DATA RE-USE AND THE PROBLEM OF GROUP IDENTITY

Leslie P. Francis and John G. Francis

ABSTRACT

Re-using existing data sets of health information for public health or medical research has much to recommend it. Much data re-purposing in medical or public health research or practice involves information that has been stripped of individual identifiers but some does not. In some cases, there may have been consent to the re-use but in other cases consent may be absent and people may be entirely unaware of how the data about them are being used. Data sets are also being combined and may contain information with very different sources, consent histories, and individual identifiers. Much of the ethical and policy discussion about the permissibility of data re-use has centered on two questions: for identifiable data, the scope of the original consent and whether the re-use is permissible in light of that scope, and for de-identified data, whether there are unacceptable risks that the data will be re-identified in a manner that is harmful to any data subjects. Prioritizing these questions rests on a picture of the ethics of data use as primarily about respecting the choices of the data subject. We contend that this picture is mistaken; data re-purposing, especially when data sets are combined, raises novel questions about the impacts of research on groups and their implications for individuals regarded as falling within these groups. These impacts suggest that the controversies about de-identification or re-consent for re-use are to some extent beside the point. Serious ethical questions are also raised by the inferences that may be drawn about individuals from the research and resulting risks of stigmatization. These risks may arise even when individuals were not part of the original data set being re-purposed. Data re-use,

AU:1

Studies in Law, Politics, and Society, Volume 73, 143 167

ISSN: 1059-4337/doi:10.1108/S1059-433720170000073004

re-purposing, and re-combination may have damaging effects on others not included within the original data sets. These issues of justice for individuals who might be regarded as indirect subjects of research are not even raised by approaches that consider only the implications for or agreement of the original data subject. This chapter argues that health information should be available for re-use, information should be available for use, but in a way that does not yield unexpected surprises, produce direct harm to individuals, or violate warranted trust.

Keywords: Justice; health data; public health; privacy; research; group identity

INTRODUCTION

Re-using existing data sets has much to recommend it. Data are re-used when they are collected for one use and then used a second time. They are re-purposed when the second use has a different aim than the first. Much data re-purposing in medical or public health research or practice involves information that has been stripped of individual identifiers but some does not. In some cases, the data were originally collected for research purposes and there was consent to further research; this consent may have been quite limited or very open-ended and the question is whether it covered the new research. In other cases, the data were not collected for research but for medical treatment or for some other purpose such as the individual's personal interests or shopping activities; this collection may or may not have involved any kind of consent to further use with other aims. Data sets are also being combined and may contain information with very different sources, consent histories, and individual identifiers.

Much of the ethical and policy discussion about the permissibility of data re-use has centered on two questions: for identifiable data, the scope of the original consent and whether the re-use is permissible in light of that scope, and for de-identified data, whether there are unacceptable risks that the data will be re-identified in a manner that is harmful to any data subjects. Prioritizing these questions rests on a picture of the ethics of data use as primarily about respecting the choices of the data subject. It also assumes that, once de-identified, information is no longer in important senses "about" the subject and that its use becomes ethically problematic only if it can be re-linked to the subject. While controversies about consent and de-identification surely matter, in this chapter, we argue that they are by no means the only, or perhaps even the most important, questions to ask about data re-use.

Instead, data re-purposing, especially when data sets are combined, raises novel questions about the impacts of research on groups and their implications for individuals regarded as falling within these groups. These impacts suggest that the controversies about de-identification or re-consent for re-use are to

some extent beside the point. Serious ethical questions are also raised by the inferences that may be drawn about individuals from the research and resulting risks of stigmatization. These risks may arise even when individuals were not part of the original data set being re-purposed. Data re-use, re-purposing, and re-combination may have damaging effects on others not included within the original data sets. These issues of justice for individuals who might be regarded as indirect subjects of research are not even raised by approaches that consider only the implications for or agreement of the original data subject.

The primary information we consider in this article is information originating in clinical research and in electronic health records (EHRs). We refer to this as "health information," although noting that a great deal of information about health may originate elsewhere, for example, in wearable devices worn for recreational purposes. Health information collected within the original confines of treatment or research is information that many people regard as especially sensitive and that in fact can be quite damaging if it is misused. It is information that has been collected within relationships of trust and expectations of confidentiality. It is also information that may be regarded as accurate in a way that other information about health is not (whether or not this is actually true; it is at least information that has been vetted in some more expert way than information collected or reported by individuals themselves.) So it is not surprising that attention has been directed to the role of informed consent for new uses of these data. Our concern, however, is that the ethical issues raised by uses of this information are not well delineated by the scope of informed consent as it is typically understood.

We should also note at the outset that we are not at all opposed to data re-use. To the contrary, one concern that we have about the focus on informed consent and de-identification is distortion of the ability to use information to improve health care and public health. On our view, information should be available for use, but in a way that does not yield unexpected surprises, produce direct harm to individuals, or violate warranted trust.

We begin by describing the scope and benefits of data re-purposing. We then turn to an examination of the strengths and weaknesses of consent and de-identification strategies. Next, we outline ways in which data re-purposing may affect not only original data subjects but also others who are similar to them in relevant ways. We conclude by arguing that these effects indicate problems of social justice in data use that are not touched upon by consent or de-identification strategies.

THE SCOPE AND APPEAL OF RE-PURPOSING DATA

It is fair to say that massive amounts of re-purposing of health information are occurring today. Here, we describe four types: the use of information initially

collected for one research study in a different study; the use of information initially collected for medical treatment for medical research; the use of information initially collected either for research or for treatment for public health purposes; and the use of information initially collected either for research or for treatment for commercial purposes. We conclude this section with a description of the precision medicine initiative (the PMI) and its vision for a data source that can be put to many different uses.

At present, two different federal regulatory regimes apply to these kinds of information. For information collected for research, the primary regulatory regime is the federal Common Rule (45 C.F.R. Part 46).[1] The Common Rule defines human subjects research to include only research with living human beings that collects personally identifiable information (45 C.F.R. § 46.102(f)). It thus does not apply to research involving medical records of persons who have subsequently died, even if the information might be quite sensitive and have implications for family members of these patients. It requires informed consent for research that does involve human subjects, even if the research involves only data and does not require contact with the individual (45 C.F.R. § 46.111(4)). Informed consent may be waived; however, if the committee responsible for reviewing the research determines that the research could not practicably be carried out without the waiver and the information, the research involves no more than minimal risk, the waiver will not adversely affect the rights or welfare of subjects, and if appropriate the subjects will be given additional information after conclusion of the study (45 C.F.R. § 46.116(d)). Among these rights of subjects is the protection of confidentiality. Waivers are granted frequently (e.g., Northwestern, 2016).

For most information originally collected for medical treatment or payment, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule governs (45 C.F.R. Part 164). The Privacy Rule, like the Common Rule, does not apply to information that does not identify individuals, but it sets out strict standards for de-identification (45 C.F.R. 164.514(b)). Unlike the Common Rule, the Privacy Rule applies to information about individuals who have died, up until 50 years after their deaths (45 C.F.R. 164.502(f)). Thus, if informed consent is required for use of this information, it must be sought from personal representatives who may be difficult to identify or find. In addition to de-identified information, which it does not cover, the Privacy Rule also permits use of information in a "limited dataset." This is a data set that excludes all potential identifiers except ZIP Codes and dates (including data of birth and date of treatment). These data sets may be used only for research or for public health and their use must be covered by a data use agreement that protects confidentiality (45 C.F.R. § 164.514(e)). Limited data sets are of some utility; they allow, for example, investigation of potential correlations with age or environmental conditions. They do not, however, contain the kind of information that would allow linkages to other data sets, such as information collected from social media sites or devices worn by patients to collect information

about their daily activities. Other uses of information from EHRs for research require what HIPAA calls an "authorization" from patients, which must include a description of the specific purpose of the research and its potential end point (which may just be a vague "at the end of the research") (45 C.F.R. § 164.508(c)). Waivers of the authorization requirement are permitted for use of information in research, however. To grant a waiver, a review board must find that the research presents no more than minimal risk to privacy because identifiers will be protected from improper use and destroyed as soon as they are no longer needed, that there are adequate written assurances that the information will not be re-used except for other research permitted under a waiver, and that the research could not practicably be carried out without the waiver and use of the information (45 C.F.R. § 164.512(i)).

A persistent difficulty has been that these two regulatory structures do not fit together seamlessly. They apply to different groups of people, have different rules about de-identification, and somewhat different standards for waivers. A further concern of researchers is that the rules are cumbersome, cause delays, and make it difficult to share information or combine it with other data sets (Nass, Levit, & Gostin, 2009). Five years after implementation of the Privacy Rule, the Association of Academic Health Centers protested that it had erected major barriers to research involving medical records. These barriers included difficulties in sharing information among centers in ways that might be particularly damaging to translational research and research involving genetic information (AAHC, 2008). Since 2011, Department of Health and Human Services (HHS) has engaged in a protracted process to amend the Common Rule, about which more below.

AU:2

Information Originally Collected in Medical Research and Re-Used in Research

As clinical trials are structured today, it is not unusual for information collected for one research study to be made available for use in another. The appeal of re-purposed data analysis includes cost savings with the use of an existing data set. When the data were collected for and have been used in research, the expectation is that the data set is of proven quality given its past and presumably successful use. An additional attraction is the possibility that the re-use of the original data set may open up new research connections between the original project and novel research initiatives. As more and more is learned about disease etiology, it is becoming apparent that some diseases that have been grouped together based on clinical presentation are not at all the same, whereas others that have seemed different may be more related than originally believed. Achieving this kind of understanding for cancer is one of the goals of the PMI, for example. On the other hand, information

may be needed for the new study that was not collected for the first one, so there may be need to link the original data set to either newly collected data or other relevant data sets.

As one example of data re-use, consider studies under the umbrella of SWOG (originally the Southwest Oncology Group). SWOG is a network of physicians, health care institutions, and designated cancer centers, in the United States and internationally. It typically has over 20,000 patients enrolled in active trials; since its origin, it has enrolled over 200,000 patients. And, "SWOG also manages a biorepository of 600,000 specimens with associated clinical data assets that are routinely used by other researchers, resulting in exciting new discoveries long after trials are completed" (SWOG, 2016). As an example, SWOG has samples and records from two large prostate cancer prevention trials which includes long-term follow up of healthy men and men with prostate, lung, colon, and other cancers. The SWOG data are available for cancer researchers and for researchers interested in other conditions. Researchers using the SWOG data and specimens must sign data use agreements meeting SWOG policies. These policies include informed consent from the patient at the time of sample collection. Sample informed consent forms available online state that the samples may be used for research about cancer or other diseases, that the information will not be sold, that it may be linked to information from medical records, and that researchers will not be given personally identifiable information. Statements about data re-use are open-ended and limited only to use in research even though patients were originally entered into research about cancer and may have believed that any planned future research would also involve cancer. SWOG policies have exceptions to the requirement of informed consent if the sample banking is retroactive (i.e., takes place after the study has been concluded) or involves samples originally collected for nonresearch purposes, as long as the collection otherwise meets the HHS criteria for waiver of informed consent. Samples are stored in a manner that does not allow direct identification of patients; electronic databases containing patient information also must not have names and other information that can directly identify patients or be linked to other databases that could identify patients. The SWOG tissue bank does keep identifiers, but they must be firewalled from individual researchers. The system has limited access and does not allow direct contact with patients; it also prohibits communication of any research data to patients or their physicians except for protocols that are expressly to be used for treatment decisions. IRB approval is required for any nonexempt studies using SWOG samples or data (SWOG, 2016a). As a reminder, federal regulations governing research with human subjects do not apply to deceased persons or to entirely anonymous materials; studies using these materials are exempt, as described above.

SWOG practices are typical of current approaches to the re-use of information originally collected for research. They do not require consent at all for fully de-identified information. They are open-ended about future uses, stating

only that these uses will be limited to research. Some might regard further research use as sufficiently similar to the purpose of the original data collection; others might judge that it is re-purposing if the information is used for research on very different medical conditions such as schizophrenia, or for research on costs of care or other social factors. SWOG policies allow linkage of information from medical records but do not give future researchers information that could allow them to know who individual patients are. These practices follow the currently operative rules governing research with human subjects, but may require changes with adoption of the proposed amendments to the Common Rule, described below.

Information Originally Collected in Medical Treatment and Re-Purposed for Research

The widespread use of EHRs has made it far easier to re-purpose information collected in clinical care for medical research. Such research includes studies of the safety and efficacy of treatment, cost-effectiveness or comparative effectiveness of treatment, drug drug interactions, rare side effects, and so on. It also includes observations of patterns and linkages that are only possible with large sets of clinical data. Methodologies are under development to allow retrospective analysis of these large data sets to answer questions that might otherwise have required expensive and lengthy randomized clinical trials (Safran, 2014). Medical centers routinely using patient data in research inform their patients in their privacy policies about the possibility that information in their medical records may be used in research; the information is then used under waivers of authorization (e.g., UCSF, 2016).

As outlined above, current federal regulations impose some constraints on re-purposing clinical data for research. De-identification is one way around these constraints, but it limits the utility of the information and, as we discuss below, may not meet all ethical concerns. Waivers of the informed consent and authorization requirements are the most common strategy employed when researchers want to use information that has not been de-identified. These waivers focus primarily on the need for the information, the great difficulty involved in re-contacting patients if a waiver is not granted, the possibility that the data will be biased if certain groups of patients refuse consent, and the strategies researchers will use to protect confidentiality. Waivers also impose ongoing constraints on data re-purposing for research, in the form of data use agreements and limitation of the approval to the study in question. Complex constellations of data use agreements are currently in place for large-scale data-sharing arrangements for research purposes.

As noted above, HHS began the process of revamping the Common Rule in 2011, with publication of an Advance Notice of Proposed Rule-making

designed to adjust research protections to research risks (HHS, 2011). After receiving and analyzing voluminous comments, HHS published the NPRM for changes in the Common Rule in September 2015 (HHS, 2015). The changes, not yet adopted, would require consent to be given for any re-use of biospecimens originally collected for research, even if they do not contain identifying information. This consent could be very broad, simply indicating that the specimen may be used in any future research. It would thus not give individuals any concrete ideas about what kind of research might be possible in the future using samples drawn from them in identifiable or de-identified fashion. The revamping would also exempt from review re-use of identifiable information collected for nonresearch purposes, such as the medical records linked to tissue samples or medical records used in research more generally, as long as there was general notice to patients that the information might be used in research. No further specification of types of research would be required. Confidentiality would still need to be protected, however. It is fair to say that the NPRM proposed revisions view the primary risks of information re-use to be patient identification (hence, the concern with biospecimens, which as they contain patient genetic information may be ineluctably identifiable) and confidentiality, rather than the further issues we raise below.

AU:3

Information Originally Collected for Research or Clinical Care and Re-Used for Public Health

State public health departments and the Centers for Disease Control and Prevention receive a great deal of information originating in clinical care. Under HIPAA, this information may be transferred for public health purposes as authorized by statute without patient authorization (45 C.F.R. § 164.512 (b)). Information transferred should be limited to the minimum necessary required (45 C.F.R. § 164.502(b)). All states have tumor registries; these registries are supported by the National Cancer Institute and are compiled into databases that allow research about cancer incidence, prevalence, and treatment (NIH 2016). These data can be linked to other data such as Medicare claims data registries, allowing vast possibilities for research. Data from the registries have been used to identify groups of people at higher risk for cancer, such as organ transplant recipients or women whose mothers took DES during pregnancy (Smith, White, Weir, Peipins, & Thompson, 2012). Data from cancer registries have also been combined with data from state birth defect registries to yield findings such as significantly increased cancer risks among children born with certain birth defects (Carozza, Langlois, Miller, & Canfield, 2012). These resources are rich and tremendously valuable and illustrate the possibility and importance of unexpected findings.

Once information has been transferred for public health purposes, it is no longer subject to HIPAA. Public health departments may need to conform to other federal regulations, such as if they conduct federally funded research with the information. But state freedom of information act laws also may allow public access to information. For example, communities may want to identify cancer clusters and analyze possible associations with environmental exposures. In one case, a local newspaper in southern Illinois sought state cancer registry data to investigate a possible cluster of childhood neuroblastoma; the state objected that there were sufficiently small numbers of children with the cancer in question to allow identification if the data were released. The Illinois Supreme Court construed the public's right to freedom of information broadly and held that the state had not shown that the information would "tend" to lead to the discovery of private information, despite the testimony of an expert in re-identification that she could identify some patients accurately using registry information combined with other publicly available information.

Information originally collected for research may also be subject to state law reporting requirements such as reports of infectious disease diagnoses or reports of abuse. This is considered a risk of research and patients are generally told about the possibility of these risks in the informed consent process. Information thus transferred to health departments is no longer subject to the rules protecting research subjects, however. One way for researchers to try to protect against requests for disclosure from public health or from law enforcement is to seek a federal "Certificate of Confidentiality," but it is unclear how much protection these give (Wolf, Patel, Williams, Austin, & Dame, 2013). AU:4

Many state public health authorities also sell data sets, subject to applicable state law restrictions. State health authorities may operate on limited public funds and use revenue from data sales to finance some of their operations. The data sold are typically de-identified and subject to data use agreements intended to prevent re-identification. The data may be valuable for research, analysis of health trends, reviews of healthcare utilization and costs, insurance underwriting, understanding the relation between health and other community variables, and developing various metrics for rating community life, among other purposes.

Information Originally Collected for Research or Clinical Care and Re-Used for Commercial Purposes

Many large health care centers aggregate and monetize their patient data. For example, the Cleveland Clinic and the Geisinger Health System have engaged in such enterprises. Explorys is an IBM data analytics product that is a spinoff from the Cleveland Clinic; it has data on over 50 million people and provides solutions for problems such as management of at-risk populations and

measurement of costs of care (IBM, 2016). xG Health Solutions uses Geisinger data to provide analytics for other health care systems seeking to identify areas where utilization or costs are too high, or where they can improve care quality and efficiency (xG, 2016).

Under the HITECH Act amendments to HIPAA in 2009, special patient authorization is required for the sale of identifiable protected health information (HITECH Act § 13405(d)(4)). This provision was motivated by charges that data aggregators were assembling large data banks of information regarding patients from prescription records that could then be sold to insurance companies for underwriting, with the result that some patients would pay more for their coverage (Nakashima, 2008). Authorization is not required, however, if the information sold is de-identified. Particularly valuable databases are prescription records from pharmacies and pharmacy benefit management companies that are de-identified as to patients but that contain the identity of the prescribing physician. These data allow pharmaceutical companies to track provider prescribing behavior so that they can tailor advertising or identify providers who might be willing to enroll patients in clinical trials. De-identified patient information from EHRs has also been used in this way. Although several states tried to ban the practice of selling pharmaceutical records de-identified as to patients for commercial use, believing that it contributed to inflated drug prices, the US Supreme Court held that the bans violated the First Amendment protection of freedom of expression (*Sorrell v. IMS Health*, 131 S.Ct. 2653 (2011)). Data from the Pew Foundation indicate that when sensitive information such as health information is involved, consumers are especially concerned about data re-use and the extent of data retention; they are more willing to share information with health care providers they trust and they consider the benefits they may obtain in deciding whether or not to share information (Rainie & Duggan, 2016). These findings suggest that the public concerns about data re-use are not limited to identifiable information, a point to which we return later.

The Precision Medicine Initiative and Information Re-Use

The PMI is a highly ambitious effort to create a cohort of over a million volunteers to investigate the molecular, environmental, and behavioral aspects of disease. It aims to develop an understanding of important variations among patients that will enable targeting therapeutic or other interventions to maximize success in treatment and prevention of disease. The information collected about cohort participants will be extensive: blood and possible other tissue samples, information from EHRs, a baseline physical exam, insurance claims, mobile health devices, participant surveys, and other sources. And cohort

participants will be expected to agree to be re-contacted over time to participate in a variety of more specific research studies (PMI Working Group, 2015).

The PMI thus is designed to permit data re-use on a grand scale. As individuals are enrolled in the cohort, there will be an overall promise of what it may achieve but no precise information about how the data will be used, how frequently individuals will be re-contacted, what studies will be of interest, how long the data will be valuable, whether other data will be needed and combined with the types of data sought initially, and what will ultimately be learned. Any consent at enrollment therefore must perforce be highly general, based on whatever parameters can reasonably be anticipated. But these parameters may change as more is learned and unanticipated connections are identified.

At the outset, the Obama White House announced principles of privacy and trust for the PMI, which included transparency. Five aspects of transparency were highlighted:

> dynamic information sharing to ensure that all PMI participants remain adequately informed throughout all stages of participation;
> communication of information about how, when, and where samples will be stored; generally how data will be used, accessed, and shared; types of studies for which the individual's data may be used; the goals, potential benefits, and risks of participation, including risks of inappropriate use or compromise of the information about participants; the privacy and security measures that are in place to protect participant data, including notification plans in the event of a breach; and the participant's ability to withdraw from the cohort at any time, with the understanding that consent for research use of data included in aggregate data sets or used in past studies and studies already begun cannot be withdrawn;
> information about data protection and rules of governance;
> prompt notification of any data breaches; and
> published summaries of research findings (White House, 2015).

These transparency considerations recognize that any initial information for participants will need to be quite general. They thus require ongoing communication with participants about what is being done and what has been learned. Putting these requirements into practice will involve far more robust means of communication with research subjects than has been common practice. One of the aims of the PMI is to enroll and engage very diverse participants, which will augment the challenges of ongoing communication. The PMI working group's recommendations about development of the cohort emphasize the importance of returning their data to participants but does not explain other aspects of continuing communication such as information about what is being learned through the PMI (PMI Working Group, 2015).

FAIR INFORMATION PRACTICES AND DATA RE-USE: INDIVIDUAL NOTICE AND CONSENT

Individual notice and consent has been one of the primary methods for addressing data re-purposing. The first comprehensive statement of what are now called Fair Information Practices (FIPs) occurred in a report for the US Department of Health, Education and Welfare published in 1973 (Gellman, AU:5 2016). The highly influential report, "Records, Computers and the Rights of Citizens," stated five basic principles for data protection. One of the principles explicitly prohibited data re-purposing without consent: There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent. Among other principles were that databases should not be secret and that data subjects should be given rights to correct any misinformation in them.

Notice and choice is pervasive in US data protection policies, as described above. It is reflected in the Common Rule emphasis on the informed consent of subjects to participation in research. It is also reflected in the HIPAA requirement that patients must authorize the use or disclosure of their information for specific research studies unless a waiver is given. The wealth of privacy policies and "I agree" buttons on internet websites also illustrate the reach of notice and choice.

European privacy law has presumptively strict provisions about consent for data re-purposing, but with exceptions that permit uses similar to those permitted in the United States. The new Data Protection Regulation (European Parliament, 2016), to go into effect in 2018, provides that data processing is unlawful without consent of the data subject to the specific purpose of the processing unless certain other conditions are met, among them performance of a task carried out in the public interest (Art. 6 §1(a), (e)). To decide whether it is permissible to re-use information without further consent, the processor must consider any links between the original purpose of the data collection and the purposes of the further processing, the context of the processing including links between the data subject and the processor, whether the data fall into particularly sensitive categories, the benefits of the processing to the data subject, and any appropriate safeguards (Art. 6 §4). For example, the standard ways in which websites for sales over the internet interact with customers initial permission to collect information and use it in future transactions, including to suggest products that the consumer might like and to advertise is a permissible re-use, unless the information falls into a sensitive category or the customer refuses. Consent must be clearly stated and the data subject must have the authority to withdraw consent at any time (Art. 7). Real-time consent is necessary for information in specified categories of sensitivity, including health information and information about sexuality (Art. 9 §1). However, there are exceptions to the consent requirement for using sensitive data in the interests of

public health, health care quality, archiving data in the public interest, and scientific or historical research, among others (Art. 9 §2 (i), (j)). Similar exceptions are provided to the much-discussed right to erasure of data that are no longer needed for their original purpose (Art. 17 §3(c), (d)). Finally, like data protection rules in the United States, the EU Data Protection Regulation applies only to identifiable personal information (Art. 1, § 1 Art. 4(1)). Thus, it is fair to say that in practice the EU Data Protection Regulation resembles research and public health practice in the United States, although it imposes more stringent consent requirements on re-purposing data in sensitive categories outside of these exceptions.

Such notice and choice approaches following FIPs place decisions about how data are to be collected and used in the control of original data sources, not data users. In theory, this is the approach that most respects individual choice. In practice, it may be far less successful in achieving goals of active engagement in these choices. Individuals may pay very little attention to the choices they make and may be affected by various cognitive biases (Cate, 2010). Notice and choice, moreover, typically occur at the point in time when information is originally collected. It has been construed to permit the kind of general agreement described above, such as the use of information gathered in one research study for any future research, or the use of information gathered in treatment for any research purposes. But there may never be follow up about what happens to the information in the future research, particularly if data sets are de-identified or merged (Ohm, 2014). At the point information is originally collected, the collector may have certain plans for the use of the information and the participants may have certain expectations about how the information will be used but researcher's plan may change as new connections become apparent or new opportunities become available (Hoofnagle, 2016). Syndromic surveillance and related statistical methods noticing a pattern the significance of which was not anticipated antecedently may reveal insights that no one could have predicted in advance or discussed in an informed consent process (Francis, Battin, Jacobson, & Smith, 2009). In short, notice and choice assumes a data world that is frozen in time. Although consent to original participation in the PMI will likely be similarly general, the PMI's commitment to following up with participants is unusual and may make consent and the ability to withdraw it more meaningful.

Flaws in notice and choice reach significantly beyond the quality of individual choice. The approach in terms of separate individual decisions also presents collective action problems. Consent the second time around for data re-use may prove elusive to obtain. Individuals may have died, moved, or be difficult to locate. Individual choices about data collection and re-use may result in disparities in the information that is available for valuable social purposes such as public health or medical research. Rothstein and Shoben (2013), supporters of new consent for when data are re-purposed, argue that such bias is insignificant and worth risking to further ethically responsible research. However, one study indicates that the difficulties in re-contacting people may be significant and more

severe than the likelihood that people will actually refuse consent. In this study, of lung cancer, researchers sought to re-contact participants to ask whether the data could be included in the federal database of genotypes and phenotypes (dbGaP). A quarter of the initial subjects had died; over a third of the remaining subjects could not be located; and there were significant differences in location rates by age, race, and gender, but there was little difference in consent rates for those who could be located (Cote, Harrison, Wenzlaff, & Schwartz, 2014).

At least one important study indicates that even if individuals would be willing to consent to data re-purposing, they would like to have at least the opportunity for some say in the matter. In the words of this study, of patients being asked to submit information to dbGaP, participants would be "glad you asked" (Ludman et al., 2010). This may have been an especially sensitive study for participants: it involved genetic information, and the study concerned the possibility of genetic links to Alzheimer's disease. The extent to which participants in certain groups refuse consent or cannot be found may depend on levels of trust in the research enterprise itself that depend in turn on perceived benefits and harms to the group. The follow up urged by the Obama White House for the PMI seeks to address this question of trust. These benefits or harms may or may not be associated with the presence or absence of informed consent on the part of original group participants in the research. Another collective action problem with individual notice and choice is that individuals may want data uses that are only available if others make similar choices to share their data but they may not be in a position to communicate with or to influence others in making their decisions about what data to share or to suppress.

DE-IDENTIFICATION

If consent for a variety of reasons is hard to come by, then the preferred alternative for finding a way to re-purpose data without consent is to transform identified data into de-identified data. The idea is that if participants in the data set cannot be traced to the study, even though data from them will be used in the study, the obligation to seek consent is moot. Primary objections to this strategy include the limited utility of de-identified information, especially if data sets are to be linked, and the risks of re-identification.

Many privacy advocates decry the possibility that data may be re-identified. But risks of re-identification are a matter of significant controversy. Estimating these risks depends on judgments about what other data sets are likely to be available and what recipients of the information can be anticipated (el Emam, Jabbouri, Sams, Drouet, & Power, 2006; Malin, Benitez, & Masys, 2011; Ohm, 2010). Re-identification risk also depends on the expertise of the re-identifier. To the best of our knowledge, there are no reported court decisions in which

people sued for damages claiming harm from re-identification of information concerning them.[2]

Re-identification is not the only concern of privacy advocates; however, Rothstein (2010) argues that it violates autonomy to use individuals' information without their consent, even in a de-identified data set. Others argue that de-identified information can stigmatize individuals if it can be attributed to a group and it is known that an individual is a member of that group (Vinterbo, 2011). Such stigmatization can occur even to members of the group who were not included in the original study.

Schwartz and Solove (2011) take an intermediate position, rejecting as reductionist the view that there are only two categories of personal data that have implications for privacy and consent: identified and de-identified data. They argue for a third category, identifiable data. These are data that are not at present identified, but that could be at some future point, perhaps if other data sets were to become available that contained information that could be used to establish linkages. We share their concern; protections that do not take into account implications about the possibilities for data re-use are at best lacking in imagination and at worst irresponsible. However, we think their argument points to a broader concern about the possibilities with de-identified data that these data could be used to draw problematic inferences, even when individuals have not actually been identified.

Consider as an example the much-discussed story of the re-purposing by researchers at Arizona State University of information originally collected from members of the Havasupai tribe for the study of diabetes. These data de-identified were used for further studies of schizophrenia, inbreeding, and tribal migration patterns, topics that were highly disfavored by the tribe. The tribe sued, and Arizona State eventually settled, despite the fact that the use of re-purposed, de-identified information met the standards of the Common Rule. Commentators have argued that this case demonstrates the importance of including data re-purposing with de-identification within the purview of informed consent. (Mello & Wolf, 2010; Rothstein, 2011). The case has also AU:6 been used to suggest exploring alternative consent models, such as tiered consent under which some participants agree to a wider scope of subsequent uses than others, or consulting the community in addition to seeking consent from individual participants (Mello & Wolf, 2010). These suggestions extend the model of informed consent rather than moving beyond it to address the problems with data re-purposing to which we now turn.

INDIRECT SUBJECTS IN RESEARCH

At times, research may affect others than the individuals who are its primary focus. Contagious disease research presents particularly good examples of the

possibilities others might be indirect subjects of research. One way is that information may be collected about them, albeit not directly from them. For example, research about questions such as the efficacy of treatment as prevention as a strategy for lowering rates of HIV transmission may wish not only to collect data about the subjects using the intervention but also about their sexual partners. Some such research might seek to enroll the partners as subjects as well and seek their informed consent. But other research strategies might never involve contacting the partners to obtain their consent or collect data directly from them. For example, the study might collect information about their sexual partners from the direct participants and then compare this information with reports to the health department of positive HIV tests. These strategies might be especially appealing if there were concerns that partners could not be found or would not agree to study participation. Nonetheless, the second strategy has direct implications for the partners. Participants in the study may change their behavior if they believe they are less likely to transmit disease because of their participation in the study, despite what they have been told about the study's experimental nature. And information about the partners is being used in the study, information derived from the direct study participants and from health department records although not directly from the partners themselves.

Studies involving contagious diseases may also affect others even if no information is collected about them. Consider testing of a new vaccine. The goal of the study may be to ascertain the extent of immunity created in subjects receiving the vaccine. Despite warnings that the vaccine is experimental, participants in the research may believe that they are better protected against the disease. They may thus behave in ways that increase the risks to others of catching the disease.

In an earlier paper, one of us argued that research ethics should pay attention to risks to indirect participants (Francis, Battin, Botkin, Jacobson, & Smith, 2006). Study designs should attempt to minimize risks, for example, by limiting and protecting the data collected about third parties. Consent processes with direct participants should inform them of any risks to third parties, such as possibilities of contagion. If risks are sufficiently high, indirect participants should be informed of them and perhaps even included in the informed consent process.

The forms of data re-purposing described above may present analogies to the situation of indirect subjects in contagious disease research. Information from direct study participants may include information about others not participating directly in the research. For example, study participants may be asked to give medical histories involving identifiable information about their family members. Medical records also may contain information about others, including family members. Although the original research or treatment would not have been about the others mentioned in the record, there is nothing to prevent further research being about them. This is particularly true if the original data sets are combined with other data sets containing information about these

individuals. For example, the original data set might have involved treatment or research of parents during the prenatal or perinatal period. With identifiable information, later research might link these records to further records about outcomes for the children. Or, research or treatment originally involving individuals with certain cancers might later be linked to other data sets such as state tumor registries to investigate whether there are familial patterns and whether other factors might be involved in phenotypic expression within families. If data sets are de-identified to the point of limited data sets, and linked to other data sets about individuals in the same location or with similar dates of treatment, inferences about disease patterns may be drawn. These patterns might in turn be used to draw inferences about others from the same location or with the same dates of treatment — even if these others were never included in the original data sets. Similar problems may arise through the re-use and re-combination of data sets that are sufficiently de-identified as not to allow results to be connected directly to particular individuals by name. Instead, the results could be connected if they reveal powerful and disturbing relationships between variables. For example, de-identified information could show that someone who lived in a particular area of the country, was of a given race or ethnic background and sex, and had purchased certain products, was highly likely to have an asymptomatic infection. No one would need link the de-identified information that had been used in the study to particular study individuals to infer that someone who met all of the variables was at high infection risk.

The power of data re-use, especially when data sets are combined, is that it affords possibilities that the later research might actually include people as subjects who were never direct subjects of the original data collection. It may also allow further inferences to be drawn about people who were never associated with the original research. While these possibilities are not exactly contagion — no one gets sick because others participated in research — they do pose potential risks for others that cannot be addressed by consent of the original research participants, even if it extends to the further data use. For example, the results of the Arizona State University research on the Havasupai applied to any tribal members who were not participants in the study, just as much as it applied to participants. The conclusions about tribal in-breeding or about migration patterns were not conclusions limited to the individual tribal members who participated in the study; they were conclusions about anyone who is a member of the tribe. This is not to deny the importance of these kinds of linkage studies that can be performed through data re-use and re-combination. The knowledge that people might have latent infections is surely valuable, as the experience with asymptomatic Hepatitis C indicates. The point instead is that risks of stigma from such uses of information must also be addressed. These risks of stigma may result from the discovery of patterns that could not have been anticipated in advance and that may not be associated with groupings based on standard demographic variables such as race or income level. It is to the concerns of justice raised by these possibilities that we now turn.

GROUPS AND IDENTITIES

Research ethics pays significant attention to groups identified as "vulnerable" in the sense that they are more likely to be subject to harm or that they are less likely to be able to protect themselves from harm. The Common Rule, for example, has special protections for pregnant women, fetuses, and neonates (45 C.F.R. Part 46, Subpart B); it also urges review boards to include additional safeguards to protect the "rights and welfare" of groups likely to be vulnerable to coercion: children, prisoners, pregnant women, mentally disabled persons, or economically or educationally disadvantaged persons (45 C.F.R. §111(b)). The Declaration of Helsinki § 20 provides that any group or individual susceptible to increased risks of incurring additional harm may only be involved in research if the research cannot be carried out in less vulnerable populations and the group stands to benefit from the research (WMA, 2013). The European Regulation on clinical trials (which only applies to interventional research), Art. 10, lists special considerations for minors, incapacitated subjects, pregnant or breastfeeding women, and participation by specific groups of subjects where expertise is needed (European Parliament, 2014). These efforts to provide special protection have come under criticism for unjustified paternalism, for vague definitions of vulnerability, and for the risk that they may unduly limit the possibilities of some groups to participate in the generation of medical knowledge relevant to them (e.g., Gennet, Andorna, & Elger, 2015; Whitney, 2014). The Declaration of Helsinki provision, aimed primarily at research in impoverished countries and strengthened to require actual benefit to the group in 2013, has been criticized for failure to say more about what benefits might be required and how they may be assured (Malik & Foster, 2016).

These formal statements about groups stop short of actually identifying any particular types of groups, other than prisoners, pregnant women, people with disabilities, and people who are educationally or economically disadvantaged. These are people who primarily fall in groups thought to need particular protection to ensure that consent is voluntarily given. They rely on the assumption that certain demographic groups often race, religion, or ethnicity are added into the mix can be identified in advance as potentially at increased risk. The recommended approach is then to put in place special protections, typically for the informed consent process. In the latest version of the Declaration of Helsinki, the approach is also to require the assurance of benefits to the communities from which subjects supposed to be vulnerable are drawn.

This approach reflects a standard way of thinking about groups in US social scientific inquiry. Conventionally, the focus is on groups that are given or claim to fall into a given demographic category. This may be an ethnic or racial identity, a religious or cultural identity, or a socioeconomic class.

Many of these groups are identified in census categories that have implications for public policy and in some cases for civil rights protection. Some initial research may focus on these groups, for example, HIV research in African-American males. Even this research may have spillover effects on others who are identified as members of the group but who were not participants in the research. These results may be particularly problematic if the group has been subject to disparagement or marginalization by other, often larger and better positioned groups in the society. For example, social scientific reports that draw repeated attention to relationships between rates of incarceration, patterns of education, and per capita income may be construed as applicable to all members of a group that may only have common skin color or some other attribute that is not linked to much else save the negative judgment of others outside the group. The stigma associated with groups believed to be carriers of communicable diseases that are frightening but little understood may result in members of such groups being ostracized. Examples include Haitians during the early days of the HIV epidemic HIV was even characterized as "the Haitian disease" at one point and travelers returning from Mexico after supposed identification of a serious form of H1N1 influenza there.

Further questions about the treatment of groups may arise with data re-purposing. As with the Havasupai example, later research questions may have far different aims than the original consent. The kind of very general consent to "future research" that is often requested does not place limits on these changes, as long as the re-use is for research. New research may also identify unexpected constellations of factors defining groups that have not heretofore been considered "vulnerable" or even groups at all. Identification of such novel connections is a core aim of the PMI. Other examples might be groups identified with exposures: Legionnaires or veterans of particular wars or travelers to Brazil who might have been bitten by mosquitoes carrying the Zika virus. Neither consent nor re-identification strategies provide particularly good protection in such cases. Consent fails because aims, risks, and benefits cannot be described in advance, nor can the individuals who may be affected. De-identification fails because the problem is not that stigma will be attached to individuals through their identities but that it will attach due to their group membership. We thus disagree with the common assumption in the United States and the EU that, once a data set has been successfully de-identified, secondary data analysis may go forward as privacy has been protected.

Another dimension should be considered in reviewing the adequacy of de-identified data protection: the importance assigned to group identification in the new research agenda that is proposed. If the analysis of re-purposed data will focus on a group that is composed in part or entirely of individuals who may be identified with the group that is the subject of the new research, implications for the group must be considered. These include the concepts of autonomy and dignity that are closely associated with protecting privacy, if

members of the group may be subject to the disapproving gaze of others for what they do. For example, members of a group associated with sexually transmitted diseases may be chastened for having unprotected sex even if they had no awareness of or involvement in the study identifying the risk. Travelers to an area newly identified with Zika may be criticized for their failure to use birth control during the trip even if they had no reason to know that the disease had spread to the area. If the connections with group membership are not carefully described, or if the supposed connection is overblown, even more individuals may be subject to inferences that hold them in disdain. There may be economic consequences as well, such as forms of red-lining in insurance or sales if bank default rates statistically suggest a much higher default rate for one group over another demographic group. In such cases, individual privacy is not directly violated in the sense that individuals are included in studies without their consent or that they are picked out from the group. But the values that lie behind privacy are weakened nonetheless.

Health research is an area where distinctions between individuals and groups matter clearly. A group that in comparison to other groups has disproportionately shown higher rates of a severe communicable disease or a genetically transmitted disease that may cut life short may be criticized, stigmatized, or worse. The identified group that is perceived as unusually susceptible may see individual members of a group experience discrimination or rejection socially and economically. These identifications may track traditionally recognized vulnerable groups, as the Chinese in Chinatown in San Francisco were identified with bubonic plague at the turn of the 20th century. But it may track other connecting factors identified in the research: neighborhood, work space, travel locations, or even entirely new interactions among environmental or genetic or other factors. These identifications are part of the promise of big data and of initiatives such as the PMI. But their implications require attention as a matter of justice.

In this regard, we recommend asking these new questions about research involving re-purposed data:

(1) Is the research likely to generate findings about a group that may be deleterious to people identified with the group, whether or not data from them are entered into one or more of the data sets involved in the research? In such situations, research findings may deepen disadvantage (Wolff & deShalit, 2007).

(2) Are the findings in (1) novel with respect to one or more groups? We identify this as a separate question because it may be that the information about the group is already generally known. For example, research about migration patterns for other tribes might make the findings about the Havasupai less novel and thus less likely to be as deleterious as they otherwise would be. Indeed, such research about migration patterns has become common.

On the other hand, research about inbreeding within the tribe would be novel, even if similar studies have been done with other groups.

(3) Is the research likely to generate associations that identify new groups in ways that might be stigmatizing or disadvantaging? Are these associations ones that individuals are unlikely to have anticipated and that might be disadvantageous to them? Answering this question requires frequent re-evaluation, if research questions change or if novel patterns become apparent. These patterns of course may be very valuable in identifying disease risks, but their potential for stigmatization should also be considered, not to halt the research but to raise questions about how research results can best be communicated to avoid deleterious effects when possible. Other aspects of this may be the need to try to communicate the information to individuals who may be identified with the group, both so that they may be aware of the possible risks and so that they may learn about potential benefits of the research.

(4) Are there offsetting benefits for individuals involved in the research as well as for individuals who might be associated with the research? If not, is it justifiable to include the individuals in the research or to report findings that may reveal possible implications about them? The Declaration of Helsinki requires that research with vulnerable populations be offset by benefits for them. Here, what we have in mind is the possibility that new vulnerabilities may be created by research using re-purposed data in the enormously powerful ways that are available today. Red-lining in insurance or sales for groups newly identified as at risk is an example of these possibilities. The Declaration has in mind impoverished populations across the globe that have been used for pharmaceutical trials and the possibility of providing some health care for them in return.

What we suggest here instead is along the lines of the Obama White House transparency principles for the PMI: that there should be ongoing efforts to inform people of how data uses are producing findings that are important to them. Unfortunately, the initial steps in implementing the PMI have not yet addressed concrete strategies for this. But there are several ways this might be done. One is publishing regular descriptions, in lay language, of research findings with highlights about those for whom they are most important. Another is sending reports of relevant research results to individuals for whom they might make a difference. The PMI is somewhat unique in that it preserves the possibilities for re-contacting individuals because of plans to collect longitudinal data. Data sets with identifiable information may well contain electronic addresses of individuals whose data are involved; messages to them could include suggestions that they contact others for whom the information might also be useful. Even when data have been de-identified as to data subjects, research and treatment information may contain contact information for the original researcher or treating physician. Communications

to them could help get the word out to people from whom the data were originally collected. Such communication, if interactive, could also help alert the researchers re-using the data about likely concerns of those from whom the data originated and others like them.

CONCLUSION

Re-use of data originally collected for research or treatment has enormous potential. As the PMI recognizes, this potential is magnified when data sets can be combined and when data can be collected over time. The standard concerns raised about these uses are that they may violate individual consent or that they might directly damage individual subjects through re-identification. In this chapter, we have argued that these may not in fact be the most important concerns, and that issues about stigmatization or disadvantage may be present even for individuals who were not included in the original data sets. As novel patterns are identified, these issues may not track conventional understandings of vulnerable population groups. We have suggested a number of strategies that might be used to communicate both risks and benefits so that research may be conducted in a way that is just both to those whose data are involved and to others who may be affected by the research.

NOTES

1. In September, 2015, a Notice of Proposed Rule-making for revisions of the Common Rule was posted (80 Fed. Reg. 53933, https://www.gpo.gov/fdsys/pkg/FR-2015-09-08/pdf/2015-21756.pdf). As of this writing, no final rule had been issued. It remains to be seen whether de-regulatory efforts of the Trump administration will affect data use and the regulations governing it.

2. To reach this conclusion, we conducted searches of the Westlaw allcases database for de-identif! Information; de-identifying information; re-identif! Information, ("re-identif! information") & "de-identified information"; research study & "de-identified information"; re-identified & individuals & "de-identified information"; and cases citing §164.514.

ACKNOWLEDGMENTS

We are grateful to Alexis Juergens for assistance with the research for this chapter. We are also grateful for support from the S. J. Quinney College of Law fund for excellence in faculty research and teaching.

REFERENCES

AU:8

Association of Academic Health Centers (AAHC). (2008). *HIPAA creating barriers to research and discovery*. Retrieved from http://www.aahcdc.org/policy/reddot/AAHC_HIPAA_Creating_Barriers.pdf

Carozza, S. E., Langlois, P. H., Miller, E. A., & Canfield, M. (2012). Are children with birth defects at higher risk of childhood cancers? *American Journal of Epidemiology*, *175*(12), 1217 1224.

Cate, F. H. (2010). Protecting privacy in health research: The limits of individual choice. *California Law Review*, *98*(6), 1765 1804.

Cote, M. L., Harrison, M. J., Wenzlaff, A. S., & Schwartz, A. G. (2014). Re-contacting participants for inclusion in the database of Genotypes and Phenotypes (dbGaP): Findings from three case-control studies of lung cancer. *Genome Medicine*, *6*(7), 54.

Department of Health and Human Services (HHS). (2011). Human subjects research protections: Enhancing protections for research subjects and reducing burden, delay, and ambiguity for investigators. *Federal Register*, July 26, *76*(143), 44512 44531.

Department of Health and Human Services (HHS). (2015). Federal policy for the protection of human subjects. *Federal Register*, September 8, *80*, 53931 54061.

el Emam, K., Jabbouri, S., Sams, S., Drouet, Y., & Power, M. (2006). Evaluating common de-identification heuristics for personal health information. *Journal of Medical Internet Research*, *8*(4), e28.

Engels, E. A., Pfeiffer, R. M., Fraumeni, J. F., Jr., Kasiske, B. L., Israni, A. K., Snyder, J. J., … Lin, M. (2011). Spectrum of cancer risk among US solid organ transplant recipients. *Journal of American Medical Association*, *306*(17), 1891 1901.

European Parliament. (2014, April 16). *Regulation (EU) No. 536/2014 of the European Parliament and of the Council*. Retrieved from http://ec.europa.eu/health/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf

European Parliament. (2016, April 27). *Regulation (EU) 2016/679 of the European Parliament and of the Council*. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

Evans, B. (2011). Much Ado about data ownership. *Harvard Journal of Law & Technology*, *25*(1), 69 130.

Francis, L. P., Battin, M. P., Botkin, J., Jacobson, J. A., & Smith. C. B. (2006). Infectious disease and the ethics of research: The moral significance of communicability. In M. Hayry, T. Takala, & P. Herissone-Kelly (Eds.), *Ethics in biomedical research: International perspectives* (pp. 135 149). Amsterdam: Rodopi.

Francis, L. P., Battin, M. P., Jacobson, J. A., & Smith, C. B. (2009). Syndromic surveillance and patients as victims and vectors. *Journal of Bioethical Inquiry*, *6*(2), 187 195.

Gellman, R. (2016, June 17). *Fair information practices: A basic history*. Retrieved from http://bobgellman.com/rg-docs/rg-FIPShistory.pdf

Gennet, E., Andorna, R., & Elger, B. (2015). Does the new EU Regulation on clinical trials adequately protect vulnerable research participants? *Health Policy*, *119*(7), 925 931.

Hoofnagle, C. J. (2016). *Letter to the FTC*. Retrieved from https://www.ftc.gov/system/files/documents/public_comments/2015/09/00003-97142.pdf

IBM. (2016). *Our purpose: We unlock the power of BIG DATA to improve healthcare for everyone*. Retrieved from https://www.explorys.com/about-us.html

Ludman, E. J., Fullerton, S. M., Spangler, L., Trinidad, S. B., Fujii, M. M., Jarvik, G. P., … Burke, W. (2010). Glad you asked: Participants' opinions of re-consent to dbGaP data submission. *Journal of Empirical Research on Human Research Ethics*, *5*(3), 9 16.

Malik, A. Y., & Foster, C. (2016). The revised Declaration of Helsinki: Cosmetic or real change? *Journal of the Royal Society of Medicine*, *109*(5), 184 189.

Malin, B., Benitez, K., & Masys, D. (2011). Never too old for anonymity: A statistical standard for demographic data sharing via the hipaa privacy rule. *Journal of the American Medical Informatics Association*, *18*(1), 3 10.

Manson, N. C., & O'Neill, O. (2007). *Rethinking informed consent in bioethics*. Cambridge: Cambridge University Press.

Mello, M. M., & Wolf, L. E. (2010). The Havasupai Indian tribe case—Lessons for research involving stored biologic samples. *New England Journal of Medicine*, *363*, 204 207.

Nakashima, E. (2008). Prescription data used to assess consumers; records aid insurers but prompt privacy concerns. *Washington Post*, August 4. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2008/08/03/AR2008080302077.html

Nass, S. J., Levit, L. A., & Gostin, L. O. (Eds.). (2009). *Beyond the HIPAA privacy rule*. Washington, DC: National Academies Press.

NIH. (2016). *National and State Cancer Registries*. Retrieved from http://epi.grants.cancer.gov/registries.html

Northwestern Office for Research, IRB Institutional Review Board. (2016). *Waiver or alteration of requirements for obtaining informed consent*. Retrieved from https://irb.northwestern.edu/process/new-study/informed-consent/waiver-or-alteration-requirements-obtaining-informed-consent

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, *57*, 1701 1777.

Ohm, P. (2014). Changing the rules: General principles for data use and analysis. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, big data, and the public good: Frameworks for engagement* (pp. 96 112). New York, NY: Cambridge University Press.

Precision Medicine Initiative (PMI) Working Group. (2015, September 17). *The precision medicine initiative cohort program—Building a research foundation for 21st century medicine*. Retrieved from https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf

Rainie, L., & Duggan, M. (2016, January 14). *Privacy and information sharing*. Retrieved from http://www.pewinternet.org/2016/01/14/privacy-and-information-sharing/

Rothstein, M. A. (2010). Is deidentification sufficient to protect health privacy in research? *American Journal of Bioethics*, *10*(9), 3 11.

Rothstein, M. A., & Shoben, A. B. (2013). Does consent bias research? *American Journal of Bioethics*, *13*(4), 27 37.

Safran, C. (2014). Reuse of clinical data. *Yearbook of Medical Informatics*, *9*(1), 52 54.

Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review*, *86*, 1814 1894.

Secretary's Advisory Committee on Automated Personal Data Systems. (1973, July). *Records, computers, and the rights of citizens*. Retrieved from https://epic.org/privacy/hew1973report/default.html

Smith, E. K., White, M. C., Weir, H. K., Peipins, L. A., & Thompson, T. D. (2012). Higher incidence of clear cell adenocarcinoma of the cervix and vagina among women born between 1947 and 1971 in the United States. *Cancer Causes and Control*, *23*(1), 207 211.

SWOG. (2016). *About Us*. Retrieved from http://swog.org/Visitors/AboutUs.asp

SWOG. (2016a). Policy Memorandum No. 39, Acquisition, maintenance and use in research of tissue and other biologic patient specimens. Retrieved from http://swog.org/Visitors/Download/Policies/Policy39.pdf

UCSF Medical Center. (2016). *How we may use and disclose your medical information*. Retrieved from https://www.ucsfhealth.org/hipaa/how_we_may_use_and_disclose_medical_information_about_you/#9

Vinterbo, S. (2011, May 12). *Are there limits to privacy preserving sharing of data?* Retrieved from http://www.ncvhs.hhs.gov/wp-content/uploads/2014/05/110512p1.pdf

White House. (2015, November 9). *Precision medicine initiative: Privacy and trust principles*. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/finalpmiprivacyandtrustprinciples.pdf

Whitney, S. N. (2014). The shell game: How institutional review boards shuffle words. *Journal of Translational Medicine*, *12*, 201.

Wolf, L. E., Patel, M. J., Williams, B. A., Austin, J. L., & Dame, L. A. (2013). Certificates of confidentiality: Protecting human subject research data in law and practice. *Minnesota Journal of Law, Science & and Technology*, *14*(1), 11 87.

Wolff, J., & deShalit, A. (2007). *Disadvantage*. Oxford: Oxford University Press.

World Medical Association (WMA). (2013). *WMA Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects*. Retrieved from http://www.wma.net/en/30publications/10policies/b3/

xG Health Solutions. (2016). *Population analytics*. Retrieved from http://xghealth.com/population-analytics/

UNCITED REFERENCES

Engels et al. (2011); Evans (2011); Manson and O'Neill (2007); Secretary's Advisory Committee on Automated Personal Data Systems (1973)

AUTHOR QUERY FORM

Emerald	**Book: SLPS-V073-3611449** **Chapter: CH006**	**Please e-mail or fax your responses and any corrections to:** **E-mail:** **Fax:**

Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

Disk use

Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article

☐ Scanning the artwork

Bibliography
If discrepancies were noted between the literature list and the text references, the following may apply:

☐ **The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.**

☐ ***UNCITED REFERENCES:*** **This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.**

Queries and/or remarks

Location in Article	Query / remark	Response
AU:1	We have followed US spelling throughout this chapter. As per US spelling, all "re-" should be a single word. Please confirm whether we can change all hyphenated "re" words as a single word.	
AU:2	Please check the sentence "Since 2011, HHS has engaged in…" for clarity.	
AU:3	Please provide expansions for NPRM and DES.	

AU:4	The citation of reference Wolf et al., 2012 in text has been changed to Wolf, Patel, Williams, Austin, and Dame, 2013 to match with the reference list. Please check.	
AU:5	The citation of reference Gellman, 2015 has been changed to Gellman, 2016 to match with the list. Please check.	
AU:6	Reference Rothstein, 2011 is cited in the text, but not provided in the list. Please provide the reference details.	
AU:7	Please check the phrase “associations ones.”	
AU:8	References "Engels et al. (2011), Evans (2011), Manson and O’Neill (2007), and Secretary’s Advisory Committee on Automated Personal Data Systems (1973) " have not been cited in the text. Please clarify as to where they should be cited.	