

## BRIEF REPORT: Health Services Research in The Privacy Age

Rachel Hess, MD, MSc,<sup>1</sup> Karen Matthews, PhD,<sup>2,3</sup> Melissa McNeil, MD, MPH,<sup>1</sup> ChungChou H. Chang, PhD,<sup>1,4</sup> Wishwa Kapoor, MD, MPH,<sup>1</sup> Cindy Bryce, PhD<sup>1,5</sup>

<sup>1</sup>Division of General Internal Medicine, Center for Research on Health Care, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; <sup>2</sup>Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; <sup>3</sup>Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA; <sup>4</sup>Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA; <sup>5</sup>Department of Health Policy and Management, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA.

**BACKGROUND:** Strengthened regulations concerning privacy of health information are affecting large-scale health outcomes research.

**OBJECTIVE:** To create a data collection system that would facilitate outcomes research, avoid selection bias, and fulfill obligations to protect privacy.

**DESIGN:** We created a web-based system that uses touch-screen computer technology for longitudinal collection of data. The system provides access to information in deidentified form, enables it to be linked to health services and outcomes data, and allows patients to join a research registry project (RRP) and be placed on a prospective subject list (PSL).

**PARTICIPANTS, MEASUREMENTS, AND RESULTS:** Pilot testing in 86 consecutive patients who were seen at a large, urban, university-based general medicine practice and had a mean age of 50 years showed that 81 patients had no difficulty, 5 had some difficulty, and none had considerable difficulty using the computer technology to complete a health survey. No patients refused to complete the survey and all patients completed the entire survey. Forty-seven (55%) joined the RRP and 42 of these 47 (89%) joined the PSL. RRP participants were less likely than RRP nonparticipants to be divorced or widowed ( $P=.03$ ) and less likely to have hypertension ( $P=.03$ ) but had no other significant differences in sociodemographic or clinical characteristics. PSL participants did not differ from PSL nonparticipants.

**CONCLUSIONS:** The new system ensures privacy and appears to facilitate research recruitment and avoid selection bias.

**KEY WORDS:** health services research; computerized data collection; Health Insurance Portability and Accountability Act; research registry; privacy.

DOI: 10.1111/j.1525-1497.2005.0227.x

J GEN INTERN MED 2005; 20:1045–1049.

In 1996, Congress passed the Health Insurance Portability and Accountability Act (HIPAA) to protect the personal health information (PHI) of individuals.<sup>1</sup> Accompanying regulations indicate that health care data must be stripped of all identifiable PHI, including an individual's name, social security number, zip code, and dates of office or hospital visits,<sup>2</sup> unless the individual provides written informed consent to use this information. The regulations can lead to selection bias in health outcomes research because individuals who consent to participate in research may differ from those who refuse. In one study, for example, individuals who consented were older,

were more likely to be male and white, and had poorer physical functioning.<sup>3</sup>

Instituting the new regulations offered us the challenge to create a system that would facilitate outcomes research while fulfilling ethical and legal protections outlined by HIPAA. Ideally, the system would streamline longitudinal data collection regarding health and health status; centrally store data in a deidentified form; enable data to be linked, without identifiable information, to health services and health outcomes information; enable patients to permit use of their PHI for research purposes; allow patients to identify themselves as potentially willing subjects for future research studies; and function seamlessly within clinical practice.

While researchers are allowed to use PHI without HIPAA authorization in special cases, such as obtaining Institutional Review Board (IRB) waiver or in circumstances when the work is preparatory to research, our approach for meeting the challenge was to invite each of our patients to join a research registry prospectively and be placed on a list of potential research participants. A registry is "an observational, nonexperimental database designed to reflect current patterns of practice without influencing the treatments or interventions prescribed."<sup>4</sup> Most registries collect data regarding a particular disease, but they may also provide data on health-related quality of life (HRQOL).<sup>4</sup> As primary care physicians managing individuals with a wide variety of disorders, we designed our registry around HRQOL.

In this article, we describe the system constructed and implemented by the University of Pittsburgh Division of General Internal Medicine. We discuss its acceptance by patients and physicians and its potential contribution to research and clinical care.

## METHODS

### System Design

We created a web-based data collection system, the Functional Assessment System Tablet (FAST). This system allows patients to use touch-screen computers to complete routine questionnaires that are currently administered in our outpatient practice and elicit information about health, marital status, education, substance use, social support, and HRQOL (RAND-36).<sup>5</sup> A summary of the responses is provided to the clinician at the time of the visit and placed in the patient's electronic medical record. The technology allows us to vary

---

*The authors have no conflicts of interest to declare for this article or this research.*

*Address correspondence and requests for reprints to Dr. Hess: Center for Research on Health Care, 230 McKee Place, Suite 600, Pittsburgh, PA 15213 (e-mail: hessr@upmc.edu).*

---

*Received for publication June 13, 2005*

*and in revised form June 16, 2005*

*Accepted for publication June 16, 2005*

time intervals for questions, skip questions that are irrelevant for particular patients, and score instruments in real time. The computers are wirelessly networked, letting patients move through the office while completing questionnaires and thereby minimizing the impact on office flow.

The FAST identifies individuals but automatically segregates their PHI in a separate, password-protected database (Fig. 1). This allows longitudinal follow-up, via an electronic honest broker, while storing the majority of data in a deidentified manner. All databases are firewall secured.

The FAST questionnaire asks patients whether they would like to participate in our HRQOL research registry project (RRP) and to join a prospective subject list (PSL) and be notified about studies for which they may be eligible. By affirmatively answering these questions, the patient is “approaching” the research staff, thus removing the obligation for HIPAA authorization.

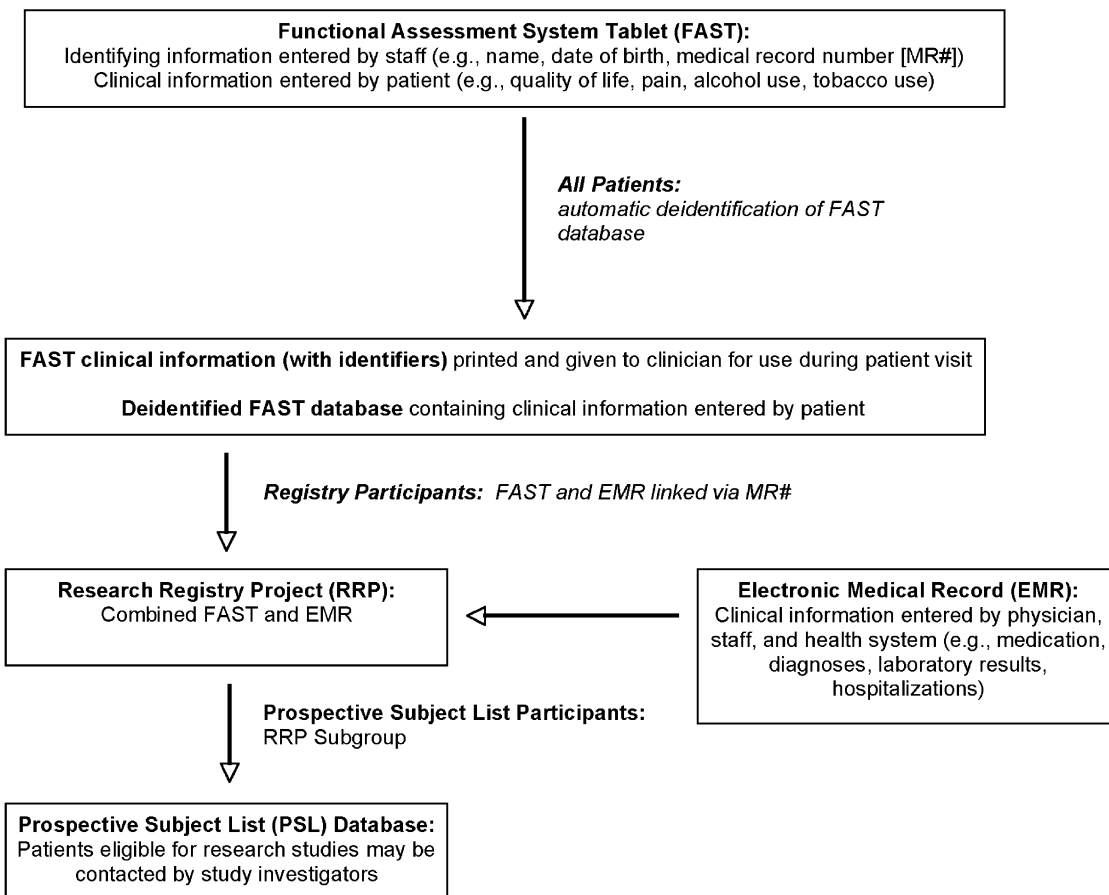
Signing informed consent for participation in the RRP permits investigators to create a future database with the subject’s complete medical record and FAST data in order to “investigate the impact of HRQOL and the presence, absence, risk, or treatment of comorbid medical conditions and symptoms on health services utilization and disease outcomes.” Specific reference is made in both the protocol and consent form as to the inclusion of HIV status and treatment as well as mental health issues. Signing informed consent to participate in the PSL gives investigators permission to query the RRP

electronically for eligibility criteria of potential subjects but otherwise respect PHI. PSL participants also consent to allow individual investigators to approach them regarding participation in a specific study. Patients not participating in the PSL may be contacted via usual channels (e.g., approached by personal physician, advertisement).

Investigators who wish to use the RRP must obtain approval from both the IRB and the RRP oversight committee. To use the PSL, the proposal must be additionally approved by the Division’s clinical oversight committee. Investigators must be affiliated or collaborating with a member of the Division or the Center for Research on Health Care, all named coinvestigators on the project. All data are controlled by the Division and administered and maintained by the Division’s Data Center.

### System Implementation

We developed the RRP and PSL as research projects and, after receiving University of Pittsburgh Institutional Review Board approval, pilot tested the system between September 2003 and May 2004 in our large, urban, university-based general medicine practice. Two physicians and 3 medical assistants in our practice agreed to help the principal investigator (PI). When consecutive patients presented for physician appointments, the assistant or PI gave each patient a touch-screen computer, explained how to complete the computerized questionnaire, and remained available to answer queries concern-



**FIGURE 1.** Relationship of Functional Assessment System Tablet (FAST), research registry project (RRP), and prospective subject list (PSL). All patients complete the FAST. FAST information is stored in a deidentified database. Patients may consent to participate in the RRP, in which case FAST data are linked to the EMR to create the RRP. RRP participants who agree to participate in the PSL are placed in the PSL database.

**Table 1. Sociodemographic and Clinical Characteristics of the 86 Patients Who Completed the Functional Assessment System Tablet (FAST)\***

Characteristic	Number (Percent)
Age (y), mean $\pm$ SD	50.0 $\pm$ 15.6 (range, 19 to 84)
Gender	
Women	76 (88)
Marital status	
Married or in committed relationship	44 (51)
Single	26 (30)
Divorced	8 (9)
Widowed	8 (9)
Highest educational attainment	
At least some high school	9 (10)
Some college	25 (29)
Completion of college	17 (20)
Some graduate school	35 (41)
Hazardous drinker <sup>†</sup>	14 (16)
Current smoker	8 (9)
Comorbid medical conditions	
Hypertension	43 (50)
Heart attack	5 (6)
Congestive heart failure	3 (3)
Depression	31 (36)
Lung disease	8 (9)
Arthritis	27 (31)
Cancer	7 (8)
Diabetes	10 (12)
RAND-36 MHC score, mean $\pm$ SD	44.4 $\pm$ 11.5 (range, 15 to 63)
RAND-36 PHC score, mean $\pm$ SD	45.9 $\pm$ 10.7 (range, 20 to 60)
Pain score on a scale of 1 (none) to 10 (worst ever), mean $\pm$ SD	2.7 $\pm$ 2.1 (range, 1 to 9)
Research registry project participant	47 (55)
Prospective subject list participant	42 (49)

\*Values are numbers (percent) unless otherwise indicated. Because of rounding, percentages may not all total 100.

<sup>†</sup>Hazardous drinking in women is defined as at least 4 drinks in a day or 7 drinks per week. Hazardous drinking in men is defined as at least 5 drinks in a day or 14 drinks per week.

SD, standard deviation; MHC, mental health composite; PHC, physical health composite.

ing computer use and noted the time needed to finish the questionnaire.

At the end of the questionnaire, the patient was asked via computer whether they wished to join the RRP and PSL. If the patient agreed, the computer printed a consent form that would be reviewed together by the PI and patient and then signed after the office visit. The final survey question was: "Did you have trouble using the computer to answer these questions?" Patients could respond that they had no difficulty, some difficulty, or considerable difficulty. Patient responses were summarized and immediately printed for the physician to review during the patient encounter. At the end of each session, medical assistants and physicians were asked to comment on concerns or issues that arose during the day.

## Data Analysis

We used direct observation to estimate the time needed to complete the questionnaire. We summarized patient characteristics, using means for continuous variables and frequencies for categorical variables. We used the Fisher exact test, Wilcoxon rank-sum test, and Student *t* test to examine data concerning computer-use difficulties and RRP and PSL participation. We used STATA version 7.0 (Stata Corp., College Station, TX) for all analyses.

In addition to analyzing quantitative data, we asked medical assistants to describe computer-related problems encountered during each session, and we asked physicians to comment on the usefulness of the computer-generated information. We summarize their comments below.

## RESULTS

We pilot tested the FAST in 86 consecutive patients, whose sociodemographic and clinical characteristics are shown in Table 1. No patient refused to complete the FAST. Patient age ranged from 19 to 84 years. Similar to the patient panels of the participating physicians, 88% of participants were women. Half were married or in a committed relationship, and over 60% had at least a 4-year college degree. Patients had a variety of medical conditions. On the RAND-36, patients' mean mental health composite (MHC) and physical health composite (PHC) scores were consistent with those found in similar patient populations.<sup>6</sup>

Most patients completed the questionnaire in  $\leq 15$  minutes. Only 1 patient needed assistance with more than 2 FAST screens before feeling comfortable using the computer independently. Every patient completed the entire questionnaire, and most finished it before their physician encounter.

Of the 86 patients, 81 reported no difficulty using FAST, 5 reported some difficulty, and none reported considerable difficulty. As shown in Table 2, patients who were older, had higher pain scores, or had lower PHC scores were more likely to report some difficulty using FAST ( $P=.058$  for age, 0.001 for pain, and 0.010 for PHC scores).

Forty-seven patients (55%) joined the RRP. Of these, 42 (89% of RRP members, 49% overall) also joined the PSL. Because we could access deidentified information for all FAST completers, we could compare characteristics of participants and nonparticipants. RRP participants were less likely than RRP nonparticipants to be divorced or widowed and were less likely to have hypertension ( $P=.030$  for each) but showed no significant differences in age, gender, educational attainment, alcohol or tobacco use, self-reported depression and arthritis, or RAND-36 scores. There were no significant differences between PSL participants and PSL nonparticipants.

Medical assistants reported that FAST altered their workflow by requiring them to answer patients' computer-related questions and to collect computers from patients. These responsibilities increased the workload by about 1–2 minutes per patient. Physicians found FAST information easy to understand and useful for patient care. In particular, the MHC scores were helpful for identifying patients at risk for depression.

## DISCUSSION

The new privacy regulations strengthen the protection of human subjects, which is critical to ethical research. Yet, they also make it more difficult to conduct retrospective studies, especially studies requiring the use of identifiers to link multiple data sources, and they make it more complicated to recruit subjects without introducing selection bias.

We believe that a web-based data collection system such as FAST offers a viable solution to these challenges, because it fulfills 5 of the 6 criteria that we identified initially. The system (1) streamlines the longitudinal collection of health and health

**Table 2. Comparison of Patients Reporting No Difficulty and Patients Reporting Some Difficulty in Completing the Functional Assessment System Tablet (FAST)\***

Characteristic	Patients Reporting No Difficulty (n=81)	Patients Reporting Some Difficulty (n=5)	P Value
Age (y), mean ± SD	49.3 ± 15.7	61.8 ± 7.9	.058 <sup>†</sup>
Gender			.47 <sup>‡</sup>
Men	9	1	
Women	72	4	
Marital status			.58 <sup>‡</sup>
Married or in committed relationship	40	4	
Single	25	1	
Divorced or widowed	16	0	
Highest educational attainment			.65 <sup>‡</sup>
At least some high school	8	1	
Some college	23	2	
Completion of college	17	0	
Some graduate school	33	2	
Pain score on a scale of 1 (none) to 10 (worst ever), mean ± SD	2.5 ± 1.9	6.4 ± 1.9	.001 <sup>†</sup>
Comorbid arthritis			.18 <sup>‡</sup>
No	57	2	
Yes	24	3	
RAND-36 MHC score, mean ± SD	44.8 ± 11.5	37.0 ± 9.3	.10 <sup>†</sup>
RAND-36 PHC score, mean ± SD	46.6 ± 10.5	34.0 ± 4.9	.01 <sup>†</sup>

\*Ease of use response options were no difficulty, some difficulty, and considerable difficulty. No one reported considerable difficulty.

<sup>†</sup>Results of Wilcoxon rank-sum test.

<sup>‡</sup>Results of Fisher exact test.

SD, standard deviation; MHC, mental health composite; PHC, physical health composite.

status data by decreasing the need for manual data entry; (2) provides access to data in a deidentified form; (3) enables the data to be linked to health services and health outcomes information; (4) allows patients to consent to the use of their identifiable information by joining the RRP; and (5) allows patients to indicate their willingness to be contacted regarding future research by participating in the PSL. We are continuing to work with our staff to ensure that the system functions seamlessly within our clinical practice.

Other investigators have designed computerized systems to identify potential subjects to investigators.<sup>7</sup> We believe that our system, utilizing the FAST for clinical data collection, the RRP for outcomes research, and the PSL to allow patients to volunteer to be contacted regarding research provides the patient with greater autonomy in the process. Additionally, it allows us to utilize information collected at a previous time point for current research endeavors, including patient recruitment.

By undertaking systematic comparisons of the RRP participants and the overall FAST population, we can use FAST to quantify selection bias in a research sample and better assess its significance. Through the PSL, we can define a volunteer population from which to recruit for future research. In our study, 55% of patients joined the RRP, and 89% of RRP members (49% overall) joined the PSL. These successful participation rates compare favorably with rates reported in other studies<sup>8-10</sup> but can likely be improved. In the pilot version of FAST, when patients were asked to join the RRP and PSL, they could only answer yes or no. Adding a "maybe" response will allow patients to obtain additional information before deciding. Repeating the RRP and PSL invitations at subsequent office visits and educating physicians about our project may also enhance participation. Repetitive invitation does raise the risk of patients feeling coerced, and we are following this closely. There have been no complaints to date.

Patients in our study had little difficulty using FAST, consistent with earlier studies that showed good acceptance of

touch-screen technology in primary care and subspecialty settings.<sup>11</sup> Additionally, we had no incomplete data, corroborating other work demonstrating that computer-based data collection decreased the amount of missing data.<sup>11</sup>

While we observed few differences between RRP participants and nonparticipants, there are many unmeasured variables yet to be explored. These may include personality type, length of time enrolled as a patient in the practice, trust in the health care system, and more severe illness. We will explore these questions as we move forward with the project.

As with most pilot studies, our study involved a small number of patients, physicians, and staff, making it difficult to generalize our results. However, we are beginning large-scale deployment and will continue to track differences between participant and nonparticipant groups as well as patient difficulties in using the system. We also plan to assess formally how our computer-based survey affects office procedures and health care delivery.

Our new system was developed in the tradition of practice-based research. By allowing physicians to continue with crucial health outcomes research while respecting patients' ethical and legal rights, we believe that the system has the potential to improve health care services on the local level and add to the body of knowledge on the national level.

*During the time of this work, Dr. Hess was supported by a Veteran's Administration Special Fellowship in Women's Health.*

*This work was presented in part at the 2004 Society of General Internal Medicine and Academy Health Meetings.*

## REFERENCES

1. Health Insurance Portability and Accountability Act. 1996.
2. **Department of Health and Human Services.** Standards for privacy of individually identifiable health information: proposed rule. The Federal Register. 1999;64:59918-60065.

3. **Woolf SHL, et al.** Selection bias from requiring patients to give consent to examine data for health services research. *Arch Fam Med.* 2000;9:1111-8.
4. **Kennedy L, Craig AM.** Global registries for measuring pharmacoeconomic and quality-of-life outcomes: focus on design and data collection, analysis and interpretation. *Pharmacoeconomics.* 2004;22:551-68.
5. **Hays RD, Morales LS.** The RAND-36 measure of health-related quality of life. *Ann Med.* 2001;33:350-7.
6. **Hays RD.** RAND 36: Health Status Inventory. 1st ed. San Antonio, TX: The Psychological Corporation; 1998:126.
7. **Guinn J, Durski K.** A real-time tracking, notification, and web-based enrollment system for emergency department research. *Acad Emerg Med.* 2004;11:1245-8.
8. **Matthews KA, Kelsey SF, Meilahn EN, Kuller LH, Wing RR.** Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. *Am J Epidemiol.* 1989;129:1132-44.
9. **Sowers M.** SWAN: a multicenter, multiethnic, community-based cohort study of women and the menopausal transition. In: Lobo RA, Kelsey J, Marcus R, eds. *Menopause: Biology and Pathobiology.* San Diego: Academic Press; 2000:672.
10. **Pradhan AD, Manson JE, Rossouw JE, et al.** Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the women's health initiative observational study. *JAMA.* 2002;288:980-7.
11. **Bliven BD, Kaufman SE, Spertus JA.** Electronic collection of health-related quality of life data: validity, time benefits, and patient preference. *Qual Life Res.* 2001;10:15-22.

## 29<sup>th</sup> Annual Meeting

April 26-29, 2006  
Weston Bonaventure Hotel and Suites  
Los Angeles, California

### Submission Deadlines:

Precourses, Workshops and Interest Groups: October 17, 2005  
Abstracts, Vignettes and Innovations: January 12, 2006

Check the SGIM website ([www.sgim.org/am06](http://www.sgim.org/am06)) for more information!