A. Holzinger  
VO 709.049 Medical Informatics  
13.01.2016 11:15-12:45  
Lecture 11  
Biomedical Data:  
Privacy, Safety, Security, Data Protection  
a.holzinger@tugraz.at  
Tutor: markus.plass@student.tugraz.at  
http://hci-kdd.org/biomedical-informatics-big-data

Learning Goals: At the end of this 11th lecture you ...  
- are able to determine between privacy, safety and security;  
- know the famous IOM report “Why do accidents happen” and its influence on safety engineering;  
- have a basic understanding of human error and are able to determine types of adverse events in medicine and health care;  
- have seen some examples on how ubiquitous computing might contribute to enhancing patient safety;  
- got an idea of the principles of context-aware patient safety;  
- saw a recent approach about pseudonymization for privacy in e-health;  
- are aware of the security characteristics of the popular personal health records;

Schedule
- 1. Intro: Computer Science meets Life Sciences, challenges, future directions  
- 3. Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)  
- 4. Biomedical Databases: Acquisition, Storage, Information Retrieval and Use  
- 5. Semi structured and weakly structured data (structural homologies)  
- 6. Multimedia Data Mining and Knowledge Discovery  
- 8. Biomedical Decision Making: Reasoning and Decision Support  
- 9. Intelligent Information Visualization and Visual Analytics  
- 10. Biomedical Information Systems and Medical Knowledge Management  
- 11. Biomedical Data: Privacy, Safety and Security  

Keywords of the 11th Lecture
- Adverse events  
- Anonymization  
- Context aware patient safety  
- Faults and Human error  
- Medical errors  
- Privacy  
- Pseudonymization  
- Privacy aware machine learning  
- Safety and Security  
- Swiss-Cheese Model of human error  
- Technical dependability

Advance Organizer (1/3)
- Acceptable Risk = the residual risk remaining after identification/reporting of hazards and the acceptance of those risks;  
- Adverse event = harmful, undesired effect resulting from a medication or other intervention such as surgery;  
- Anonymization = important method of de-identification to protect the privacy of health information (anonym: re-identification);  
- Authentication = to verify the identity of a user (or other entity, could also be another device), as a prerequisite to allow access to the system; also: to verify the integrity of the stored data to possible unauthorized modification;  
- Confidentiality = The rule dates back to at least the Hippocratic Oath: “Whatever, in connection with my professional service, or not in connection with it, I see or hear, in the life of man, which ought not to be spoken of abroad, I will not divulge, as reckoning that all such should be kept secret”;  
- Data protection = ensuring that personal data is not processed without the knowledge and the consent of the data owner (e.g. patient);  
- Data security = includes confidentiality, integrity, and availability of data, and helps to ensure privacy;  
- Hazard = the potential for adverse effects, but not the effect (accident) itself; hazards are just contributory events that might lead to a final adverse outcome;  
- Human fallibility = addresses the fundamental sensory, cognitive, and motor limitations of humans that predispose them to error;

Advance Organizer (2/3)
- k-Anonymity = an approach to counter linking attacks using quasi-identifiers, where a table satisfies k-anonymity if every record in the table is indistinguishable from at least k – 1 other records with respect to every set of quasi-identifier attributes; hence, for every combination of values of the quasi-identifiers in the k-anonymous table, there are at least k records that share those values, which ensures that individuals cannot be uniquely identified by linking attacks;  
- Medical error = any kind of adverse effect of care, whether or not harmful to the patient, including inaccurateness, incompleteness of a diagnosis, treatment etc.;  
- Nomen nescio (N.N.) = used to sign an anonymous non-specific person;  
- Patient safety = in healthcare this is the equivalent of systems safety in industry;  
- Personally-identifying information = can be used to connect a medical record back to an identified person;  
- Prevention = any action directed to preventing illness and promoting health to reduce the need for secondary or tertiary health care; including the assessment of disease risk and raising public health awareness;  
- Privacy = [dis pro. “pri.” – UK pro. “pri.” – from Latin privatus “separated from the rest”], is the individual rights of people to protect their personal life and matters from the outside world;  
- Privacy policy = organizational access rules and obligations on privacy, use and disclosure of data;
Slide 11-1 Key Challenges

- Sensitive, Personal Health Data
- Mobile solutions, Cloud solutions
- Primary use of Data
- Secondary use of Data for Research
- In the medical area ALL aspects require strict

Privacy, Safety, Security and Data Protection!


Slide 11-2 We start with thinking about safety first ...


Slide 11-3 Exposure of catastrophes - associated deaths

- Safety = any protection from harm, injury, or damage;
- Data Protection = all measures to ensure availability and integrity of data
- Privacy = (US pron. “prai . . .”); UK pron. “pri . . .”; from Latin: privatus “separated from the rest”, are the individual rights of people to protect their personal life and matters Confidentiality = secrecy (“ärztliche Schweigepflicht”)

- **Availability** = \( p(x) \) that a system is operational at a given time, i.e. the amount of time a device is actually operating as the percentage of total time it should be operating;
- **Reliability** = the probability that a system will produce correct outputs up to some given time;
- **Security** = (in terms of computer, data, information security) means protecting from unauthorized access, use, modification, disruption or destruction etc.;
- **Dependability** = the system property that integrates such attributes as reliability, availability, safety, survivability, maintainability (see slide 11-22);

![Availability](http://www.ares-conference.eu)

**Slide 11-5 The famous report “Why do accidents happen”**

![Dependability](http://www.ares-conference.eu)

**Slide 11-6 The impact of the “To err is human” IOM study**

![Reliability](http://www.ares-conference.eu)

**Slide 11-7 Research activities stimulated by the IOM report**

![Security](http://www.ares-conference.eu)

**Slide 11-8 Deaths from medical error (2009) …**
Slide 11-9 Medical Error Example: Wrong-Site Surgery

Integration of a correct surgery site protocol into a daily patient care model is a useful step in preventing occurrences of wrong site dermatologic surgery.

A. Holzinger 706.049

Slide 11-10 Deal with errors: Eindhoven Classification Model


A. Holzinger 706.049

Slide 11-11 Adverse event reporting and learning system

3 Modules: AERFMI = Adverse Events Reporting Forms in Medical Imaging AEKMI = Adverse Events Knowledge Manager in Medical Imaging AEKMI = Adverse Events Knowledge Manager in Medical Imaging Rodrigues et al. (2010)

A. Holzinger 706.049

Slide 11-12 Re: Framework for understanding human error


A. Holzinger 706.049

Slide 11-13 Reason (1997) Swiss Cheese Model


A. Holzinger 706.049

Slide 11-14 Risk management - FAA System Safety

Note: Now just definitions, refer to risk management in Lecture 12

A. Holzinger 706.049
Slide 11-15: Improving Safety with IT – Example Mobile


A. Holzinger 11/01/2015

Slide 11-16: Enhancing Patient Safety with ubiquitous devices


A. Holzinger 11/01/2015

Slide 11-17: Security Problems of ubiquitous computing

1) Protection precautions:
   1) vulnerability to eavesdropping,
   2) traffic analysis,
   3) spoofing and denial of service.
   4) Security objectives, such as confidentiality, integrity, availability, authentication, authorization, nonrepudiation and anonymity are not achieved unless special security mechanisms are integrated into the system.
2) Confidentiality: the communication between reader and tag is unprotected, except of high-end systems (ISO 14443). Consequently, eavesdroppers can listen in if they are in immediate vicinity.
3) Integrity: With the exception of high-end systems which use message authentication codes (MACs), the integrity of transmitted information cannot be assured. Checksums (cyclic redundancy checks, CRCs) are used, but protect only against random failures. The writable tag memory can be manipulated if access control is not implemented.

Holzinger, A. & Norskov (2008)

A. Holzinger 11/01/2015

Slide 11-18 Clinical Example: Context-aware patient safety 1/2


A. Holzinger 11/01/2015

Slide 11-19 Clinical Example: Context aware patient safety 2/2


A. Holzinger 11/01/2015

Slide 11-20 Patient Safety

- (1) measuring risk and planning the ideal defense model,
- (2) assessing the model against the real behavior of professionals, and modifying the model or inducing a change in behavior when there are gaps,
- (3) adopting a better micro- and macro-organization,
- (4) gradually re-introducing within the rather rigid, prescriptive system built in steps 1–3 some level of resilience enabling it to adapt to crises and exceptional situations.
Types of adverse events in medicine and care


Types of faults: Design – Physical – Interaction


A Two-Tiered System of Medicine


Toward a strategic view on safety in health care


Data ...
Requirements of an electronic patient record

Anonymization: Personal data cannot be re-identified (e.g. k-Anonymity)
Pseudonymization: The personal data is replaced by a “pseudonym”, which allows later tracking back to the source data (re-identification)

Pseudonymization of Information for Privacy

Three-Layer Based Security Model

Identifiers (PKI)
- Private Key
- Public Key
- Certificate

Data Owner (Patient)
- Personal data
- Non-personal data
- Access
- Consent

Beneficiary (HCP)
- Personal data
- Non-personal data
- Access
- Consent

Three-Layer Based Security Model

Note: Similar to authorization, a user affiliation requires that both the patient as data owner and the trusted relative as affiliated user are authenticated at the same workstation. Consequently, both user identifiers are transferred to the pseudonymization server where they are encrypted with both the users’ inner symmetric keys. The patient’s inner private key is also encrypted with the relative’s inner symmetric key, and all elements are stored in the pseudonymization metadata storage as affiliation relation.


Example:
http://healthbutler.com/
https://www.healthcompanion.com

Example for component relationships

Machine Learning and Data Privacy ...

- Lawfulness and fairness
- Necessity of data collection and processing
- Purpose specification and purpose binding
- There are no "non-sensitive" data
- Transparency
- Data subject's right to information correction, erasure or blocking of incorrect/legally stored data
- Supervision (= control by independent data protection authority) & sanctions
- Adequate organizational and technical safeguards
- Privacy protection can be undertaken by:
  - Privacy and data protection laws promoted by government
  - Self-regulation for fair information practices by codes of conducts promoted by businesses
  - Privacy-enhancing technologies (PETs) adopted by individuals
  - Privacy education of consumers and IT professionals


Example: Differentially Private Kernel Learning

(a) Interactive Model
(b) Semi-Interactive model
(c) Non-interactive Model
Simplest Privacy Metric

- The larger the set of indistinguishable entities, the lower probability of identifying any one of them

“Hiding in a crowd”

Anonymity set \( A \)

\[ A = \{ (s_1, p_1), (s_2, p_2), \ldots, (s_n, p_n) \} \]

- \( s_i \): subject who might access private data
- \( p_i \): probability that \( s_i \) accessed private data
- \( w_i \): weight of the attribute assumes the \( i \)-th possible value

Example: Entropy

- Remember: Entropy measures the randomness (uncertainty) – here private data
- Violator gains more information -> entropy decreases!
- Metric: Compare the current entropy value with its maximum value and the difference shows how much information has been leaked

Privacy loss \( D(A,t) \) at time \( t \), when a subset of attribute values \( A \) might have been disclosed:

\[ D(A,t) = H^*(A) - H(A,t) \]

- \( H^*(A) \): the maximum entropy
- \( H(A,t) \): entropy at time \( t \)
- \( w_i \): weights capturing relative privacy "value" of attributes

Effective Anonymity Set Size

- Effective anonymity set size is calculated by

\[ L = |A| \sum_{i=1}^{n} \min p_i \frac{1}{|A|} \]

Maximum value of \( L \) is \( |A| \) iff \( p_i = 1/|A| \)

L below maximum when distribution is skewed skewed when \( p_i \) have different values

Deficiency:

- \( L \) does not consider violator’s learning behavior

Example: Anonymization of Medical Data

87% of the population in the USA can be uniquely re-identified by Zip-Code, Gender and date of birth

Anonymization of Patient Data

- **K-Anonymity** ... not fully protected against attribute disclosure
- **L-Diversity** ... extension requiring that the values of all confidential attributes within a group of \( k \) sets contain at least \( l \) clearly distinct values
- **t-Closeness** ... extension requiring that the distribution of the confidential attribute within a group of \( k \) records is similar to the confidential attribute in the whole data set

Three Examples of Freeware

- **Argus**: [http://neon.vb.cbs.nl/casc](http://neon.vb.cbs.nl/casc)
- **ARX**: [http://arx.deidentifier.org](http://arx.deidentifier.org)
- **sdcTable**: [http://cran.r-project.org/web/packages/sdcTable/](http://cran.r-project.org/web/packages/sdcTable/)
Privacy Aware Machine Learning for Health Data Science

- Production of Open Data Sets
- Design of Synthetic data sets
- Privacy preserving ML, DM & KDD
- Data leak detection
- Data citation
- Differential privacy
- Anonymization and pseudonymization
- Securing expert-in-the-loop machine learning systems
- Evaluation and benchmarking

Slide 11-45 Future Outlook

- Privacy, Security, Safety and Data Protection are of enormous increasing importance in the future.
- Trend to mobile and cloud computing approaches.
- EHR are the fastest growing application which concern data privacy and informed patient consent.
- Personal health data are being stored for the purpose of maintaining a life-long health record.
- Secondary use of data, providing patient data for research.
- Production of Open Data to support international research efforts (e.g. cancer) without boundaries.
- Data citation approaches are needed for full transparency and replicability of research ...

Sample Questions (1)

- What is the core essence of the famous IOM report "Why do accidents happen"?
- What is a typical ultrasafe system – what is an example for a high risk activity?
- Which influence had the IOM report on safety engineering?
- What are the differences between the concepts of Privacy, Security and Safety?
- Why is privacy important in the health care domain?
- How do you classify errors when following the Eindhoven Classification Model?
- Please describe the basic architecture of a adverse event reporting and learning system?
- What is a typical example for medical errors?
- Please, explain the Swiss-Cheese Model of Human Error!

Sample Questions (2)

- What factors does the framework for understanding human error include?
- Which possibilities does ubiquitous computing offer to contribute towards enhancing patient safety?
- What different types of risk does the FAA System Safety Guideline explain?
- Ubiquitous computing offers benefits for health care, but which genuine security problems does ubiquitous computing bring?
- How can mobile computing device help in terms of patient safety?
- What is a context-aware patient safety approach?
- How can we describe patient safety both quantitatively and qualitatively?
- What is technical dependability?
- Which types of technical faults can be determined?

Sample Questions (3)

- What types of adverse events can be discriminated in medicine and health care?
- How is the safety level (measurement) defined?
- Which factors contribute to ultrasafe health care?
- What are the typical requirements of any electronic patient record?
- Why is Pseudonymization important?
- What is the basic idea of k-Anonymization?
- What is a potential threat of private personal health records?
- Please describe the concept of a personal health record system!
- How would you analyze personal health record systems?
- What does a privacy policy describe?
- Which ethical issues are related to quality improvement?
Appendix: Advances in patient safety are hampered by...

http://www.dh.gov.uk/en/Managingyourorganisation/Informationpolicy/Informationpolicyparadigmshift/index_en.htm (European Commissioner’s Office in the UK)

http://ec.europa.eu/justice/data/privacy_small_en.htm (European Commission Protection of private personal data)

http://www.dsk.gv.at/(Österreichische Datenschutz Kommission)

http://www.nap.edu/openbook.php?isbn=0309068371 (National Academy Press, To err is human)

http://medical-dictionary.thefreedictionary.com (medical dictionary and thesaurus)


Appendix: Example for a simple warning message


Appendix: Example for trust policies in HIS networks


Appendix: Some Useful Links

http://www.nap.edu/openbook.php?isbn=0309068371 (National Academy Press, To err is human)

http://medical-dictionary.thefreedictionary.com (medical dictionary and thesaurus)

http://www.ico.gov.uk (Information Commissioner’s Office in the UK)

http://ec.europa.eu/justice/data/privacy_small_en.htm (European Commission Protection of private personal data)

http://www.dsk.gv.at/(Österreichische Datenschutz Kommission)

http://videolectures.net/kdd09_mohammed_adhdsb (Anonymizing Healthcare Data: A Case Study on the Blood Transfusion Service)


Appendix: Example of new threats to health data privacy

A real-world example of cross-site information aggregation: The target patient “Jean” has profiles on two online medical social networking sites (1 and 2). By comparing the attributes from both profiles, the adversary can link the two with high confidence. The attacker can use the attribute values to get more profiles of the target through searching the Web (3) and other online public data sets (4 and 5). By aggregating and associating the five profiles, Jean’s full name, date of birth, husband’s name, home address, home phone and cell phone number, two email addresses, occupation, medical information including lab test results are disclosed!

Slide 11-40 Security and Privacy of some PHR’s

1) Privacy Policy
- 0. The Privacy Policy is not visible or not accessible.
- 1. The Privacy Policy is accessed by clicking one link.
- 2. The Privacy Policy is accessed by clicking two or more links.

2) Data Source
- 0. Not indicated.
- 1. User.
- 2. User healthcare provider.
- 3. User and his/her healthcare providers.
- 4. User, other authorized users and other services/programs.
- 5. Self-monitoring devices connected with the user.

3) Data Management
- 0. Not indicated.
- 1. Data user.
- 2. Data user and his/her family data.

4) Access management
- 0. Not indicated.
- 1. Other users and services/programs.
- 3. Other users.
- 4. Other users, healthcare professionals and services/programs.

---

5) Access audit
- 0. No.
- 1. Yes.

6) Data access without the end user's permission
- 0. Not indicated.
- 1. Information related to the access.
- 2. De-identified user information.
- 3. Information related to the access and de-identified user information.
- 4. Information related to the access and identified user information.

7) Security measures
- 0. Not indicated.
- 1. Physical security measures.
- 2. Electronic security measures.
- 3. Physical security measures and electronic security measures.

8) Changes in Privacy Policy
- 0. Not indicated.
- 1. Changes are notified to users.
- 2. Changes are announced on home page.
- 3. Changes are notified to users and changes are announced on home page.
- 4. Changes may not be notified.

9) Standards
- 0. Not indicated.
- 1. HIPAA is mentioned.
- 2. System is covered by HONcode (HON = Health on the Net).
- 3. HIPAA is mentioned and system is covered by HONcode.

---

Legend: PL = Privacy policy location; DS = Data source; SM = Data managed; AM = Access management; AA = Access audit; DA = Data accessed without the user’s permission; SM = Security measures; CP = Changes in privacy policy; S = Standards

---

Tapp et al. (2009) Quality improvement in primary care: ethical issues explored.
International Journal of Health Care Quality Assurance, 22, 1, 8-29.