Ethics, Confidentiality, etc.
& Concluding Thoughts

Peter Szolovits
6.872/HST.950

---

Treatment of Human Subjects: The Belmont Report
Ethical Principles and Guidelines for the Protection of Human Subjects of Research

- Balancing (societal) benefits vs. (individual) risks
- History of abuses
  - Nazi “experiments” ⇒ Nuremberg code
  - Tuskegee syphilis study

---

Nazi Medical Experiments

- Freezing / Hypothermia
- Genetics
- Infectious Diseases
- Interrogation and Torture
- Killing / Genocide
- High Altitude
- Pharmacological
- Sterilization
- Surgery
- Traumatic Injuries

A cold water immersion experiment at Dachau concentration camp presided over by Professor Ernst Holzlöhner (left) and Dr. Sigmund Rascher (right). The subject is wearing an experimental Luftwaffe garment

---

Tuskegee Syphilis Experiment

- 1932-1972 experiment to study natural progression of disease
- 399 African-American sharecroppers w/ syphilis
- failed to treat even after penicillin was shown to be an effective treatment in 1940's

http://en.wikipedia.org/wiki/Tuskegee_syphilis_experiment
Practice & Research

- The term “practice” refers to interventions that are designed solely to enhance the well-being of an individual patient or client and that have a reasonable expectation of success.
- The term “research” designates an activity designed to test an hypothesis, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge.
- Research and practice may be carried on together when research is designed to evaluate the safety and efficacy of a therapy. ... if there is any element of research in an activity, that activity should undergo review for the protection of human subjects.

From the Belmont Report

Basic Ethical Principles

- Respect for Persons
- Beneficence
- Justice

Respect for Persons

- Each person is an autonomous agent, capable of deliberation about personal goals and of acting under the direction of such deliberation
- Persons with diminished autonomy are entitled to protection: e.g., children, physically or mentally disabled, prisoners.
- Requires Informed Consent
  - Adequate information
  - Voluntary participation

(Inform ed Consent)

- Study involves research, purpose of research, duration, procedures, what is experimental?
- Foreseeable risks and discomforts
- Possible benefits to participants or others
- Alternative procedures that might be beneficial
- How confidentiality will be maintained
- For research involving more than minimal risk, what compensations and treatments may be available, and where to get further information
- Participation is voluntary; no penalty for refusal

From the Belmont Report

http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm
**Beneficence**

- Do no harm
  - one should not injure one person regardless of the benefits that might come to others
- Minimize risk to participants
- Maximize possible benefits
  - to society
  - but, research subjects may not benefit directly
- Some tradeoffs are unavoidable

*From the Belmont Report*

**Justice**

- Varied views of equal treatment
  - equal share
  - individual need
  - individual effort
  - societal contribution
  - merit
- Select participants fairly
- Distribute benefits fairly

*From the Belmont Report*

**Enforcement: The Common Rule**

- Applies to all US Government funded projects involving human subjects
- Institutional Review Boards (IRB) review and must approve all such proposed research; responsible to protect subjects
- yearly review of research protocols, informed consent, training of researchers, etc. Criteria of Belmont Report.
- expedited review for research involving “no more than minimal risk”; consent may be waived
- exemptions for educational research, food quality research, and retrospective research on public or de-identified data
- IRB's also responsible for protection of confidentiality
- MIT's IRB is the Committee on Use of Humans as Experimental Subjects (COUHES)

*http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm*
Informational Harm
(most relevant to our class)

- Details are covered by HIPAA (see below)

2011 Proposed Revisions to
“Common Rule”
Not (yet?) enacted

- Problems:
  - attention(major risks) = attention(minor risks)
  - Virtually no enforcement
  - Burdensome to researchers
  - Vagueness allows bad behavior

Proposed “Common Rule” Changes

- Risk-based protections
  - Make social science research easier; reduce reporting and re-review requirements of minimal risk studies; define categories of excused studies
- Unifying review of multi-center studies under one IRB
- Simplify, shorten consent forms; include any financial relations of investigators; open questions about whether separate consent is required for different studies on the same data or samples
- Require more uniform data protection rules; merge in HIPAA approaches
- Data collection to improve oversight; extend regulations to cover all research with any Federal connection, harmonize regulations

### 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?
Legitimate Concerns (some may be ameliorated by ACA)

- Difficulty getting insurance
  - “Individual insurers may deny you coverage based on your medical history if it includes:
    - Use of prescription drugs to treat anxiety, depression or a physical condition, including Ativan, Klonipin, Paxil, Prozac, Serzone, Zoloft, Xanax and Wellbutrin.
    - Counseling for anxiety, depression, grief or an eating or sleep disorder. Even if you briefly sought counseling as a way to cope with the Sept. 11 terrorist attacks, you could be denied individual health insurance, according to researchers with Georgetown’s Health Privacy Project.” (MSN, March 9, 2004)
- Medical Information Bureau
  - Data on all applicants for private life insurance in past 7 years

Additional Legitimate Concerns

- When employer pays insurance premiums, you may lose your job
- Self-insured companies
- Small employers facing “experience rated” policies
- Non-employment discrimination based on health
- Adoption
- Politics
- Social stigma
Security

- Integrity of data
  - No unauthorized modifications
  - No “dropped bits”
- Availability
  - Natural disaster
  - Adversary attack
  - Inadequacy of backup, fail-over
- Enforcement of confidentiality policies

De-Identification

Identifiable

- **HIPAA**: 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, date of birth, zip code, gender, race, profession
  - “any other unique identifying number, characteristic, or code”
  - “actual knowledge that the information could be used … to identify”
- 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
- **Limited Data Set**: allows inclusion of dates, full zip codes, but requires limited data use agreements

Sweeney’s Cambridge

- 1997 Cambridge, MA voting list on 54,805 voters
  - Name, address, ZIP, birth date, gender, …
- Combinations that uniquely identify:
  - Birth date (mm/dd/yy) 12%
  - BD + gender 29%
  - BD + 5-digit ZIP 69%
  - BD + 9-digit ZIP 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
- Genetics makes this immensely more problematic
  - Think Gattaca

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
- T. Lasko: spectral anonymization
  - Build a model of data that captures the $n$-th order statistics of the distribution
  - Synthesize “fake” patients from that distribution
- J. Ghosh: detailed modeling of data
  - Build a Bayesian Network model that captures the dependencies among data
  - Synthesize “fake” patients or directly use the model

De-Identification of Text

- Techniques very much like what we discussed in natural language processing to extract information from narrative text
- Use to identify the HIPAA-prohibited categories, but more broadly e.g., hard to tell patient name from doctor’s; eliminate both
- Residual problems:
  - “his brother, the star pitcher, came to visit”

Authorship

- Eugene Braunwald example of bad biomedical habits
  - “Dr. Braunwald has over 1000 publications in peer-reviewed journals.”
  - John Darsee, brilliant young investigator in Braunwald’s lab, faked data in dozens of publications
  - “Titular” co-author on at least 7 papers
- Harvard reaction
Unwanted and Stealth Authorship

Rofecoxib (Vioxx) Study

- For the publication of clinical trials, documents were found describing Merck employees working with medical publishing companies to prepare manuscripts and subsequently recruiting external, academically affiliated investigators to be authors.
- Recruited authors were frequently placed in the first and second positions of the authorship list.
- For the publication of scientific review papers, documents were found describing Merck marketing employees developing plans for manuscripts, contracting with medical publishing companies to ghostwrite manuscripts, and recruiting external, academically affiliated investigators to be authors.
- Recruited authors were commonly the sole author on the manuscript and offered honoraria for their participation.
- ... only 50% (36 of 72) of review articles published either a disclosure of Merck sponsorship or any financial compensation from the company.

http://jama.ama-assn.org/cgi/content/full/299/15/1800

Quo Vadis?

- ACA + ARRA/HITECH ⇒ Universal EMR Adoption
  - Poor implementations
  - Drug names, interaction alerts
  - Delays due to input demands, overwhelming volume, copy/paste—usability
  - Data held hostage, poor exchange standards
  - Regulatory “friction” coming
  - Computation vs. Communication
- Christensen’s “Prescription”:
  1. Precision diagnostics
  2. Specialized, highly focused treatment
  3. Social network to care for chronic disease
- Genetics should help with 1 & 2.
Do Good

- ...and do well!

- Don’t forget Subject Evaluations!
  - [http://web.mit.edu/subjectevaluation](http://web.mit.edu/subjectevaluation)
  - until Monday, 12/16, 9am
  - Please write comments and suggestions