A Data Capture Framework for Large-scale Interventional Studies with Survey Workflow Management

Shiqiang Tao, PhD[1], Ningzhou Zeng, BS[2], Xi Wu, BS[2], Xiaojin Li, MS[3], Wei Zhu, BS[3], Licong Cui, PhD[1,2], GQ Zhang, PhD[1]
[1]Institute of Biomedical Informatics, University of Kentucky, Lexington, KY
[2]Department for Computer Science, University of Kentucky, Lexington, KY
[3]Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH

Abstract

Capturing high-quality survey data is an arduous process for large-scale and extensive interventional studies. This paper presents the architecture, interface design, and an innovative form generation engine of a system called RE-Form: Refactorized Electronic Web Forms. REForm provides researchers the capability to design and manage surveys and the flexibility to organize them in a customizable workflow. REForm has been designed, implemented, pilot-tested and deployed for an NCI-funded interventional study IMPACT. It enables IMPACT to design and conduct a baseline survey of 95 questions with 662 options, a post-intervention survey of 82 questions with 574 options, six Ecological Momentary Assessment (EMA) surveys including four questions and 26 options each, and a follow-up survey containing 15 questions and 125 options. Feedback designed in the post-intervention survey showed that more than 94 percent of IMPACT participants highly endorsed the design and usability of the surveys created using REForm.

Introduction

High-quality data capture plays a critical role in interventional studies. Researchers often employ online surveys to obtain information about participants before and after interventions to perform comparative analysis. Studies[1–4] show that online survey as a research tool emerged from the 1980s[5] has been widely used in scientific studies nowadays. Products such as Survey Monkey[6], Google Form[7], and Research Electronic Data Capture (REDCap)[8] can assist researchers to conduct online surveys.

However, researchers typically use the online survey as a toolkit separately from study interventions[9–11]. This separation results in unnecessary data management work such as redundant storage of study participants information, manual randomization of study subjects, and mapping between survey takers and intervention takers. The situation gets worse for personalized studies in which interventions are assigned based on user answers to the online survey. Another limitation with current survey systems is the lack of workflow control. Surveys are treated as independent from each other, which produce burdens for both researchers (survey givers) and study participants (survey takers), especially for large-scale interventional studies. Large-scale intervention studies can have more than ten surveys with hundreds of questions, involve thousands of participants, and last for years. Researchers need to have an extra mechanism to track the status of each study participant and send them the correct survey. Study participants can easily get lost in the procedure and drop out of the study. In addition, current survey tools as mentioned above do not provide adequate support for large surveys with about one hundred questions. Such large surveys require convenient navigation, real-time data storage, and intuitive progress indicator. Insufficient support for these features compromises user experience and discourages them from continuing the future steps in the study.

To address these challenges, we introduce **R**efactorized **E**lectronic Web **Form**s (REForm), a web application that designs for data capture of large-scale web-based interventional studies with online surveys. REForm provides an interactive interface for researchers to build their surveys and allows them to customize a workflow to organize and deliver both interventions and surveys; Also, researchers can feed in rules to automate tasks such as study subjects randomization and personalized intervention assignment; REFrom consists of a unique set of features for supporting large-scale surveys: it features a progress bar to provide page-wise progress and indicate the status of partial completion or full completion; it eases user burden in dealing with large surveys by combining a mechanism of fast page navigation with specially designed progress bar; it provides real-time data entry saving and storage as a built-in feature, preventing potential data loss that could occur in long surveys.

We implemented REForm and tested it for the NCI-funded **IM**proving Communications with **P**atients **A**bout **C**linical **T**rials (IMPACT), an online educational program aimed at improving oncology nurse engagement in patient decision making about clinical trials. IMPACT utilizes REForm to build and deliver surveys to participants, perform random-

ization automatically, and assign personalized tailored interventions. During the launching of IMPACT since July 30, 2016, 390 oncology nurses have enrolled and participated in the IMPACT program, producing 510 completed surveys and providing positive feedback on REForm confirming our design objectives.

1 Background

IMPACT is an NCI-funded online educational program about clinical trials designed by Case Western Reserve University and developed by University of Kentucky in consultation with the Oncology Nursing Society (ONS). Clinical trials are necessary to develop new cancer prevention and treatment strategies. Despite the encouragement from government and organizations, participation in clinical research studies remains poor. Less than 10% of cancer patients take part in a clinical trial[12–14]. Oncology nurses share a therapeutic relationship with their patients and are estimated to have twice as much contact with patients as their physician[15, 16]. Given the close involvement of oncology nurses in supporting patient decision making about care and treatment[17], and potential barriers to discussing clinical trials with patients[18], IMPACT is proposed to be developed, implemented, and evaluated as a theory-driven, interactive, web-based, tailored educational program to improve oncology nurse engagement in patient decision making about clinical trials.
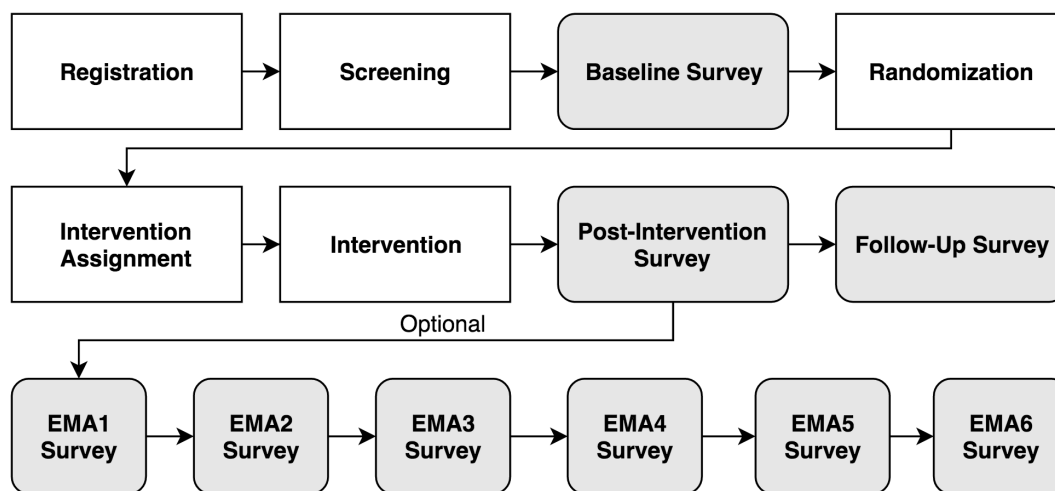

Figure 1. Survey Workflow for IMPACT.

IMPACT is the primary use case of REForm, as it is a large-scale interventional study with a goal to recruit thousands of participants and it requires participants to go through the procedure (illustrated in Figure 1) including nine surveys. Participants first need to register in the REForm system and answer the screening questions. Then they take the baseline survey after passing the screening. Based on screening questions and some questions in the baseline survey, participants are randomized into two groups: experimental group and control group. This step is called randomization in interventional studies. IMPACT employs stratified randomization, a randomization scheme where strata are constructed based on user characteristics and randomization is performed separately within each stratum. Stratified randomization can ensure the balance of group assignment on the various combinations of the characteristics. IMPACT prepares two categories of educational interventions: one is a library of videos and the other is plain reading materials. The experimental group receives videos as the intervention while the control group receives reading materials. Additionally, participants of experimental group received a personalized, specific number of videos according to their answers to the baseline survey. All participants take a post-intervention survey. After the post-intervention survey, participants can choose to take the six optional Ecological Momentary Assessment (EMA) surveys. In the end, IMPACT delivers a follow-up survey to participants 3 months after the completion of the post-intervention survey. The entire procedure constructs a complex workflow. Hence it is important for REForm to support the design and control of workflow.

2 Methods

The REForm system is built with Ruby on Rails [19] web development framework in a rapid and iterative development style. We developed REForm in close consultation with users employing the Web-Interface-Driven Development (WIDD) methodology[20]. WIDD enables us to elicit requirements and ensure usability more effectively. In the follow-

ing subsections, we describe the design and implementation of REForm.

2.1 Survey Authoring

In REForm, a **survey** consists of multiple **page**s and each page can have multiple **question**s. "Question" is the key data model for REForm, as it is the basic element of a survey. The "Question" model consists four parts of information: content, options, question type, and branching logic. Content describes what a question asks for; options are the pre-defined answers to the question for users to select; question type determines how the question is presented; and branching logic states if the question is dependent on users' answers to other questions.

The left part of Figure 2 shows the interface for creating a survey. A survey has a title and a description. The description includes directions about how to complete the survey. A survey has a default first page (or page 1) as shown in the right part of Figure 2. Survey authors can add new pages using the button "Add New Page" at the bottom.

Figure 2. Left: The form to create a new survey; Right: Default page 1 of a survey and the button to add a new page.

Figure 2 (right part) also shows the button to add new questions to the page. REForm supports seven question types in : 1) "Single Choice", 2) "Multiple Choice", 3) "Text Single", 4) "Text Multiple", 5) "Group Single", 6) "Group Multiple", and 7) "Matrix". These types directly determine the question's visual presentation and layout. Figure 3 presents the examples of each type of questions. The questions numbers in Figure 3 are corresponding to the numbers mentioned above. "Single Choice" and "Multiple Choice" questions allows participants to select single and multiple answers. "Text Single" and "Text Multiple" questions takes free text as input. REForm also supports "Group Single" and "Group Multiple" questions in which options are categorized into different independent groups. Question 7 demonstrates the "Matrix" type of questions which have a set of sub-questions sharing the same options.

2.2 Intervention and Survey Workflow Design

REForm provides an interactive interface for survey authors to customize a workflow as shown in the left part of Figure 4. Each step of a workflow is a built survey or prepared intervention. Before each step, survey author can create a paragraph of so called "Pre-Message", which provides welcome language and instructions for study participants takers to complete the step. Similarly, there is an optional post message at the end of each step that can provide appreciation language and summary information about user's current status and brief introduction of the next steps. The workflow also supports reminders for each step. Survey authors can add one or more reminders to a step. The form displayed on the right part of Figure 4 shows an example of how to create a reminder for a survey (e.g., baseline survey). The data in the form is interpreted as: REForm will send a message (content details are in the screenshot) to a user's contact if the user does not complete baseline survey 7 days after it is started.

2.3 Randomization

Our randomization algorithm (see Figure 5) can automatically categorize survey participants into different groups based on configuration rules using stratified randomization. Researchers need to configure two parameters: randomization questions and group assignment ratio. Selection of randomization questions tells the algorithm the combinations of user characteristics and the group assignment ratio is used to precompute the group assignment table. In a group assignment table, each row is an assignment rule that specified the group name: interventional or control and the study participant id that is assigned to the group. We state that a row is not occupied if the participant id in that

Figure 3. Combined display of different types of questions supported by REForm.

Figure 4. Left: Workflow management interface of REForm; Right: The form to create reminders.

row is empty. Algorithm (in Figure 5) pre-computes the group assignment tables. Each table has the same name as the stratum. In this algorithm, we demonstrate a scenario using Gender and Race as the characteristics to construct strata and the ratio of participants in experimental group and control group is equal. Gender has two values: Male (M) or Female (F) and Race has two values: White (W) and Non-White (N) so that we have for strata: MW (Male and White), MN (Male and Non-White), FW (Female and White), and FN (Female and Non-White). We enter these four strata and the expected max number of participants in each stratum as the input parameters and obtain four group assignment tables as output.

After group assignment tables are created, group assignment is straight forward: we can easily find the stratum for each new participant, then look up in the related group assignment table to get the next un-occupied row, and assign this participant to the group listed in the row.

```
 1: Input: Strata S:[MN, FW, MN, FN], Maximal Number of Participants in Each Stratum Max
 2: Output: Group Assignment Tables

 3: for each s ∈ S do
 4:     Create group assignment table s'
 5:     while t < Max do
 6:         a = [1, 2].Random
 7:         insert a rows into s' with group name as intervention and participant id as empty
 8:         insert a rows into s' with group name as control and participant id as empty
 9:         t+ = a * 2
10:     end while
11: end for
```

Figure 5. Precompute Group Assignment Tables for Stratified Randomization.

2.4 Personalized Interventions Assignment

Personalized interventions or tailored interventions are a subset of the entire intervention library selected based on the user information captured by baseline survey. Our intervention assignment algorithm (see Figure 6) generates a tailored set of interventions for each participant in the experimental group.

Before introducing the intervention assignment algorithm, suppose we have an intervention library **L**, which can be a number of images or videos. Each item of the intervention library is noted as **i**. We also have a set of mapping rules noted as **R** between question answers and intervention items.

In the video assignment algorithm (see Figure 6), it takes a participant's survey answers and mapping rules as input and output a subset of intervention library assigning to the participant.

```
 1: Input: Baseline Survey Answers A, Intervention Library L, and Mapping Rules R between A and L
 2: Output: a set of selected interventions: S

 3: Create a empty set S
 4: for each a ∈ A do
 5:     Find related mapping rules r from R
 6:     for each r' ∈ r do
 7:         i = r'.getIntervention()
 8:         add i into S
 9:     end for
10: end for
```

Figure 6. Intervention Assignment Algorithm.

2.5 Large-scale Survey Support

For large-scale or long surveys, the following three features are important: real-time data storage, page-wise progress indicator, and fast page navigation. Large surveys take more time to complete and they increase the chance of unexpected events such as power outage and web browser not responding. Therefore, real-time data storage can save survey once any of these events happen. In a large-scale survey, participants are more likely to encounter some questions they are not sure about and would like to skip and return to it later. This requires the survey system to be able to mark that page and allow participants to get back to that page instantly if they want, e.g., with one mouse click. In the following section we describe how REForm addresses these challenges.

Real-time Data Storage. REForm employs AJAX (asynchronous JavaScript and XML) technique to save selections without reloading the entire page. Once a mouse click or keyboard stroke is detected when a user makes an selection, request is sent to server; then server receives the request, processes it, and save the selection to database. Server then sent back two responses: one is to update the progress bar if needed, and the other is to show some questions with dependencies if that selection makes those questions' dependencies satisfied.

Page-wise Progress Indicator. Figure 7 shows the progress bar specially designed for REForm. Each circle represents

Table 1. Setting of tests for comparative evaluation between REForm and REDCap.

Name	Question Type	Number of Questions	Number of Options
Test 1	Single Choice	5	4
Test 2	Multiple Choice	5	4
Test 3	Text Field	5	1
Test 4	True/False	5	2
Test 5	Matrix	8	8
Test 6	Mixed types including all above	22	1 to 11 (in each question)

one survey page. Each circle has three possible colors: white, yellow, and green. White means that none of the questions is answered on that page; yellow means questions on the page is partially answered; and green means all questions are answered on the related page. With the implemented real-time data storage feature, this progress bar is also updated at real time.

Figure 7. Progress bar for REForm: it combines page-wise progress indication and fast page navigation functions.

Fast Page Navigation. In addition to functioning as survey progress indicator, the progress bar is also used for fast navigation. Each circle is a hyper link to the related page. A user can click the circle to go to that page directly instead of using previous or next button to navigate page by page.

2.6 Evaluation Method

We evaluated two key aspects of the REForm system: **(1)** the ease of use to build surveys from the perspective of survey authors; and **(2)** the intuitive and effective support for large-scale surveys from the perspective of study participants.

For **(1)**, we invited three data scientists with years of experience of using and building web application as evaluators to do a comparative evaluation between REForm and REDCap. All evaluators we chose had no experience with REForm survey building interface and little experience with REDCap. The evaluation consisted of six tests as described in Table 1. The measure of performance is the time required to complete the tests. Less time cost indicates better performance. In addition to the time cost, we also recorded evaluator's feelings about these two systems' features during the process of doing the tests. Each of tests 1 to 5 was to be implemented with single question type with the purpose to compare REForm and REDCap's the capability to build specific types of questions. The five selected questions are common within REForm and REDCap. These short tests are also intended for the evaluators to get familiar with both systems. Test 6 required evaluators to build the demographic section of baseline survey of IMPACT mentioned above with the input provided as plain text files.

For **(2)**, we designed two questions in the IMPACT post-intervention survey:

(a) How satisfied are you with the information presented to you?
   1. Not at all satisfied; 2. A little satisfied; 3. Moderately satisfied; 4. Very satisfied; 5. Extremely satisfied

(b) The web site is easy to navigate?
   1. Not at all; 2. Very little; 3. Some what; 4. Quite a bit; 5. A great deal

The post-intervention survey were sent to 3000 participants via email and 132 of them responded the two questions. Survey participants' answers to these two questions reflected REForm's performance for supporting large-scale surveys. More details are discussed in section 4.

3 Results

For IMPACT study, we used REForm to deploy a baseline survey consisting of 95 questions and 662 options, a post intervention survey of 82 questions and 574 options, six EMA surveys of four questions with 26 options, and one follow up survey consisting of 15 questions and 125 options. REForm construct a workflow involving these steps as:

1) baseline survey, 2) intervention, 3) post-intervention survey, 4) EMA 1 to 6 survey, and 5) follow-up survey. Each step has a "pre-message" and "post-message". We also built two consecutive reminders for each step which are sent at specific time intervals according to the study design. Participants receive surveys and intervention automatically and in the order specified by the workflow.

Participation invitations were sent out to 3000 target users via email and 390 of them registered in REForm within 9 weeks since its launching in July 2016, while 358 of them consented to participate and 264 passed the screening and became eligible. 238 of the eligible participants completed baseline surveys and 150 of them finished their interventions. After that, 132 study participants completed their post-intervention surveys. As of September 22, 2016, EMA1, EMA2, EMA3, and EMA4 have 75, 62, 6, and 3 completions respectively.

3.1 Evaluation

The evaluation was designed to measure REForm's performance from both the survey author's and the survey participant's perspectives.

First, we evaluated REForm's ease of use for survey authors to create online surveys. We prepared 6 survey building tests with details showing in Table 1 and invited three evaluators to build these tests using REForm and REDCap respectively. Figure 8 shows the time spent by each evaluator to build questions for the six tests. REForm and REDCap shared similar performance for tests from 1 to 4. REDCap had significant advantage in supporting matrix question type in test 5. However, it is interesting to find that in test 6, two out of the three evaluators reported better performance using REForm than REDCap. Evaluators' explanation was that REDCap had better support for building a single type of questions, as it inherits the last question's information when adding a new one. But this advantage did not exist in test 6 where questions were in multiple types and the information inheritance produced extra work needed when question type changes. In addition, evaluators also mentioned REForm supports the following features that REDCap does not have: 1) REForm provides more flexibility, e.g. it allows an "other" option with a text box as the last option for user provide more information. REDCap needs to build another question to achieve the same functionality; 2) REDCap does not have group question types so that users have to build two separate questions to have the same functionality of one group question in REForm; and 3) REForm supports inline style formatting of question content while REDCap does not. Overall, evaluators found that REForm had more intuitive interface and were more lightweight and easier to learn. REDCap, as a widely used survey tool, has its advantages reflected in function completeness such as question preview and the better support for the matrix type of questions.

Second, we evaluated REForm's support for large-scale surveys, which can better assist survey participants to complete their surveys. We received 132 responses for the two questions displayed in Figure 9 from the completed post-intervention surveys. For question (a), 129 of 132 participants (97.7%) rated moderately satisfied or better (answers 3, 4, and 5) with the information presented to them and 97 out of 132 (74.5%) rated very satisfied or better (answer 4 and 5). Answers to question (b) show that 124 out of 132 (93.9%) rated the REForm survey to be very easy to navigate (answers 4 and 5). No participant thought it was hard to navigate (answer 1 and 2).

4 Discussions

REForm provides an interactive survey authoring interface that supports multiple question types, inter-question, and inner-question branching logic. REForm allows researchers to integrate surveys and intervention in a comprehensive workflow. REForm also innovates a unique set of features providing assistance for answering large-scale surveys. With these contributions, REForm addresses the main data capture challenges of large-scale interventional studies.

Our system enhances the current survey systems in four ways: (1) integration of intervention with surveys keep study data in one single data repository, eliminating unnecessary data integration tasks; (2) comprehensive workflow management automates the survey and intervention delivery including configurable "Pre-Message", "Post-Message" and reminders, largely reducing the administration burden for survey; (3) specially designed widgets help study participants navigate and answer large surveys, which requires less user effort and encourages them to complete the entire study procedure; (4) and with user-composed rules, REForm can automate research procedures like the randomization and the personalized intervention assignment. REForm is a live and evolving system. With the increasing number of question types and better practices, it will be able to better handle current limitations like building questions of "Matrix" type. Currently, we have only limited number of evaluators performing the evaluation to build the survey. We plan to invite more data scientists to perform systematic comparative evaluations.
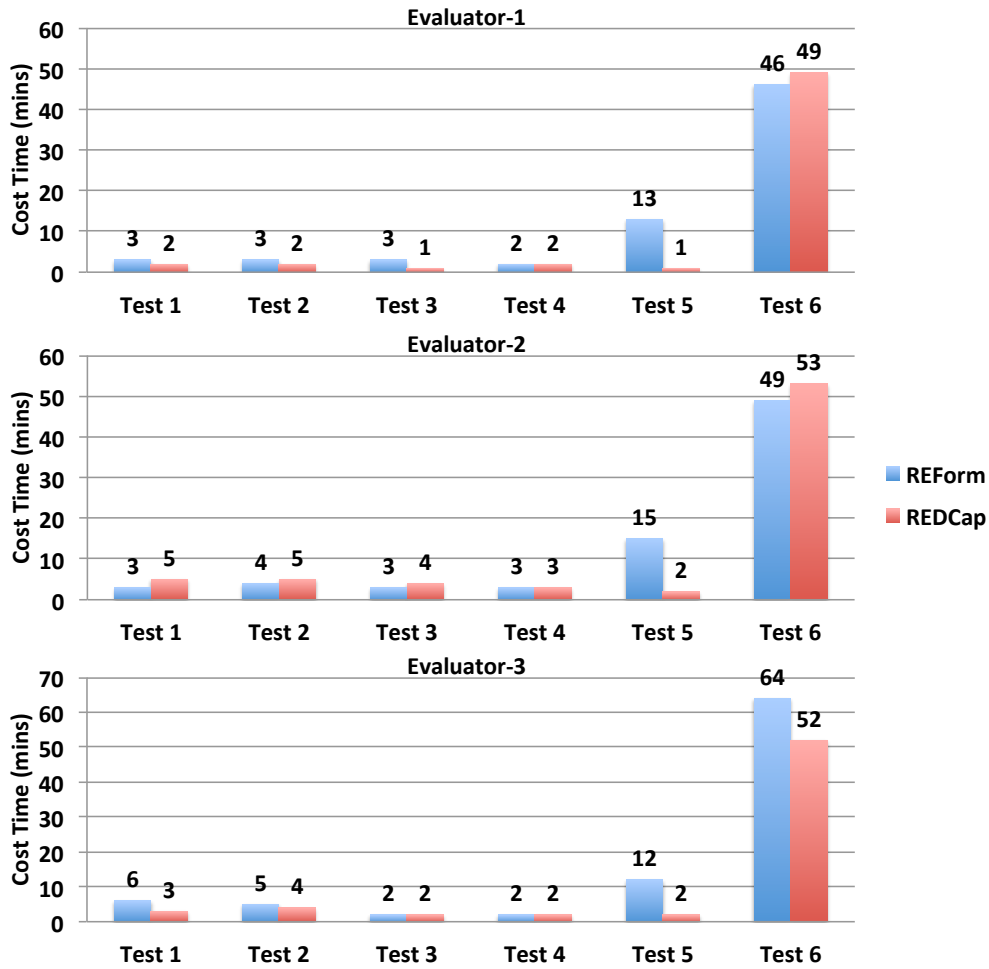
Figure 8. Comparative evaluation between REForm and REDCap by time cost to build survey questions in six tests.
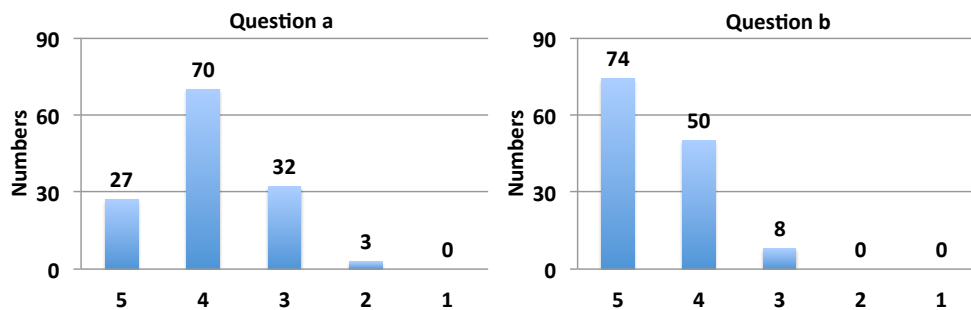

Figure 9. Answers to evaluation questions (a) and (b) from IMPACT participants. X-axis shows the selected options for question (a) and (b) and Y-axis shows the number of times those options are selected.

5   Conclusion

This paper introduced REForm, a web application created to solve the data capture challenges of large-scale interventional studies. REForm provides survey authoring interface and supports integration of online surveys with interventions in a comprehensive workflow. REForm also provides sufficient support for large-scale surveys by implementing features such as real-time input data saving, page-wise progress indication, and fast page navigation. REForm has been deployed for IMAPCT study since July 2016 and initial user feedback showed that it met the design objectives.

References

[1] Evans, Joel R., and Anil Mathur. "The value of online surveys." Internet research 15.2 (2005): 195-219.

[2] Cobanoglu, Cihan, Bill Warde, and Patrick J. Moreo. "A comparison of mail, fax and web-based survey methods." International journal of market research 43.4 (2001): 441-452.

[3] Greenlaw, Corey, and Sharon Brown-Welty. "A comparison of web-based and paper-based survey methods testing assumptions of survey mode and response cost." Evaluation Review 33.5 (2009): 464-480.

[4] Sills, Stephen J., and Chunyan Song. "Innovations in survey research an application of web-based surveys." Social science computer review 20.1 (2002): 22-30.

[5] Schonlau, M., Fricker, R.D. Jr and Elliott, M.N. (2001), "Conducting Research Surveys via E-Mail and the Web", Rand Corporation, Santa Monica, CA.

[6] https://www.surveymonkey.com/, Accessed on September 22, 2016.

[7] http://www.google.com/forms, Accessed on September 22, 2016.

[8] Harris, Paul A., et al. "Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support." Journal of biomedical informatics 42.2 (2009): 377-381.

[9] Fleming, Geoffrey M., et al. "A multicenter international survey of renal supportive therapy during ECMO: the Kidney Intervention During Extracorporeal Membrane Oxygenation (KIDMO) group." ASAIO Journal 58.4 (2012): 407-414.

[10] Simonsen, Sara. "Advancing public health through the use of REDCap (Research Electronic Data Capture) software to promote healthy behaviors in women from 5 culturally diverse Utah communities." 143rd APHA Annual Meeting and Exposition (October 31-November 4, 2015). APHA, 2015.

[11] Fleisher, Linda, et al. "Application of best practice approaches for designing decision support tools: the preparatory education about clinical trials (PRE-ACT) study." Patient education and counseling 96.1 (2014): 63-71.

[12] Lara, Primo N., et al. "Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment." Journal of Clinical Oncology 19.6 (2001): 1728-1733.

[13] Lara, Primo N., et al. "Evaluation of factors affecting awareness of and willingness to participate in cancer clinical trials." Journal of Clinical Oncology 23.36 (2005): 9282-9289.

[14] Virani, Shamsuddin, et al. "Barriers to recruitment of rural patients in cancer clinical trials." Journal of Oncology Practice 7.3 (2011): 172-177.

[15] Murthy, Vivek H., Harlan M. Krumholz, and Cary P. Gross. "Participation in cancer clinical trials: race-, sex-, and age-based disparities." Jama 291.22 (2004): 2720-2726.

[16] Townsley, Carol A., Rita Selby, and Lillian L. Siu. "Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials." Journal of Clinical Oncology 23.13 (2005): 3112-3124.

[17] Miller, Suzanne M., et al. "The relationships among knowledge, self-efficacy, preparedness, decisional conflict, and decisions to participate in a cancer clinical trial." Psycho-Oncology 22.3 (2013): 481-489.

[18] Umutyan, Ari, et al. "Overcoming barriers to cancer clinical trial accrual." Cancer 112.1 (2008): 212-219.

[19] http://rubyonrails.org/, Accessed September 22, 2016

[20] Tao, Walter, et al. "Web-Interface-Driven Development for Neuro3D, A Clinical Data Capture and Decision Support System for Deep Brain Stimulation" HIS(2016): in press.

[21] http://project-redcap.org/, Accessed September 22, 2016