Big Brother?
Quality and Privacy Issues in Health Care Data Mining

SINI: July 2008
Reading and self-tutoring links at http://onlineinformatics.com
Data mining is now a fact of life with broad application that includes diverse domains of weather prediction, fraud detection, cancer survival prediction, and even analyzing your spending patterns at the grocery store! What can we do to assure data quality and protect patient privacy when we mine clinical databases and use those results to make decisions? This presentation will explore the answers. (expert)
Pre-Reqs for this Talk

- Not designed for informatics novices.
- Also not designed for data mining experts!
- Easier to understand if some knowledge (not covered):
  - **Basic statistics**
    - variance, mean, Bayesian frequency, cluster, regression, association, prediction, outlier, model development and testing
  - **Databases**
    - fields, records, constraints, data structures, relational tables,
  - **Informatics**
    - integration, interoperability, data integrity, computer security, heuristics, human factors,
- Everyone is welcome to stay!
Patients increasingly lack trust

Systemic problem confirmed

Use vs Disclosure (PHI)

Secondary data uses

Experiments

Advocacy

Technology risks and solutions

Defining data quality

Data quality methods

Data quality issues

Data quality and patient care quality

Patients increasingly lack trust

Systemic problem confirmed

Use vs Disclosure (PHI)

Secondary data uses

Experiments

Advocacy

Technology risks and solutions

SINI: July 2008
Big Brother: Data Mining Quality and Privacy Issues

Http://onlineinformatics.com

linda.goodwin@duke.edu - copyright 2008 - re-use with written permission
Big Brother: Data Mining Quality and Privacy Issues

Opportunities
- Automated exploration and discovery
- Prediction and classification
- Explanation and description

Value Chain: Motivation

Steps in the process

Domain expertise

Data quality
- Patient privacy

Health care data mining issues

Others

Model development and testing

Technical and mathematical issues

Interpretations
- Interesting and/or meaningful

Application of Results

Optimization
- Overfitting
- Validation

Relevance

Interesting and/or meaningful

Health care data mining issues

Patient privacy

Others

SINI: July 2008
Big Brother: Data Mining Quality and Privacy Issues

Http://onlineinformatics.com
Dogbert Consults

You need to do data mining to uncover hidden sales trends.

If you mine the data hard enough, you can also find messages from God.

...Sales to left-handed squirrels are up... and God says your tie doesn't go with that shirt.
Data Mining Opportunities

• Automated exploration and discovery
  – Lead levels of children in Uruguay
  – TB vaccines based on genomic data

• Prediction and classification
  – Prediction of liver fibrosis in pts with chronic hepB
    • Hepatogastroenterology. 2008 Jan-Feb;55(81):preceding table of contents.
  – Diabetes (comparison of predictive models)
    • Diabetes Care. 2008 May 28
  – Pharmacogenomics of lipid metabolism

• Explanation and description
  – Developing a brain “atlas”
  – DNA polymorphisms in post-trauma spinal cord injury
Data Mining Opportunities -

• Automated exploration and discovery
  – CA SB1096 (would allow pharmacies to sell customers’ prescription data)
  – Global gene mining (pharma industry)
    • Toxicol Appl Pharmacol. 2005 Sep 1;207(2 Suppl):679-83

• Prediction and classification
  – Pharmacy and/or prescription practices
    • Mod Healthc. 2007 May 14;37(20):32.

• Explanation and description
  – Insurance claim “frequent flyers”
    • Stud Health Technol Inform. 2006;122:903.Click here to read
  – Profits before people (stress)
    • J Epidemiol Community Health. 2007 Oct;61(10):862-70.Click here to read
Data Mining: Motives

- Abundance of data - data explosion problem!
  - Naisbett: We are drowning in information, but starving for knowledge!
  - Ozbolt's "data cemeteries"
- Data mining extracts knowledge from data

"Well, your teacher is wrong. We are not living in the Information Age. We are living in the Misinformation Age."
Data Mining: Motives

• Available tools
  – Affordable hardware
  – Increasingly usable software
    • Machine learning and AI
    • Statistics
  – Database management systems and warehouses

• Knowledge management
  – Evidence-based practice
  – Outcomes analysis

• Health care competition
  – Cost data
  – Safety data
  – Quality data
Health Care Data Mining Motives

- Improved Care
  - Diagnosis
  - Treatment
  - Surveillance and early detection
  - Outcomes
- Improved Service
  - Quality improvement
  - Safety monitoring
  - Benchmarks and scorecards
- Profit$
The Value Chain

Data
- Patient data
- Staff data
- Cost data
- Safety Data
- Quality data

Information
- Bob is diabetic with A1c of 8.4
- 3 nurses’ licenses will lapse next week
- Dr Jones performed 4 TAH last month
  - Those 4 patients average LOS = 1.2 days

Knowledge
- Diabetic A1c below 7.0
- Ave LOS for TAH is 2.0 days
- Nurses cannot provide patient care without active license

Decision
- Change Bob meds and educate about diet and exercise
- Analysis of Dr Jones techniques to share with other surgeons
- Follow up with nurses for renewal
Applications, operations, techniques

- Cross selling
- Quality control
- Customer retention
- Fraud prevention

- Predictive modeling
- Link analysis
- Database segmentation
- Deviation detection

- Classification
- Clustering
- Associations
- Value prediction
Evolution

- **1960s:**
  - Data collection, database creation, network DBMS

- **1970s:**
  - Relational data model, relational DBMS

- **1980s:**
  - RDBMS, advanced data models (extended-relational, OO, etc.)

- **1990s—2000s:**
  - Data mining and data warehousing, multimedia databases, and Web databases
Data Mining Disciplines

- Data Mining
- Database Technology
- Statistics
- AI and Machine Learning
- Information Science
- Computer Science
- Other Disciplines
What Is Data Mining?

• Data mining
  – Knowledge Discovery in Data = KDD
  – Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data

• Other definitions
Similar but Different View

- Data Sources
- Data Extraction and Cleaning
- Aggregated Data
- Selection and Preprocessing
- Warehouse
- Prepared Data
- Patterns & Models
- Knowledge
- Data Mining
- Interpretation and Evaluation

\[ p(x) = 0.02 \]
Yet Another Model

- Prettier perhaps but not all that accurate
- Does accurately depict the notion of iteration in data mining
Yet Another Model

- Task Understanding
  - Determine Task Objectives
  - Assess Clinical Problem
  - Determine Data Mining Goals
  - Produce Project Plan

- Data Understanding
  - Collect Initial Data
  - Describe Data
  - Explore Data
  - Verify Data Quality

- Data Preparation
  - Select Data
  - Clean Data
  - Construct Data
  - Integrate Data
  - Format Data

- Modeling
  - Select Modeling Technique
  - Generate Test Design
  - Build Model
  - Assess Model

- Evaluation
  - Evaluate Results
  - Review Process
  - Determine Next Steps
  - Assess Model

- Deployment
  - Plan Deployment
  - Plan Monitoring & Maintenance
  - Produce Final Report
  - Review Project
Roles in the KDD process

- **Domain Experts**
  - Business Analysis

- **Mining Specialists**
  - Data Analysis
  - Data Gathering
  - Data Preparation
  - Data Mining
  - Result Interpretation
  - Business Application
  - Business Feedback

- **Data Administrator**

- **Business Goal**
Data Mining Methods

- Case-based reasoners
- Clustering algorithms
- Decision trees (rules)
- Genetic algorithms
- Market basket analysis
- Neural networks
- Rough sets and fuzzy logic
- Statistical techniques
- Support vector machines
- Visualization tools
- Others
Steps in the process

- Understand/learn the application domain
- Data collection and extraction
- Data cleaning and preprocessing
- Data reduction and transformation
- Choose the mining algorithm/method(s)
- Data mining search for patterns of interest
- Interpretations of results
- Application and use of discovered knowledge (evidence-based practice)
Health care data mining issues

- Domain expertise!
- Data quality
- Patient privacy

Not covered today:
- Model building and testing
- Model overfitting
- Technical and mathematical issues
- Meaningful/relevant results and application
- Others…
THE DATA MINER

EUREKA! I FOUND A CORRELATION.

WHEN YOU’RE ON VACATION, ALL YOUR EMPLOYEES TELECOMMUTE.

THEY DO?

AND 100% OF ALL EXPENSE VOUCHERS ARE SIGNED WHEN YOU’RE OUT SICK.

WE HAVE VOUCHERS?
Issues in the process

**Patient Privacy and/or Data Quality Issues***

* Understand/learn the application domain
* Data collection and extraction
* Data cleaning and preprocessing
* Data reduction and transformation

Choose the mining algorithm/method(s)

* Data mining search for patterns of interest
* Interpretations of results
* Application and use of discovered knowledge
Multiple methods – ROC curves

Pre-LMP 10 Weeks 20 Weeks 30 Weeks

Linear

Neural
Multiple methods – ROC curves

Receiver Operating Characteristics

- Sensitivity (True Positive Rate)
- 1-Specificity (False Positive Rate)

- Chance
- LRDS1
- NNDS1
- LRDS2
- NNDS2
Data Preprocessing

- **Data cleaning**
  - Explore missing values and noisy data, identify or remove outliers, resolve inconsistencies, de-identification

- **Data integration**
  - Integration of multiple databases or files

- **Data transformation**
  - Normalization and aggregation

- **Data reduction**
  - Reduce data volume but with same or similar analytical results

- **Data discretization**
  - Part of data reduction but with particular importance for numerical data
Data Preprocessing Activities

Data Cleaning
- [water to clean dirty-looking data]
- [‘clean’-looking data]
- [show soap suds on data]

Data Integration

Data Transformation
- \(-2, 32, 100, 59, 48\) → \(-0.02, 0.32, 1.00, 0.59, 0.48\)

Data Reduction
- T1, T2, T3, T4, ..., T2000 → T1, T3, A1, A3, ..., A115
Clinical data in the real world is dirty
  – incomplete
  – Inconsistent
  – noisy
  – redundant

• Data mining quality depends on data quality!
Discretization

- Three types of attributes:
  - Nominal — values from an unordered set
  - Ordinal — values from an ordered set
  - Continuous — real numbers
- Discretization divides the range of a continuous attribute into intervals
  - Some classification algorithms only accept categorical attributes.
  - Reduce data size by discretization
  - Prepare for further analysis
Remember…

• Data extraction and preprocessing comprises the majority of the work in clinical data mining!

• Clinical/domain expertise is critically important in preprocessing decisions…
Potential patient privacy and data quality issues exist at almost every step in the data mining process.
- Data preprocessing is a big issue, since it consumes at least half (or more) of the time and resources in clinical data mining and includes:
  - Data cleaning and data integration
  - Data reduction and feature selection
  - Discretization
- Clinical/domain expertise is critically important in data preprocessing decisions
Data Quality Overview

Defining data quality
- "fit for use" (Tay and Ballou, 1998)
- Multiple articles and authors, 1995-96

Data quality methods
- Manual
- Automated
- Enter Sub-topic
- Accuracy
- Completeness
- Inconsistency
- Redundancy
- Timeliness

Data quality issues
- Collection
- Preprocessing
- Quality analysis
- Patient outcomes

Data quality and patient care quality
- Failure to understand user needs
- Poorly defined, non-existent, or non-implemented DQM plans
- IOM (2002) 98,000 deaths per year
- Eckerson (2002) $600 billion in the US per year (not just health)

Remember!

SINI: July 2008
Big Brother: Data Mining Quality and Privacy Issues

Http://onlineinformatics.com

linda.goodwin@duke.edu - copyright 2008 - re-use with written permission
Defining data quality

• "fit for use" (Tayi and Ballou, 1998)
• Multiple articles and authors, 1995-96)
  – Accuracy
  – Completeness
  – Consistency
  – Timeliness
Data quality methods

- Manual
- Automated
Data quality issues in health care

• Accuracy
• Completeness
  – Missing variables
  – Missing values
• Inconsistency
  – Contradictory
  – Variances
  – Multiple terms
Data quality issues in health care

- Accuracy
- Completeness
- Inconsistency
- Metrics
  - What and how to measure data quality
- Redundancy
- Standardization
  - terminology and data standards
- Temporal and longitudinal
- Timeliness
- Value and Volume
  - Large volume but little added value
Data quality issues in health care

- Accuracy
- Completeness
  - Missing variables
  - Missing values
- Inconsistency
  - Contradictory
  - Variances
  - Multiple terms
- Metrics (lacking) to measure data quality
- Redundancy
- Standardization (lack of)
- Temporal and longitudinal
- Timeliness
- Value
- Volume
Data quality and patient care quality

• Failure to understand customer needs
  – Internal customers
  – External customers
• Poorly defined, non-existent, or non-implemented DQM plans
• IOM (2002) 98,000 deaths per year
• Eckerson (2002) $600 billion in the US per year (not just health)
Clinical data is predictably flawed
  - Incomplete, redundant, inconsistencies, errors, etc.

Data quality issues exist at almost every step in the data mining process
  - Data preprocessing seeks to optimize data quality for analyses

So why do we care about patient privacy?
  - Can’t we just de-identify data before mining it and not worry about privacy issues?
Remember!

- Good decisions require good (quality) data
Patients increasingly lack trust
Withhold information
Fail to seek needed care
Provide false id
Refuse to participate in research
GAO report confirms a problem
IOM report confirms a problem
AHIMA reports a problem
Just a few or thousands of examples

Systemic problem confirmed

Use vs Disclosure (PHI)
HIPAA
DOJ lack of prosecution

Secondary data uses
Experiments
Google Health
Others

ANA Ethics Code

Advocacy
- Http://www.govhealthit.com/specials/peel/

Technology risks and solutions
Pick a number that reflects your opinion of HIPAA

1. Personally have no idea whether HIPAA protects patients’ privacy or not
2. HIPAA protects patients’ privacy pretty well
3. HIPAA mostly gutted patients’ privacy protections in the USA
Privacy Laws - 2008

• No constitutional right to privacy
  – No comprehensive federal privacy law
  – Patchwork of state laws that supercede HIPAA and are anything but standardized or consistent between various states

• Medical/health records are less protected than credit reports
Use of PHI

• Sharing, application, utilization, examination, or analysis of PHI
• WITHIN the “organization”
Disclosure of PHI

- The release, transfer, access, or divulging of PHI
- to an OUTSIDE person or entity.
**PHI: 18 data points (HIPAA)**

<table>
<thead>
<tr>
<th>1. Patient name</th>
<th>11. Certificate and/or license numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Address</td>
<td>12. Vehicle identifiers and serial numbers</td>
</tr>
<tr>
<td>3. Dates</td>
<td>13. Web URLs</td>
</tr>
<tr>
<td>4. Telephone numbers</td>
<td>14. Internet IP addresses</td>
</tr>
<tr>
<td>5. Fax numbers</td>
<td>15. Device identifiers and serial numbers</td>
</tr>
<tr>
<td>6. Email addresses</td>
<td>16. Biometric identifiers including finger and voice prints</td>
</tr>
<tr>
<td>7. Social security number</td>
<td></td>
</tr>
<tr>
<td>8. Health plan or beneficiary numbers</td>
<td></td>
</tr>
<tr>
<td>9. Medical record numbers</td>
<td></td>
</tr>
<tr>
<td>10. Account numbers</td>
<td></td>
</tr>
</tbody>
</table>

Not PHI: Gender, Dx, Rx, Procedures, etc.
Believe it or not!

- Most consumers are unaware of the volume and extensiveness of data sharing in health care
- Most health care providers are also limited in their awareness
Privacy in Health Care

- Cost-benefit?
- Risk?
- Issues?
- Status?
- Data Mining Considerations?
Privacy in Health Care: Cost-Benefits

- Costs of HIT – USA
  - IHC – savings in transcription paid for approx $2500 per MD for EHR
  - Outpt MD office EHR estimated at $25-45,000 per physician + annual 12-20%
  - First Consulting Group study
    - Hospital systems estimated at $17,500 per bed + annual 19%
      - 500 beds = $7.9 million + annual $1.35 million
    - RAND estimates $63,000 per bed + 30%

- (Claimed) Benefits of HIT
  - Improved quality
  - Cost savings

Risk?
Issues?
Status?
Privacy in Health Care: Risk

- Risk?
  - Increased access increases risk
    - Curiosity motives
    - Profit motives
  - Technology offers better security
    - If we use it
    - Security breaches have greater consequences
Example health care data flow diagram
Privacy in Health Care: Issues

- **Issues?**
  - Providing care and maintaining privacy are often competing goals in health care
  - Sharing data should help lower costs and improve quality of care (but minimal evidence thus far)
  - Technology is the easy part! People and policy are hard…
Privacy Protection Issues

- Increase in managed care
- Proliferation of HIT
- Human genome mapping
- Increased demand for health data
- Commercial use of health data
Wicked question

• Can we use sensitive data to optimize cost and quality without compromising patient privacy?

• Short answer:

  • No!
More complex/long answer

• Maybe…
  – Consider compromising a little bit of privacy
  – To improve quality and cost substantially
  – Using secure technologies

• But this has not and is not happening

• Future predictions and projections do not offer much hope for improvement
Privacy in Health Care

- Cost-benefit?
- Risk?
- Issues?
- Status?
Inside the Fence: Legal Users of Your Medical Records
Secondary Users of Health Information

- Drug marketers
  - In 2005, the AMA made more than $44 million from the sale of physician data, much of it prescribing data
- Public assistance programs
- Law enforcement agencies
- Courts
- Companies such as Medical Information Bureau (mib.com)
- Others
Personal health information is for sale

Table 1: Sample Data Elements for Commercial and Medicare Databases

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Medical Information (Inpatient and Outpatient)</th>
<th>Health Plan Features</th>
<th>Financial Information</th>
<th>Drug Information</th>
<th>Enrollment Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>Admission date and type</td>
<td>Coordination of benefits amount</td>
<td>Total payments</td>
<td>Generic product ID</td>
<td>Date of enrollment</td>
</tr>
<tr>
<td>Age</td>
<td>Principal diagnosis code</td>
<td>Deductible amount</td>
<td>Net payments</td>
<td>Average wholesale price</td>
<td>Member days</td>
</tr>
<tr>
<td>Gender</td>
<td>Discharge status</td>
<td>Copayment amount</td>
<td>Payments to physician</td>
<td>Prescription drug payment</td>
<td>Date of disenrollment</td>
</tr>
<tr>
<td>Employment status and classification (hourly, etc.)</td>
<td>Major diagnostic category</td>
<td>Plan type</td>
<td>Payment to hospital</td>
<td>Therapeutic class</td>
<td></td>
</tr>
<tr>
<td>Relationship of patient to beneficiary</td>
<td>Principal procedure code</td>
<td>Payments—total admission</td>
<td></td>
<td>Days supplied</td>
<td></td>
</tr>
<tr>
<td>Geographic location (state, ZIP Code)</td>
<td>Secondary diagnosis codes (up to 14)</td>
<td></td>
<td>National drug code</td>
<td>Refill number</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>Secondary procedure codes (up to 14)</td>
<td></td>
<td>DRG</td>
<td>Therapeutic group</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Length of stay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Place of service</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provider ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quantity of services</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Deb Peele, MD - privacyrights.org
Medicare and Medicaid data is for sale

To address the need for better data on privately insured Americans, Thomson Medstat created the MarketScan® data collection. Since its creation, MarketScan has been expanded to include data on Medicare and Medicaid populations as well, making it one of the largest collections of claims-based patient data in the nation. MarketScan data reflect the real world of treatment patterns and costs by tracking millions of patients as they travel through the healthcare system, offering detailed information about all aspects of care. Data from individual patients are integrated from all providers of care, maintaining all healthcare utilization and cost record connections at the patient level.

Source: Deb Peele, MD - privacyrights.org
Prescriptions For Sale

• **Nex2, Inc. (Sold to United Healthcare in 2002):**
  In stealth-mode, Nex2 built what are arguably the largest, near-realtime drug history databases in the world, with **over 200 million Americans’ five-year running drug histories online** (over 12 TB total). The databases are updated every 24 hours by every retail pharmacy in America via the PBMs... [these] prescription profiles acting as a powerful surrogate for the medical record itself.

  Source: Deb Peele, MD - privacyrights.org
De-identified not always private

• 1997 – Dr Latanya Sweeney (CMU)
  – Using publicly available data (voter list), she re-identified the MA governor's medical record where name, address, and SSN had been removed
    • She cross-referenced date of birth, gender and ZIP. In fact, 87% of the population of the United States can be uniquely identified with these data points!

• Post 9/11
  – Strong advocate for safety as well as privacy (citizens do not need to give up their privacy)
  – http://privacy.cs.cmu.edu/people/sweeney/
Only a third of U.S. adults say they trust health plans and government programs to maintain confidentiality all or most of the time.
One in five **Americans** believe that a health care provider, insurance plan, government agency, or employer has improperly disclosed personal medical information. Half of these people say it resulted in personal embarrassment or harm.
Gallup Poll (2000)

- 77 percent of all Americans feel their personal health information privacy is very important, and
- 84 percent said they were very concerned or somewhat concerned that personal health information might be made available to others without their consent.
- Only 7 percent said they are willing to store or transmit personal health information on the Internet, and only 8 percent felt a Web site could be trusted with such information.
Gallup Poll (2000)

- 90 percent said they would trust their doctor to keep their personal health information private and secure, and 66 percent said they would trust a hospital to do the same.
- Only 42 percent said they would trust an insurance company, and 35 percent would trust a managed care company.
Harris/Westin 2006 poll: EHR National Program

- only 26% were aware of a national EHR effort
  - awareness highest -- as expected -- among better-educated, higher-income, and online-using
  - rather surprising -- given extensive mass media coverage
- 55% believe EHR can decrease frequency of medical errors significantly
  - was 62% in 2005
- 60% believe EHR can reduce healthcare costs significantly
  - was 73% in 2005
- 68% believe EHR can improve patient care by reducing unnecessary tests and procedures
  - was 73% in 2005
- 62% believe “The use of Electronic Medical Records makes it more difficult to ensure patients’ privacy”
  - was 67% in 2005
“The privacy of personal medical records and health information is not protected well enough today by federal and state laws and organizational practices.”

- 58% agree that privacy is not protected
- 42% disagree

- HIPAA and its (lack of) enforcement has eroded confidence in national health privacy protection
  - Widely-reported medical data breaches
  - Lack of DOJ enforcement
Patients lack trust

- GAO report confirms a problem
- IOM report confirms a problem
- AHIMA report confirms a problem
- Thousands of URL examples - breach
Patients lack trust

- Withhold information
- Fail to seek needed care
  - Personal evidence
- Provide false id
  - Personal evidence
- Refuse to participate in research
  - Http://www.cchconline.org/what_privacy.php
In a 2000 survey of Fortune 500 companies, only 38% responded that they do not use or disclose employee health information for employment decisions.

2006 Memorandum
Wal-Mart’s Board of Directors
• “Redesign benefits and other aspects of the Associate experience, such as job design, to attract a healthier, more productive workforce.”
Fear is Justified: Breaches

- 4 million patient records improperly disclosed Jan-May 2008
  - 9000 records on laptop stolen from locked car (Tennessee)
  - 66 paper mental health records disappeared (Alabama)
  - 88,000 records on stolen computer (New York)
  - 12,000 scanned records on stolen physician office computer (Pennsylvania)
  - Etc.
- 285 breaches (all IDTR)
- Over 16 million total privacy breaches
Patient privacy

• Examples of breach
  – Find links at http://onlineinformatics.com
    • Deliberate - prescription mining
    • Deliberate - unauthorized access
      – The majority of breach remains with health care workers who have legitimate access but use it inappropriately
    • Unintended/accidental - disclosures
  • Litigation
  • Conclusion
Patient privacy

- Examples of breach
- Litigation
  - Find links at http://onlineinformatics.com
  - HIPAA Complaint Process
    - Number of complaints filed
    - Number of complaints prosecuted
  - State/Civil/Tort Cases
- Conclusion
Patient privacy

• Examples of breach

• Litigation

• Conclusion
  – HIPAA was a step in the right direction but does not adequately protect patient privacy
Fear feeds behaviors

15% of American adults say they have done something out of the ordinary to keep medical information confidential.

California HealthCare Foundation, national poll, January 1999
Privacy-protective Behaviors

- Paying out-of-pocket
- Doctor-hopping
- Giving inaccurate or incomplete information
- Asking a doctor not to write down certain health information or to record a less serious or embarrassing condition
- Avoiding care altogether
Data Quality and Patient Trust

Avoid health care

Privacy protective

Data Quality

-3  -2  -1  0  1  2  3

0.1% 2.2% 13.6% 34.1% 34.1% 13.6% 2.2% 0.1%
Privacy-protective Behaviors

- Impact on individual patient
- Impact on provider
- Impact on health system
- Impact on the nation’s health
IHE and RHIO and NHIN

- All major data sharing efforts struggle with privacy issues and data quality
  - Nationwide Health Information Network (NHIN)
  - Regional Health Information Organizations (RHIO)
  - Integrating the Healthcare Enterprise (IHE)
Where are we headed?

• given 42+% of public feeling potential privacy risks outweigh potential EHR benefits?

• and patient rights groups and privacy advocates calling for new privacy rules in Congress?

• and push-back already happening in UK, where 53% of public and 52% of GPs oppose the UK national EHR plan, in an organized campaign?
Advocacy

- ANA Ethics Code
- Http://www.govhealthit.com/specials/peel/
Experiments

- MSN and Cleveland Clinic
- Google Health
- Others
A poll

Pick a number that reflects your opinion of HIPAA

1. Personally have no idea whether HIPAA protects patients’ privacy or not
2. HIPAA protects patients’ privacy pretty well
3. HIPAA was designed to protect PHI held by “covered entities”
   1. Doesn’t apply to other orgs/agencies
   2. Almost no enforcement of existing regs
4. HIPAA mostly gutted patients’ privacy protections in the USA
Take Aways

• Unless patients trust us to protect their sensitive health/medical data
  – Data quality issues will proliferate
  – Data mining results will be frequently meaningless
  – Many health care decisions will be based on flawed data

• Patient privacy issues are hard to solve!

• Ignoring health consumer privacy wishes has worked, thus far, but will increasingly fail
  – Increasing public awareness
  – eHR will only succeed, long term, with the public’s trust
    • Public trust will not survive without solid privacy protections
A challenge: refocus our goals

- Do the “right” thing
  - “Golden” rule
- People before profits…
When all is said and done, will our health care records be used to heal us or reveal us?

(Donna Shalala, U.S. Secretary of Health and Human Services)
Questions?

linda.goodwin@duke.edu

http://onlineinformatics.com

(resources tab)