INTERNET BASED DATA COLLECTION FOR ANALYSIS WITH SAS® USING FRONTPAGE® 98

Robert Workman, Southern Connecticut State University
Richard F. Pless, Independent Consultant

Abstract
This paper describes a data collection and analysis project designed to monitor blood product usage by hemophilia patients. Data collection was accomplished using Internet based forms created with Microsoft FrontPage® 98 and located on a PC running the Microsoft Personal Web Server. SAS® procedures used included INFILE with the DSD option, which allows comma-delimited files with consecutive missing values to be read correctly and system calls to redirect output to the end user’s e-mail account.

The Application
This application began as a corollary to a home health care study currently being implemented at the Yale University Hemophilia Treatment Center. The aim of the Yale project is to use P.C. based video teleconferencing to assess bleeding episodes in the patient’s home thereby reducing unnecessary emergency room visits. While designing the project and determining what data was needed to evaluate the study, we concluded that home infusion information (where patients or their families administer drugs to control bleeding) was crucial to judging the project’s results. This application was designed to collect this crucial home infusion information enabling us to judge the efficacy of the larger tele-medicine project.

Healthcare Data Considerations
Without attempting to cover all possible Internet-based healthcare data collection concerns in this paper, we do address the most glaring security concern. There are two basic intrusion threats from an information systems perspective: attacks against stored data, and reading data during transmission. Web security is far beyond the scope of this paper and information on this complex topic can be found relatively easily on the Internet.

Our primary defense against intrusion in this project was to separate a patient’s identifying information from their clinical data. Knowing that ‘someone’ has a heart condition without knowing that ‘someone’ s identity does not breech patient confidentiality. Knowing that ‘John Smith’ has a heart condition is confidential information which requires safeguards. By simply replacing the patient name with a randomly chosen id number, we can protect patient confidentiality without having to manually edit CGI scripts to include encryption.

Hemophilia Home Care Program Treatment Log

Patient ID

Reason for Treatment
☐ Surgery Related ☐ Injury ☑ Spontaneous ☐ Other

Time Between Start of Bleeding and First Treatment
Less than one hour ☐ 1 to 3 Hours ☑ More than 3 hours ☐

Bleeding Site ☑ Shoulder ☑ Factor ☑ Number of Units ☑ Lot Number

Days of work or school missed ☑

Submit  Clear Data

Figure 1. The Hemophilia Home Care Program Treatment Log.
Figure 2. FrontPage® Forms Properties Dialog Box

1. Injury, Less than one hour, Ankle, fIX, 3, 12, 0, Submit Query
2. Surgery-related, 1 to 3 hours, Elbow, fVIII, 2, 3, 1, Submit Query
2. Other, 1 to 3 hours, Elbow, fVIII, 1, 22, 2, Submit Query
3. Spontaneous, 1 to 3 hours, Shoulder, fIX, 2, 132, 0, Submit Query
2. Injury, Less than one hour, Elbow, fVIII, 3, 321, 1, Submit Query

Figure 3. Sample comma delimited output.

Data Output Options

Figure 1 shows the data entry form proposed for use by the hemophilia patients. FrontPage® 98 supports a number of form output options. Data output options are selected in the FrontPage® Forms Properties Dialog Box an example of which is shown in Figure 2. As may be seen from the dialog box output options include comma delimited and e-mail output. When comma delimited output is used data is saved on the server's hard disk. Sample comma delimited output data is shown in Figure 3. When e-mail output is selected responses to the form are e-mail to a selected recipient. Figure 4 shows an example of e-mail output.

```
IDNumber: 3
Treatment: Surgery-related
StartOfTreatment: Less than one hour
BleedingSite: Elbow
Factor: fVIII
Units: 2
Lot: 345
DaysMissedSchool: 1
Submit: End of Record
```

Figure 4. Sample E-Mail output.

```/* FRONTPAGE Data Collection*/
/* June 12, 1998 */
libname sasdata "c:\98SUGI";

data sasdata.patient;
infile 'c:\98SUGI\result2.txt' delimiter =',' missover
DSD;
length ptid $4 reason $15 time $17 site $8 factor $5 units
4 lot $6 missed 4 garbage $12;
input ptid reason time site factor units lot missed
garbage;
drop garbage
run;```
**Figure 5.** SAS® Program.

**SAS® Analysis Of Comma Delimited Data**

Analyzing the data set generated by the CGI script was straightforward using the DSD option as demonstrated in Figure 5 above. Simply using the infile option, `delimiter = ','` causes the SAS® program to incorrectly read the remainder of that record, shifting all of the fields to the right of the missing value to the left.

Adding the DSD option after specifying the comma delimiter tells SAS® not to skip successive delimiters but to assign a null value to that field. One word of caution, when using the DSD option the data must be delimited using only commas. If the fields are then in quotes and further separated by commas the DSD option will fail and you will receive a fatal error.

**Reporting the results of SAS® Analysis**

The goal of this project is to provide summary statistical information to the Nurse Practitioner responsible for the patients. We assume that our end user has some knowledge of Windows 95, but does not have the time to organize the data on a weekly basis. The first option we considered was a three computer setup involving a web server, another machine running SAS® and finally the Nurse Practitioner’s P.C. The data would be collected and encrypted on the web server, transferred to the server running SAS® for decryption, analysis and re-encryption. The results would then be e-mailed to the Nurse Practitioner’s P.C. While this would be a useful approach if the three functions (web server, SAS®, and e-mail client) were separated across different operating systems (e.g. UNIX and Windows) it is not the optimal approach in a network running SAS® on Windows 95.

The main advantage of implementing a FrontPage® Web Server with SAS® is the fact that they can be integrated on a single machine to provide powerful data collection and statistical analysis capabilities to a user with no knowledge of either package. After booting the web server and writing the SAS® program, maintenance would be minimal, and training would little more than instructing the user how to run SAS® in batch mode and where to find the output.

**Summary and Conclusion**

We have presented a preliminary model for collecting data via a FrontPage® Server and analyzing it with SAS®. We plan to revise this form and the SAS® program to collect more specific information and add additional analysis. Another refinement we are currently investigating is an inventory monitoring capability. A SAS® program set to run automatically each day will flag those patients running low on anti-coagulants and generate an e-mail notifying the Nurse Practitioner and the hospital blood bank. We are also examining more secure ways of storing information on the web server, namely keeping the data set and output encrypted when not in use. While this program is not yet ready for implementation, we believe that further development will produce an application beneficial to hemophilia care providers as well as any provider monitoring patients receiving home healthcare.

-------------------

Richard F. Pless, Independent Consultant
1953 West Foster #3, Chicago, IL 60640,
rpless@yahoo.com

Robert Workman, Department of Computer Science,
Southern Connecticut State
University 501 Crescent St., New Haven, CT 06515,
workman@scsu.ctstateu.edu
Associate Professor Computer Science
Develop and teach courses in Internet applications and management, and office productivity.

Acknowledgements

Allison A. Sakara, Nurse Coordinator, Yale University
Hemophilia Treatment Center

References