Big medicine, big data, big ethics

Prof. Dr. med. B. Elger, 24.06.16

Plan of the presentation

Aim: Provide an overview of the most recent discussion of ethical issues concerning Big Data research.

➤ Definition: what is Big Data?

➤ Examples of Big Data research in the health sciences

➤ Ethical and legal issues in Big Data research

Universität Basel   2

Plan of the presentation

➤ **Ethical and legal issues in Big Data research**

• Privacy, confidentiality, anonymisation

• Respect for participant autonomy: informed consent in Big Data health research

• Conflicts of interests

• Justice: new digital divides

• Is there a Big Data exceptionalism or should traditional research regulations apply to Big Data health research?

Universität Basel   3

Definition
What is Big Data?

Big vs. small

➤ Big science

➤ Big data vs. small or little data

➤ Big data, big ethics?

➤ Big ethics (consortia ethics) vs. small ethics

Definition

➤ "The three key features of Big Data are volume, velocity and variety":

– **Volume:** "vast amount of data generated and available for storage (e.g. 1.15 billion Facebook users worldwide)"

– **Velocity:** "the speed at which this data is created"

– **Variety:** "there are different kinds of data (medical, sociodemographic, financial and virtually any field), and in addition there is a variety of ways in which Big Data can be processed"

Casañas i Comabella C, et al. Social media and palliative care research. BMJ Supportive & Palliative Care 2015;5:138–145. doi:10.1136/bmjspcare-2014-000708 139

Universität Basel   6

Slide 1

Universität Basel | **Definition** | UNIVERSITÉ DE GENÈVE

- "Large collection of disparate data sets that, taken together, can be analyzed to find unusual trends."
- "Big Data is a promising technology with the potential to identify previously unsuspected or unproven associations between disparate data points."
- "Health information may be gleaned from unusual sources — beyond the usual clinical records, research data, and reports about drugs and medical devices. To find associations between everyday activities and health, Big Data researchers have used, among other things, consumer purchasing records and social media."

Rothstein MA. Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics. The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics 2015;43(2):425-9

Universität Basel   7

Slide 2

Universität Basel | **Definition** | UNIVERSITÉ DE GENÈVE

- "There has been much written about Big Data's promises and possibilities as well as its limitations and unrealistic expectations."
- "Is there is anything unique about research using Big Data? Is there a 'Big Data exceptionalism?'"

Rothstein 2015

Universität Basel   8

Slide 3

Universität Basel | UNIVERSITÉ DE GENÈVE

Examples of
big data research in the health
sciences

Slide 4

Universität Basel | **a. Predictive analytics** | UNIVERSITÉ DE GENÈVE

- "Predictive analytics models are already being deployed to help identify in real time high-risk patients", e.g. at risk for cardiopulmonary arrest.
- "In the near future, models based on machine learning (that is, ones using a computer that can learn from data instead of requiring additional programming) will be able to instantaneously consider the risk of all patients in a hospital, their individual therapeutic goals and preferences, hospital staffing (including staff members' experience and performance), resource constraints, and external conditions such as whether other hospitals are diverting patients in the emergency department in the case of a disaster."
- "The model could then advise hospital administrators on whom to admit to the ICU and how to staff it."

Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood) 2014;33(7):1139-47

Universität Basel   10

Slide 5

Universität Basel | **b. Administrative data from medical records** | UNIVERSITÉ DE GENÈVE

SUPPLEMENT ARTICLE

"Big Data" Versus "Big Brother": On the Appropriate Use of Large-scale Data Collections in Pediatrics

- "Discussions of "big data" in medicine often revolve around gene sequencing and biosamples."
- "It is perhaps less recognized that administrative data in the form of vital records, hospital discharge abstracts, insurance claims, and other routinely collected data also offer the potential for using information from hundreds of thousands, if not millions, of people to answer important questions."

Currie J. "Big data" versus "big brother": on the appropriate use of large-scale data collections in pediatrics. Pediatrics 2013;131 Suppl 2:S127-32.

Universität Basel   11

Slide 6

Universität Basel | **b. Administrative data from medical records** | UNIVERSITÉ DE GENÈVE

- "For example, large data sets can be used to uncover the effects of exposures that may have small effects on individuals but large cumulative effects on populations."
- "They may also be used to identify subgroups in which there are effects (compared with clinical studies in which it may be difficult to recruit enough members of different subgroups to be able to identify significant differences) or to study rare conditions."
- "Third, administrative records can be used to track individuals over time, and so are ideally suited to measuring the long-term impacts of health conditions or interventions."

Currie 2013

Universität Basel   12

c. Social media and palliative care research

➤ "A recent review of palliative care research revealed that 39% of studies excluded patients that were considered as too ill by the healthcare professionals."

➤ "Use of SM allows a much more direct, fluid communication with the population of interest."

➤ "People near the end of life are often hard to reach populations in healthcare research. However, research has shown that a vast majority of seriously ill people would be happy to be asked to participate in research. Furthermore, policymakers may lack information from dying people, simply because they are considerably difficult to recruit."

Casanas i Comabella 2015

Universität Basel 13

---

c. Social media and palliative care research

➤ "When utilising the traditional methods of recruitment, screening of patients based on the inclusion criteria is usually undertaken by the medical team. For researchers based in the academic context, not only the identification of the patients but also the first contact with the potential participants is usually performed by the relevant clinicians.

➤ Recruitment in SM is based on self-referral process."

Casanas i Comabella 2015

Universität Basel 14

Research using social media

➤ "Studies have attempted to predict important health trends, including the HealthMap Ebola tracker, Google Flu Trends, and Twitter Influenza Surveillance."

Rothstein 2015

Universität Basel 15

Ethical and legal issues

Is there a Big Data exceptionalism or should traditional research regulations apply to Big Data health research?

Privacy, confidentiality, anonymization

Benefits vs. risks

Universität Basel — UNIVERSITÉ DE GENÈVE

➢ The use of Big Data (including administrative data from medical records) for health research raises the traditional ethical question:

– How should one weigh the benefits of the research against the risks to subjects?

Currie 2013

Universität Basel 19

Privacy, confidentiality, anonymization

Universität Basel — UNIVERSITÉ DE GENÈVE

Risk of harm: not physical harm, but related to loss of privacy

➢ **Consequences:** stigmatization, discrimination (health insurance, employers)

Universität Basel 20

Privacy, confidentiality, anonymization

Universität Basel — UNIVERSITÉ DE GENÈVE

Group harm:

Internet research:

➢ "Research findings that reveal uncomfortable information about groups will become the next hot topic in research ethics, e.g. researchers who use Twitter are able to tell uncomfortable truths about specific groups of people, potentially with negative effects on the researched group" (Zwitter 2014).

Predictive analysis:

➢ "Equitable Representation: For instance, the data could be used to identify vulnerable high-risk, high-cost patients and exclude them from care.

➢ "Such concerns could be mitigated by the use of community engagement boards, whose members would advise modelers as they acquired data and designed models" (Cohen et al. 2014)

Zwitter A. Big data ethics. Big Data & Society. July–December 2014: 1–6
Cohen et al, 2014,

Universität Basel 21

Intangible harm

Universität Basel — UNIVERSITÉ DE GENÈVE

Tangible vs. intangible harm

➢ "The argument has been made that codes of research ethics, from Nuremberg to Helsinki to the Common Rule, were only designed to address the risk of physical injuries to human subjects."

➢ What is the significance of "intangible infringements on autonomy, privacy, dignity, and similar interests", may they be "balanced against the societal interest in scientific discovery"?

➢ "Katz, perhaps the greatest authority on research ethics, said 'I want to emphasize . . . the centrality of dignity, not physical injuries, in any appraisal of the ethics of research.' The regulation of research is intended to protect against both physical and dignitary harms."

Rothstein 2015

Universität Basel 22

US: Anonymization, de-identification

Universität Basel — UNIVERSITÉ DE GENÈVE

Protection against harm:

➢ The Health Insurance Portability and Accountability Act (HIPAA) privacy rule (1996) allows the **use of de-identified patient-level data without consent** and accepts two methods of de-identification.

– "First, eighteen defined identifiers, including the patient's city name and email address, could be removed. However, doing so reduces the predictive power of the data set.

– Second, an "individual with appropriate expertise" can declare that "the risk that data can be reidentified is 'very small.'" The objective standards for making such a declaration are still being established."

Cohen et al, 2014

Universität Basel 23

US: Anonymization, de-identification

Universität Basel — UNIVERSITÉ DE GENÈVE

Inform about research and re-identification risks?

➢ "Even a robust de-identification process does not make re-identification impossible, as shown by a recent study on re-identifying people from genomic sequence data.

➢ With enough time and money, someone could re-identify some of the formally de-identified patient-level data.

➢ However, few people will have the resources or motivation to engage in deliberate re-identification.

➢ Still, data breaches remain possible. Therefore, we recommend that collectors of predictive analytics data also notify all patients that the data gathered on them in the course of regular health care may be used in de-identified form in predictive analytics models."

Cohen et al, 2014

Universität Basel 24

Switzerland LHR/HFG

LRH Article 2 al. 2: Champ d'application

Elle [LRH] ne s'applique pas à la recherche pratiquée
a. […]
b. sur du matériel biologique anonymisé;
c. sur des données liées à la santé qui ont été collectées anonymement ou anonymisées.

Universität Basel 25

Switzerland LHR/HFG

LRH Art. 33 Réutilisation de données personnelles non génétiques liées à la santé

1 Les données personnelles non génétiques liées à la santé peuvent être réutilisées à des fins de recherche sous une forme non codée lorsque la personne concernée ou, le cas échéant, son représentant légal ou ses proches ont donné leur consentement éclairé.

2 Les données personnelles non génétiques liées à la santé peuvent être réutilisées à des fins de recherche sous une forme codée lorsque la personne concernée ou, le cas échéant, son représentant légal ou ses proches ne s'y sont pas opposés après avoir été informés.

Universität Basel 26

Switzerland: Anonymization, de-identification

**Ordonnance relative à la recherche sur l'être humain à l'exception des essais cliniques
(Ordonnance relative à la recherche sur l'être humain, ORH)**

Art. 26 Codage

1 Le matériel biologique et les données personnelles liées à la santé sont réputés correctement codés au sens de l'art. 32, al. 2, et 33, al. 2, LRH lorsqu'ils sont qualifiés d'anonymisés dans l'optique d'une personne qui n'a pas d'accès au code.

2 Le code doit être conservé par une personne qui est désignée dans la demande et n'est pas impliquée dans le projet de recherche, séparément du matériel biologique ou des données personnelles.

Universität Basel 27

Switzerland: Anonymization, de-identification

Ordonnance relative à la recherche sur l'être humain (ORH)

Art. 27 Conditions de décodage

Le matériel biologique codé et les données personnelles liées à la santé codées peuvent uniquement être décodés si:

a. le décodage est nécessaire pour prévenir un risque immédiat pour la santé de la personne concernée;

b. une base légale pour le décodage existe; ou que

c. le décodage est nécessaire pour garantir les droits de la personne concernée, notamment le droit de révocation.

Universität Basel 28

Respect for participant autonomy

➢ Informed consent in Big Data research

Consent

➢ Big data re-raises the same questions as 15 years ago the are of biobanking:

Is different and more comprehensive research regulation needed?

Universität Basel 30

Abandoning consent?

Universität Basel — UNIVERSITÉ DE GENÈVE

➢ How can the correct balance be reached of not impeding research, but offering sufficient protection from harm.

➢ There is a lot of talk about harm, but the evidence is not clear. It is important to document harm (however: hidden harm might never be recognized, as we don't know which information employers and insurance companies used to exclude people for "other reasons")

➢ "Many entities, including the Institute of Medicine (IOM), have concluded that the HIPAA's requirements often seriously impede research without adequately protecting patients' privacy. The IOM recommends replacing HIPAA with an oversight regimen that focuses on desired outcomes rather than on prescriptive regulation, that mandates strict data protection standards, that involves legal penalties for re-identification of data, and that applies to all users of health data rather than only to HIPAA-covered entities" (Currie 2013).

Universität Basel 31

The debate about Consent

Universität Basel — UNIVERSITÉ DE GENÈVE

➢ "It could be asserted that non-interventional research ought to be deregulated — even research currently subject to the Common Rule. That was the main recommendation of an Institute of Medicine (IOM) publication in 2009" [IOM: *Beyond the Privacy Rule: Enhancing Privacy, Improving Health through Research*].

➢ "The IOM report draws a clear distinction between "interventional research" and research that is 'exclusively information based'."

➢ It recommends: "abolishing informed consent for all information-based research, including research using individually identifiable health information".

➢ "The IOM report justified its recommendation by invoking a broad utilitarian argument: 'If society seeks to derive the benefits of medical research in the form of improved health and health care, information should be shared for the greater good, and governing regulations should support the use of such information, with appropriate oversight'."

Rothstein 2015, Currie 2013

Universität Basel 32

Consent

Universität Basel — UNIVERSITÉ DE GENÈVE

No informed consent needed if individuals are non-identifiable

➢ "We recommend that model developers be allowed to use patient data that have already been collected without explicit consent, provided the developers comply with federal regulations regarding research on human subjects and privacy of health information. Under the Common Rule that governs such research, the informed consent of human subjects is not required if researchers cannot identify the individuals whose data are being analyzed."

➢ "This would be similar to telling patients when they enter a hospital or a physician's practice that their records may be used for quality improvement."

Cohen et al. 2014

Universität Basel 33

Consent

Universität Basel — UNIVERSITÉ DE GENÈVE

Non-identifiable = ?

Universität Basel 34

US versus Europe

Universität Basel — UNIVERSITÉ DE GENÈVE

● **NBAC (USA 1999) and Council of Europe (2006)**
  ■ Reversibly anonymised
    = coded data to which Researchers don't have access (HFG: verschlüsselt)
    → **= Identifiable** (link exists)

● **Office for human research protection (USA 10.8.04)**
  ■ Reversibly anonymised
    → **= Non-identifiable** (but: link exists)

US versus Europe

Universität Basel — UNIVERSITÉ DE GENÈVE

● **NBAC (USA 1999) and Council of Europe (2006)**
  ■ Reversibly anonymised
    = coded data to which Researchers don't have access (HFG: verschlüsselt)
    → **= Identifiable** (link exists)
    Helsinki: « human subject research »

● **Office for human research protection (USA 10.8.04)**
  ■ Reversibly anonymised
    → **= Non-identifiable** (but: link exists)
    Helsinki: **Not** « human subject research »

Predictive analytics

Universität Basel — UNIVERSITÉ DE GENÈVE

- ➤ "Patients are not aware that their data is being used for predictive analysis.
- ➤ As an analogy, patients are generally unaware if their physicians are using computerized decision aids to guide treatment.
- ➤ Predictive analytics falls between two well accepted models regarding consent. On the one hand, predictive analytics resembles clinical research, in which explicit consent is usually required. On the other hand, it also resembles quality assurance or quality improvement activities, in which consent is not generally required."

Cohen et al. 2014

Universität Basel 37

Social media - internet driven research

Universität Basel — UNIVERSITÉ DE GENÈVE

Opinion 1:

- ➤ "If Big Data researchers use publically available sources, then there is no research ethics problem, but accessing and using private sources of data is another issue."

Rothstein 2015

Universität Basel 38

Social media - internet driven research

Universität Basel — UNIVERSITÉ DE GENÈVE

Opinion 2:

- ➤ Publicly available data: but do users anticipate data use for research?
- ➤ "Users are often not aware of data protection policies, although they are explained in the Terms and Conditions of these platforms.
- ➤ There is a subtle difference between internet forums and other SM platforms, with regards to ethical concerns. For platforms such as YouTube or Twitter, the user is aware that his or her data is in the public domain: when someone uploads a video to YouTube, they know that it is going to be made available to everyone wanting to watch it. The same does not apply to internet forums."
- ➤ Guidelines for researchers are needed.

Casanas i Comabella et al. 2015

Universität Basel 39

Social media - internet driven research

Universität Basel — UNIVERSITÉ DE GENÈVE

Guidelines:

- ➤ "The General Medical Council (GMC) and the British Medical Association (BMA) have recently issued comprehensive guidelines for healthcare professionals on the use of SM and its impact on the patient-doctor relationship.
- ➤ The British Psychological Society (BPS) and the American Psychological Association (APA) have recently started addressing the need for guidelines on internet mediated research, with BPS having published the comprehensive Ethics Guidelines on Internet mediated research.
- ➤ The guidelines use the principles of the BPS Code of Human Research Ethics, which include respect for the autonomy and dignity of persons; scientific value; social responsibility; and maximising benefits and minimising harm.
- ➤ It is crucial that local ethics committees are aware of these guidelines and facilitate discussion involving not only researchers, but also service users."

Casanas i Comabella et al. 2015

Universität Basel 40

Social media - internet driven research

Universität Basel — UNIVERSITÉ DE GENÈVE

- ➤ Examples for present solutions in Switzerland:
  - – Sociology department (Unige), Faculty of psychology (Unibas) evaluate research protocols in their faculties that do not require approval by Research ethics committees (HFG).

Universität Basel 41

Conflicts of interest

Universität Basel — UNIVERSITÉ DE GENÈVE

Universität Basel — **Conflicts of interests** — UNIVERSITÉ DE GENÈVE

➢ **Problem:** "Many groups that develop models [for predictive analysis] have a conflict of interest between promoting transparency and protecting their intellectual property and profits when they sell their models to health care institutions. Thus, scientific peer review and independent validation are desirable."

Cohen et al. 2014

Universität Basel 43

Universität Basel — **Conflicts of interests** — UNIVERSITÉ DE GENÈVE

➢ **Problem:** Social media companies may have conflicts of interest when it comes to granting access to data (commercial value of data).

Universität Basel 44

Universität Basel — UNIVERSITÉ DE GENÈVE

Justice: new digital divides

Universität Basel — **Digital devides** — UNIVERSITÉ DE GENÈVE

Data production (whose data are studied)

➢ "While the use of the internet and SM has become mainstream in many high-income countries during the past decade, there are still millions of people that cannot access this technology.

➢ This may be due to lack of internet connection or equipment, or to a lack of skills, and has been linked to age and level of education.

➢ This inequality in terms of internet access is known as the digital divide."

Currie 2013

Universität Basel 46

Universität Basel — **Digital devides** — UNIVERSITÉ DE GENÈVE

Data access (who is allowed to study the data)

➢ "Much of the enthusiasm surrounding Big Data stems from the perception that it offers easy access to massive amounts of data."

➢ "But who gets access? For what purposes? In what contexts? And with what constraints?"

➢ "Only social media companies have access to really large social data."

➢ "An anthropologist working for Facebook or a sociologist working for Google will have access to data that the rest of the scholarly community will not'."

➢ "Some companies restrict access to their data entirely; others sell the privilege of access for a fee; and others offer small data sets to university-based researchers."

Boyd D, Crawford K. Critical questions for big data. Information, Communication & Society Vol. 15, No. 5, June 2012, pp. 662–679

Universität Basel 47

Universität Basel — UNIVERSITÉ DE GENÈVE

Is there a Big Data exceptionalism or should traditional research regulations apply to Big Data health research?

➢ Conclusions

Conditions

Strong human rights based tradition of research ethics:

➢ Abandoning consent for all non-interventional studies (including Big Data) is controversial.

➢ Given present human rights based privacy and data protection traditions in Europe, it seems unlikely that there will ever be consensus to abandon the traditional concept of deontological research ethics. Research cannot be justified using simply broad utilitarian arguments that data research profits society.

➢ Appropriate oversight mechanisms need still to be found.

➢ There is no obligation to participate in research: "Individuals ought to have the ultimate right to decide whether to participate in research" (Rothstein 2015).

Universität Basel 49

Conclusions

Autonomy based and pragmatic reasons to maintain trust:

➢ "Autonomy is the main interest of research subjects in their biological specimens and health records — both identifiable and unidentifiable. …many individuals strongly believe that their biological specimens and health records 'belong to them.'

➢ There have been numerous surveys of patients and their caregivers on this issue and they all say the same thing. If individuals are asked for their permission to perform research using their biological specimens and health information, they will overwhelmingly agree.

➢ On the other hand, if their specimens and information are used without their knowledge and consent, they will consider it a serious breach of trust."

Rothstein 2015

Universität Basel 50

Conclusions

➢ "Despite unique aspects, such as its data sources, scale, and open access provisions, the ethical issues surrounding Big Data are similar to those involving traditional biomedical research.

➢ Without question, the regulation of research can be improved in many ways.

➢ The development of new analytical tools, however, such as Big Data, should not serve as a catalyst for abandoning foundational principles of research ethics…respect for persons, beneficence, justice".

➢ More discussions are needed to ensure protection of individuals and groups in social media research.

Thank you for your attention!

Rothstein 2015

Universität Basel 51