

# Protecting Patients' Privacy in Database Research

**Dr. Zoltán Alexin, PhD.**  
**University of Szeged,**  
Department of Software Engineering  
H-6720 Szeged Árpád tér 2.  
e-mail: [alexin@inf.u-szeged.hu](mailto:alexin@inf.u-szeged.hu)  
<http://www.inf.u-szeged.hu/~alexin>



<http://www.futureict.eu>

TeleMediCare 2013, III. International Conference on TeleMedicine & TeleCare, 2nd October, 2013. Desio, Italy

# Who am I

- Mathematician, PhD. on application of machine learning algorithms to text analysis
- Created and supervised several R&D projects on this topic
- Since 2004, I began studying privacy issues
- Member of a regional medical research ethics committee
- Contributed in the Association on Fair Data Processing
- Member of the presidential board of the Hungarian Data Protection Society
- Blogger ([www.magyarorszag.hu](http://www.magyarorszag.hu))
- I have cases before Civil Courts, Hungarian Constitutional Court, European Commission, ECtHR on fundamental questions of medical data processing
- Achievement: excluding data from the National Health Insurance Fund database related to unsubsidized care events, ethics approval of medical research projects without intervention, and other minor results

# Content

- Benefits and dangers of database research
- Purposes, for which medical data can be processed
- The difference between administration and research
- Legal and ethical requirements for processing without consent
- What is anonymization and how should be done?
- An insight to Hungarian regulation
- Summary

# Benefits and dangers of database research



- Benefits of database research
  - Extracting knowledge from large number of cases
  - Exploiting wealth of unused data for public good
  - The cost efficiency of different treatments can be compared
  - Discovering correlations between symptoms
  - Discovering drug interactions
- Dangers of database research to privacy of data subjects
  - The privacy and dignity of the patients might be compromised
  - Discrimination upon health status or adherence (good or bad patient)
  - Patients avoid physicians or will not tell important facts about themselves

# Purposes, for which medical data can be processed

- Primary
  - For the prevention, diagnosis or treatment of data subject or of his family members
- Secondary
  - Administration of medical institutions (Quality Control, i.e. Assessment, Assurance, Improvement)
  - Public health (prevention and improvement)
  - Medical research (compatible with any other purposes)
- Ternary
  - Legal cases, courts, police, national security, public security

# Legal bases for data processing for medical research purposes

- Council of Europe, Committee of Ministers, Recommendation No. R (97) 5 on the Protection of Medical Data Art 4.3.
- Obligatory data processing by law
  - EU 95/46/EC Data protection directive, Preamble (34), „where important reasons of public interest ... for research”
- Data processing is permitted by law
  - EU 95/46/EC Art 7. e) „processing is necessary for the performance of a task carried out in the public interest”
- Data subject consents to data processing
- In all cases there must be **appropriate safeguards!**

# Administration and research

## What is the difference?

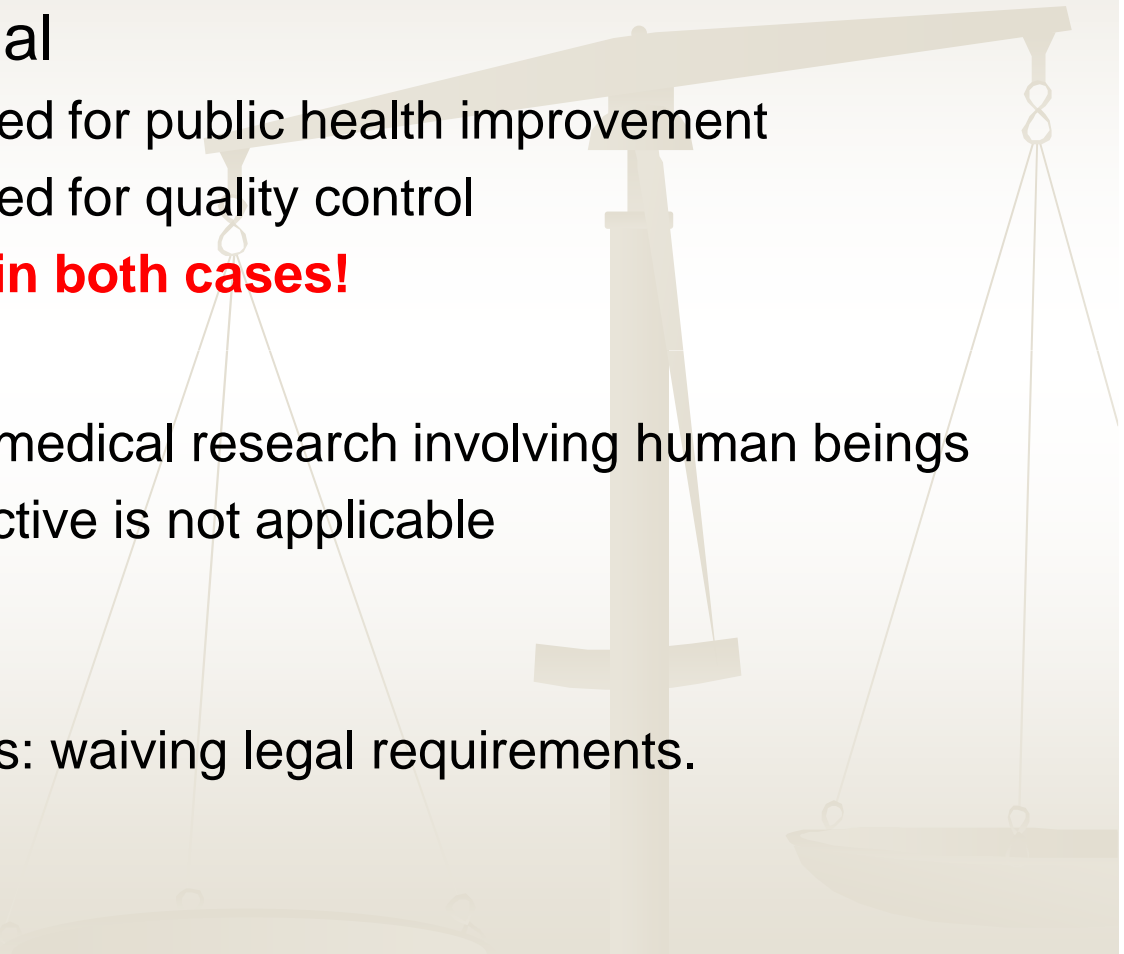
- Committee on the Role of Institutional Review Boards in Health Services Research Data Privacy Protection, National Academy Press, Washington, D.C. 2001: Protecting Data Privacy in Health Services Research, page 11.
- The following are characteristics of projects using HSR (Health Science Research) methods that are **research**, not Quality Assessment or Quality Improvement:
  - It explores previously unknown phenomena.
  - It collects information beyond that routinely collected for the patient care in question.
  - It compares alternative treatments, interventions, or processes.
  - It manipulates a current process.
  - The results are expected to be published for general societal benefit.



# Ethical requirements

- World Medical Association – Helsinki Declaration, Ethical Principles for Medical Research Involving Human Subjects
  - Written protocol, approved by independent ethics committee
  - Publication of approved research projects
  - Scientifically qualified leader (PhD)
  - Informed consent, without pressure (It has been modified in 2008: *There may be situations where consent would be **impossible** or **impractical** to obtain for such research **or would pose a threat to the validity of the research**. In such situations the research may be done only after consideration and approval of a research ethics committee.*)
  - Research project must be beneficial for the research subjects, they must be informed
  - Right to object (in advance as well)
  - Right to withdraw consent.

# Can legal and ethical requirements be waived?



- Data are clearly personal
  - Say that data is processed for public health improvement
  - Say that data is processed for quality control
  - **No consent is needed in both cases!**
- Data are not personal
  - The processing is not a medical research involving human beings
  - The data protection directive is not applicable
- **Let's go to anonymize!**
- The goal of anonymization is: waiving legal requirements.

# Questions on anonymization

- Anonymized information:
  - HIPAA CFR 45 164.514: does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual
  - 95/46/EC Art 2. a) not identifiable person is the one who cannot be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;
- What if data are pseudonymised (personal identifiers are replaced by numbers and the correspondence is retained)?
  - Those are not anonymous (EDPS statement, 2013. 15th March)
- What if all natural personal identifiers are removed but data remain identifiable?
  - What is the acceptable risk? (Physicians: 33%, mathematicians: less 0.1%)

# Mathematical concepts

- Generalization: reducing the accuracy of data, replacing date of birth with year of birth; or quantizing numeric data
- Quasi-identifier: it is not an identifying characteristic but combining with other quasi identifiers could identify a person (e.g. ZIP, date of birth, gender).
- Measure of identifiability equals to probability of re-identification
  - $\text{Pr}(\text{id}) = \# \text{identifiable\_people} / \# \text{subjects\_in\_the\_database}$
- k-Anonymity: for all combinations of quasi identifiers, there are at least k indistinguishable records in the database. The re-identification risk is less than  $1/k$ .
  - A table is 3-anonymous: for all sets of quasi-identifiers there are at least 3 records in the database.

# What do people think of it?

- Survey of Californian Healthcare Foundation, 2010. (ca. 1950 people)
- In the United States 5-10% of the people are using PHR (Personal Health Record).
- The survey tried to detect the reasons why do potential users keep themselves off from this new service.
- 75% of the non-PHR users have concerns about the privacy of their medical information.
- When sharing medical information a majority of all respondents express discomfort (42%) or uncertainty (25%) with their health information being shared with other organizations even if their name, address, date of birth and social security number were not included.
- On the question “If your doctor had a computer system that is capable of sharing medical information (without name, date of birth, social security number) would there be anything that you would not tell?” – 50% told no, 49% told yes, or not sure, and 1% did not answer.

# Current state in Hungary

- Data are forwarded to national medical databases by the force of the law, without informing data subjects
- Data are processed always for administration, never for research and therefore ethical approval is never asked
- Local data (at hospitals or clinics) are processed for research purposes by approval of a REC but consent is never needed. RECs are incompetent in data protection.
  - The law exposes all existing data for research without consent, and providing right to object.
- They do not understand maths. Recently an authorization was given on using ZIP, date of birth, gender data for health government bodies
  - The re-identification risk is 78%, in countryside or among elderly is above 90%.

# Summary

- All stakeholders agree in regulating medical research
- Medical researchers want free hand when doing research
- Ethical norms are elaborated by doctors not listening to lawyers and mathematicians
- Laws are made by lawyers, doctors trying to express their interests and views
- Nobody understands mathematics
- Patients are never asked.



**Thanks for the attention!**