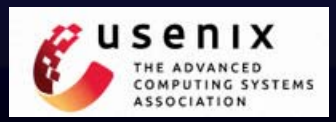
Of Codes, Genomes, and Electronic Health Records: It's Only Sensitive If It Hurts When You Touch It

Daniel Masys, M.D.
Affiliate Professor
Biomedical and Health Informatics
University of Washington
Seattle, WA



Usenix HealthSec Workshop
August 6, 2012

Topics

- A brief history of confidentiality and information security in healthcare: Hippocrates to HIPAA to HITECH
- Security vulnerabilities in healthcare settings
- What genomic data adds to the issues
- Models for medical information access
- Why is this so hard to do?
- Your job...

“What I may see or hear
in the course of treatment
or even outside of the treatment
in regard to the life of men,
which on no account one must spread
abroad, I will keep to myself
holding such things
shameful to be spoken about.”

- Hippocrates

HIPAA/HITECH Rules

(Health Insurance Portability and Accountability Act of 1996 as amended by Health Information Technology for Economic and Clinical Health Act of 2009)

- Secretary of HHS issued regulations for medical data privacy and security in 1999.
- “Covered entities” had to be in compliance with Privacy Rule effective April, 2003, small health plans by April 2004
- Compliance with HIPAA Security Rule for electronic systems containing Protected Health Information (PHI) was required by April, 2005
- HITECH amendments in 2009 added enforcement provisions, and patient access to electronic health data in electronic form

HIPAA, not HIPPA :-)



"Misspelling is not a violation of the Rule"

Director, US Office of Civil Rights

Speaking at UCSD, February 2003

Overview of effects of HIPAA/HITECH Privacy Rules

- Gives individuals the right to:
  - A written notice of information practices from health plans and providers
  - Inspect and copy their Protected Health Info
  - Obtain a record of disclosures
  - Request amendments to their medical records
  - Have reasonable requests for confidential communications accommodated
  - Request restrictions on uses and disclosures
  - Complain about violations to the covered entity and to HHS

Overview of effects of HIPAA/HITECH Privacy rules

- Requires covered entities to:
  - Make a good faith effort to get signed acknowledgement of information practices related to Protected Health Information (PHI) used in treatment, payment and operations (TPO)
  - Obtain authorization for special additional uses of PHI
  - Designate a privacy official
  - Develop policies and procedures (including receiving complaints)
  - Provide privacy training to their workforce
  - Develop a system of sanctions for employees who violate the entity's policies
  - Meet documentation requirements
  - Implement appropriate administrative, technical, & physical safeguards to protect privacy

The ‘spirit’of HIPAA/HITECH

- Protected Health Information (PHI = person identifiable) must be managed with the same attention to consent for use, access control, and documentation of actions performed as are applied to physical objects such as tissue.
- Access to PHI is based on the general principle of “need to know” and “minimum necessary” rather than professional role

HIPAA Security Rule

Compliance required in 2005

Few large healthcare institutions fully compliant in 2012

Security Rule Overview

- Affects HIPAA Covered Entities that maintain Protected Health Information (PHI) in electronic form
- Directs CE's to 'develop, implement, maintain, and document' security measures, and keep them current.

Security Rule: Basic Concepts

- Scalable: burden relative to size and complexity of healthcare organization
- Not linked to specific technologies, and anticipates onoing changes in technology
- Unlike Privacy Rule, affects only electronic information
- Applies security principles well established in other industries

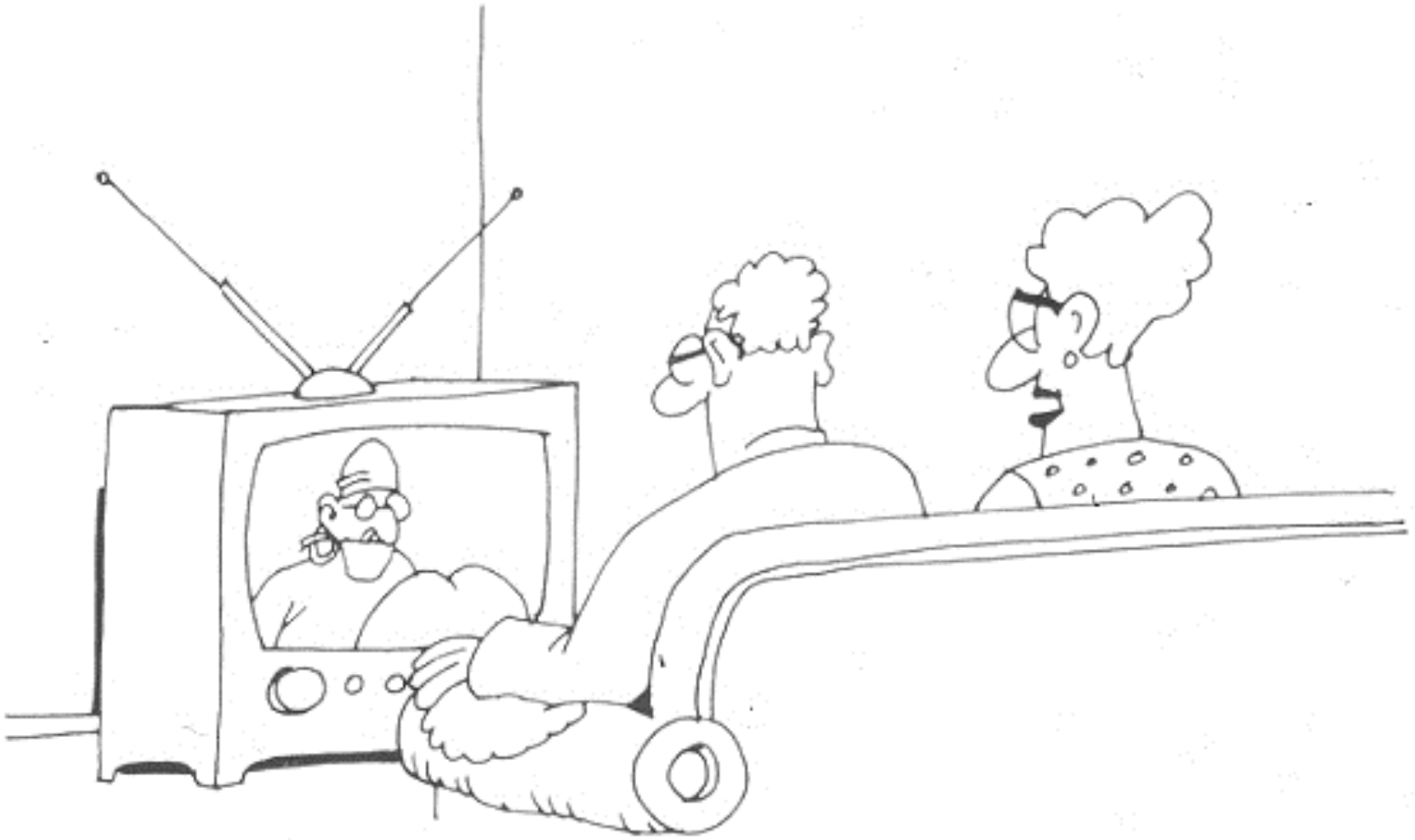
HIPAA Security Rule Functional areas

- Information Availability
- Protection against unauthorized:
  - Access
  - Alteration
  - Deletion
  - Transmission
- Monitoring (audit trails)

Covered entities are required to:

- Assess potential risks and vulnerabilities
- Protect against threats to information security or integrity, and against unauthorized use or disclosure
- Implement and maintain security measures that are appropriate to their needs, capabilities and circumstances
- Ensure compliance with these safeguards by all staff


ETTA ©1999 FORT WORTH STAR-TELEGRAM
HULME
WE CHECKED YOUR CONFIDENTIAL MEDICAL RECORDS ON THE INTERNET. CHEESE AND ANCHOVIES WOULD BE BAD FOR YOU, SO WE LEFT THEM OFF.
PIZZA
PIZZA

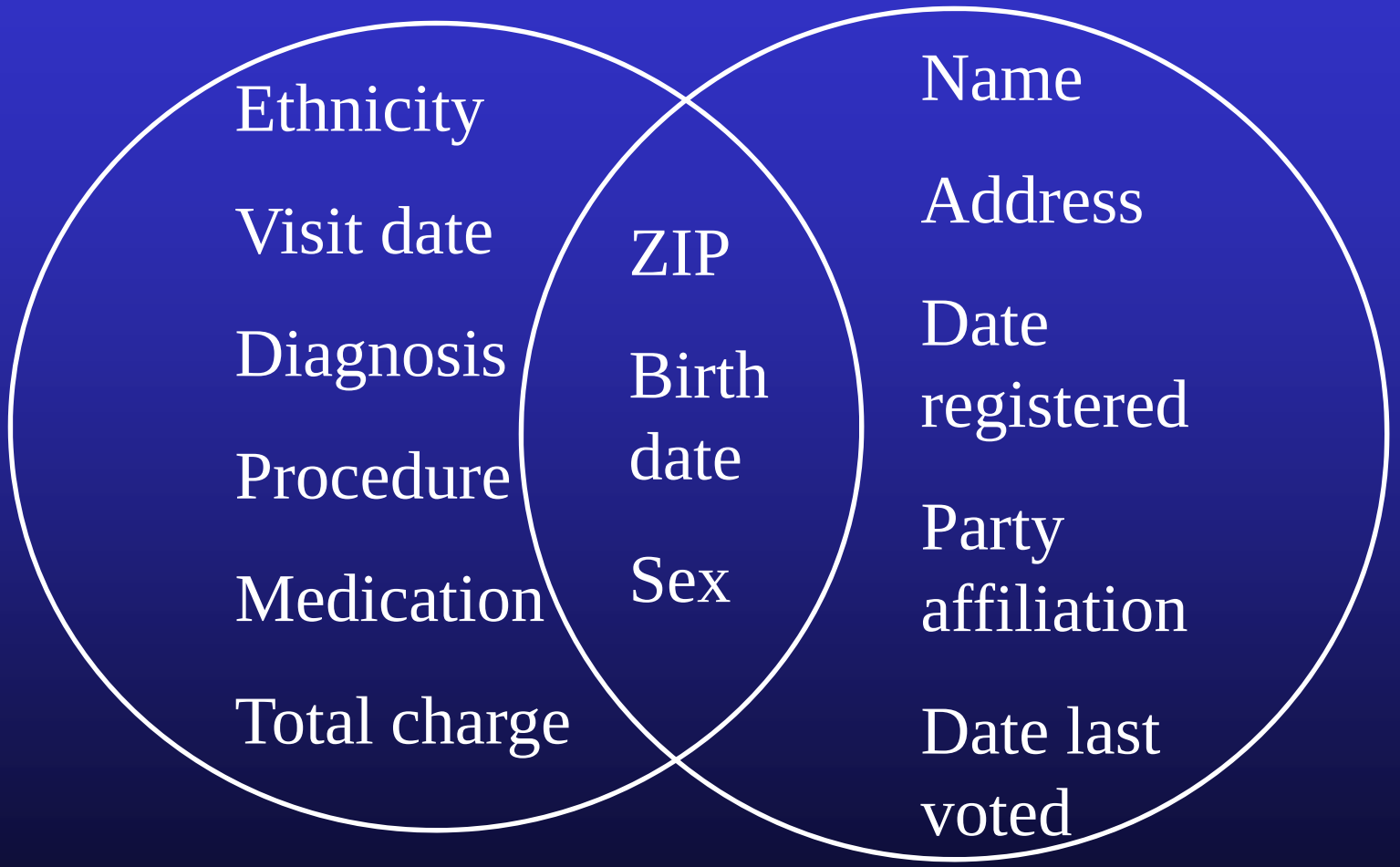


"Oh look, Henry, it's our neighbor having his annual proctoscopic exam!"

Security Vulnerabilities in Healthcare Settings

- Unintentional disclosures
- Well-intentioned but inappropriate employee behavior
- Disgruntled employees
- Self-insured employers
- ? Competitors
- VIP patients
- Hackers
- Data mining

Data mining as confidentiality threat



Latanya Sweeney, MIT, 1997

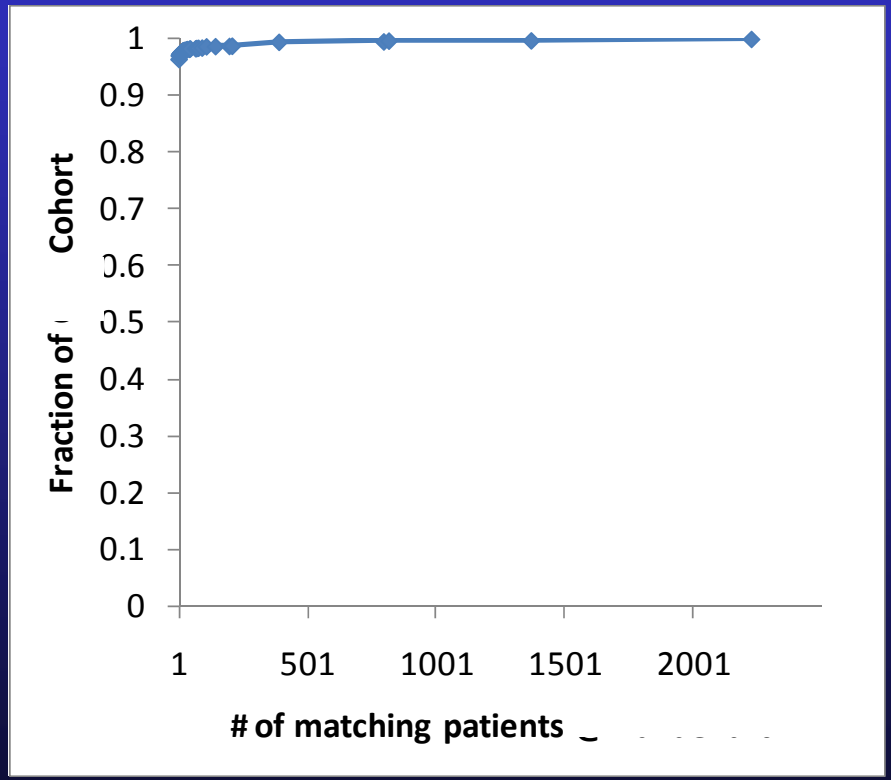
Uniqueness in Cambridge voters

Birth date alone	12%
Birth date & gender	29%
Birth date & 5-digit ZIP	69%
Birth date & full postal code	97%

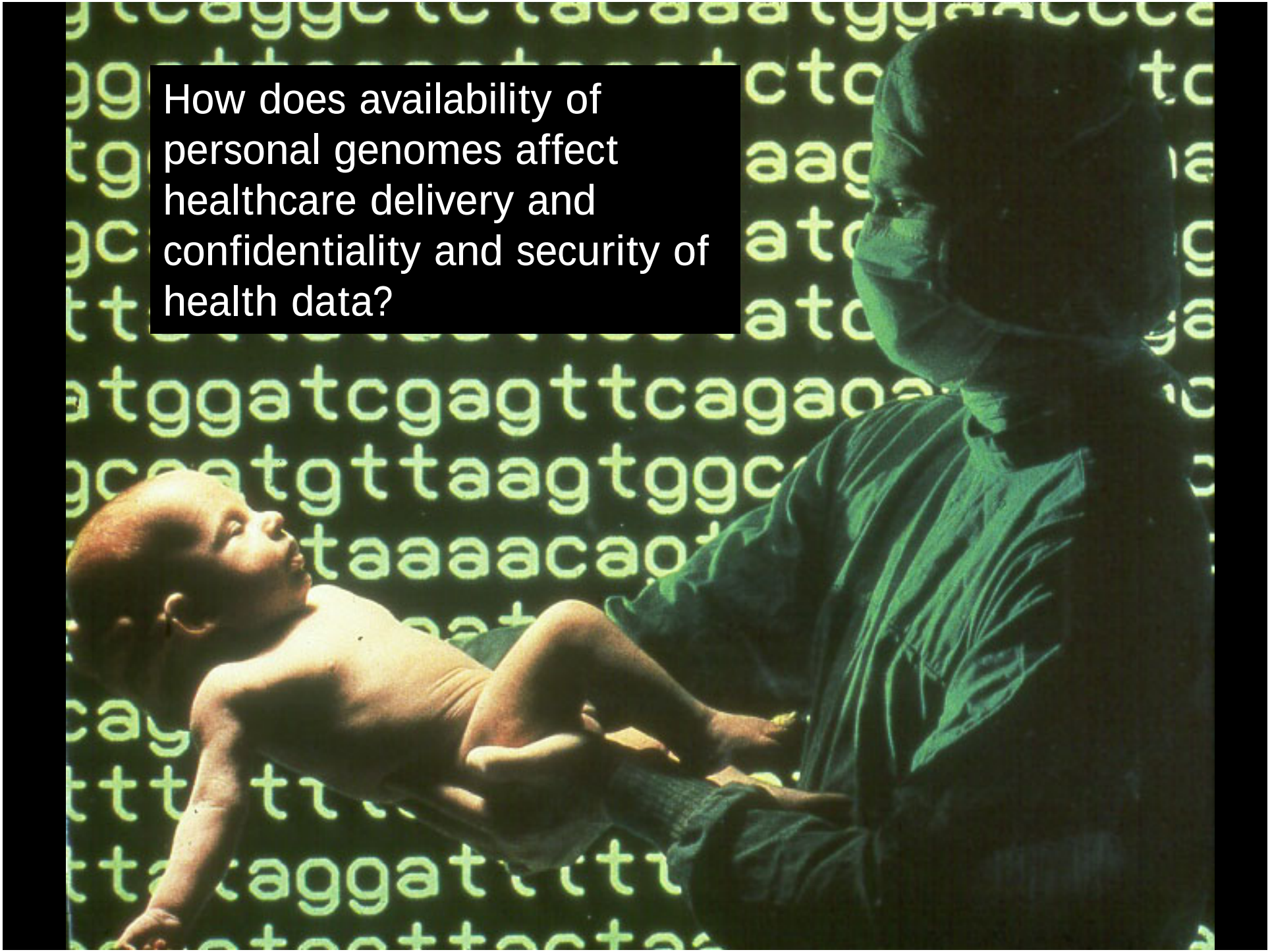
Birth date includes month, day and year.
Total 54,805 voters.

The richness of 'traditional' EHR data: Leveraging Diagnosis Billing Codes to establish uniqueness*

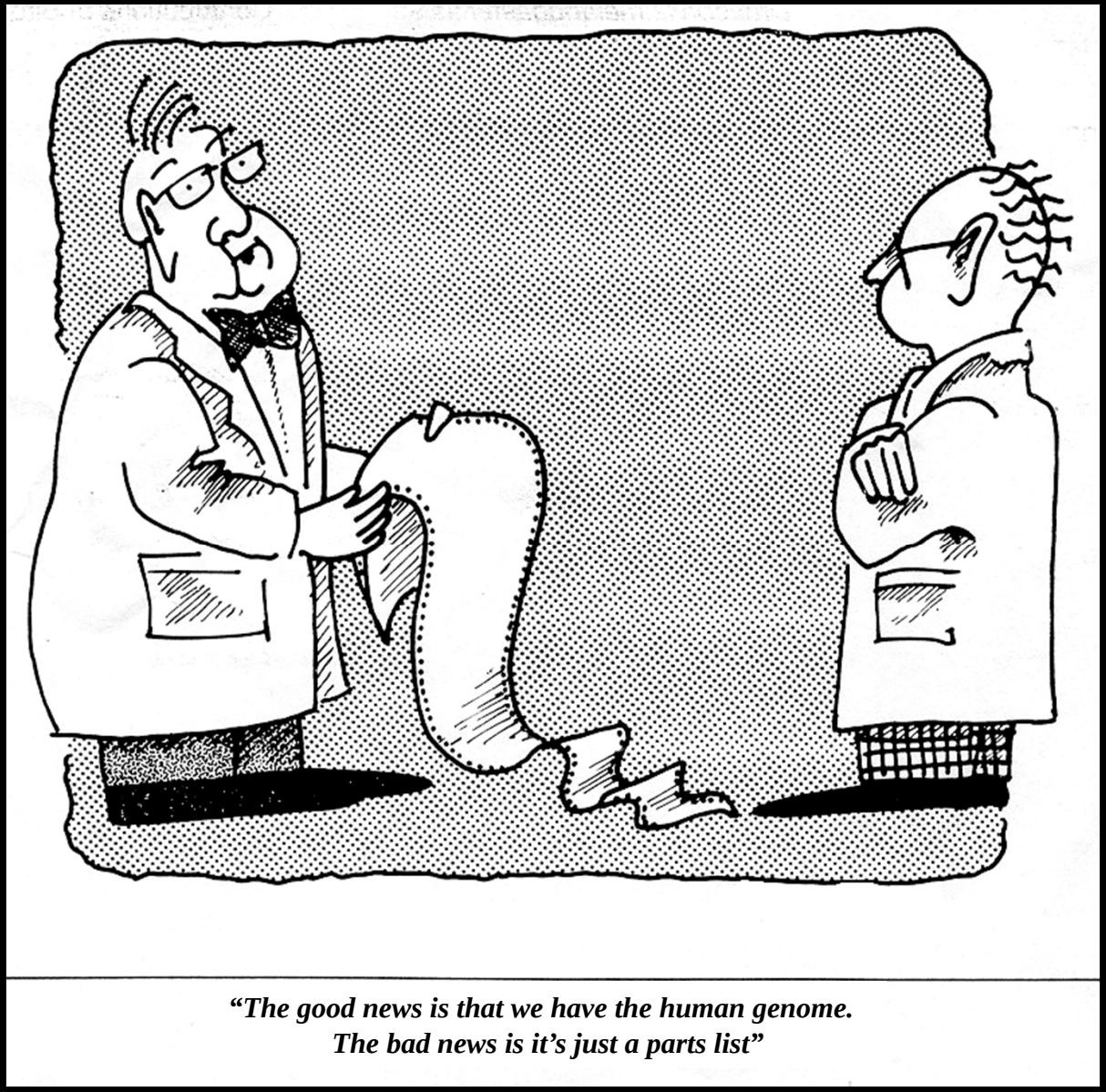
- Cohort: ~2500 patients in a genome association study
- Each individual in the cohort has set of ICD-9 codes
- Evaluated for "distinctiveness" with respect to entire EMR population (1.5 million)
- ~97% of individuals are unique



*Loukides G, Denny J, & Malin B. Do clinical profiles constitute privacy risks for research participants? *AMIA Fall Symposium.* 2009.


How does availability of personal genomes affect healthcare delivery and confidentiality and security of health data?

The Genome Sequence is at hand...so?



"The good news is that we have the human genome.
The bad news is it's just a parts list"

The Vision

- Molecular and clinical biomarkers for health conditions individuals either have or are susceptible to
  - Includes traditional healthcare history, physical findings, diagnostic imaging, standard clinical laboratories
  - Increasingly: large volumes of molecular data
    - Structural genomics: DNA in residence (~22,000 genes)
    - Functional genomics: genes switched on (1-2% active)
    - Proteomics (400,000 proteins from 22,000 genes)

The Vision, cont'd

- Pharmacogenomics
  - "The right dose of the right drug for the right patient at the right time"
  - Drug development:
    - Avoid drugs likely to cause side effects
    - Re-investigate "back-burner" drugs
    - Develop entirely new drugs targeting fundamental disease processes



"Here's my sequence..."

New Yorker, 2000

Firefox

New low-cost DNA decoding machine - ...

yourlife.usatoday.com/health/story/2012-01-11/C

Google

USA TODAY Your Life | Get Home Delivery | Mobile

Home | News | Travel | Money | Sports | Life

Your Life | Health | Fitness & Food | Family & Parenting | Sex & Relationships | Your Look | Blogs

Google

Learn more on how to stay safe on the Internet.

Company announces low-cost DNA decoding machine

Updated 1/11/2012 9:13 AM

Comment | Recommend 98 | Tweet 75

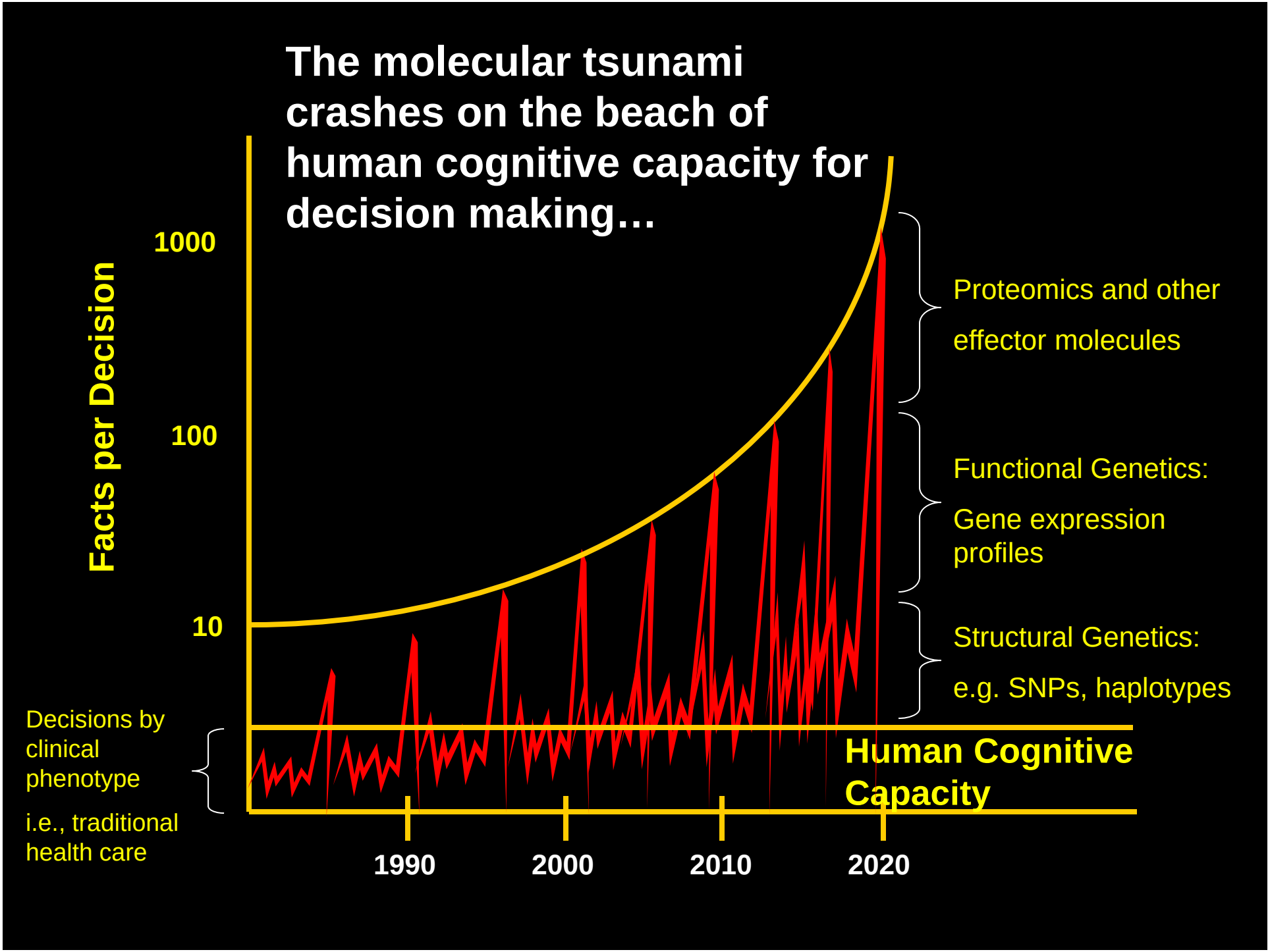
NEW YORK – A biotechnology company announced it has developed a machine to decode a person's DNA in a day for $1,000, a long-sought price goal for making a person's genome useful for medical care.



PR NEWSWIRE

Life Technologies Corp. said Tuesday it was taking orders for the technology, which it expects to deliver in about a year. The Carlsbad, Calif., company said three major research institutions had already signed up for the $149,000 machine: the Baylor College of Medicine, the Yale School of Medicine and the Broad Institute of Cambridge, Mass.

The machine is a sequencer, meaning that it lets scientists identify the sequence of the 3 billion chemical building blocks that make up a person's DNA. Since the


The molecular tsunami crashes on the beach of human cognitive capacity for decision making…
Facts per Decision
1000
100
10
Proteomics and other effector molecules
Functional Genetics: Gene expression profiles
Structural Genetics: e.g. SNPs, haplotypes
Decisions by clinical phenotype
i.e., traditional health care
Human Cognitive Capacity
1990
2000
2010
2020

Realities of 2012

- Our ability to acquire person-specific DNA data far exceeds our understanding of its meaning
- Genetic data conclusively explains the basis for only a tiny set of the 8000+ diseases of humans and responses to therapy
- As a result DNA data acquired now will likely need to be re-interpreted many times over in the future as DNA science unfolds

General observations about clinical genomics

- Genomic data is the current poster child for complexity in healthcare
- No practitioner can absorb and remember more than a tiny fraction of the knowledge base of human variation
- Therefore, computerized clinical decision support is the only effective way to insert genomic variation-based guidance into clinical care

Clinical genomics effects on health data privacy and security

- Very small amounts of DNA data confer uniqueness in globally-sized populations
- Personal molecular data volumes are large but roughly equivalent to current digital medical imaging
- Unlike images, DNA data not amenable to lossy compression

Properties of High Assurance health data systems

- Availability - when and where needed
- Authentication -a person or system is who they purport to be (preceded by Identification)
- Access Control - only authorized persons, for authorized uses
- Confidentiality - no unauthorized information disclosure
- Integrity - Information content not alterable except under authorized circumstances
- Attribution/non-repudiation - actions taken are reliably traceable

Why is this so hard?

1. The nature of biomedical data

The nature of biomedical data

- Variable levels of sensitivity; “sensitive” is in the eye of multiple beholders, and highly context-dependent
- No bright line between person-identifiable and “anonymous” data
  - So inherently rich in attributes that re-identification potential never reaches zero
- Genome as Future Diary: An individual’s medical data may have implications for other family members who have much different values and preferences, and for future generations

Why is this so hard?

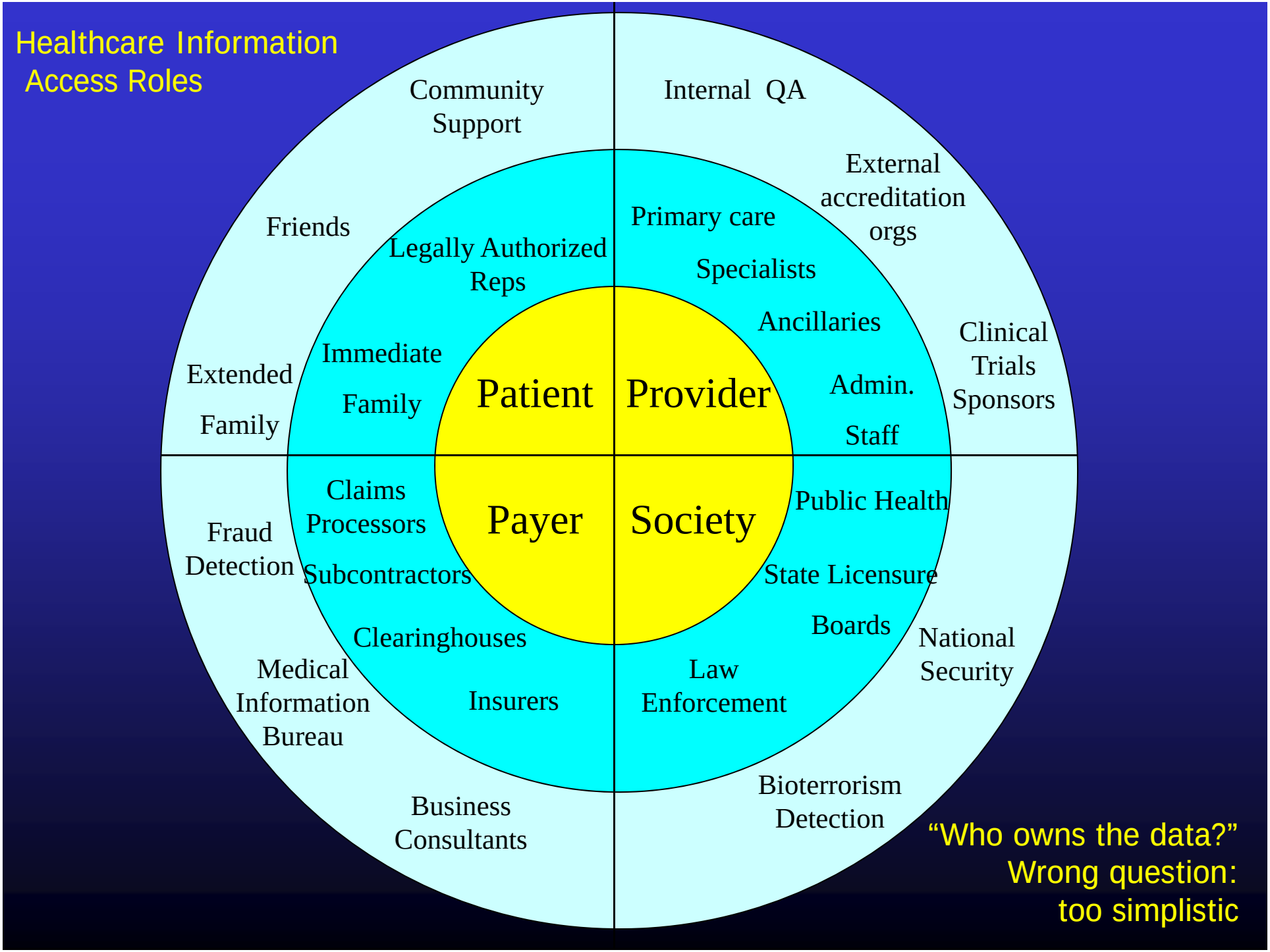
1. The nature of biomedical data
2. Complex interpersonal and organizational roles with respect to data

Complex roles: entities with justifiable (and variable) rights to medical data

- First order role definitions:
  - Provider, Patient, Payer, “Society”
- Second order:
  - Providers: primary vs. consultant provider, ancillary support staff
  - Patient: self, family, legally authorized reps
  - Payer: billing staff and subcontractors, clearinghouses, insurers
  - Society: public health agencies, state medical boards, law enforcement agencies

Complex roles: entities with justifiable (and variable) rights to medical data

- Third order:
  - Providers: internal and external QA entities (peer review, Joint Commission), sponsors of clinical research
  - Patient: community support groups, personal friends
  - Payers: fraud detection (Medical Information Bureau), business consultants
  - Society: national security, bioterrorism detection


Healthcare Information Access Roles
Community Support
Internal QA
External accreditation orgs
Friends
Legally Authorized Reps
Primary care
Specialists
Ancillaries
Clinical Trials Sponsors
Extended Family
Immediate Family
Patient
Provider
Admin. Staff
Claims Processors
Payer
Society
Public Health
Fraud Detection
Subcontractors
State Licensure
Boards
Clearinghouses
National Security
Medical Information Bureau
Insurers
Law Enforcement
Business Consultants
Bioterrorism Detection
"Who owns the data?"
Wrong question:
too simplistic

Why is this so hard?

1. The nature of biomedical data
2. Complex interpersonal and organizational roles with respect to data
3. Patients who wish to exercise control over access to their data seldom understand the implications of their decisions
4. Personal preferences regarding data access change, sometimes suddenly

Why is this so hard?

1. The nature of biomedical data
2. Complex interpersonal and organizational roles with respect to data
3. Patients who wish to exercise control over access to their data seldom understand the implications of their decisions
4. Personal preferences regarding data access change, sometimes suddenly
5. “Privacy Fundamentalism” – irrational political forces (“Nothing about me without me”) block efficient systems approaches

Why is this so hard?

1. The nature of biomedical data
2. Complex interpersonal and organizational roles with respect to data
3. Patients who wish to exercise control over access to their data seldom understand the implications of their decisions
4. Personal preferences regarding data access change, sometimes suddenly
5. “Privacy Fundamentalism” – irrational political forces (“Nothing about me without me”) block efficient systems approaches
6. Differing perceptions of risk and benefit

Desiderata for electronic consent in healthcare

E. Coiera et. al., J. Am Med Informatics Assoc, 2004

1. Permits access to health data by checking that patient consent exists for the information requests, using methods that check for explicit, inferred or implied consent
2. Should allow access to patient information to those who have been explicitly permitted by a patient

Desiderata for electronic consent in healthcare, cont'd

E. Coiera et. al., J. Am Med Informatics Assoc, 2004

3. Should never allow access to patient information by those explicitly denied access by the patient
4. Should allow access to patient information to individuals determined to have inferred or implied consent based on their clinical roles, responsibilities, or clinical circumstance

Desiderata for electronic consent in healthcare, cont'd

E. Coiera et. al., J. Am Med Informatics Assoc, 2004

5. Does not endanger patient safety by denying access to information by clinically approved individuals when consent is indeterminant
6. Does not impede clinical work by clinically approved individuals, when consent is indeterminant

Desiderata for electronic consent in healthcare, cont'd

E. Coiera et. al., J. Am Med Informatics Assoc, 2004

7. Has security safeguards to prevent access by circumventing consent checking mechanisms
8. Minimizes the number of requests made to clinicians and patients to avoid disruption of clinical care or the private lives of individuals

Desiderata for electronic consent in healthcare, cont'd

E. Coiera et. al., J. Am Med Informatics Assoc, 2004

9. Does not require expensive or burdensome infrastructure

Authors' Observation: criteria are in conflict with one another, and no single model performs well against all 9 criteria

Models for e-consent

E. Coiera et. al., J. Am Med Informatics Assoc, 2004

1. General consent = “opt in”. Patient accepts all provider policies (Notices of Information Practices). Most common current model.
2. General consent with specific denial. Patient accepts provider policies but denies consent for a) particular information or b) particular parties’ access or c) disclosure for particular purposes

Models for e-consent

E. Coiera et. al., J. Am Med Informatics Assoc, 2004

3. General denial with specific consent = Paitent denies all access except for consent for a) particular information or b) particular parties' access or c) disclosure for particular purposes
4. General denial = "opt out". Each new episode of care requires explicit consent. (Likely scenarios for opt out: psychiatric care, drug rehab, sexually transmitted disease treatment).

Implementation: e-Consent objects

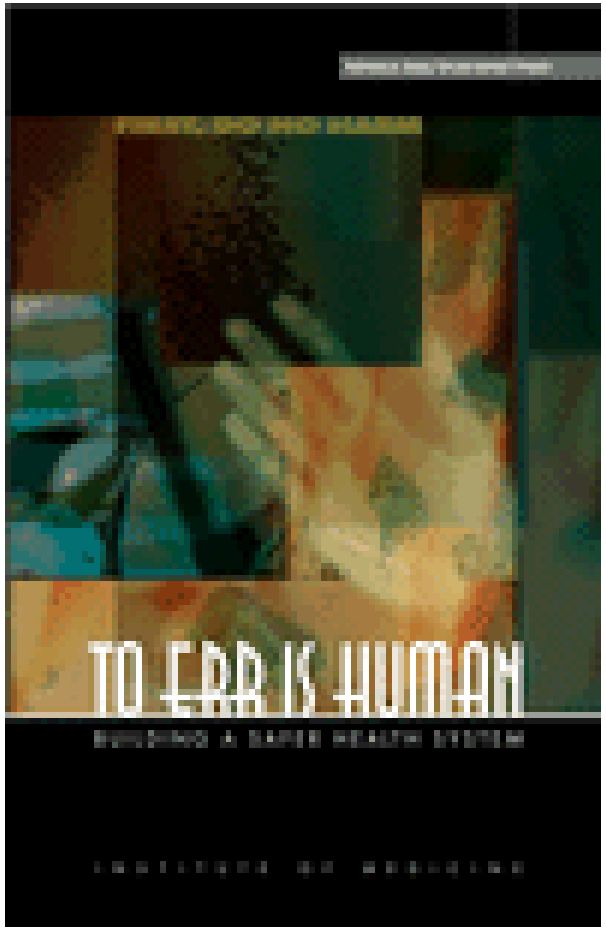
Rights management wrappers associated with clinical information that record the assertion:

Access to (information)
by an (entity)
for a (purpose)
in a (context)
is {consented to | denied }

Could attach to specific facts, episodes of care, or complete medical record

Putting Health Information Security into Perspective

- The perennial fervor and paranoia related to health information security is sometimes marked by “irrational exuberance”
- Data available to date suggests that breaches of confidentiality in healthcare usually cause either no apparent harm or some personal psychological harm, while inaccessibility of healthcare data causes preventable medical errors, up to and including death



Kohn L, et al. Committee on Quality of Health Care in America.

To Err is Human: Building a Safer Health System.

Institute of Medicine, Dec 1999

Medical Errors

- Between 44,000-98,000 preventable deaths each year in hospitals
- Injury rates from 2.9% (general med-surg) to 46% (ICU settings)
- 7th leading cause of death in US
- Underestimates due to:
  - Injury thresholds for reporting
  - Errors had to be documented in clinical record

Medical Errors

- Majority of errors do not result from individual recklessness, but from flaws in health system organization (or lack of organization).
- Failures of information management are common:
  - illegible writing in medical records
  - lack of integration of clinical information systems
  - inaccessibility of records
  - lack of automated allergy and drug interaction checking

Properties of High Assurance health data systems

- Availability - when and where needed
- Authentication -a person or system is who they purport to be
- Access Control - only authorized persons, for authorized uses
- Confidentiality - no unauthorized information disclosure
- Integrity - Information content not alterable except under authorized circumstances
- Attribution/non-repudiation - actions taken are reliably traceable

Putting Health Information Security into Perspective: My Premise

- If ‘keeping the bad guys out’ causes even a single additional death due to inaccessibility of information to authorized providers, patients and their families, we have failed to achieve a proper perspective on health information security

Your Job as a Usenix community member

1. Rage against the machine: don't design systems that give information management advantages to healthcare organizations while increasing health risks to patients and families
2. Remain True to Hippocrates: *Primum Non Nocere* – First Do No Harm.

Primum Non Nocere: Don't Make This Worse

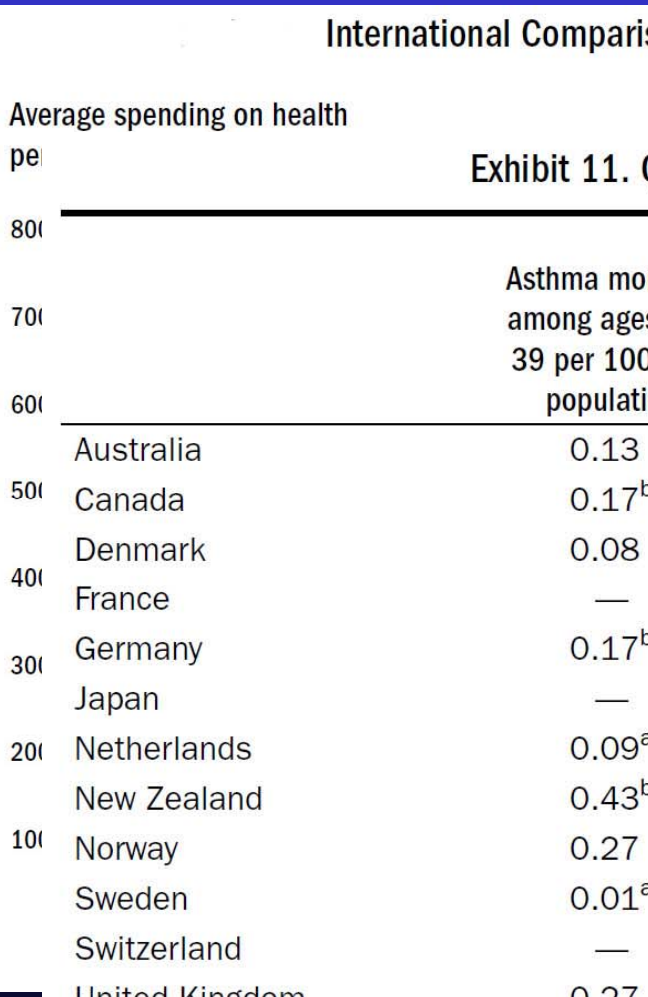


Exhibit 11. Quality Indicators in Select OECD Countries, 2009

	Asthma mortality among ages 5 to 39 per 100,000 population	Diabetes lower extremity amputations per 100,000 population	In-hospital fatality rate within 30 days of admission per 100 patients[c]		
			Acute myocardial infarction	Ischemic stroke	Hemorrhagic stroke
Australia	0.13	11.0	3.2	5.7	17.2
Canada	0.17[b]	9.5	3.9	6.3	20.6
Denmark	0.08	18.1	2.3	2.6	16.4
France	—	12.6[b]	—	—	—
Germany	0.17[b]	33.7	6.8	4.0	13.8
Japan	—	—	9.7[a]	1.8[a]	9.7[a]
Netherlands	0.09[a]	12.0[b]	5.3[b]	5.7[b]	22.5[b]
New Zealand	0.43[b]	7.0	3.2	5.4	21.1
Norway	0.27	9.9	2.6	2.8	11.6
Sweden	0.01[a]	5.7	2.9[b]	3.9[b]	12.8
Switzerland	—	7.4[a]	4.5[a]	—	14.8[a]
United Kingdom	0.27	4.8	5.2	6.8	19.3
United States	0.40[b]	32.9[a]	4.3[a]	3.0[a]	21.0[a]
OECD Median	***0.09***	***9.9***	***4.6***	***4.9***	***19.3***

