Application of Information Technology ■

Database Design to Ensure Anonymous Study of Medical Errors: A Report from the ASIPS Collaborative

WILSON D. PACE, MD, ELIZABETH W. STATON, MSTC, GREGORY S. HIGGINS, BS, DEBORAH S. MAIN, PHD, DAVID R. WEST, PHD, DANIEL M. HARRIS, PHD

Abstract Medical error reporting systems are important information sources for designing strategies to improve the safety of health care. Applied Strategies for Improving Patient Safety (ASIPS) is a multi-institutional, practice-based research project that collects and analyzes data on primary care medical errors and develops interventions to reduce error. The voluntary ASIPS Patient Safety Reporting System captures anonymous and confidential reports of medical errors. Confidential reports, which are quickly de-identified, provide better detail than do anonymous reports; however, concerns exist about the confidentiality of those reports should the database be subject to legal discovery or other security breaches. Standard database elements, for example, serial ID numbers, date/time stamps, and backups, could enable an outsider to link an ASIPS report to a specific medical error. The authors present the design and implementation of a database and administrative system that reduce this risk, facilitate research, and maintain near anonymity of the events, practices, and clinicians.

■ **J Am Med Inform Assoc.** 2003;10:531–540. DOI 10.1197/jamia.M1339.

The study and reduction of medical errors have become a major theme within health care. Two recent Institute of Medicine reports,[1,2] recent changes in Joint Commission on Accreditation of Healthcare Organization (JCAHO) requirements,[3] and congressional action to increase funding for research on medical errors have increased interest in identifying more effective ways to collect, catalog, and analyze medical error reports. Prominent among recommendations for reducing errors are identifying and learning from medical errors and near misses through both mandatory and voluntary reporting systems.[1,4]

Applied Strategies for Improving Patient Safety (ASIPS) is a multi-institutional, practice-based project designed to collect, codify, categorize, and analyze data on medical errors occurring in primary care offices and to develop interventions to reduce those errors. The ASIPS project, a three-year research study funded by the Agency for Healthcare Research and Quality (AHRQ), includes a voluntary Patient Safety Reporting System (PSRS) that captures both anonymous and confidential reports of medical errors. Information about medical errors gathered through the PSRS will be combined with our analyses of malpractice insurance claims and state Medicaid claims data to determine the types of errors that occur in primary care and to aid in the development of interventions to reduce those errors.

The PSRS database/administrative system is designed to collect error reports and manage data while maintaining confidentiality of highly sensitive information. Confidentiality is maintained by de-identifying reports and eliminating elements within the database that would facilitate linking a report within the PSRS to a specific event identified by other means, such as through the patient's medical record. This report presents the legal reasons for protecting confidentiality of our data as well as others' experiences designing databases for similar purposes. We then present the design objectives and implementation of a database and administration system that facilitates research activities while protecting (1) the identity of practices, clinicians, and staff who report medical errors and (2) the detail of the events themselves, should the database be viewed by people outside the research team.

Background

The PSRS provides clinicians and office staff a vehicle for voluntarily reporting patient safety events, including both near misses and errors that lead to patient harm (we use the collective term *medical errors*). A person who believes she or he knows of a medical error can report the event via automated telephone hotline, through a Web site, or in paper form. The PSRS allows users two methods to make reports: confidential (identifiable but held in confidence) and anonymous (completely unidentifiable). For confidential reports, ASIPS participants provide their name and phone number

Affiliations of the authors: Department of Family Medicine, University of Colorado Health Sciences Center, Aurora, Colorado (WDP, EWS, GSH, DSM, DRW); The CNA Corporation, Alexandria, Virginia (DMH).

The authors acknowledge the contributions of all ASIPS participants, especially those who have entrusted them with their stories of medical errors. The authors thank Thomas J. Miyoshi of the Colorado AHEC System and George "Ed" Fryer, Jr., PhD, of the Robert Graham Policy Center for their analyses of the database's security features. Preliminary information concerning the database and data safety processes was presented at the North American Primary Care Research Meeting, New Orleans, Louisiana, October 2002. Funding for this study was provided by the Agency for Healthcare Research and Quality, grant # 1U18HS011878-01, Wilson D. Pace, MD, principal investigator.

Correspondence and reprints: Wilson D. Pace, MD, Department of Family Medicine: UCHSC at Fitzsimons, PO Box 6508 - Mail Stop F496, Aurora, CO 80045-0508; e-mail: <wilson.pace@uchsc.edu>.

Received for publication: 01/29/03; accepted for publication: 04/23/03.

and answer one free-text question that presents a minimal description of the incident (Fig. 1). Members of the study staff then collect more detailed information during a follow-up telephone interview. For anonymous reports, participants make a one-time report during which they select broad categories to identify their role (e.g., physician, nurse, staff) and their practice type (residency practice, community health center, rural office, etc.). They then answer up to four free-text questions and up to six multiple-choice questions about the event and the patient(s) involved.

Because we cannot follow-up with additional clarifying questions, anonymous reports typically provide less detail about an event than do confidential reports. Anonymous reports neither allow us the chance to probe for additional information nor, typically, provide enough data to understand root causes of the error. Confidential reports are advantageous because during the follow-up visit, interviewers familiar with medical errors in primary care offices can probe for secondary information, such as details about office systems, training, and event-specific environmental issues that reporters rarely elaborate on spontaneously. Furthermore, those reporting seem to prefer the confidential reporting form because of the shorter time required to complete the form.

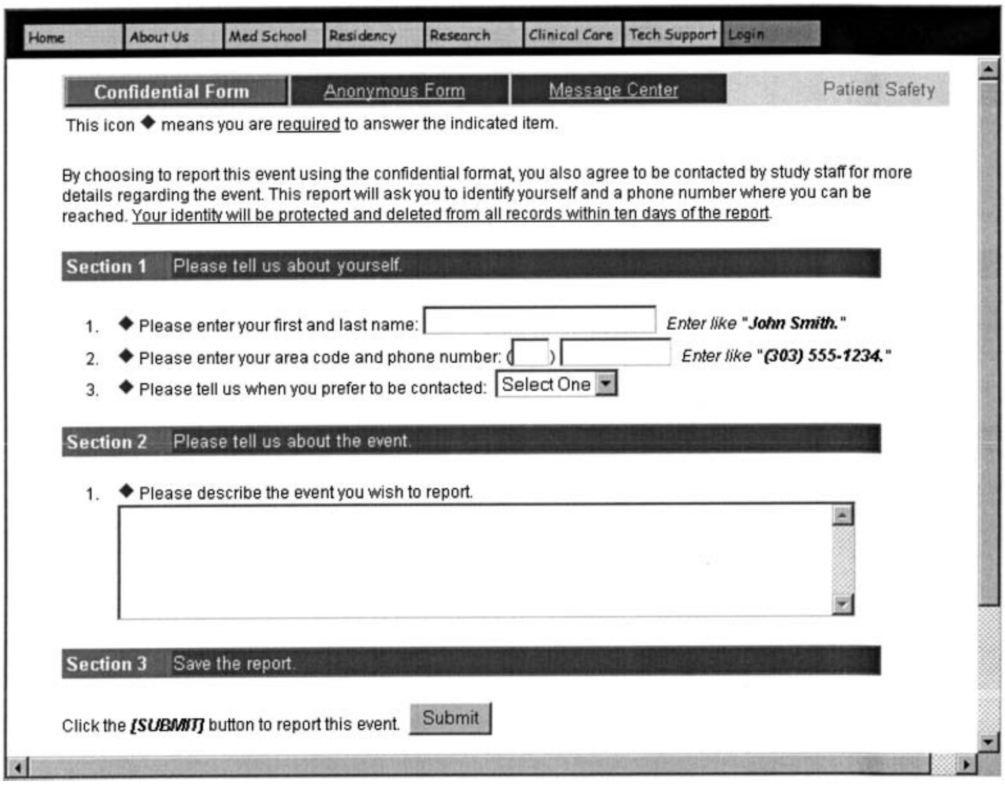
Unlike other voluntary reporting systems,[5] we did not issue a private login to participants that would allow them to remain anonymous and still provide more detailed information. We were not convinced that people who fear being identified would be any more trustful of a private handle or key that we provide. We also believed that any login system would further complicate the process of completing a report, reducing the number of reports we receive.

If one assumes that the primary reason for collecting and analyzing medical errors is to develop systems for error reduction or mitigation, then anonymous reports may be generally less helpful than confidential reports. This belief is supported by the experience of other safety reporting systems; for example, those heading the Federal Aviation Administration's Aviation Safety Reporting System feel so strongly that confidential reports provide superior information that their system will not accept anonymous reports.[6]

Threats to Confidentiality of the PSRS Data

Clinicians and office staff using a medical error reporting system need to feel that they can make a report without fearing disclosure, which may lead to discipline or malpractice lawsuits. While we follow stringent electronic data security processes to protect against unauthorized access to ASIPS report data, we have no control over the legal discovery of report data through subpoena.

Many efforts to reduce errors are part of quality improvement processes within health care organizations. The information gathered for quality improvement cannot be subpoenaed for a lawsuit in many states. State quality assurance laws (also called *peer review legislation*) usually protect organizational quality improvement databases from legal discovery. These laws do not protect quality improvement databases, however, if organizations share their data with anyone outside the organization. Thus, the legal protection does not apply to projects such as ASIPS, in which multiple organizations are contributing medical error reports to a single database that crosses those organizational boundaries. Furthermore, if an organization shares quality improvement data for a specific medical error, that organization's internal quality improve-



This icon ◆ means you are required to answer the indicated item.

By choosing to report this event using the confidential format, you also agree to be contacted by study staff for more details regarding the event. This report will ask you to identify yourself and a phone number where you can be reached. Your identity will be protected and deleted from all records within ten days of the report.

Section 1 Please tell us about yourself.

1. ◆ Please enter your first and last name: Enter like "*John Smith.*"
2. ◆ Please enter your area code and phone number: ( ) Enter like "*(303) 555-1234.*"
3. ◆ Please tell us when you prefer to be contacted:

Section 2 Please tell us about the event.

1. ◆ Please describe the event you wish to report.

Section 3 Save the report.

Click the *[SUBMIT]* button to report this event.

Figure 1. ASIPS confidential reporting Web form.

ment information concerning that case may also be subpoenaed.[1] Thus, the risk of having someone identify a specific report within a shared database, such as the ASIPS database, has more to do with the loss of the organization's internal data protection than with the details included in the shared database.

For ASIPS and other patient safety improvement projects that cross organizational boundaries, the ability to protect all medical error databases from subpoena becomes a troublesome issue. To date, most multi-institutional medical error collection systems have focused on collecting anonymous reports,[5,7] presumably to reduce risk if legal discovery were to occur.[8] ASIPS was developed with the belief that more detailed and useful information could be derived from confidential reports; thus, we needed to examine the legal threats to discovery that the ASIPS project might face.

By being a multi-institutional research project housed within the University of Colorado, the ASIPS database is not protected from subpoena by Colorado's peer review legislation, which protects quality improvement activities as long as the information remains within the individual institution conducting the quality improvement activity.[9] Federal laws appear to offer more protection but remain essentially untested. The ASIPS project is funded by AHRQ; thus, our research database is technically protected from subpoena by the statutory confidentiality provision of AHRQ's authorizing legislation[10]; however, this protection has not yet been tested through litigation.[11] To date, no attempt has been made to subpoena research data protected through AHRQ (AHCPR) legislation, so the protection remains untested in court. Besides disclosure by subpoena, research data may be available under the Freedom of Information Act if collected with federal support and used to justify changes in law.[12]

As of this writing, two federal bills (H.R. 663 and S. 720 of the 108th Congress) are under consideration by Congress that would protect medical error reports (and associated organizational patient safety information) from unintended discovery. Both bills propose to modify Title IX of the Public Health Service Act by designating patient safety information as privileged and confidential and not discoverable in connection with a civil or administrative proceeding.

Because current laws cannot absolutely guarantee the confidentiality of the information contained in the ASIPS database nor protect the organizations that participate in ASIPS, one of our primary objectives was to design a system that makes this information as untraceable as possible, without sacrificing the valuable detail gained by confidential investigation of medical errors. The goal was to create a database that would not allow outsiders to positively identify a safety event detailed in our database, should they gain access to our data, to protect both the reporters and their organizations. Creating a database that would allow confidential reporting while ensuring rapid and effective de-identification of report data turned out to be more difficult than first thought. Very little help was found in the database design literature.

Others' Experiences in Creating Confidential or Anonymous Databases

The best known error reporting system that collects confidential reports and then de-identifies them is the Federal Aviation Administration's critical incident reporting system (known as the Aviation Safety Reporting System or ASRS), maintained by NASA.[13] The ASRS database fields, as described in the NASA Reference Publication 1114,[14] provide very specific information concerning the event being reported. Although the name of the reporter is removed from the database, it is highly likely that someone with access to the ASRS database and basic information about an aviation safety event, such as date, time, type of plane, and location of event, could match the ASRS data to the event in question. In fact, it is likely that many reports could be matched to the reporter based on the reporter's role within the event, such as pilot or controller. The ASRS database, however, is protected by federal law so that it cannot be used in disciplinary actions.[15] Therefore, the threat of matching reports with known events was not a concern to the ASRS developers. Organizations lacking such legal protections cannot use the ASRS database model to develop a reporting system that protects medical errors from possible re-identification based on reports.

In the health services research arena, investigators typically create a de-identified dataset for study by taking a set of identifiable health data and removing or transforming specific elements.[16] The Health Insurance Portability and Accountability Act (HIPAA) identifies a range of data elements that must be deleted or transformed to consider a data set "de-identified."[17] Thus, we expected the literature to contain approaches for creating a secure database that would maintain untraceable data. While we found several reports that explore mathematical approaches to preserving critical relationships within databases while prohibiting re-identification of the original data element[18,19] and a number of reports that discuss network security issues that must be considered,[20–23] we could find no literature that addresses the issues of maintaining anonymity in a real time, de novo database.

Design Objectives

Balancing our needs as researchers of medical errors with our participants' need for protecting their confidentiality, we embarked on an iterative design process to develop a system that could store and manage the many data elements required for studying medical error reports. We identified three major areas as significant threats to maintaining a database of de-identified reports: (1) various date/time relationships between reported events, (2) tracking of reports at the practice level, and (3) data security. Within these three areas we identified nine specific requirements of the database:

Date/Time Relationships

1. Link all information pertaining to a single event in a manner that does not identify the time (absolute or relative) the event was reported or the place it occurred without losing any internal event chronology.
2. Manage the data collection without time/date markers.
3. Capture taxonomic codes for events, identifying the version of the coding taxonomy used for each event without creating a relative time/date sequence.
4. Allow updates to records over time and record when an update was made to a specific record for data analysis.
5. Capture change of number and type of events reported over time without applying a time/date stamp to any specific event.

Track Types of Reports at the Practice Level

6. Identify practices that have reported and have not reported specific types of events, without revealing which specific event they reported.

Data Security

7. Protect the data from unauthorized viewing within our own institution.
8. Protect the database from possible data loss without maintaining copies of identifying information.
9. Create no paper trail for any step in the process except for the initial event report when reported via paper.

Typical design features of relational databases often defeat the intent of the above requirements. Even anonymous reports may not be immune from linkage back to a particular individual or institution if common database markers link a report to a specific or relative time. Various data management and data protection systems create absolute or relative date/time relationships within the database or between various copies of the database. Removing classic data tracking elements, such as time/date stamps and serial numbers, creates issues with the management of data flow as well as issues in the desired linkage of selected information, such as the types of patient safety reports made by a given practice. Furthermore, typical data protection procedures can defeat attempts to delete data elements or mask reporting relationships.

System Description

We have spent 18 months developing and refining a database to handle collection, storage, and analysis of medical error reports, along with an integrated Web-based administrative system to handle the workload. The features of the ASIPS database and administrative system are summarized in Table 1.

The ASIPS database currently meets all design objectives we set, according to two outside analyses performed by database experts. In the following paragraphs, we discuss each objective and describe how we addressed the issue. Along the way, it may be helpful to refer to Figure 2, which shows the essential steps and data elements related to a confidential report as it goes through our system from initial report by a participant to total de-identification, to analysis and storage.

Table 1 ■ Design Features of the ASIPS Database and Administrative System

Database features
- Contact information and medical error information are stored in separate databases
- Reports are assigned a random 10-digit identifier
- Contact information is automatically deleted at a predetermined time after report submission
- Coding process creates database objects that describe types of events to be tracked over time and place
- Database objects are tracked to practice name and general time once a critical mass of similar objects exists in the database, but no specific event can be traced to a time or location
- Data can be transferred to a third party for analysis; updated fields are used to track changes without any time/date stamping

Administrative system features
- All administration and data management tasks are handled through Web interface (leaving no paper trail)
- Administrative tasks related to processing reports are labeled and embedded into each event report to manage workflow

1. The database must link all information pertaining to a single event in a manner that does not identify the time (absolute or relative) the event was reported or the place it occurred without losing any internal event chronology.

Internal chronology of the event (that is, what happened first in the chain of events leading to the medical error) is always maintained in the text of the report. Generally, people who make a report clearly state what happened first, second, and so on. We do not remove these references because they are crucial to understanding the cascade of events that led to an error. Instead, we focus on removing any references that could identify the time (absolute or relative) that the event was reported. Furthermore, the database maintains full referential integrity between tables that describe an event.

Solving the absolute and relative time issue initially appeared straightforward: we use a randomly generated ten-digit number as the event identification number (primary key). While this solution worked well, without careful attention to detail, serial numbers invariably show up in tables linked to the primary event table. Serial numbers in subsidiary tables create a relative time/date stamp for the whole report. Some of the subsidiary tables do not actually need a table-specific primary key; instead, the random event ID serves to link rows to other tables; thus, we completely dropped the serial number to solve this issue. Other tables must contain their own primary key, as well as be linked to the main table (containing the original event information). This linkage is particularly needed for tables that contain parent–child relationships between data elements. These parent–child relationships do not imply temporal relationships between data fields; rather, they indicate a hierarchical relationships in our taxonomy that must appear in the database.

Because we use a random digit ID in the main table of the database, we chose not to use another random digit to serve as the primary key for the subsidiary table, because using multiple random digits can create problems with the parent–child relationships. If an internal reflexive data type is used to define the parent–child relationship, then the child must always have a primary key number greater than the parent, which cannot be guaranteed with a random digit as the primary key for the subsidiary table. Our solution to the parent–child problem is to use two fields to create a binary primary key: the random event primary key and a second "serial" field that restarts at one for each event ID. This approach organizes all data elements in the subsidiary table in a hierarchical manner while maintaining the random order of event recording (Fig. 3). This solution can be adopted for any table that must maintain referential integrity and contains multiple rows of data that need to be specifically ordered.

2. Manage data collection without time/date markers.

The data management process for reports to ASIPS involves up to ten steps, many of which are time-limited (Fig. 4). Any given event report may move through these steps in many

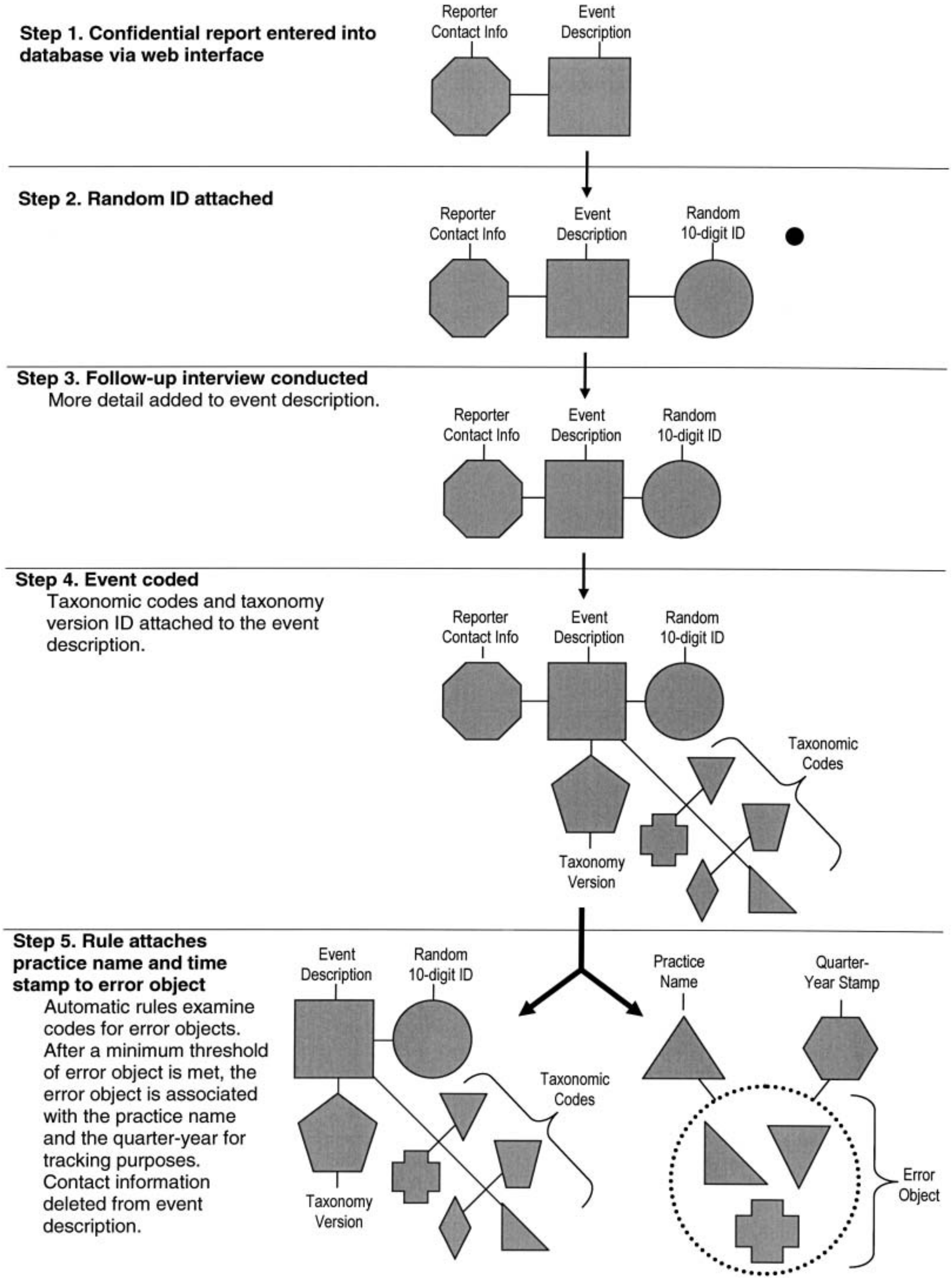
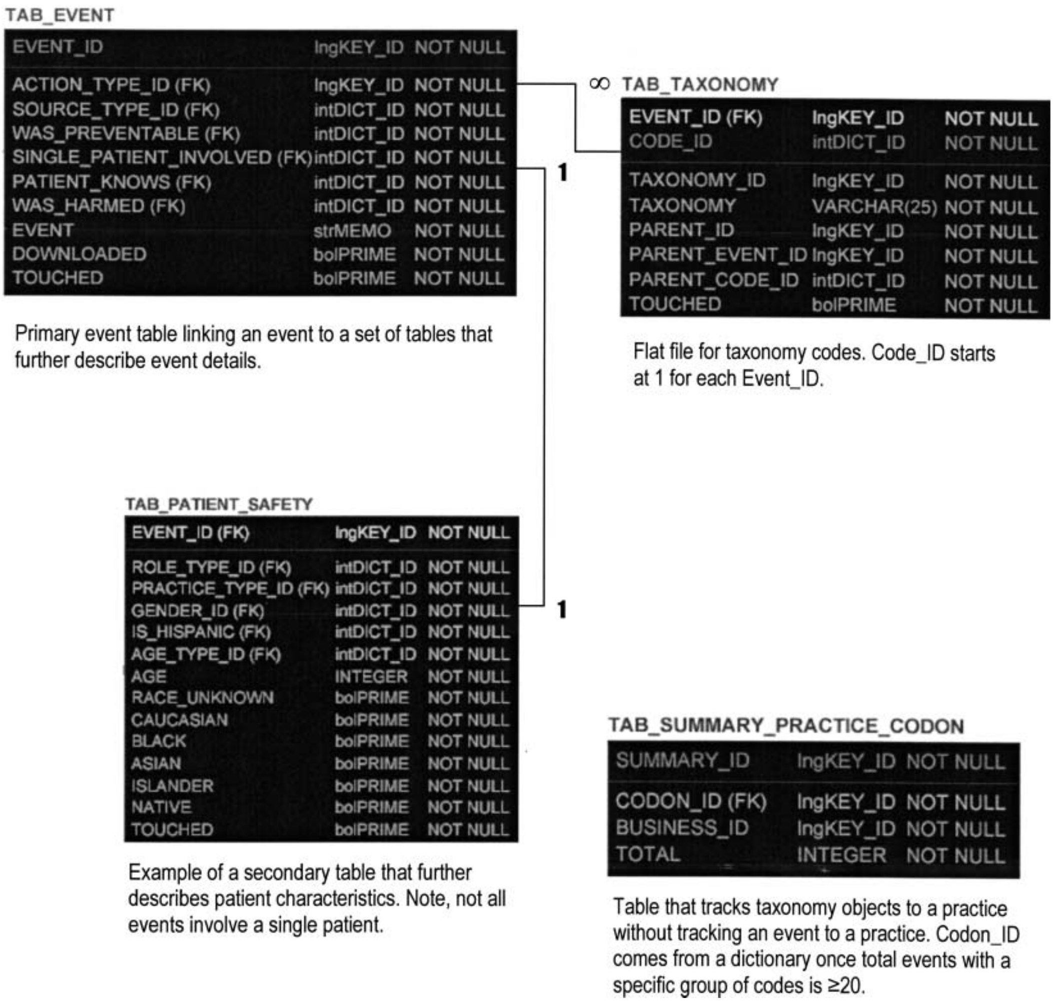

Step 1. Confidential report entered into database via web interface
Reporter Contact Info
Event Description
Step 2. Random ID attached
Reporter Contact Info
Event Description
Random 10-digit ID
Step 3. Follow-up interview conducted
More detail added to event description.
Reporter Contact Info
Event Description
Random 10-digit ID
Step 4. Event coded
Taxonomic codes and taxonomy version ID attached to the event description.
Reporter Contact Info
Event Description
Random 10-digit ID
Taxonomic Codes
Taxonomy Version
Step 5. Rule attaches practice name and time stamp to error object
Automatic rules examine codes for error objects. After a minimum threshold of error object is met, the error object is associated with the practice name and the quarter-year for tracking purposes. Contact information deleted from event description.
Event Description
Random 10-digit ID
Taxonomic Codes
Taxonomy Version
Practice Name
Quarter-Year Stamp
Error Object

Figure 2. Schematic of linking tables for error code objects to practices and time frames.

different pathways. To handle this complex process, we created a series of data flow dictionaries and flags. Each event report in the database has a pointer that indicates where it is in the data collection, cleaning, coding, and review process. When a time/date stamp is necessary, it is kept in a separate, or staging, database and deleted after a preset period. Confidential information will be deleted automatically after a specified period, but may be deleted earlier when all data are collected. While this tracking system creates a short-lived, relative time/date stamp, once the event is fully coded, no history of when or who handled the event is maintained.

3. The database must capture taxonomic codes for events, identifying the version of the coding taxonomy used for each event without creating a relative time/date sequence.

Each event report is coded for analysis using a medical error classification system modified for primary care. There currently is no recognized standard for medical error coding. In fact, there are few published taxonomies, and those that have been published are either specific to a small branch of medicine[24] or broadly based on industrial literature.[25] The Institute of Medicine is scheduled to complete a report on



TAB_EVENT

EVENT_ID	lngKEY_ID	NOT NULL
ACTION_TYPE_ID (FK)	lngKEY_ID	NOT NULL
SOURCE_TYPE_ID (FK)	intDICT_ID	NOT NULL
WAS_PREVENTABLE (FK)	intDICT_ID	NOT NULL
SINGLE_PATIENT_INVOLVED (FK)	intDICT_ID	NOT NULL
PATIENT_KNOWS (FK)	intDICT_ID	NOT NULL
WAS_HARMED (FK)	intDICT_ID	NOT NULL
EVENT	strMEMO	NOT NULL
DOWNLOADED	bolPRIME	NOT NULL
TOUCHED	bolPRIME	NOT NULL

Primary event table linking an event to a set of tables that further describe event details.

TAB_TAXONOMY

EVENT_ID (FK)	lngKEY_ID	NOT NULL
CODE_ID	intDICT_ID	NOT NULL
TAXONOMY_ID	lngKEY_ID	NOT NULL
TAXONOMY	VARCHAR(25)	NOT NULL
PARENT_ID	lngKEY_ID	NOT NULL
PARENT_EVENT_ID	lngKEY_ID	NOT NULL
PARENT_CODE_ID	intDICT_ID	NOT NULL
TOUCHED	bolPRIME	NOT NULL

Flat file for taxonomy codes. Code_ID starts at 1 for each Event_ID.

TAB_PATIENT_SAFETY

EVENT_ID (FK)	lngKEY_ID	NOT NULL
ROLE_TYPE_ID (FK)	intDICT_ID	NOT NULL
PRACTICE_TYPE_ID (FK)	intDICT_ID	NOT NULL
GENDER_ID (FK)	intDICT_ID	NOT NULL
IS_HISPANIC (FK)	intDICT_ID	NOT NULL
AGE_TYPE_ID (FK)	intDICT_ID	NOT NULL
AGE	INTEGER	NOT NULL
RACE_UNKNOWN	bolPRIME	NOT NULL
CAUCASIAN	bolPRIME	NOT NULL
BLACK	bolPRIME	NOT NULL
ASIAN	bolPRIME	NOT NULL
ISLANDER	bolPRIME	NOT NULL
NATIVE	bolPRIME	NOT NULL
TOUCHED	bolPRIME	NOT NULL

Example of a secondary table that further describes patient characteristics. Note, not all events involve a single patient.

TAB_SUMMARY_PRACTICE_CODON

SUMMARY_ID	lngKEY_ID	NOT NULL
CODON_ID (FK)	lngKEY_ID	NOT NULL
BUSINESS_ID	lngKEY_ID	NOT NULL
TOTAL	INTEGER	NOT NULL

Table that tracks taxonomy objects to a practice without tracking an event to a practice. Codon_ID comes from a dictionary once total events with a specific group of codes is ≥20.

Figure 3. Database structure for selected tables with the ASIPS database.

data standards for medical error taxonomies in late 2003, but this report is unlikely to include a model taxonomy.

We use a modification of the Victoroff taxonomy, which has been used by one of ASIPS's major partners, an insurance company, to code malpractice claims for three years. Given our desire to analyze PSRS data in conjunction with insurance data, we chose to adopt and modify the Victoroff taxonomy for our research. The Victoroff taxomony includes another set of codes that describe procedures and medications involved using other standardized systems. The taxonomy has embedded hooks for all major coding systems so that CPT, ICD, NDC, SNOMED, and even product serial numbers can be linked into the system depending on the user's preference. For ASIPS, we use the taxonomy's hooks for the International Classification of Primary Care (ICPC), as it is appropriately primary care-based.

Like many other error classification systems, the Victoroff/ASIPS taxonomy is a work in progress that is modified from time to time based on research notes captured at the time of coding and qualitative analysis of event textual data. Thus, in addition to capturing the error taxonomy codes and the embedded ICPC codes, the database/administrative system also needs to track the evolving taxonomy version used to code a particular event report. The system needs to record the following for each event: (a) the version of the taxonomy used for coding, (b) the actual codes assigned to the event, (c) an unpredictable set of parent–child relationships between taxonomy codes, including multiple children, and (d) the ICPC and the ICPC Drug classification codes related to the event.

To record these pieces of information, we created a single, flat table. Each row contains the binary primary key (random digit ID + repeated code ID) as described above, a taxonomy field, and a parent–child relationship field. The taxonomy field also holds the taxonomy version code and codes for ICPC and ICPC Drug, which are distinguishable by the parent taxonomy code. Version control was a significant problem: using *version 1*, *version 2*, and so on, creates a relative time stamp. We finally solved this problem by using another random number identifying each version with the current version signified with a descriptive prefix. A full set of all past taxonomy versions are sent to the analysis team so that each coded event can be matched to the correct version. With the exception of the currently used version, there is no way to determine the order of previous versions. The ASIPS project has used four versions of the Victoroff taxonomy and will

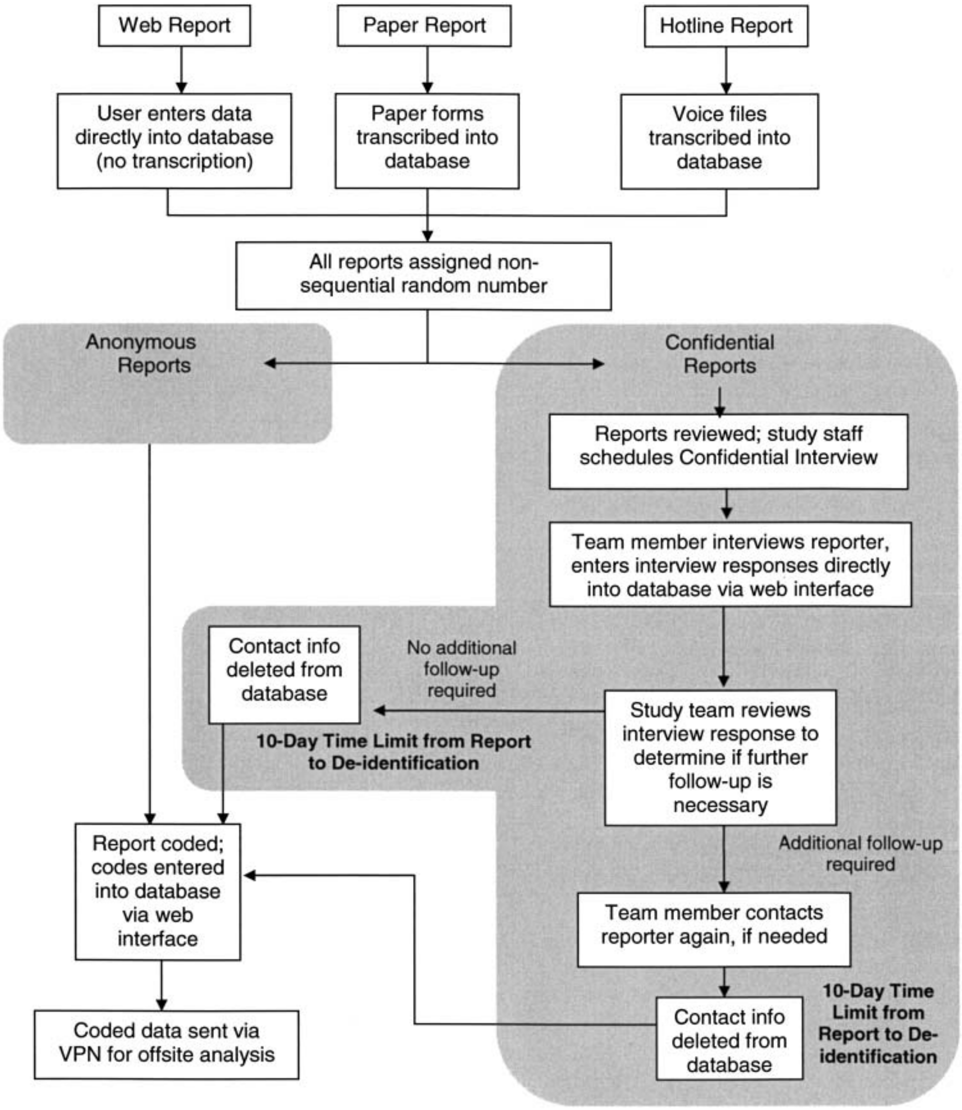

Web Report
Paper Report
Hotline Report
User enters data directly into database (no transcription)
Paper forms transcribed into database
Voice files transcribed into database
All reports assigned non-sequential random number
Anonymous Reports
Confidential Reports
Reports reviewed; study staff schedules Confidential Interview
Team member interviews reporter, enters interview responses directly into database via web interface
Contact info deleted from database
No additional follow-up required
Study team reviews interview response to determine if further follow-up is necessary
10-Day Time Limit from Report to De-identification
Report coded; codes entered into database via web interface
Additional follow-up required
Team member contacts reporter again, if needed
Coded data sent via VPN for offsite analysis
Contact info deleted from database
10-Day Time Limit from Report to De-identification

Figure 4. The data management process for reports to ASIPS involves many steps, most of which are time-limited. An event report may move through these steps in many different pathways.

continue to use our current version for the remainder of the project—15 months.

4. Allow updates to records over time, including updates to specific records already marked as complete and previously transferred for data analysis without using time/date stamps.

To solve this issue, we developed a series of flags. A flag is automatically set to track when a completely coded event report is transferred to the ASIPS analysts (who do not have access to the whole database for security reasons). The flagging system also flags reports that have been updated as part of internal quality control and that need to be transferred to the analysis staff again. The flags are reset each time an event has a code altered and each time it is transferred for analysis. Thus, the data analysis team can rapidly update their copy of the datasets using append and update queries without having to match thousands of codes by looking for additions, deletions, or updates.

5. The database must capture change of number and type of events reported over time without applying a time/date stamp to any specific event, *and*

6. The database must identify practices that have reported and have not reported specific types of events, without revealing which specific event they reported.

The use of confidential reports allows the ASIPS project to potentially understand which offices have reported selected types of events. This information is very helpful in identifying practices that may have solutions to a particular type of error (as indicated by their not reporting that type of error). While we would not assume that a lack of reporting equals a lack of errors, we would like to see if this is the case. Additionally, understanding which offices have not reported selected types of events may indicate practices that need education about recognizing a particular type of medical error. Similarly, because a major goal of ASIPS is to create interventions to reduce errors, it is important to understand the rough frequency of reports or the nature of reported events before and after interventions are implemented to help determine effectiveness. This type of tracking would be simple if we maintained time/date stamps. Thus, the desire to understand which practices have and have not reported a particular error conflicts with the de-identification process.

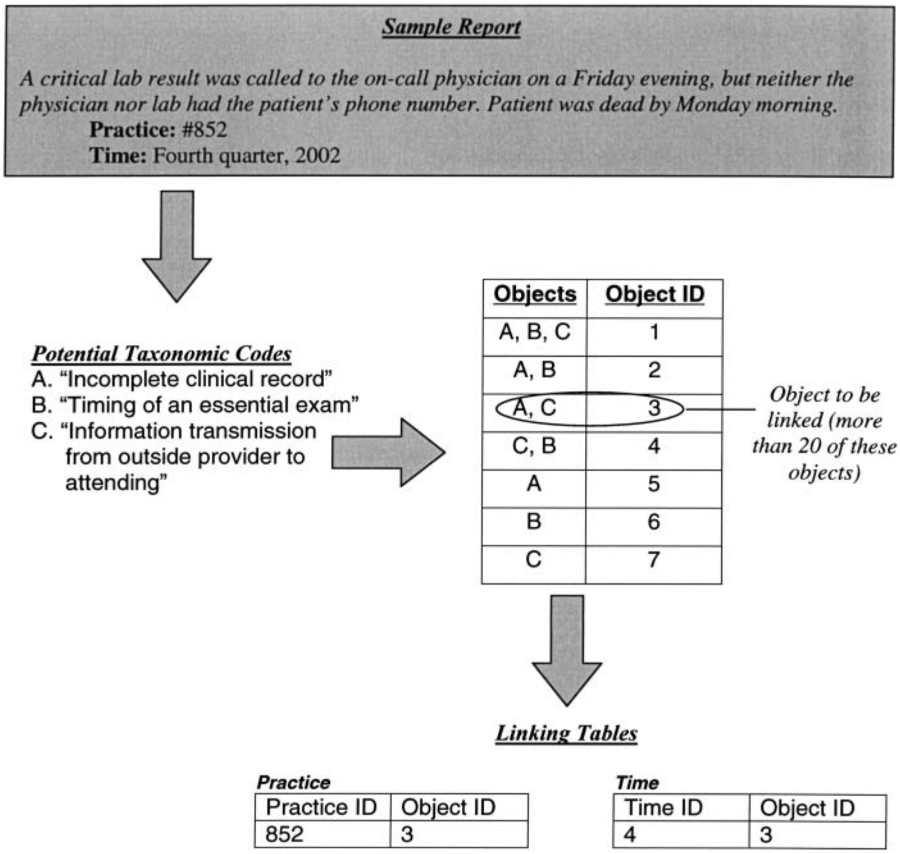
Design objectives 5 and 6 were solved in a similar fashion. The ASIPS solution to this problem is to begin tracking practices that report certain event types and the time frames of the reports after a threshold number of events of this nature have been reported. The error taxonomy codes that describe an event are best conceptualized as objects; that is, sets of codes that appear together. Thus, to track error types, the database must assemble and count each coded event as a *set of objects* (Fig. 5). Time frames are broken into large lumps of time, such as quarters of a year. Like objects are counted, and, once the threshold is reached, all future similar objects have a practice code and general time frame attached in different tables. (We elected to count reports after a threshold is reached because one is less likely to be able to match a known event to a single report among many similar types of reports recorded in our database.) The event taxonomy object and the practice code or time frame code are stored without any other identifying data as simple linking tables (Fig. 2 step 5 and Fig. 5). This process meets the location and time frame tracking criteria without linking any single event to a practice or time.

7. Data must not be viewable by others within the University of Colorado Health Sciences Center campus or within our own research enterprise.

The ASIPS data are stored on a dedicated, secure server with no access outside the institutional firewall and limited internal access. Each directory is further limited to a subset of users using standard Microsoft security tools. The ASIPS database is encrypted using a symmetric block cipher cryptography. Block ciphers are cryptographic algorithms that operate on 64-bit blocks of plain text. Some algorithms use fix-length keys; for others the key length may vary. We use RC2 block cipher encryption developed by RSA Data Security, Inc. RC2 is a variable-key-length block cipher; however, when using the Microsoft Base Cryptographic Provider, the key length is hard-coded to 40 bits. When using the Microsoft Enhanced Cryptographic Provider, the key length is 128 bits by default and can be in the range of 40 to 128 bits in 8-bit increments. A randomly generated 128-bit RC2 key is created and stored in the registry of the Web server. This key is used by Web pages to encrypt or decrypt data at the field level. One must be an authenticated Web user to be able to read the data in an unencrypted format.

8. Protect the database from possible data loss without maintaining copies of identifying information.

We addressed the tension between our need to protect data from catastrophic loss and our need to not maintain



Sample Report

A critical lab result was called to the on-call physician on a Friday evening, but neither the physician nor lab had the patient's phone number. Patient was dead by Monday morning.

Practice: #852
Time: Fourth quarter, 2002

Potential Taxonomic Codes

A. "Incomplete clinical record"
B. "Timing of an essential exam"
C. "Information transmission from outside provider to attending"

Objects	Object ID
A, B, C	1
A, B	2
A, C	3
C, B	4
A	5
B	6
C	7

Object to be linked (more than 20 of these objects)

Linking Tables

Practice

Practice ID	Object ID
852	3

Time

Time ID	Object ID
4	3

Figure 5. Simplified representation of error code objects from a coded event.

copies of identifying information by backing up only the nonidentifiable data. The routine processes of data protection—such as nightly, weekly, and monthly backups; hard drive imaging; and off site data storage—can lead to keeping extra, unsecured copies of identifying confidential information. For example, reporters' identities must be stored for ten days to allow us to follow-up on medical error reports. We bypass usual data protection for this staging database: we perform a nightly image, only, of the ASIPS confidential contact information database, overwriting the image file each night. Thus, temporary contact information is more vulnerable to data loss, but information intended for deletion is not stored on several different tapes with varying shelf lives. While we created a second database for contact information that is linked and imaged but not written to a tape backup, an alternative solution could be setting up the database so that selected tables within a database are not written to the tape backup system. If the issue of backups is not considered and discussed with both the database developers and the network support staff, efforts to de-identify information can be negated through any one of the (usually desirable) data protection mechanisms discussed above.

The complete deletion of confidential contact information is not the only concern posed by routine data protection activities. The creation of a series of backup copies of the database creates another set of time/date stamps. By addressing each copy of the database in sequence it would be easy for someone outside the ASIPS project, if given access to our system through court order, to determine when a particular event was recorded into the database, thus, establishing a close proximity to the event occurrence date. To deal with this proximal time stamp, the ASIPS project has adopted a limited data backup protocol. An image of the ASIPS database is encrypted a second time (described above) and imaged nightly. This image is overwritten each night. Weekly database backup files are kept for three weeks, and monthly we ship a copy of the database off site. The weekly backups are overwritten each month. Thus, only a very limited time sequence can be re-established from our backup copies. This two-month time frame is likely to pass before any legal inquires for a particular event are filed. Thus, an individual searching tape backups to establish a date of an event is likely to find the event present in all backup tapes. This system degrades our ability to recover lost data, but we believe it is a useful trade-off for increased security.

9. The whole process must not create a paper trail for any step in the process except the initial event report for those individuals choosing the paper reporting method.

Moving the entire process to an electronic system solved this issue. With the exception of the initial paper reports, which are shredded immediately after data entry, there are no paper copies of reports to be accidentally left lying around, none to be legally discovered, and none to be lost.

Status Report

The ASIPS reporting system and administration database has handled more than 500 reports of medical errors, resulting in over 8,000 recorded taxonomy codes. The day-to-day management of the data collection, review, and coding runs smoothly. The Web management tool handles the complex tasks of communicating work assignments, directing data collection, and tracking each report through the process. Project staff members receive from the system their work assignments for confidential report follow-up, including specific interview questions developed from initial review of the report. Staff members often are not present during the assignment of follow-up activities, but cases assigned to them show up under their work list based on login. Follow-up interviews for confidential reports are directed by a browser-based set of data collection screens that guide data collection and provide for direct data entry.

To assess our success with de-identifying reports in our database, we sent a copy of the empty database to two data experts to determine if they could create a relative order of events or link an event to a practice. They both stated it was not possible.

The ASIPS project has received no requests for release of data to third parties. Even if the ASIPS database proves immune to discovery, it remains unknown whether the act of sharing specific patient safety data makes an organization legally vulnerable to having its internal quality improvement data subpoenaed. Our attorneys believe that a claimant's legal team would have to prove that ASIPS received a report about a *particular* event. It is most improbable that our participants report *every* medical error they know of, so it would be very difficult to make the case that we definitely received a report about a particular event. If one could make that case, our participating organizations would lose their quality improvement data protection. Because it is nearly impossible to make that case, their data should be safe from legal discovery.

Nonetheless, the issue of confidentiality remains one of significant concern to some participating organizations. Despite the efforts described above to safeguard report information, one organization involved in ASIPS will only permit their employees to file anonymous reports. In a recent survey of clinicians and staff participating in the ASIPS project, 20% identified safety of the reports as a concern.

Discussion

The collection, follow-up, coding, and analysis of medical errors are important steps in designing improved medical care systems.[26] Studying errors across multiple organizations yields more generalizable findings concerning the types, frequency, and consequences of errors, as well as the effect of a given intervention. In most states, activities that cross institutional boundaries cannot use peer review legislation to protect the collected information. The federal legislation designed to protect research data collected in federally funded studies provides unproven protection from disclosure. Thus, it is imperative that database designers direct their attention to features that support or hinder the true de-identification of reported medical errors.

Developing databases that track events and changes in events over time without revealing a time sequence for a given event is atypical but important in the current medical-legal climate of the United States. Legal opinions vary, but it is possible the current safety efforts of national organizations could lead to unwanted disclosure of medical errors. The reporting of sentinel events to JCAHO, drug safety events to the U.S. Pharmacopeia, and the emerging efforts of the National Quality Initiative to collect multi-institutional data all have the potential to lead to undesired disclosure of information either from the reporting system or from court-ordered disclosure of internal institutional investigations. Unwanted disclosure of information purportedly held in confidence could have a chilling effect on the use of reporting systems, with effects reaching beyond medical event reporting systems.

Selected states, for example, California and Oklahoma, have expanded peer review legislation such that multi-institutional reporting systems can be included within peer review.[1] Colorado recently passed similar legislation, although initial legal opinions indicate the ASIPS project would still not be protected under the revised statute. Whether research can be incorporated into peer review–protected systems also is debatable. Until federal legislation specifically protecting patient safety reporting systems is passed and tested in the courts, we believe it is important for developers and operators of such systems to pay attention to the issues discussed in this report.

The role of patient safety reporting systems is yet to be fully delineated. In general, reporting systems are known to collect only a small percentage of all medical errors. Thus, safety reporting systems may not be sensitive to changes in actual rates of events. Anonymous systems have limitations for root cause analysis and best practices research that the ASIPS system overcomes. The goals of the ASIPS system—including tracking types of events to particular practices (institutions) and general time frames—are unlikely to be unique to the ASIPS project. We have presented an approach to solving the issue of de-identifying reports that may apply to other error reporting systems. Here we have described selected problems with the use of standard relational database designs and have highlighted features of the ASIPS system that overcome some of these problems in the hopes of stimulating a dialogue that will further work in the area.

References ■

1. Kohn LT, Corrigan J, Donaldson MS (eds). Institute of Medicine. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press, 2000.
2. IOM Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.
3. Joint Commission on Accreditation of Healthcare Organizations. 2002–2003 Comprehensive Accreditation Manual for Ambulatory Care. Oakbrook Terrace, IL: JCAHO, 2003.
4. Quality Interagency Coordination Task Force. Report to the President. Doing What Counts for Patient Safety: Federal Actions to Reduce Medical Errors and Their Impact. Washington, DC: QICTF, 2000.
5. Dovey SM, Meyers DS, Phillips RL Jr, et al. A preliminary taxonomy of medical errors in family practice. Qual Safety Health Care. 2002;11:233–8.
6. Connell L. Administrator of the NASA-run Federal Aviation Administration's Aviation Safety Reporting System. Personal communication, 2003.
7. Makeham MA, Dovey SM, County M, Kidd MR. An international taxonomy for errors in general practice: a pilot study. Med J Aust. 2002;177(2):68–72.
8. Bhasale AL, Miller GC, Reid SE, Britt HC. Analysing potential harm in Australian general practice: an incident-monitoring study. Med J Aust. 1998;169(2):73–6.
9. Quality management functions—confidentiality and immunity. Title 25, Article 3, Number 109. Colorado Revised Statutes: Title 25, Article 3, Number 109.
10. Anonymous. The Public Health Service Act. 42. U.S.C.: 42:299c-3(c).
11. Merewitz SG. Statutory Confidentiality Protection of Research Data. Memorandum from Susan Greene Merewitz, Senior Attorney for the Agency for Healthcare Research and Quality. <http://www.ahrq.gov/fund/datamemo.htm>. Accessed Oct 25, 2002.
12. Office of Management and Budget. Circular A-110. 1993. <http://www.whitehouse.gov/omb/circulars/a110/a110.html>. Accessed Oct 25, 2002.
13. NASA. Aviation Safety Reporting System Web Site. <http://asrs.arc.nasa.gov>. Accessed Oct 25, 2002.
14. Reynard WD, Billings CE, Cheaney ES, Hardy R. The development of the NASA Aviation Safety Reporting System. Reference Publication 1114 ed. NASA Scientific and Technical Information Branch, 1986.
15. Anonymous. Federal Aviation Regulations (FAR). Title 14 Code of Federal Regulations: 91.25.
16. Behlen FM, Johnson SB. Multicenter patient records research: security policies and tools. J Am Med Inform Assoc. 1999;6: 435–43.
17. Anonymous. Health Insurance Portability and Accountability Act of 1996. 104–191. Public Law: 104-191.
18. Ohrn A, Ohno-Machado L. Using Boolean reasoning to anonymize databases. Artif Intel Med. 1999;15:235–54.
19. Herting RL Jr, Barnes MR. Large scale database scrubbing using object oriented software components. Proc AMIA Symp. 1998:508–12.
20. Biskup J, Bleumer G. Cryptographic protection of health information: cost and benefit. Int J Biomed Comput. 1996;43(1-2): 61–7.
21. Bowen JW, Klimczak JC, Ruiz M, Barnes M. Design of access control methods for protecting the confidentiality of patient information in networked systems. Proc AMIA Symp. 1997: 46–50.
22. Carroll ET, Wright S, Zakoworotny C. Securely implementing remote access within health information management. J AHIMA. 1998;69(3):46–50; quiz 51-2.
23. Conner VW. Patient confidentiality in the electronic age. J Intraven Nurs. 1999;22:199–202.
24. Trustees of Columbia University in the City of New York. Medical Event Reporting System (MERS). Available at: http://www.mers-tm.net. Accessed Jul 2, 2003.
25. Van Vuuren W, Shea CE, van der Shaaf TW. Development of an Incident Analysis Tool for the Medical Field. EUT Report. Eindhoven, The Netherlands: Eindhoven University of Technology, 1997.
26. Leape LL, Lawthers AG, Brennan TA, Johnson WG. Preventing medical injury. QRB Qual Rev Bull. 1993;19:144–9.