An Adaptable, Transportable Web-Based Data Acquisition Platform for Clinical and Survey-Based Research

Jane Turpin; Richard Rose, PhD; Bryan Larsen, PhD

Given the need for multicenter clinical trials to support evidence-based treatments, the authors hypothesized that the process of gathering large amounts of data from disparate clinical sites could be facilitated through direct input of clinical and survey data through a Web interface. A series of data collection instruments was created and published as Web pages to support a clinical study performed at Des Moines University Osteopathic Medical Center.

The most challenging tool to implement was the visual analog scale, which required special programming. Specific Web pages allowed research study participants to input their own data, while other Web pages were restricted to use by the investigator for inputting clinical and laboratory observations. Data from these sources were automatically combined in a single spreadsheet by the Web administrator in a manner that maintained the confidentiality of participants. Subsequently, the system was tested from a remote site (Chicago), and data were captured at Des Moines University Osteopathic Medical Center. This test involved a variety of data including the visual analog scale.

Although this system provided a facile method for collecting and analyzing a large amount of data in almost real time, with a demonstrable savings of time and money, the authors believe the more important use for this data collection system is in large multicenter clinical trials. Hence, the authors commend its use for support of large outcomes studies for osteopathic researchers engaged in interinstitutional efficacy studies.

Research within the osteopathic medical profession has become a focal point for colleges of osteopathic medicine, osteopathic postgraduate training initiatives (OPTIs), and professional organizations and has even become a candidate for funding through the National Center for Complementary and Alternative Medicine at the National Institutes of Health. The demand for research data is increasing and those engaged in developing such data increasingly need more efficient ways of collecting and analyzing results of their studies. Particularly in multicenter outcomes studies, the timely collection of objective data will be the hallmark of research that gives credence to new methods and treatments. The ability to gather, collate, and analyze large quantities of clinical and laboratory data will be vital to the timely completion and evaluation of multicenter research studies.

Investigators who have participated in pharmaceutical company-sponsored projects are familiar with the voluminous case record forms that the physician-investigator or his or her staff must complete. These forms can require dozens of individual entries on 20 or more pages depending on the study. At intervals during the research study, a clinical research associate visits the research site and pores over these forms looking for missing entries and errors and requires the investigator or staff to find the correct information and sign off on changes. The task of entering this paper data into a database, however, is handled by a team of data managers at the pharmaceutical company, and the data can be made available to the investigator in summary form in a timely fashion.

Investigator-initiated projects within a college of osteopathic medicine or an OPTI may require as much record-keeping as the large industry-sponsored research programs. Many clinician-investigators pressed by clinical duties lack the time to create and maintain clinical research records and the funding for personnel to handle data collection.

As part of a research study done at Des Moines University Osteopathic Medical Center, we developed a data collection tool that we demonstrated to be applicable to future studies at our institution and at other institutions. The system was implemented and tested by a third party to provide a clearer sense of how this tool might be useful to single-center or multicenter studies.

Methods
A manufacturer of nutraceutical compounds engaged Des Moines University Osteopathic Medical Center to evaluate a hypothesis that a specific plant protein is able to alter hunger,
The participants (history), and the investigator entered information about the participants’ height, weight, blood glucose, and CCK into a separate record (physical).

To maintain confidentiality, each participant was given a unique number printed on a card attached to a lanyard worn during the time they participated in the study. Each time data were entered, the respondent had to use the number on the card to complete the data entry screen. When all of the questions were answered on a particular data collection form, the respondent activated a “submit” button, a “thank you” screen replaced the data collection form, and the respondent had no further access to that form. A blank data collection form subsequently appeared and the new screen was ready for the next time point. The respondent could activate a “submit” button only if all required fields had been completed. If the volunteer failed to answer a required question, a message of “please answer [unfilled field]” flashed and the respondent could then finish and submit the form. Each time the “submit” button was activated, a copy of the data was transmitted to the principal investigator by e-mail so that, in addition to the data being recorded on our server, an archive copy of the data was in the possession of the principal investigator. We created six different forms that were presented as Web pages (Table).

This study was conducted in a room normally used for student testing, so the computers were in carrels, which precluded research participants viewing other individuals’ responses. Because new blank forms were generated after each response, the volunteers did not have to use the same workstation each time they entered data.

We kept paper copies of all the questionnaires in the event we encountered a participant who was so computerphobic that he or she could not complete the form on-line. The paper forms also served as backup in case of power failure or loss of Web service.

**Results**

This initial study resulted in the electronic submission of 4960 data items and 2016 copies of data by e-mail to the principal investigator’s address. These e-mails were copied to a compact disc for permanent, unalterable, archival storage of raw data.

Initially, there was concern that differing levels of computer literacy might influence the quality of self-reported information. This was found not to be the case. The Web design was simple and straightforward, with most questions answered by the participants (history), and the investigator entered information about the participants’ height, weight, blood glucose, and CCK into a separate record (physical).

The most daunting part of the study was to determine the relative satiety of individuals during the study. We enrolled 40 participants who would make three visits (once to test placebo, twice to test two different doses of the active compound). Satiety was gauged by responses to questions about hunger, fullness, desire to eat, and thirst. According to the literature, these observations are typically made on a visual analog scale without scale markings. The respondent marks on a 10-cm scale the degree of hunger, fullness, desire to eat, or thirst, and the investigator measures the distance along the scale with calipers and records it on paper or in a database. The respondents in our study rated their feelings of satiety at 15-minute intervals for up to 4 hours after a test dose and test meal. We calculated that, for the measures of satiety alone, we would need to evaluate a minimum of 7200 visual analog scales in addition to other data we intended to collect. If just one minute were needed to measure the visual analog scale and enter the result into a database, 120 person-hours would be required to input this portion of the data alone. At a modest $15/hour for data entry service, the cost of this part of the study would be $1800 and would not preclude occasional data entry errors.

We elected to use a direct data entry method with a computer interface to capture both participant-generated and investigator-generated data. Questions about food and exercise habits, general health, and medications were answered by the participants (history), and the investigator entered information about the participants’ height, weight, blood glucose, and CCK into a separate record (physical).

To maintain confidentiality, each participant was given a unique number printed on a card attached to a lanyard worn during the time they participated in the study. Each time data were entered, the respondent had to use the number on the card to complete the data entry screen. When all of the questions were answered on a particular data collection form, the respondent activated a “submit” button, a “thank you” screen replaced the data collection form, and the respondent had no further access to that form. A blank data collection form subsequently appeared and the new screen was ready for the next time point. The respondent could activate a “submit” button only if all required fields had been completed. If the volunteer failed to answer a required question, a message of “please answer [unfilled field]” flashed and the respondent could then finish and submit the form. Each time the “submit” button was activated, a copy of the data was transmitted to the principal investigator by e-mail so that, in addition to the data being recorded on our server, an archive copy of the data was in the possession of the principal investigator. We created six different forms that were presented as Web pages (Table).

This study was conducted in a room normally used for student testing, so the computers were in carrels, which precluded research participants viewing other individuals’ responses. Because new blank forms were generated after each response, the volunteers did not have to use the same workstation each time they entered data.

We kept paper copies of all the questionnaires in the event we encountered a participant who was so computerphobic that he or she could not complete the form on-line. The paper forms also served as backup in case of power failure or loss of Web service.

**Results**

This initial study resulted in the electronic submission of 4960 data items and 2016 copies of data by e-mail to the principal investigator’s address. These e-mails were copied to a compact disc for permanent, unalterable, archival storage of raw data.

Initially, there was concern that differing levels of computer literacy might influence the quality of self-reported information. This was found not to be the case. The Web design was simple and straightforward, with most questions answered by

---

**Table**

<table>
<thead>
<tr>
<th>Web Page</th>
<th>User</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prescreening form</td>
<td>Study participant</td>
<td>To record brief medical history before the study</td>
</tr>
<tr>
<td>Qualification form</td>
<td>Investigator</td>
<td>To record patient inclusion and exclusion criteria and physical findings (height, weight) before the study</td>
</tr>
<tr>
<td>First form of the day</td>
<td>Study participant</td>
<td>To ask if any medications were being used at the time of the study and if individual was fasting</td>
</tr>
<tr>
<td>Every-15-minute form</td>
<td>Study participant</td>
<td>To determine hunger, fullness, thirst, and desire to eat as well as to report on adverse events</td>
</tr>
<tr>
<td>Glucose report form</td>
<td>Investigator</td>
<td>To report blood glucose observations</td>
</tr>
<tr>
<td>Cholecystokinin report form</td>
<td>Investigator</td>
<td>To report laboratory finding on cholecystokinin</td>
</tr>
</tbody>
</table>
either clicking radio buttons or moving a slider on the visual analog scale (Figure). The prescreening form served as a kind of training exercise in the Web data entry process. On each study day, we had a “first form of the day” (completed at 8 AM) on which we asked respondents to fill out satiety ratings as practice for the forms they would complete repeatedly after their noon meal. This provided reinforcement for use of the system. We also placed the appropriate Web page on the computers before the study volunteers arrived, eliminating the need for respondents to navigate the Web to find the proper questionnaire.

Our initial success in implementing this Web-based research data collection system prompted us to consider whether it would be useful for multicenter studies. We created a remote test that would evaluate all aspects of the data input system. We established a remote center in Chicago where data was input to the central research server in Des Moines. One author (B.L.) created two sets of information: one to simulate medical chart notes with nine specific items of data for each patient and one to simulate a 10-item patient history. Another author (R.R.) interpreted these narratives and input the data from the remote location as a physician or patient might. Thus, 190 data items were returned from R.R. to B.L. via the Web data capture system. The Web data capture forms were designed to require the remote investigator (R.R.) to input data in the form of yes/no buttons, pull-down menus, Likert scale, visual analog scale, and free response area to test all data input modalities. When data input by the remote investigator (R.R.) were submitted, a copy of the data was transferred to the principal investigator (B.L.) in Des Moines by e-mail and the same data were also placed in a spreadsheet maintained by the Web administrator.

The data received were compared to the narratives originally created to see if the system worked. The remote investigator submitted 10 physician reports and 10 patient reports, comprising 190 individual items of information. During this test, the remote investigator reported a problem with the data input screens, which we subsequently corrected. The remote investigator also asked for clarification on information regarding two imaginary patients, suggesting that whenever
this system is used to collect data from remote sites, investigators should do one or more trial runs to ensure that all potential users understand the interface and that all functions are working properly. Of the 190 separate data items input by the remote investigator, only one was in error. A white blood cell count was reported with a final zero missing and, had we included an error-checking algorithm to flag abnormal laboratory results, the remote investigator would have been prompted to review that information.

In view of this test, we believe that in a short time, a Web data entry system for almost any clinical study could be written, tested, and implemented. We discovered, however, that someone other than the designer of the Web data collection instrument should do the testing because the intent of the designer is not always apparent to the users of the product.

Discussion
While this Web-based data acquisition system was designed for one particular study, it was immediately apparent that its use could go well beyond our initial intention. The inclusion of visual analog scales is useful not only for our study, but for other purposes, such as quality of life studies and pain scales. One could envision the use of radio buttons, pull-down menus, free text space, and visual analog scales or Likert scales for educational research, patient-reported clinical research, and even investigator-reported information. Our experience has shown that by recording glucose and CCK values onto the secure Web site, the data can easily be wedded to the study subject records through relational databases.

The creation of the visual analog scale was complicated and required special programming. We were fortunate to have the expertise necessary within our institution. We are now able to easily write Web-based questionnaires that are attractive and easy for participants and investigators to use. Placement of the questionnaires on a Web site would facilitate access and data entry from large collaborative studies and, with appropriate programming, could allow near real-time analysis of data for continuous monitoring of patient enrollment. In prospective studies that are not blinded, data could be analyzed as it arrives on the server and the investigator could be alerted as soon as sufficient statistical power is gained. Some quality assurance is achieved through automated processes, such as not allowing submission if data are missing, but some human review of records for accuracy and completeness will still be required.

By no means is our Web-based data acquisition system the only method of collecting clinical data in consistent electronic form. The development of a uniform reporting tool for osteopathic examinations has been an innovation of the Louisa Burns Osteopathic Research Center, and electronic SOAP (subjective, objective, assessment, and plans) notes are also under development. These could be transmitted among research sites electronically; however, these projects may not include all of the data that might be desirable for clinical, educational, or outcomes research. Many other research projects that do not specifically involve osteopathic examination may be envisioned by osteopathic physicians and residents within our OPTI programs. The Web-based data acquisition program we developed has an open architecture that will allow almost limitless options for data capture, and we anticipate using it extensively for future studies.

We think this system could be a useful tool for simultaneous data collection in multicenter studies. After the creation of a study protocol and identification of clinical sites and a lead investigator at each site, the type of data to be collected under the protocol would be carefully considered by investigators. Generally the data to be collected would come from different sources (research participant, investigator, laboratory) and/or be collected at different times or in different locations (hospital versus clinic, acute versus convalescent). It would be most logical to submit data at the time and place of observation. Therefore, multiple data collection instruments (Web pages) would be needed to accommodate the various data sources. The research staff, in consultation with the Web site administrator, would then be ready to construct individual pages that would be used to elicit clinical information. The type of data to be collected (discrete, continuous, numeric, text, limited in range) will determine the type of data collection instrument to be created. The investigator plays a key role in deciding if data fields are optional or mandatory, as the Web design can prompt the inclusion of missing data. Web pages for use by patients must be designed to be especially user-friendly and succinct.

After creation of the individual pages, a test of the functionality of the data collection instrument (both for patients and investigators) and functionality of the related parts of the system (server reliability, e-mail data verification, arrival of uncorrupted data on the server) should be evaluated. When functionality and reliability are confirmed, investigators may be assigned special codes to authenticate their entries.

As the study begins to acquire patients, the Web administrator can monitor data as it arrives and can prepare interim summaries of recruitment success for individual study sites, demographic characteristics of participants, and a near real-time evaluation of adverse events and therapeutic efficacy. The principal investigator could check on sites that have low enrollment or not meeting targets for inclusion of adequate numbers of specific subgroups. Interim reports would be available in electronic form and maintained by the principal investigator, transmitted to subinvestigators by e-mail, or placed on a secure Web site for viewing by all authorized users.

When the study is closed to recruitment, the Web administrator can deactivate the site so no additional entries can be made. An attractive feature of this system is that almost as soon as the study is closed, the data can be reviewed by the investigators, greatly enhancing the preparation of reports and the development of inferences.
Conclusion
We have reported our experience with a Web-based information system that permits direct collection patient history data, physician observations, and other information such as laboratory data, with near real-time collation of data from different sources. While this system proved exceedingly valuable in completing our clinical research project that generated large data sets, we view it as a useful tool in collecting data in large multicenter clinical trials. In a test of remote data entry, the Web-based data collection system performed effectively.

Acknowledgment
The clinical study that enabled us to develop the Web-based data collection system was funded by Kemin Foods, LC, of Des Moines, Iowa. Some results of that clinical investigation have been reported in the literature.6

References