart of events[13].

Directly below the diagramming interface, the methodology, raw data, and results can be viewed as a data grid. The variables appear on the left side, with one variable per row. Thus, the rows give an inventory of variables that are mentioned in the clinical trial study. The columns of the data grid correspond to different nodes in the process model. Any numerical data generated is placed along the same row as the variable and is beneath its corresponding event. Thus, each cell in the data grid is associated with an event node from the flow diagram and variable and the cell itself corresponds to the specifications or characterization of a variable for an experimental procedure of a group of patients for which the node refers to in the flow chart. A cell can show (1) summary statistics, and (2) all the patient values for a given variable, if available (Figure 3). We are working towards the customization of visualizing the data from individual cells. We are also developing methods to derive useful information from the data within related cells (e.g., points in Kaplan-Meier diagram where curves cross).

Population Statistics

Survival Rate					
Control		Low Dose		High Dose	
Median	14.9	Median	11.6	Median	14.9
Range	0.2-57.0	Range	0.2-56.8	Range	0.9-57.8
Time	Survival Rate*	Time	Survival Rate*	Time	Survival Rate*
0	1.0	0	1.0	0	1.0
10	0.52	10	0.68	10	0.71
20	0.31	20	0.35	20	0.38
30	0.18	30	0.27	30	0.20
40	0.17	40	0.21	40	0.08
50	0.08	50	0.15	50	0.05

* Values are estimates based on figure

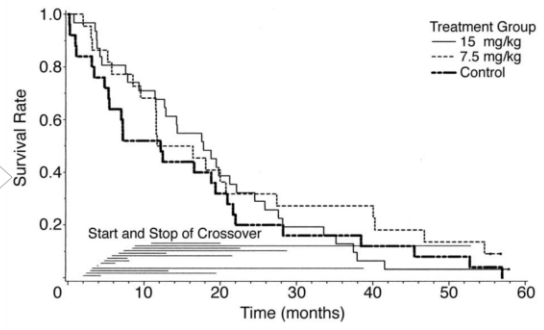


Figure 3. (Left) Example of data stored within a cell. Data is sorted by population and organized as individual data or population statistics. (Right) Graph generated by data. In this example, a Kaplan-Meier graph is drawn[13].

The panel for statistical methods provides a visual inventory of all the tests performed. Each test is listed with its corresponding inputs from the data grid, the test statistic, output statistics such as a p-value, and a statement of

significance (Figure 4). For the purpose of the usability study, participants were presented with a mockup of the visualization in Microsoft Visio for each published report created; the final version is being implemented as an interactive application using Java.

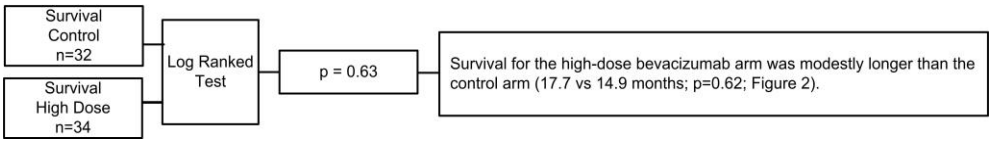


Figure 4. Statistical method visualization. Inputs include experimental arm name, sample size, and variable. This is followed the statistical test within the box. Outputs include a quantitative measure of statistical significance, and a statistical statement.

Overview of Usability Study

The usability study had three objectives: (1) test the ease of interpretation and time saved against the status quo (i.e., the published report), (2) test the intuitiveness of the visualization through reproducibility of the visualization, and (3) elicit preferences from a targeted user group.

Paper cohort: Clinical trials were randomly chosen based on a PubMed search using the keywords "EGFR", "lung cancer", "non-small cell lung cancer", and "clinical trial" (Table 2). The search yielded 38 published reports which were different from the set of reports used to develop the original visualization. For the initial scope of this study, we randomly selected three papers that met the criteria of being a clinical trial about NSCLC involving EGFR mutations to assess time spent and accuracy. The fourth report was used for the study on reproducibility.

Table 2. Randomly selected clinical trial reports on non-small cell lung cancer.

Trial number	Title of Report	Outcome variable	Sample size	Total Events	Date Published
1	Randomized Phase II Trial Comparing Bevacizumab Plus Carboplatin and Paclitaxel With Carboplatin and Paclitaxel Alone in Previously Untreated Locally Advanced or Metastatic Non-Small-Cell Lung Cancer[13]	Response Rate (RR)	99	33	2005
2	First-Line Gefitinib in Patients with Advanced Non-Small-Cell Lung Cancer Harboring Somatic EGFR Mutations[14]	Objective Response Rate (ORR) = sum of patients with confirmed complete and partial responses / number of patients treated	98	15	2008
3	EGFR expression as a predictor of survival for first-line chemotherapy plus cetuximab in patients with advanced non-small-cell lung cancer: analysis of data from the phase 3 FLEX study[15]	Overall Survival (OS)	1125	24	2011
4	First-Line Gefitinib in Patients with Advanced Non-Small-Cell Lung Cancer Harboring Somatic EGFR Mutations[16]	Objective Response Rate (ORR) = sum of patients with confirmed complete and partial responses / number of patients treated	889	17	2008

Participant recruitment: The participants were students with familiarity in biostatistics and informatics. In this usability study, students served as a proxy for potential users (i.e., biostatisticians and clinician researchers). The study for time spent and accuracy, and for eliciting preferences had 11 participants, and the study for reproducibility had 6 participants. All participants were asked in a pre-questionnaire about the length and fluency of their experience with clinical trial reports, non-small cell lung cancer, and statistics.

Task definition to assess time spent and accuracy: For the first objective, the participants used the visualization to answer questions demonstrating their comprehension of the clinical trial and recorded the time required to answer each question. Questions were divided into two types: comprehension to assess whether the individual is able to synthesize evidence from the published report, and information retrieval (IR) to focus on locating specific pieces of evidence in the report. Comprehension questions were developed based on reporting guideline requirements[5], and

asked readers to interpret the objective and claims made in the published report. For example, one comprehension question asked: "The trial states 'This large prospective biomarker study found that patients with activating EGFR mutations derive the greatest PFS benefit from erlotinib maintenance therapy.'[16] Describe the method, numerical data, and analyses for this statement." IR questions focused on locating key information as adapted from applicable CONSORT requirements[6]. IR questions include reporting the eligibility criteria, locating the experimental arms, summarizing the methodology, and identifying the results of statistical tests. Questions of both types were presented using multiple choice, fill-in-the-blank, and short answer. All questions were reviewed by a biostatistician who was not involved in the development of the system to reduce bias in word-choice and to ensure conformance to standard guidelines and terminology. The gold standard was created by a domain expert who was given an open amount of time. Tasks were timed and graded for accuracy by determining the percentage of questions answered correctly.

Study design to assess time spent and accuracy: A two-arm randomized trial design was used with 11 participants and 3 clinical trial studies. The presentation of the trial was randomized for each participant and each participant reviewed clinical trial 1, 2 and 3 (Table 2). For each clinical trial, participants were randomized into the visualization study arm or the status quo study arm. In either study arm: (1) participants filled out paperwork (consent form, pre-test questionnaire) and received a tutorial on how to interpret the visualization based on two example questions; then (2) participants completed the usability sessions either with the status quo or visualization; finally (3) participants answered a post-questionnaire about the visualization (Figure 5).

Task definition to assess reproducibility: For the second objective, the participants were asked to generate a section of the visualization (i.e., the experimental flow) using a provided set of guidelines and a tutorial. Reproducibility was assessed through percent match of elements, nodes and variables. Elements are annotations given to a node, and can include sample size and time points. Nodes are used to label processes for recruiting a sample population and intervention processes. Variables describe and characterize a node. Example variables include randomization methods, baseline experimental procedures, and survival assessment methods. Tasks were compared with a gold standard created by a domain expert who was given an open amount of time.

Task definition to assess user preferences: A post-questionnaire was given to assess the affinity and usefulness of the visualization to gather impressions on the adequacy of its contents and to provide feedback on design, interface, and suggestions for additional functionalities. Participants answered questions using a Likert scale from one to ten, ten being very satisfied and very usable.

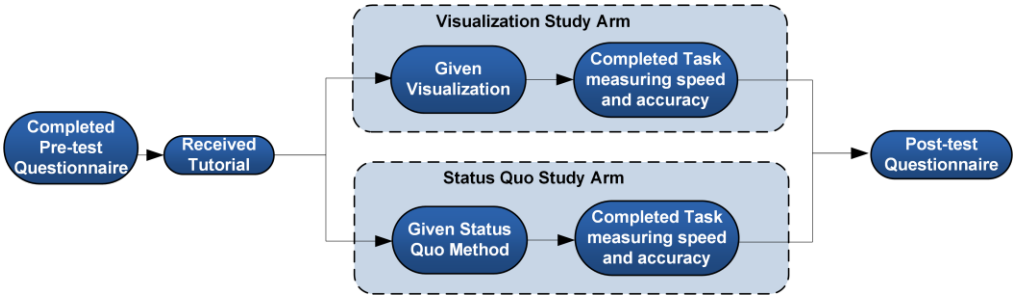


Figure 5. Study Design.

Statistical Methods

The participants used the status quo or the visualization to answer questions demonstrating their comprehension of the clinical trial and recorded the time required to answer the questions. The dependent measures of this usability study included time spent, measured in minutes; and accuracy, measured as the percentage of questions answered correctly. A weighted accuracy, combining time and accuracy, was calculated to place more emphasis on high accuracy scores that were obtained in a shorter amount of time. Descriptive statistics were computed for all measures.

$$\boldsymbol{Weighted\ Accuracy} = \frac{\mathbf{1}}{\boldsymbol{Time}} \cdot \boldsymbol{Accuracy}$$

Overall time spent and accuracy was determined by averaging over all values in each condition. Groups were conditioned on having either the status quo paper representation or the visualization. A pilot study was used to estimate the amount of time and accuracy for each task that was considered reasonable. A power calculation was performed to determine the appropriate sample size for the combination of participants and clinical trials needed. With an estimated time difference of 10 minutes (30 minutes vs. 40 minutes) and standard deviation of 8 minutes, a sample size of 12

per group would yield an 83% power with 5% significance level. With an estimated accuracy difference of 15% (70% vs. 85%) and standard deviation 17%, a sample size of 12 per group would yield an 80% power with 5% significance level. Hence, a sample size of at least 24 is needed, meaning at least 8 participants each reading 3 clinical trial reports. This is satisfied by the number of participants enrolled. A 2-sided student's t-test was used to compare (1) accuracy, (2) time spent, and (3) weighted accuracy using the visualization vs. using the status quo method.

Results

Participant Characteristics: Eleven student participants were involved in the study for time spent and accuracy, and for eliciting preferences; and six additional student participants were involved in the study for reproducibility. Participants ranged in experience from one to five years participated. All participants have read a clinical trial report before and took on average 80 minutes to read it completely. General participant characteristics are presented in Table 3.

Table 3. Characteristics of Participants. For confidence, scale values are 1 = not confident and 10 = very confident. For courses, values indicate number of college/graduate level courses.

	Time & Accuracy (n=11)		Reproducibility (n=6)	
	Mean	SD	Mean	SD
Confidence with understanding of cancer mechanisms	5.09	1.868	4.67	3.055
Confidence with knowledge on NSCLC	4.36	1.963	4.33	3.055
Confidence in understanding knowledge within reports	6.09	1.921	6.33	1.528
Confidence in understanding statistical methods	6.36	1.502	2.00	1.000
Confidence in assessing the quality of statistical tests	4.45	1.916	5.33	3.215
Courses on biology-undergraduate	3.18	2.676	6.67	2.887
Courses on biology-graduate	3.09	5.665	2.33	1.528
Courses on statistics-undergraduate	1.27	0.786	7.67	2.082
Courses on statistics-graduate	2.27	1.191	3.00	2.000

Time spent and accuracy: In this usability study, time spent and accuracy was measured using self-reported completion time, and percent questions answered correctly. Non-significant differences were found in both time and accuracy between the visualization condition and the status quo condition for each clinical trial study (Table 4). The point estimate of report #2 was shown to have decreased accuracy as compared with report #1 and #2. The accuracy can be affected due to an increase in complexity of the study design and greater amount of content for both the visualization and status quo method. The accuracy for comprehension question and for IR questions were separated for exploratory analyses. The mean accuracy for comprehension questions within one reports suggests a difference between the visualization condition and the status quo condition, favoring the visualization (data not shown). This suggests that using the visualization can increase comprehension. This trend within reports are currently being studied in an attempt to significantly increase accuracy in the visualization.

Overall accuracy was similar between the visualization and status quo, however, participants with the visualization had on average a quicker overall time than participants with the status quo (visualization 25.8 ± 10.10 minutes vs. status quo 35.8 ± 9.94 minutes; p=0.01). This suggests that information is easier to locate in a visualization than in the status quo. While the visualization provided similar accuracy, the tradeoff is a significant times savings when compared to the status quo alone. The weighted accuracy results further demonstrated that participants with the visualization had a combined quicker time and increased accuracy than participants with the status quo (visualization 3.39 ± 1.66 vs. status quo 2.10 ± 0.87, range=0.71 to 7.00, p=0.008).

Table 4. Measures of performance as a function of errors and time.

	Visualization				Status Quo			
Trial	Accuracy (%)	SD	Time (min)	SD	Accuracy (%)	SD	Time (min)	SD
1	77.3%	5.34%	27.3	10.73	68.9%	10.01%	34.5	10.63
2	58.5%	11.67%	25.0	8.86	53.8%	14.80%	34.3	9.63
3	84.5%	6.10%	24.6	12.42	78.8%	10.23%	38.2	11.20

Reproducibility: Compared with the gold standard, the visualizations created by the six participants with limited training were 86.1% similar to the gold standard with a standard error of 6.45%. Dissimilarity was due to incomplete

visualizations as opposed to visualizations that differ fundamentally in structure. Out of all errors, 32% of the dissimilarity was due to neglecting to fill in available time points; 24% was neglecting to put in baseline experimental procedures; 12% was neglecting to fill in survival assessments time points; 12 % was neglecting to put in events that can change the sample size, such as drop outs; and 20% was due to neglecting to fill in sample sizes for populations. The reasons for error and the high similarity score shows that our representation is logical and easy to understand for the participants in our usability study.

Preferences: All eleven participants preferred the visualization to the status quo, or preferred both the report and the visualization. No one preferred the status quo to the visualization. Participants rated the usefulness on average as 8.0 (7.1-8.9, where 10 is completely essential), and the satisfaction of the visualization at the current state as 7.8 (6.7-8.9, where 10 is completely satisfied). The likelihood of participants using the visualization again is reported on average as 87.2% with a standard error of 10.7%. Additional comments mentioned participants appreciated the clean and well-formatted visualization, but wanted more context, such as information covered in the introduction and discussion. Participants suggested additional development and/or functionality, such as having a summary box to better navigate a visualization that contains too many details and including the ability to distinguish significant findings visually. In general, participants found the visualization to be a good overview that assisted in understanding the clinical trial study.

Discussion

Understanding the strength and quality of clinical trials is a critical step in providing better healthcare to medical practice. Our previously developed visualization places essential information in context with a process model. A user can navigate through nodes to identify the population involved, its sample size, the procedures performed, and data that result. Because the visualization can help summarize essential elements and connect relevant elements together, it can be valuable for biostatisticians and research clinicians, who routinely access information from within clinical trial reports, and are considered our potential users. With any system, it is critical to examine usability issues pertaining to people of varying backgrounds in academia successfully accessing information needed within clinical trial reports. The goal of this study was to determine whether our previously designed visualization could be used to increase time saved and accuracy as compare to the current method, while maintaining an easy to understand format.

The results of the usability study were consistent with our intuition. We found that having the visualization required on average 28.1% less time (25.8 min vs. 35.8 min; $p=0.01$) while maintaining similar accuracy. These findings did not appear to be affected by participants' varying levels of familiarity with the statistics, clinical domain (i.e., non-small cell lung cancer) and clinical trial procedures. This suggests that having essential information placed in context of the entire experiment helps users cognitively critique and apply contributions of clinical trials on a deeper level in a more timely fashion. This enables informatics tools to query information to be used for meta-analysis and probabilistic disease modeling and assist with the difficult task of assessing the quality and usefulness of each trial. The results of the reproducibility study showed that visualizations were on average 86.1% similar to the gold standard when produced by participants. This suggests that the proposed visualization is easy to understand and apply, and logically represents essential elements in a clinical trial study. This paper presents an attempt at identifying a standardized view that follows from intuition given that users have formal training in informatics.

While all participants favored the visualization over the current method, questionnaires revealed that much work is needed to improve the satisfaction and usability of the visualization. One solution to avoid bias of a less completely documented clinical trial study is to use the visualization to supplement an individual's understanding gained from reading the status quo published report. While the study design proposed in this paper assigns participants to either the status quo or the visualization condition, in actuality, the two conditions are not mutually exclusive. This suggests that the combination of having a visualization to reference while reading the status quo published report can further help to save time and increase accuracy. One of the next steps include extending the evaluation to potential users. While the results from this study were gathered on students serving as surrogates, preliminary interviews with potential users show promise. In an unstructured interview, one biostatistics professor anecdotally noted that she liked the hybrid process model-spreadsheet for contextualizing observations and statistics.

We recognize one limitation is the current reporting of clinical trial studies, which may contain missing information. Because the visualization is designed to present the same information as the status quo, missing information can negatively affect the user satisfaction rating. Another limitation is related to the design of the task questionnaires. Because no standard list of questions exist to test comprehension of clinical trials, we tailored questions from standard reporting guidelines to determine the types of information necessary for comprehension. The final list of questions was confirmed by domain experts to determine if answering questions display understanding. We also need to address

the limitation of self-reporting (e.g., timing), which may be inaccurate. Another limitation includes the recruitment of participants and the selection of baseline assessment measures in the pre-questionnaire analysis. Participants were students that served as proxy for potential users. Classes provided a quantitative way to rank students, however, the assessment may not be accurate for non-traditional students who gained experiences outside of classes or have a time-lapse of many years between graduate and undergraduate courses. One more limitation relates to the issue with questionnaires based on the Likert scale. While Likert scales can pinpoint problem areas, they are unable to give information on the nature of the identified problem.

Conclusion

The current research addresses the usefulness of the visualization and tests its efficacy in understanding clinical trials in a timely manner. Our results suggest that the application can decrease time without sacrificing accuracy, the visualization is reproducible across multiple users, and the system is generally accepted in a targeted user group. With the abundance of clinical trial research, further work is necessary to translate the published report into a concise and informative visualization containing the same information but with more functionality. The logical structure of the results of clinical trial reports will allow computers to better assess the quality and strength of trials, extract important statistics, and assist in comparison of trials.

Acknowledgements

This work is supported by the NIH grant number LM011333. We thank James Sayre, PhD., and Grace Hyun Kim, PhD., for their helpful discussions and constructive criticisms. We also thank our domain experts, our pilot study participants, and all our usability study participants.

References

1. Urick PN. Implementing community-based standards of care. J Manag Care Pharm. 2005 May;11(4 Suppl):S11-6.
2. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocolsto published articles. JAMA. 2004 May 26;291(20):2457-65.
3. Ilic D, Tepper K, Misso M. Teaching evidence-based medicine literature searching skills to medical students during the clinical years: a randomized controlled trial. J Med Libr Assoc. 2012 Jul;100(3):190-6.
4. Bogdan-Lovis E, Fleck L, Barry HC. It's NOT FAIR! Or is it? The promise and the tyranny of evidence-based performance assessment. Theor Med Bioeth. 2012 Aug;33(4):293-311.
5. Cronin P, Rawson JV, Heilbrun ME, Lee JM, Kelly AM, Sanelli PC, Bresnahan BW, Paladin AM. How to report a research study. Acad Radiol. 2014 Sep;21(9):1088-116.
6. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Mar 23;340:c869.
7. Sim I, Carini S, Tu S, Wynden R, Pollock BH, Mollah SA, Gabriel D, Hagler HK, Scheuermann RH, Lehmann HP, Wittkowski KM, Nahm M, Bakken S. The human studies database project: federating human studies design data using the ontology of clinical research. AMIA Summits Transl Sci Proc. 2010 Mar 1;2010:51-5.
8. Sim I, Olasov B, and Carini S. The Trial Bank System: Capturing Randomized Trials for Evidence-Based Medicine. AMIA Annu Symp Proc. 2003; 2003: 1076.
9. Weng C, Wu X, Luo Z, Boland MR, Theodoratos D, Johnson SB. EliXR: an approach to eligibility criteria extraction and representation. J Am Med Inform Assoc. 2011 Dec;18 Suppl 1:i116-24.
10. Tong M, Hsu W and Taira RK. A representation for standardizing numerical data from clinical trial reports. 2012 RSNA Scientific Assembly and Annual Meeting Bioinformatics Exhibit. LL-INE1228-TUA
11. Tong M, Taira RK. Improving the accuracy of treatment descriptions in clinical trials using a bottom-up approach. Proc AMIA Fall Symp, 2012. p. 1393–1402
12. Hsu W, Speier W, Taira RK. Automated extraction of reported statistical analyses: Towards a logical representation of clinical trial literature. Proc AMIA Fall Symp, 2012. p. 350-359
13. Johnson DH, Fehrenbacher L, Novotny WF, Herbst RS, Nemunaitis JJ, Jablons DM, Langer CJ, DeVore RF 3rd, Gaudreault J, Damico LA, Holmgren E, Kabbinavar F. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. J Clin Oncol. 2004 Jun 1;22(11):2184-91.

14. Sequist LV, Martins RG, Spigel D, Grunberg SM, Spira A, Jänne PA, Joshi VA, McCollum D, Evans TL, Muzikansky A, Kuhlmann GL, Han M, Goldberg JS, Settleman J, Iafrate AJ, Engelman JA, Haber DA, Johnson BE, Lynch TJ. First-line gefitinib in patients with advanced non-small-cell lung cancer harboring somatic EGFR mutations. J Clin Oncol. 2008 May 20;26(15):2442-9.
15. Pirker R, Pereira JR, von Pawel J, Krzakowski M, Ramlau R, Park K, de Marinis F, Eberhardt WE, Paz-Ares L, Störkel S, Schumacher KM, von Heydebreck A, Celik I, O'Byrne KJ. EGFR expression as a predictor of survival for first-line chemotherapy plus cetuximab in patients with advanced non-small-cell lung cancer: analysis of data from the phase 3 FLEX study. Lancet Oncol. 2012 Jan;13(1):33-42.
16. Brugger W, Triller N, Blasinska-Morawiec M, Curescu S, Sakalauskas R, Manikhas GM, Mazieres J, Whittom R, Ward C, Mayne K, Trunzer K, Cappuzzo F. Prospective molecular marker analyses of EGFR and KRAS from a randomized, placebo-controlled study of erlotinib maintenance therapy in advanced non-small-cell lung cancer. J Clin Oncol. 2011 Nov 1;29(31):4113-20.