Patient Data Privacy in

Electronic Records

6.872/HST950

Lecture #9

Harvard-MIT Division of Health Sciences and Technology
HST.950J: Medical Computing
Peter Szolovits, PHD

Protecting

- What?
  - Privacy
    - Individual's desire to limit disclosure of personal information
  - Confidentiality
    - Information sharing in a controlled manner
  - Security
    - Protecting information against accident, disaster, theft, alteration, sabotage, denial of service, …
- Against what?
  - "Evil hackers"
  - Malicious insiders
  - Stupidity
  - Information Warfare

Privacy

- Right to be let alone; e.g.:
  - snooping on Dan Quayle by J. Rothfeder
  - outing of Arthur Ashe (HIV), Henry Hyde (adultery)
  - celebrity medical problems (Tammy Wynette, Nicole Simpson)
  - applies mostly to known individuals

Privacy in obscurity

- Right to remain unknown
- Correlation among pervasive databases:
  - census
  - marketing
  - health

Confidentiality

- Use and sharing of information by multiple users at many institutions
- Should be controlled by coherent policy
- Enforced by appropriate technology
- E.g., who may use results of your life insurance physical exam, for what purposes?

National Academy of Sciences Study, 1997

Charge to the committee:

- Observe and assess technical and non-technical mechanisms for protecting privacy and maintaining security in health care information systems.

- Identify other methods worthy of testing in health care settings.

- Outline promising areas for further research.

Committee Members

Paul Clayton, *Chair,*
Columbia Presbyterian Medical Center
Earl Boebert,
Sandia National Laboratories
Gordon DeFriese,
Sheps Ctr. for Health Services Research
Susan Dowell,
Medicus Systems Corp.
Mary Fennell,
Brown University
Kathleen Frawley,
AHIMA
John Glaser
Partners Healthcare System
Richard Kemmerer
Univ. of Calif., Santa Barbara

Carl Landwehr
U.S. Naval Research Laboratory
Thomas Rindfleisch
Stanford University
Sheila Ryan
Univ. of Rochester, School of Nursing
Bruce Sams,
Kaiser Permanente (retired)
Peter Szolovits
MIT
Robbie Trussell
Presbyterian Healthcare System, Dallas
Elizabeth Ward
Washington State Dept. of Health
Paul Schwartz (*Special Advisor*),
Univ. of Arkansas School of Law

Site Visits

Institutions Visited

- Large, urban hospital
- Integrated delivery system
- Affiliated health care system
- Community Health Info Network (CHIN)
- State health system
- Insurer

Issues Discussed

- Problems encountered
- Security and confidentiality policies
- Security mechanisms
- Effectiveness of mechanisms
- Education and training
- Disciplinary sanctions
- Needs to promote better security

Trade-offs among IT characteristics

- Critical to *improve the quality* and *reduce the costs* of health care.
- Privacy and security must be resolved if patients are to share sensitive health information with care providers.
- Protect patient privacy while ensuring that providers have legitimate access to information for purposes of care.

Privacy and Security Concerns Addressed in the Report

- Inappropriate releases of information from individual organizations
    - authorized users leaking information
    - unauthorized users breaking into systems to retrieve or alter information, or to render systems dysfunctional
- Systemic flows of information among organizations in health care and related industries

Health Information Held by Individual Organizations Can Be Protected

- **Technical practice:** A variety of practices provide effective protection in an operational environment and can be implemented with reasonable effort.

- **Policy and implementation:** Technical mechanisms must be accompanied by organizational mechanisms for developing access and release policies, training workers, and penalizing violations of policy.

- **Incentives:** Health care organizations need proper set of incentives to address privacy and security concerns.

Two Approaches to Protect Privacy

- Pre-emptive controls
  - Lock & key
  - Need to know often need pre-specified understanding of who needs what under which circumstances -- *military model*
- Retroactive controls
  - Community of trust
  - Checking up, not prevention
  - Sanctions

Threat Model

Must understand what you are protecting against:

- Nature: confidentiality, security
- Source: insider, outsider
- Means: tourist, cracker, NSA
- Information at risk
- Scale

Credible threats:

- *accidental disclosures by insiders*
- *abuse of record access privileges by insiders*
- *insider access for profit or spite*
- *unauthorized physical intruder*
- *vengeful outsider who seeks to access, damage, disrupt*

Recommended Technical Practices for Immediate Implementation

- Individual Authentication such as login IDs and passwords to ensure accountability
- Access Controls restrict access to need-to-know
- Audit Trails track all accesses to clinical information
- Protection of remote access points
- Software discipline limit ability to download, install, or copy software
- System assessment evaluate vulnerabilities
- Physical Security & Disaster Recovery

Authentication and Access

Eliminate undesirable (horrendous) current practices, e.g.,

- all doctors log in as MD
- nurses, receptionists use doctor’s account
- four-digit (or six-digit) id+password
- *all* data available to everyone
- no record of who creates, alters or destroys data
- poorly-controlled access from networks, remote sites

System and Software Discipline

- Standard workstations
    - hardware
    - approved software

- Control over networking

- Control over software installation/dissemination
    - viruses
    - network downloads
    - floppy drives

- Testing of security features

Physical Security

- Lock the computer room (wherever it may be!)

- Backups, recovery procedures
    - protect the backup data
    - test the recovery procedure

- Erase the disk when de-commissioning the computer

Recommended Organizational Practices for Immediate Implementation

- Security and confidentiality policies
- Security and confidentiality committees
- Information security officers
- Education and training programs
- Sanctions
- Improved authorization forms
- Patient access to audit logs

Policies and Governance

- Clearly stated policy:
  - Responsibility
  - Education
  - Data access
  - Guardianship
  - Associating people with their actions (identification, capabilities, temporary access, termination)
  - Enforcement
  - Testing
  - Transparency

- Governance:
- Policy-making body
- Security officer
- Buy-in
  - CIO
  - Human Resources
  - Entire community
- Education

Enforcement

- Auditing
  - Periodic sampling of access logs
  - Users ability to check
- Human Resources (Personnel)
  - Emphasize importance
  - Explicit criterion of evaluation
  - Education and training
  - Reprimand, termination for all levels of employees

Testing

sine qua non

- Monitoring and awareness
- Review of performance
- Auditing
- Tiger teams
- Published results

Recommended Security Practices for Future Implementation

- **Strong authentication**:
  - single-session passwords,
  - encrypted authentication sessions,
  - token-based authentication
- **Enterprise-wide authentication** (single logon)
- **Access validation** to ensure that retrieved information matches user's access privileges
- **Expanded audit trails**
  - alll internal accesses to information
  - global audit trails to trace secondary distribution of data
- **Electronic authentication of records**

Stronger Incentives Needed

- Strong incentives to use IT, but fewer incentives to address privacy and security issues.
  - Existing legislation is inconsistent across states; no strong federal legislation mandating protections [in 1997]
  - Sporadic violations of privacy and security have not rallied broad public interest.
- Little guidance for improving privacy and security
  - no effective standards to guide attempts to better protect health information.
  - few means of sharing information about privacy and security violations, effective ways of protecting health information

Recommended Elements of Industry Infrastructure for Privacy & Security

- Standing committee for developing and updating privacyand security standards.
  - examine security mechanisms and help establish rules governing data flows.
  - reports directly to Secretary of HHS
- Organization for gathering and sharing information about security threats, incidents, and solutions in health care.
  - similar to the computer emergency response team (CERT) for the Internet
  - seed funding from Congress

Systemic Concerns Regarding Privacy and Security

- Many concerns regarding patient privacy stem from *sharing of information* among organizations in health care industry.
- Existing data flows are largely *unregulated* and often occur *without patient consent* or knowledge.
- Possible development of a *universal patient identifier* could exacerbate such concerns.

Proposed Means of Addressing Systemic Concerns

Encourage national debate to determine appropriate balance between patient privacy and organizational needs for information

- Fair information practices (e.g., federal Privacy Act of 1974)
- DHHS should establish program to promote consumer awareness of issues and uses of health information.
- Professional societies should educate members about privacy and security issues
- DHHS should conduct studies to determine extent to which various users need patient identifiable health information
- DHHS should work with the U.S. Office of Consumer Affairs to determine way to give consumers a visible, centralized point of contact

Fair Information Practices (Federal Privacy Act, 1974)

- No secret databases that include personally identified information
  - Agencies must publish policies on all databases
- Right to see my information, with ability to correct
- Prevent data collected for one purpose from being used for another
- Agency responsible for reliability and security of data
- Right to sue re: privacy issues (such as an ombudsman).

Recommendation on Patient Identifiers

Any method used to identify patients or link patient records should:

1. be accompanied by a policy framework that identifies the kinds of linkages that violate patient privacy and that specifies legal sanctions.

2. facilitate identification of parties that link records.

3. allow unidirectional linking of information: it should facilitate linking of records based on information given by patient (such as an identifier), but prevent a patient's identity from being easily deduced from records or the identifying scheme itself.

Recommendation for Meeting Future Technological Needs

➢ establish formal *liaisons* with industry and government security working groups.

➢ support *research* in areas of particular importance to health care, but that might not be otherwise pursued.

➢ fund experimental *testbeds* to explore different means of controlling access in an operational environment.

Recommendation for Meeting Future Technological Needs

- establish formal *liaisons* with industry and government security working groups.

- support *research* in areas of particular importance to health care, but that might not be otherwise pursued.

- fund experimental *testbeds* to explore different means of controlling access in an operational environment.

Future Security Technologies of Particular Interest to Health Care

- Methods of *identifying and linking* patient records that protect patient privacy.

- Technologies for enabling patients to receive health care *anonymously*:ii pseudonyms, cryptographically generated aliases, narrative templates, smart cards.

- *Audit tools* that allow more frequent examination of audit logs to detect inappropriate accesses to information.

- Tools for rights enforcement and management to control *secondary distribution* of data

HIPAA Regulations on Individually Identifiable Health Information

Based on 45 CFR parts 160 & 164 Federal Register Vol. 65, No. 250, pp. 82462-82829, Dec. 28, 2000

Why?

- Part of Administrative Simplification section of HIPAA (Health Insurance Portability and Accountability Act of 1996 --Kennedy/Kassebaum Bill)
- 1/5 of Americans believe personal health information (PHI) has been used inappropriately
- PHI use necessary for improved quality, reduced cost
- existing protections fragmented

History of Privacy Provisions

- Congress gave itself until Aug 21, 1999 to enact legislation -- it did not do so
- Backup was that Secretary of HHS was to promulgate rules by Feb 21, 2000 -- this was extended because of 70,000 comments
- Rule promulgated Dec. 2000
- Bush administration has put it on hold, mainly because of cost complaints
- Sec. Thompson agreed to issue the rule, Apr. 2001
- Congress may legislate later, based on experience
- *work in progress*

Other simplification issues

- Standards for electronic health care transactions, including detailed data elements
    - unique health identifiers
    - providers
    - patients
    - code sets
    - security standards
    - electronic signatures
    - transfer of information among health plans
- Target date: Feb 21, 1998

Sanctions

- Civil penalties for violations of standards: $100/person/violation, max $25,000/violation/year
- Knowing violations of health identifier or deliberate disclosure:
  - $50,000 + 1 year jail
  - $100,000 + 5 years jail if under false pretenses
  - $250,000 + 10 years jail if with intent to sell, transfer or use, for commercial advantage, personal gain, or malicious harm

Principles

- Allow smooth flow of PHI for treatment, payment, related operations, public interest
- Prohibit flow of PHI for other purposes, without consent of subject
- Fair information practices
  - Allow subject to access PHI (*later, excludes psych notes*)
  - Allow subject to have records amended for errors or incompleteness
  - Allow subject to know who else uses PHI
- Require persons who hold PHI to safeguard it
  - accountable for own use and disclosure
  - legal recourse
- Minimal Necessary Use and Disclosure
  - Few limits on use for treatment, more for other functions

Limitations of HIPAA

- Responsibilities cannot follow data; therefore
- Recommendation applies to
  - Health Plans
  - Health Care Clearinghouses
  - Provides who transmit PHI electronically
- Does not apply to others who hold/process data
  - contractors, third-party administrators, researchers, public health officials, life insurance issuers, employers, marketing firms, …
- …but: Covered Entities required to contract with business associates to pass on responsibilities, along with identifies health data used "in behalf of" a covered entity
- Does not apply to paper records
  - …but: If the information was ever in electronic form, reponsibility is "sticky"
- No private right of action

Consent (before HIPAA)

- Most patients believe their private medical data may *not* be divulged without specific consent

- But, consent may effectively be forced

- But, many exemptions exist:

  - For treatment *and related purposes* (e.g, utilization review)
  - For obtaining payment
  - Emergency care, health depts., law enforcement, coroners, business operations, oversight, research

When is a nod a nod?

- *Agreement:* informal, perhaps implied, e.g., to let a consultant see clinical notes, let hospital include patient in a directory

- *Consent:* written, but often generic, e.g., on admission to hospital. This covers most health care operations

- *Authorization:* written, specific to the case. For psychiatric notes and all data uses other than health care operations. E.g., research.

- Patient may negotiate *Restrictions* on disclosure, e.g., to particular staff, family members, etc.

Uses of data by Covered Entities

- For treatment, payment, health care operations *without patient authorization*

- For public health, research, health oversight, law enforcement, use by coroners, mandatory State reporting, search warrants *without patient authorization*

- Must allow access to the subject of the records

- Must get individual consent for any other uses

Substitute regulatory protections for pro forma authorizations often used today.

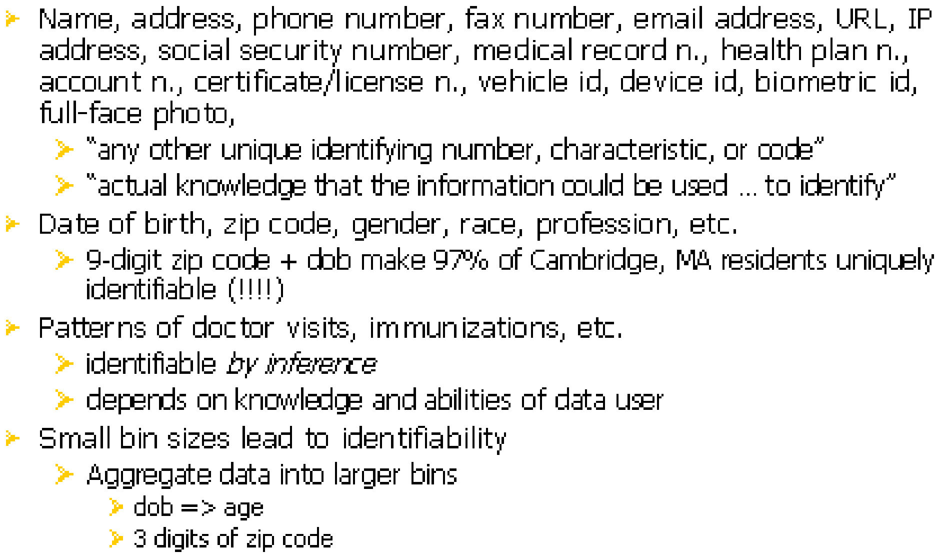
Health Care Operations

- Treatment
- Payment
- Quality assessment and improvement activities
- Review competence of professionals, organizations; conduct training; accreditation
- Insurance rating concerning existing coverage
- Auditing
- Legal proceedings
- *Added:* Business planning and development, management, general administration, fundraising, internal marketing

NOT Health Care Operations

- Marketing
- Sale, rent or barter of information
- Use in parts of organization not health-related
- Rate setting prior to subject's enrollment
- Employment determinations
- Fund raising
- Research to obtain generalizable knowledge

Identifiable



- Name, address, phone number, fax number, email address, URL, IP address, social security number, medical record n., health plan n., account n., certificate/license n., vehicle id, device id, biometric id, full-face photo,
  - "any other unique identifying number, characteristic, or code"
  - "actual knowledge that the information could be used … to identify"
- Date of birth, zip code, gender, race, profession, etc.
  - 9-digit zip code + dob make 97% of Cambridge, MA residents uniquely identifiable (!!!!)
- Patterns of doctor visits, immunizations, etc.
  - identifiable *by inference*
  - depends on knowledge and abilities of data user
- Small bin sizes lead to identifiability
  - Aggregate data into larger bins
    - dob => age
    - 3 digits of zip code

Sweeney's Cambridge

- 1997 Cambridge, MA voting list on 54,805 voters
- Name, address, ZIP, birth date, gender, iú
- Combinations that uniquely identify:
  - Birth date (mm/dd/yy)ii 12%
  - BD + genderiiii 29%
  - BD + 5-digit ZIPiiii 69%
  - BD + 9-digit ZIPiiii 97%
- Unique individuals
  - Kid in a retirement community
  - Black woman resident in Provincetown

Problem of other information

- Governor Weld's data found in Mass de-identified dataset

- Dates you visited a health care provider (over a lifetime) are probably unique

- Can be used to re-identify you if someone has both de-identified data and other data that link to identifiers

Danger of Re-identification

Protection via generalization

Computational Disclosure Control

- Make sure data cannot be traced back to a set of size <*n*
  - Generalization
  - Suppression of unique combinations
  - Account for leakage from what *has* been suppressed; e.g., back-calculating from aggregate statistics
- How to estimate external information?
- **Every** release becomes more external info.

Methods of Generalization/Suppression

- Underlying problem (find minimal generalization/suppression to achieve a level of anonymity) is NP-hard (Vinterbo)
- Mainly heuristic search over space of possible generalizations/suppressions
  - Scrub
  - Datafly
  - μ-Argus (Netherlands)
  - *k*-Similar

Sources

- *For the Record: Protecting Electronic Health Information*, National Academy Press, 1997 (http://www.nap.edu/readingroom/books/for/)
- Universal Health Identifiers:
  P. Szolovits and I. Kohane, "Against Simple Universal Health-care Identifiers," *J Am Med Informatics Assoc*, vol. 1, pp. 316-319, 1994.
- Confidentiality policy:
  D. M. Rind, I. S. Kohane, P. Szolovits, C. Safran, H. C. Chueh, and G. O. Barnett, "Maintaining the Confidentiality of Medical Records Shared over the Internet and World Wide Web," *Annals of Internal Medicine*, vol. in press, 1997. 1997(127): 138-141.
  (http://www.acponline.org/journals/annals/15jul97/mronnet.htm)
- Web implementation:
  J. Halamka, P. Szolovits, D. Rind, and C. Safran, "A WWW implementation of national recommendations for protecting electronic health information," *J Am Med Informatics Assoc*, vol. 4, pp. 458-464, 1997.
- HIPAA:
  http://aspe.hhs.gov/admnsimp/final/PvcPre01.htm