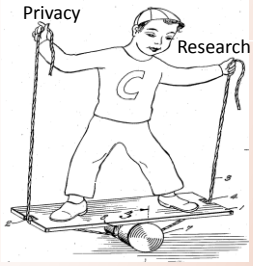
UiT
THE ARCTIC UNIVERSITY OF NORWAY

Societal benefits vs. privacy: what distributed secure multi-party computation enable?



eHelse 2015
21 - 22 April
Oslo

Kassaye Yitbarek Yigzaw
UiT The Arctic University of Norway

Outline

- Background
- Introduction
- Data utility
- De-identification
- Secure multi-party computation

Background

- Electronic health data are being widely collected
- Administrative data, e.g. census, survey, socioeconomic, and registry
- Invaluable resource to improve healthcare systems´ effectiveness, efficiencies and quality of care
- Enormous benefits for individuals and society in general

Jutte DP et al. Administrative Record Linkage as a Tool for Public Health Research. Annual Review of Public Health 2011
Geissbuhler A et al. Trustworthy reuse of health data: A transnational perspective. International Journal of Medical Informatics 2013

Data reuse opportunities

- Why are healthcare costs increasing?
- What are the comparative benefits and risks of prescription drugs?
- What is the evidence base for procedures?
- What explains variation in health care spending and use?
- How do environmental factors affect disease patterns?
- How can the health of minorities and special needs groups be improved?
- What does this mean for patients like me?

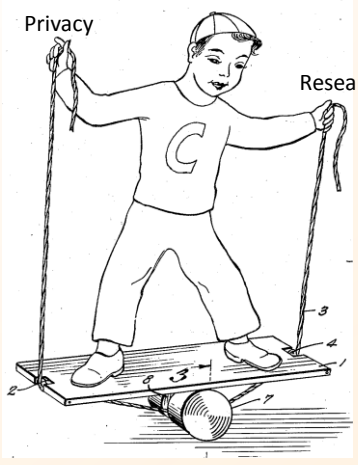
Slide borrowed from “Michael G. Kahn. Learning Health Systems From Concept to National Deployment. IEEE BHI2014 Conference”

Introduction

- Individuals privacy concerns are the main challenge
- Healthcare institutions are also concerned about their own privacy[1]
- Most ethical and legal regulations allow data reuse through:
  - Informed consent
  - Consent waiver by ethics committee (e.g. REK) (under certain conditions)
  - Data de-identification
- Secure multi-party computation (SMC)

[1]El Emam K, et al. Physician privacy concerns when disclosing patient data for public health purposes during a pandemic influenza outbreak. BMC Public Health 2011;11:454.

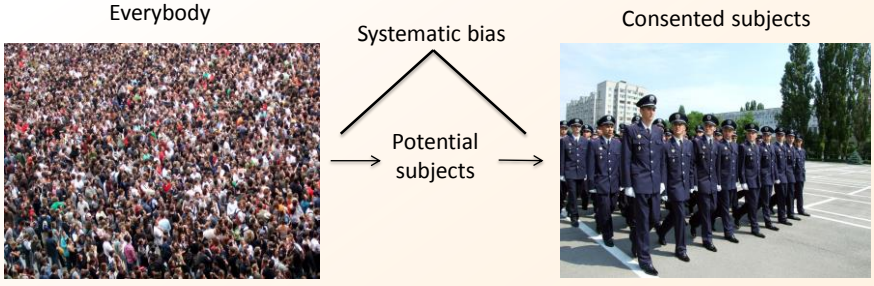
Increase privacy protection and data utility



Data utility

- Data utility is the value of a given data for research
- The data being available for research
- Analytical completeness is possible required analyses that can be done on the data
- Analytical validity is accuracy and generalizability of analyses results

Informed consent



- Decreased generalizability of analyses results to the general population

Bohensky MA et al. Data Linkage: A powerful research tool with potential problems. BMC Health Services Research 2010

Informed consent cont…

- Consent requires infeasible time and money for large public health studies
- Consent alone protects autonomy, but does not guarantees that released data will remain private[1]

[1]Taylor P. Personal genomes: when consent gets in the way. Nature 2008

De-identification

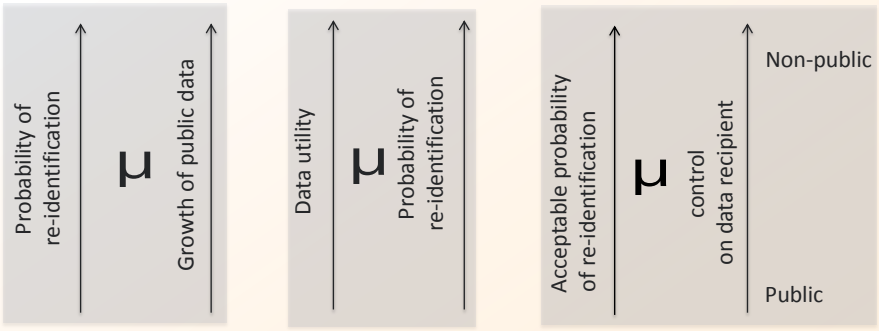
- Data de-identification methods remove or modify identifiers
- The aim is to prevent identity disclosure
- Often, there is probability of assigning correct identity (re-identification) to records
- The more quasi-identifiers are removed or modified
  - Less probability of re-identification
  - Less analytical completeness or validity

De-identification cont…

- HIPAA safe harbor removes 18 identifiers
- Limited dataset removes 16 of the 18 identifiers, except dates and some geographical data
- Safe harbor has less analytical completeness, e.g. association between treatments and health outcomes[1]
- Limited dataset has more probability of re-identification[2]

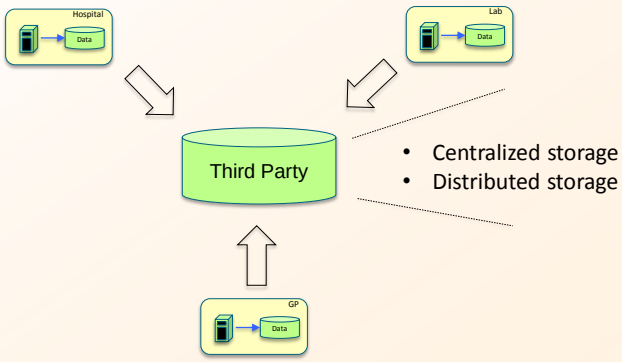
[1]Nass SJ et al. Beyond the HIPAA Privacy Rule: Enhancing privacy, improving health through research. National Academies Press; 2009
[2]Benitez K et al. Evaluating re-identification risks with respect to the HIPAA privacy rule. JAMIA 2010

De-identification cont…

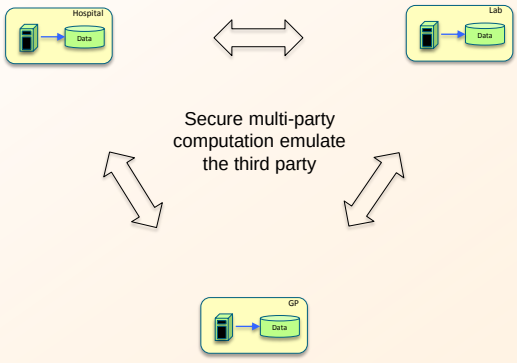


Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. Rochester, NY: Social Science Research Network; 2009
Emam KE at al. Anonymising and sharing individual patient data. BMJ 2015.
Benitez K et al. Evaluating re-identification risks with respect to the HIPAA privacy rule. JAMIA 2010

De-identified data sharing



Secure multi-party computation (SMC)



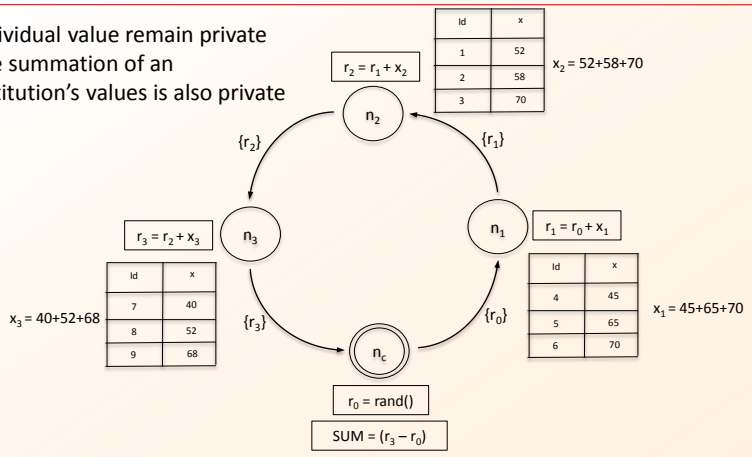
Lindell Y, et al. Secure multiparty computation for privacy-preserving data mining. Journal of Privacy and Confidentiality 2009

Secure multi-party computation (SMC)

- SMC ensures that no more private information is revealed beyond the computation output
- Each data custodian has the capability of determining what/who compute on their data

Secure summation protocol

- ✓ Individual value remain private
- ✓ The summation of an institution's values is also private



Andersen A, Yigzaw KY, Karlsen R. Privacy preserving health data processing. IEEE Healthcom, 2014

Example applications

- SMC protocol for disease surveillance[1]
- SMC protocol for logistic regression tested for detection of adverse drug events[2]

[1]El Emam K et al. A secure protocol for protecting the identity of providers when disclosing data for disease surveillance JAMIA 2011
[2]El Emam K et al. A secure distributed logistic regression protocol for the detection of rare adverse drug events. JAMIA 2013

History

- The SMC concept is pioneered by Yao (1982)
- Most SMC protocols were designed to show feasibility
- During the last decade better SMC protocols and implementations started to appear
- SMC protocols vary with privacy guaranty, efficiency and scalability
- Strong privacy is often achieved using more complex techniques, which are less efficient and scalable

Bogdanov D. Sharemind: programmable secure computations with practical applications. PhD Thesis. Tartu University, 2013.

Discussion

- SMC techniques do not modify or remove data attributes and do not have selection bias, therefore data utility is not affected
- More efficient and scalable techniques are being developed
- Protects the privacy of both individuals and health institutions
- Enable health institutions to maintain strong control over their private data
- These could encourage more individuals and institutions to allow data reuse

Acknowledgement

- PhD supervisors (Johan Gustav Bellika, Anders Andersen, and Gunnar Hartvigsen)
- Meskerem Asfaw Hailemichael
- Tromsø Telemedicine Laboratory (TTL)
- UiT The Arctic University of Norway
- Norwegian Center for Telemedicine and Integrate Care (NST)

Thank you for your attention!

Kassaye Yitbarek Yigzaw
PhD candidate
UiT The Arctic University of Norway
kassaye.y.yigzaw@uit.no

http://www.panoramio.com/photo/10889343