

Available online at [www.sciencerepository.org](http://www.sciencerepository.org)

Science Repository



## Research Article

## Use of Existing Health Data in Epidemiologic Research—Issues of Informed Consent Under Normal Circumstances and at a Time of Health Crises

Jørn Olsen<sup>1</sup>, Carsten Obel<sup>2\*</sup> and Uffe Juul Jensen<sup>3</sup><sup>1</sup>Senior consultant, Department of Clinical Epidemiology, Aarhus University, Denmark<sup>2</sup>Professor, Department of Public Health, Aarhus University, Denmark<sup>3</sup>Professor, Department of Philosophy, Aarhus University, Denmark

## ARTICLE INFO

## Article history:

Received: 19 May, 2020

Accepted: 6 June, 2020

Published: 26 June, 2020

## Keywords:

Ethics

data protection

informed consent

epidemiology

COVID-19

citizen partnership

## ABSTRACT

In epidemiologic research we study why we get sick and how we get better. To do this we frequently need large datasets on exposure, diagnoses, treatment and more. We need data often classified as sensitive and regulated by law stating a need for informed consent. We argue that modern epidemiologic research often can be done on existing data without having informed consent and without violating basic ethic principles. We also argue for a timely and fair access to data in approved project. Modern encryption technics and methods of data analyses can reduce the risk of disclosure of personal data to a level close to what we have for anonymous data. If we allow open use of administrative health data and existing research data, we will be able to produce much more information to advance disease prevention, health promotion and treatment. Epidemiologists should collaborate more with computer scientists and patient groups in developing/implementing principles for 'modern methods of data analyses'. Under a severe health crisis data are in high demand to provide the information needed to prevent deaths and diseases and often time does not permit requiring 'informed consent'. Such a situation is now plying out worldwide under the Covid-19 pandemic.

© 2020 Carsten Obel. Hosting by Science Repository.

## Background

Advances bioscience, epidemiology, information science and technology over the last 30-40 years has had impressive impact on diagnostic, therapeutic and preventive practices and much more will follow [1]. Researchers and clinicians predict more new radical breakthroughs if scientific and technological possibilities are allowed to be exploited. The development (in particular the accumulation and use of big data) has, however, given rise to not only amazing new opportunities but also ethical, legal and political dilemmas and discussions and it has added little to our ability to predict epidemics [2-6]. National, trans-national (e.g., EU) and international organizations (e.g., WHO, CIOMs) have intervened and suggested or implemented regulation to protect citizens' privacy. In many cases this may slow down the speed of new discoveries that could be of importance in our fight against diseases and premature deaths. Epidemiologists should make themselves heard in this struggle over use of existing health-related data.

It is a much too important issue to leave for lawyers and law makers alone.

If you see some diseases as the end result of a crime, where causative agents need to be identified and eliminated/reduced, you should facilitate this search for causes. You should not be blocking the steps to stop the culprit but that has too often been the case. Think about the H1N1 influenza epidemic that affected pregnant women hard. Existing data from Norway documented that pregnant women would benefit from vaccinations [7]. Unfortunately, that conclusion came late because of restrictive legal rules for use of data without informed consent. Or think about the fabricated study that wrongly indicated that measles vaccinations could cause autism. Use of existing data made it possible to show this is highly unlikely but this documentation came late and a drop in vaccination rates was a result of the scare but it could have been much worse [8].

\*Correspondence to: Carsten Obel, Professor, Department of Public Health, Aarhus University, Denmark; E-mail: [co@au.dk](mailto:co@au.dk)

To address important health questions like these we usually need personal data on health – that is, we need sensitive data since all health data are classified as “sensitive”. Using data of this type has to take into consideration three ‘first principles’:

- i. Personal health data belong to the person that gave rise to the data.
- ii. Use of personal data in research (not always in clinical practice) requires informed consent – the data owner is usually the only one who can give this consent.
- iii. We all have a right to privacy.

Access to use existing health data in epidemiologic research varies widely over time and between populations. In recent years, some countries (especially the Nordic countries) have often provided access to existing research data or health data on entire (or almost entire) populations for research and without requesting informed consent. This practice has resulted in a large number of important publications, but it has been necessary to bypass principles of right to privacy that many countries at present do not allow without informed consent.

Violating principles 1-3 by private and public authorities has sometimes been done under the belief that no one is harmed, and many may benefit. Individual rights have –with utilitarian arguments - been challenged to promote benefit for the public at large. The new rules developed by the EU on General Data Protection (GDPR) implements the probability of harm as a guide for decision making which is a valuable tool for more and still safe use.

The tension between protection of individual rights and concern for the public interest may well deepen in the future unless we find a way to keep a good track record of no disclosure of sensitive data that cause harm to individuals or sensitive groups. At the same time people expect to be informed about unwanted side effects of drugs, occupational and environmental hazards etc. and to predict new health hazards before it is too late. This calls for more and better use of the data we have.

In the following we present more arguments from an epidemiological stance to find a balance between respecting individual privacy and using our research means to promote the public interest.

- i. Use of existing health data in research is in many countries not allowed without informed consent and since data were collected for a different purpose no specific informed consent is available. Often no consent has been collected or even could have been collected for practical or technical reasons. New technologies in data protection brings new opportunities and strict rules for proper and safe data storage and data analyses should be revisited.
- ii. Accepting a “right to be forgotten” option; that is a right to be taken out of all research projects they have not given ‘informed’ consent to. Since people choosing this option benefit from those who allow use of their data in health research, the ‘opt out’ option should not be a default choice – it should be an ‘informed discontent’ option that at least requires some action and thought.
- iii. Anonymous data (data cannot identify individuals in any way) can in many countries be used for research without informed

consent but unfortunately it is difficult and expensive to make it completely impossible to identify people in existing registers. The solution to this problem may well be that most of the data analyses will be done by robots with no storage space for personal identifiers. These robots can be programmed to deliver only aggregated data with limited risk for unwanted disclosure of private information.

- iv. Often we do studies with the aim of monitoring health in entire populations and this has been seen as an instrument for the ‘Big brother watches you’ – surveillance of entire nations in order to control their behaviour. A truer picture of the research use may be ‘little sisters use data to watch and limit the power of big brother’ – to keep check and balances.
- v. We have new and better methods and we have much better data from, e.g., biological monitoring of environmental exposures, and we have detailed genetic and clinical data for an increasing number of people. The value of this information will increase over time when we add follow-up time allowing analyses of health problems that develop over decades and sometimes even over a lifetime. How exempt from informed consent is understood and used will determine if we end up in a position where we produce research in large quantities and of good quality or if we end up producing limited research results that too often may do more harm than good.
- vi. The ethical concerns and problems related to large collections and storage of personal data lie in the existence of these data, not in their use for research. If rules for use become too restricted, the risk of having these data will outweigh the potential benefits. Storage of data that are not being used is simply unacceptable. Such a storage maintains the risk of misuse without opening for the benefits of its use. We should not establish ‘data cemeteries’ and we should only keep data in few places located outside political institutions where potential misuse (including political misuse) can be monitored and stopped. One option for storage could be at patient organizations, another could be universities if they are independent and free of powerful political and economic structures.
- vii. Datamining and machine learning open up for large scale explorative, agnostic studies that will need more computer power than most of us have at present but that is just a matter of limited time before we see the full ‘firepower’ of modern epidemiology [9, 10]. All disclosure of personal data will only reach robots with no memory outside the limited time of data analysis. These data may be open to research where only the research topic need ethical approval. If these conditions are accepted as exempt from informed consent approval, the avenue for large scale studies using existing data will be open and the benefits can be substantial. Informed consent should still be the principle for de novo data collected for research based upon hypotheses and of course for data collection that carry a risk for the participants.
- viii. These considerations from our standpoint of epidemiological research is widely reflected in International Ethical Guidelines for Health-Related Research Involving Humans (CIOMS 2016).

However, the implementation of the principles formulated in the guidelines and the concrete balancing of individual rights and the public

interest is to a great extent left to ethic committees and Data Protection Agencies. We suggest that a wider engagement of patients and patient-groups in developing and implementing principles and practices is wanted and needed. Such a partnership (as recommended by, e.g., European Patients Forum) will counteract mistrust between researchers and patients (and the public in general) and it will promote research in rare diseases of limited interest for the industry. Furthering partnership between researchers and patients is especially important at a time with growing collaboration between public institutions/health care systems and private corporations.

An extensive involvement of and collaboration with patients in health research would further contribute to rehabilitate or revitalize a classical idea of health care as a common good, i.e. a practice in which each and every participant (professional, researcher, patient, relatives and citizens in general) contribute to secure and reproduce the conditions of that practice. Patients and citizens always have contributed to health care practices by supplying data but in health care as a common good, patients should - together with researchers and professionals - contribute (as free and recognized participants) to develop and implement the principles and procedures necessary to uphold and develop practice. Today this includes principles and procedures for collecting and using data. That will also include principles for setting up biobanks, repositories for biological material, since they produce a different risk related to the disclosure of unknown information that could benefit or cause harm for the person who provided the sample. We will not address this further in this paper.

### Important Decisions

We are at a crossroad where we either adopt research use of existing data or speed up the research production to new levels, or we stick to old routines and accept less research and more mistakes. How strict we use of the informed concept principle will determine which path we walk. We fear we end up with informed consent procedures that will not be taken seriously ('the rent a car overload of information consent form'), or that rules are set with no understanding of the practical conditions in epidemiology. If data can be collected without no personal risk – data exist, no invasive procedures are needed and if data can be analysed by robots existing data from existing administrative registers and existing research data should be open for research approved by ethic committees. Informed consent should not be requested if participants have an 'opt out' opportunity.

If informed consent will be required a large variety of forces of selection will be operating and be a serious source of bias in several types of studies, especially in making geographical comparisons or comparisons over time. More data and a better understanding of causal analysis have made causal conclusions justified even outside the RCT domain and the epidemiology community stands ready to answer questions like why we get sick and how do we get better in a much better way than before [10, 11].

We recognize that political systems differ across countries and legislation needs to take into consideration probabilities of harm and misuse. Many countries may fear to be left behind in a time with increasing use of artificial intelligence and be less concerned about our

3 first principles for right to privacy. Epidemiologists are hopefully ready to make safe use of existing data, also under these conditions.

### Covid-19 Pandemic

Principles are often best scrutinized under extreme condition and the ongoing covid-19 epidemic provides plenty of food for thoughts in that respect. Data in the course of this epidemic are collected, analysed and translated into policy. Often no protocols were prepared ahead of time and often time does not permit getting informed consent to use these data for a particular purpose and having no valid data collection systems in place could mean life or death for many. In the light of the importance of having data private ownership of data must wait for better times, but we should of course make use of the tools we have to protect against misuse of data. Involving patient groups in this work should be done if possible.

Data on the progress of the disease, compliance to actions aiming at reducing the rate of transmission etc. require high class surveillance systems and good and timely data collected without bureaucratic obstacles. We need data on mortality and morbidity and data on actions to reduce virus transmission, often in the form of how the social distance in risk groups could be reduced. We need to monitor the spread of the infectious agent taking into consideration, that the high-risk social groups may change over time. The infection may have started in affluent segments of the populations like skiers and cruisers. Later, other social groups became the high-risk groups.

The National Institute of health, NIH (Link) describe 4 areas of ethical concern in a pandemic 1. Equitable access to health care 2. Ethical responses to a pandemic threat related to quarantine and isolation, especially for the elderly 3. Keep health care workers safe 4. Make sure health care workers get the needed protective gear to be safe at work. If an effective drug, like Tamiflu for some types of influenza, is available but short in supply should it then be given to the most ill patients, to the ones with no co-morbidity, to those who can pay and thus support research by paying a high price for the drug etc.? If an effective vaccination becomes available should vaccination be mandatory for health care workers or others by waving informed consent.

Data on risk behaviour may be extracted from survey systems like; data collected by cameras in public places, monitoring of data from mobile phones data on gathering of groups of people or recording morbidity and mortality for these groups and much more. There will often be a need to conduct ongoing monitoring of the populations risk behaviour and their health conditions. Is informed consent a mandatory requirement for these groups or only for some?

Making prediction for the expected path of an epidemic is often very unreliable since the mathematical models are based on limited and imperfect data. These models may work well if all causal paths are included in the decision process but that exist only in utopia. Sensitivity and specificity of tests need to be known and taking into consideration.

Participation rates in surveys often have to be high since even small fractions of non-participants may bias results and could mislead policy making. Data obtained from common registrations system are treated in

many countries as a public good since data were collected using public funds. It is however a violation of one of our first principles, but benefits may be high and cost low – and often paid already. General cost benefit principles can in some situation justify violation of right to privacy if the alternative is grim.

Since some of the information needed for policy making is to be open for others to see, we recommend the right to be forgotten, to leave studies and registration systems altogether. That escape option should be open if it is not used by many. We also emphasize the need to get patient groups involved as soon as possible as it was done in the AIDS epidemic.

We have seen a number of epidemics since the days of the bird flu epidemic (AIDS, Ebola and SARS) but also COL, AMI and mental disorder that share the epidemic construction and cost many lives. Catastrophic epidemics have been rare but are possible and they write their own ethical standards.

Many of the laws that violate privacy have been accepted because they were made time limited, but these limitations may be prolonged if considered necessary. To have policy guided by evidence rather than fear and ignorance has been well accepted, most likely because the population to a large extent see their value and necessity but as the epidemic matures it is important to get back to the same ethical standards as was in place before the epidemic. That often include use of informed consent if data are stored for others to read.

## REFERENCES

1. Morten Schmidt, Sigrun Alba Johannesdottir Schmidt, Kasper Adelborg, Jens Sundbøll, Kristina Laugesen et al. (2019) The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol* 11: 563-591. [[Crossref](#)]
2. Christine S Cocanour (2017) Informed consent-It's more than a signature on a piece of paper. *Am J Surg* 214: 993-997. [[Crossref](#)]
3. N Sivanadarajah, I El-Daly, G Mamarelis, M Z Sohail, P Bates (2017) Informed consent and the readability of the written consent form. *Ann R Coll Surg Engl* 99: 645-649. [[Crossref](#)]
4. National Library of M (2009) ACOG Committee Opinion No. 439: Informed Consent. *Obstet Gynecol* 114: 401-408. [[Crossref](#)]
5. Lokesh P Nijhawan, Manthan D Janodia, B S Muddukrishna, K M Bhat, K L Bairy et al. (2013) Informed consent: Issues and challenges. (Review Article). *J Adv Pharm Technol Res* 4: 134-140. [[Crossref](#)]
6. Daniel Kotz, Wolfgang Viechtbauer, Mark Spigt, Rik Crutzen (2019) Details about informed consent procedures of randomized controlled trials should be reported transparently. *J Clin Epidemiol* 109: 133-135. [[Crossref](#)]
7. Siri E Håberg, Lill Trogstad, Nina Gunnes, Allen J Wilcox, Håkon K Gjessing et al. (2013) Risk of Fetal Death after Pandemic Influenza Virus Infection or Vaccination. *N Engl J Med* 368: 333-340. [[Crossref](#)]
8. Van Calster B, Wynants L, Fralick M, Colak E, Mamdani M et al. (2019) Machine Learning in Medicine. *N Engl J Med* 380: 2588-2590. [[Crossref](#)]
9. Chen Y, Pedersen L, Chu W, Olsen J (2007) Drug exposure side effects from mining pregnancy data. *ACM SIGKDD Explorations Newsletter* 9: 22-29.
10. Olsen J, Jensen U (2019) Causal criteria: time has come for a revision. *Eur J Epidemiol* 34: 537-541. [[Crossref](#)]
11. Tian J, Pearl J (2000) Probabilities of causation: Bounds and identification. *Ann Math Artificial Intel* 28: 287-313.