



Editorial

Data protection and epidemiological research: a new EU regulation is in the pipeline

Since the days of the Helsinki Declaration (www.wma.net), informed consent has been a cornerstone in all medical research. That is probably how it should be if you do experimental research on humans that carries a risk for the participants or collect new data, but in non-experimental research based on existing data it is not so clear cut. The Helsinki Declaration states that ‘the health of my patient will be my first consideration’ and ‘while the goal of the primary purpose of medical research is to generate new knowledge, this goal can never take precedence over the rights and interests of individual research subjects’. Most will agree on these general principles. However, if the principle of informed consent is required when using existing data, where the only risk is related to unwanted disclosure of personal data, they go too far and the term ‘research subject’ is misleading. If we are using data that already exist and the research can be done with no risk for the people under study, there may well more ethical problems in not doing the research than in doing it. We lack an ethics committee for important missed research opportunities!

Using the Helsinki Declaration uncritically to fit all types of research has not been without problems, including ethical problems. Important data collections have been destroyed because no informed consent had been given at the time of data collection. Useful information has not been used to benefit the public, sometimes because the data could have revealed unpleasant facts for those in charge of health service and medical treatment.

Informed consent is of course needed when research carries a risk for the participants, but much information is gathered without informed consent and stored in the head of the therapist and in medical files. Most countries record births, deaths and several diseases. A few countries record use of medicine, treatment procedure, diagnostic data etc. Some of this is done without having informed consent. This information is of potential value in public health and

clinical research. It is of value for the people in general and getting informed consent after data collection will be difficult, in most cases even impossible.

The existence of these data is a threat to privacy, but not necessarily in the use of the data for research. On the contrary, use of the data in research can make the benefit/risk balance positive. In any case, how rules for protecting privacy are written have a profound impact on our options for research in public health, clinical medicine, social science, economy etc. Rules for data protection are, however, often written by people with little knowledge of many of the consequences of what they write. They are usually concerned about individual rights more than collective rights and they are endorsed by politicians who may be concerned about supporting collective rights and see no benefit in opening up data that might produce critical reviews of what they do—all in all a dangerous cocktail in many countries.

The informed consent principle is also the cornerstone of the new legislation/regulation from the EU as voted for by the Parliament on 12 March 2014 (http://europa.eu/rapid/press-release_MEMO-14-186_en.htm). They ask for an explicit consent indicating that a general consent will not suffice. If it has to be ‘informed’ what does that actually mean and how long can it last? All this is also of importance for re-use of research data or use of large research data collections based on informed consent, like the Danish National Birth Cohort (DNBC) or the UK Biobank (www.dnbc.dk and www.ukbiobank.ac.uk). In the DNBC, participants were asked to provide data for health research that had been approved by one of the ethics committees in the country or by the Data Inspectorate. They were given the right to withdraw the data at any point in time and could say no to any new data collection. There is no time limit for this consent. It is considered valid until it is withdrawn. In the UK Biobank, the consent can also be

withdrawn at any point in time, but if not, it is valid even after death and mental incapacity. They also specify that research may lead to commercial development.

Making data available for research is often justified if it serves a common good and can be done without violating privacy. It is more complicated if public data are put up for sale as suggested by the Statens Serum Institut in Copenhagen. If personal data become an object not just for the common good but for private enterprise for profit, even a national tax-paid health care system will have a hard time to justify that this can be done without explicit consent. After all, ownership of data belongs to the people who provided the data. Only they can sell their own data.

But what about the use of data collected in routine health care without informed consent? As the legal text is written by the EU Parliament, these data cannot be used for research since they are 'sensitive' and all health data are classified as sensitive. The current text, which has not yet been endorsed by the Council of Ministers, opens up for exemptions which are phrased in ambiguous ways. 'Member states law may provide for exceptions to the requirement for consent for research, as referred to in Paragraph 2, with regard to research that serves a high public interest, if that research cannot possibly be carried out otherwise'; but what is meant by 'high public interest' and who will decide? Later the text states 'The data in question shall be anonymized or, if that is not possible for the research purpose, pseudonymized under the highest technical standards and all necessary measures shall be taken to prevent unwarranted re-identification of the data subjects – however, the data subject shall have the right to object at any time in accordance with Article 19'.

Writing a text on data protection that has to cover social media, bank transfers, research and much more is perhaps an impossible task. The EU Council of Ministers should consider having a separate text for research and include freedom of the press in that text.

Since data linkage is the key in most research projects, anonymous data are often of limited use, and pseudonymized data may be difficult to obtain and much of the value of 'Big Data'¹ will be lost in this process. If large-scale data have to exclude not only personal identification like names or phone numbers, but also information that in combination with other available information makes it

impossible identify people you know well, even including yourself, not much is left and the promises of Big Data are gone.

Big Data require sharing of data and sharing of data is not only threatened by strict rules for data protection that may be inspired by the wish to protect governments more than data. Full use of research data is also limited by our own reluctance to share data. The Society for Epidemiologic Research announced a Symposium at their annual meeting in Seattle in June 2014, entitled 'Data sharing: why we hate it, why we need it and ways to get there'.

In some countries health care is almost entirely financed by public sources, making it reasonable to expect that this knowledge can be used to help others. Many studies need to come from large groups, not just a few selected individuals. Epidemiologists, and others, have not explained well enough to the public what is at stake and why research on large-scale populations is needed for identifying preventable causes of diseases and for getting better health care. This research can be, and has been, done in ways that have caused no known unwanted disclosure of personal data and it may well be the safest form of medical research we have. We should, however, accept that some feel strongly intimidated by use of their personal health experience in research and we should provide an opt-out option for these people—a right to be forgotten, as stated in the EU text from the Parliament. And we should make sure we use the best possible data protection standards. The days when epidemiologists carried around personal data on their PCs should be over.

The area of Big Data carries risk but also promises. The EU Regulation on data protection may end up supporting the risks and eliminating the promises if these data sources exist but cannot be used for the public good in research.

*Jørn Olsen, Department of Public Health-Epidemiology, University of Aarhus, Bartholins Alle 2 – Building 1260, DK-8000 Aarhus C, Denmark.
E-mail: jo@ph.au.dk*

Reference

1. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA* 2014;311:2479–80