

Valerie Racine's Commentary on "Big Data & Genetic Privacy: Re-identification of Anonymized Data"

Commentary On

Case: Big Data & Genetic Privacy: Re-identification of Anonymized Data

The study by Gymrek *et al.* 2013, and others like it, generated demands for additional restrictions in database sharing policies, changes to how and what kinds of data were collected and anonymized, and worries about some of the foundational concepts in research ethics, including the notions of informed consent, privacy, confidentiality, and the nature of the researcher/clinician – subject/patient relationship. This short commentary will focus on those concepts in biomedical research ethics.

Most researchers and ethicists agree that it is important to safeguard privacy and confidentiality for patients and research subjects, but to do so in a way that does not impede scientific progress. This “sweet spot” between the competing goals of scientific research and the individual’s right to privacy is especially relevant for current genomic and genetic analyses using big data. For instance, Genome Wide Association Studies (GWAS) capitalize on correlated sets of large databases of individuals’ genetic variants to determine whether certain variants are important contributors to complex diseases or disorders. There is also much optimism about the prospects of personalized medicine, in which medical professionals would access and integrate patients’ personal genomic data into targeted and tailored treatments. The success of personalized medicine, however, requires knowledge about which sorts of treatments will be effective for certain genetic variants, which depends on genomic analyses of big data.

While there are clear potential benefits of biomedical research analyses of large sets of genomic and genetic data, that information is also particularly sensitive as it can accurately reveal subjects’ identity in the same way as social security numbers can. It can also reveal the identity of an individual’s relatives. Because of the way this

information can serve as accurate individual identifiers, some researchers have taken the notion of genetic privacy to denote a special instance of privacy (e.g. Rothstein 1997), based on the notion of “genetic exceptionalism” – i.e. “the view that being genetic makes information, traits, and properties qualitatively different and deserving of exceptional consideration” (Lunshof *et al.* 2008).

If we accept a concept of genetic privacy, based on genetic exceptionalism, then there are implications for the way we think about infringement of privacy and breach of confidentiality within the biomedical research context. For instance, Lunshof *et al.* (2008) argues that because some violations of privacy occur which are beyond the control of individuals or institutions (as in the above case scenario), they do not necessarily signal a moral failure even though those violations may cause harm in some instances. However, they note that the promise of confidentiality implies a relationship of trust and, with it, moral responsibilities on those who promise confidentiality. For that reason, a breach in confidentiality does entail a moral failure with respect to the relation of trust between the researcher/clinician and subject/patient.

These moral considerations have led research scientists and ethicists to rethink the model of informed consent that typically guides the relationships of trust between clinician/researcher and patient/subject in the biomedical context, and to reconsider what, if any, sense of privacy and anonymity should be promised to patients and research subjects.

Informed consent is typically used in cases of specific research studies. It is problematic in research that makes use of big data because it does not, and cannot, explicitly cover all future investigations, or future instances of sharing and aggregating data across research communities. Because of these elements in big data science, the traditional notion of informed consent cannot be implemented in the usual way.

Consequently, some have proposed more liberal notions of consent, such as “open,” “broad,” or “blanket” consent (Mittelstadt & Floridi 2015). These notions of consent require research participants to consent to all future research activities that makes use of their data. However, those approaches have been criticized for limiting patients’ or subjects’ autonomy (Mittelstadt & Floridi 2015; Master *et al.* 2014). An alternative proposal to the models of general consent is the notion of “tiered” consent. That notion of consent would enable patients and subjects to choose to

limit future access to their data to only some kinds of research, or to require researchers to re-consent patients and subjects for specific kinds of future research. That approach has been criticized for creating too many difficulties for researchers and the management of large databanks.

Another alternative has been to emphasize the concept of solidarity rather than consent. This approach relies on the participation of “information altruists” concerned with the public good. It is mainly concerned with how research can be pursued and harms can be mitigated, “by providing data subjects with a ‘mission statement’, information on potential areas of research, future uses, risks and benefits, feedback procedures and the potential commercial value of the data, so as to establish a “contractual” rather than consent basis for the research relationship” (Mittelstadt & Floridi 2015; Prainsack and Buyx 2013). The proposed reliance on solidarity and public sentiment has been criticized for placing undue burdens on individuals to participate in research. However, it might also serve to emphasize the ethical responsibilities of big data researchers and database managers, and encourage scientists to be more proactive in the disclosure and transparency of risks of harm that might occur as a consequence of the loss of privacy (Lunshof *et al.* 2008; Barocas & Nissenbaum 2014). In this way, genomic and genetic research dependent on large sets of data has the potential to shift the moral responsibilities of researchers from protecting the privacy of individuals to ensuring the just distribution of any benefits from the outcomes of their research (Fairfield & Shtein 2014).

The emerging concepts of consent under negotiation within this research context, and the emphasis on researchers’ duty to benefit research participants and their communities more widely as well as the research participants’ duty to contribute to the public good, are areas of ethical deliberation intended to maintain the public’s trust in the medical profession, and scientific institutions more broadly. These ethical concepts and proposals, therefore, ought to be evaluated by how well they are able to do so.