Title

Ensuring privacy in the study of pathogen genetics

Journal

The Lancet Infectious Diseases, 14(8)

ISSN

1473-3099

Authors

Mehta, Sanjay R

Vinterbo, Staal A

Little, Susan J

Publication Date

2014-08-01

DOI

10.1016/s1473-3099(14)70016-7

Peer reviewed

Published in final edited form as:
Lancet Infect Dis. 2014 August ; 14(8): 773–777. doi:10.1016/S1473-3099(14)70016-7.

Ensuring privacy in the study of pathogen genetics

Sanjay R. Mehta, MD,
Division of Infectious Diseases, University of California, San Diego, CA, USA

Staal A. Vinterbo, PhD, and
Division of Infectious Diseases, University of California, San Diego, CA, USA

Prof. Susan J. Little, MD
Division of Biomedical Informatics, University of California, San Diego, CA, USA

Abstract

Rapid growth in the genetic sequencing of pathogens in recent years has led to the creation of large sequence databases. This aggregated sequence data can be very useful for tracking and predicting epidemics of infectious diseases. However, the balance between the potential public health benefit and the risk to personal privacy for individuals whose genetic data (personal or pathogen) are included in such work has been difficult to delineate, because neither the true benefit nor the actual risk to participants has been adequately defined. Existing approaches to minimise the risk of privacy loss to participants are based on de-identification of data by removal of a predefined set of identifiers. These approaches neither guarantee privacy nor protect the usefulness of the data. We propose a new approach to privacy protection that will quantify the risk to participants, while still maximising the usefulness of the data to researchers. This emerging standard in privacy protection and disclosure control, which is known as differential privacy, uses a process-driven rather than data-centred approach to protecting privacy.

Introduction

The rapid advances in medical technology in the past decade have substantially decreased the cost and effort necessary to obtain genetic information for health-care applications. Genome-wide association studies with array technology, and rapid sequencing of host and pathogen genomes, RNA expression profiles, and microbiomes with next-generation sequencing technology are now commonplace. Individual use of host (eg, *BRCA* screening and breast cancer risk) and viral (eg, drug-resistance testing for HIV treatment) genetic data to inform medical management is regarded as routine health-care practice in many resource-rich countries. Subsequent aggregation and sharing of such data for research purposes maximises its usefulness by allowing a broader scope for the scientific community to apply ideas and skills to datasets that could be difficult for one individual or one group of

Correspondence to: Dr Sanjay R Mehta, Division of Infectious Diseases, University of California, San Diego, CA 92103–8208, USA srmehta@ucsd.edu.

Contributors
All authors contributed equally to the content and the writing of the report.

Declaration of interests
We declare that we have no competing interests.

researchers to assemble.[1–4] However, the privacy risks associated with such datasets, as recently outlined in work by Gymrek and colleagues[5] and in an accompanying commentary,[6] could be higher than previously appreciated. The investigators[5] used several publicly available host genetic databases to deduce the identity of individuals selected from the CEPH (Center for the Study of Human Polymorphisms, Paris, France) collection of reference families. Although such a risk was included in the consent process as only a remote possibility, this study showed that the privacy risks associated with so-called omics data are not yet fully defined.

With respect to infectious diseases, routine and next-generation technologies can provide rapid and fairly inexpensive identification and characterisation of a pathogen by analysis of its genome. This molecular characterisation of the pathogen can provide clues towards antimicrobial resistance and pathogenicity at the time of diagnosis, information that can greatly affect clinical decision making and outcomes. The compilation of such sequence data also offers substantial advances for the epidemiological tracking of these pathogens through the community and environment. These data are widely used in research to guide population-level intervention efforts (eg, influenza virus vaccine development[7] and the monitoring of bacterial resistance patterns[8]).

Although inferences and predictions can be made on the basis of data from individual research programmes or public health institutions, in many regions several separate programmes are collecting these types of data. Application of these methods to comprehensively collected sequence data acquired in real time has the potential for immediate and substantial public health benefits. However, the optimum application of these large-scale analyses has not yet been defined, partly because their technological development has outpaced the studies needed to define optimum implementation strategies. Particularly, the balance between the public health benefit and personal privacy risk for individuals whose genetic data (personal or pathogen) are included in such work has been difficult to delineate, since neither the true benefit nor the actual risk to the participant has been adequately defined.

This absence of equipoise with respect to risk has been a major limitation of large-scale molecular epidemiological analyses of HIV and hepatitis C virus (HCV)—studies that are needed to characterise the public health potential of network analyses for these infections. Uncertainty specifically related to the potential loss of privacy and the potential for unintended HIV or HCV status disclosure of individuals who might have transmitted the pathogen, along with concerns related to sharing sequence data, has resulted in a stalled research agenda.[9,10] In this Personal View, we examine the issues surrounding privacy protection in the specific case of HIV molecular epidemiological studies and offer a general pathway by which some of these issues could be addressed.

Molecular epidemiology of HIV

HIV is an ideal candidate for molecular epidemiological analysis because, as a single-stranded RNA virus, its intrinsically high mutation rate gives rise to essentially unique viral genetic sequences in each infected individual.[11] Phylogenetic analyses of HIV sequences

have been used extensively to identify the origins and early spread of the HIV epidemic,[12] to follow national trends of disease spread,[13,14] to assess characteristics of local epidemics,[15,16] and to identify historical outbreaks.[17] Although previous analyses have been retrospective, more recently these data are being investigated as a powerful means to track the disease, to make predictions about the dynamics of spread,[13] and to assess the effectiveness of interventions within populations.[18]

Should these methods prove viable, they could play an integral part in community-level public health prevention efforts by focusing prevention and intervention resources within subpopulations at the greatest risk of HIV transmission. Since the existing standard of care for people infected with HIV-1 in the USA involves sequencing of a portion of the *pol* gene at the time of diagnosis to determine if any baseline resistance to antiretroviral therapy is present,[19] the data necessary to do such molecular analyses are often readily available.[16] However, just as the uniqueness of each individual's virus allows for detailed molecular epidemiology, this same distinctiveness could represent a personally identifying characteristic.[9]

Genetic information

The past decade has seen a large increase in the amount of biological data that are stored electronically, particularly genetic sequence information. Large public repositories of HIV-1 sequence data are maintained and accessible online to researchers worldwide.[20] Although technology has increased the ease with which these data can be generated and centralised, efforts to develop guidelines for their use that minimise privacy risks remain absent or inadequate.

The challenge of preserving privacy associated with analysis of genetic information was addressed by US policy makers in the Genetic Information Non-discrimination Act of 2008 (GINA),[21] which was created to prevent discrimination based on genetic information. However, GINA's focused language and the explosive growth in technology and science that has occurred since it was enacted have left it with a narrow purview. Particularly, GINA defines genetic information as information about an individual's genetic tests, information about the genetic tests of the members of an individual's family, and information about manifestations of disease or disorder in members of an individual's family. GINA goes on to define genetic testing as "an analysis of human DNA, RNA, chromosomes, proteins, or metabolites that detects genotypes, mutations, or chromosomal changes".

Salient to HIV molecular epidemiology, GINA does not address community-level genetic information or non-host genetic information. Community-level genetic information refers to information that can be inferred about an individual from known genetic patterns present in members of his or her community. The increasing mobility of individuals in high-income countries has led to substantial mixing of populations between communities separated by geography, unique demographic characteristics, racial or ethnic groupings, or exposure risks, but tight-knit and sometimes closed populations remain.[22,23] GINA also does not directly address non-host genetic information—ie, the genetic information of circulating pathogens and commensal organisms that live within an individual. However, these genetic

data can be very host specific, such as the sequence of the circulating virus of an HIV-infected individual.

When this non-host genetic information is taken into account, community-level information can be even more powerful. The case study of HIV molecular epidemiology, wherein the science is outpacing regulation, shows this fact well. Routinely generated HIV viral sequence data (used to identify and characterise resistance to anti-retroviral therapy) provide great opportunities for both researchers (to study HIV epidemiology) and clinicians (to guide personal health-care decisions). However, more recently, these data are being used to elucidate putative HIV transmission clusters both locally[16,24–28] and worldwide.[14,29] Because of the nearly unique HIV sequence of each infected person, identification of sequences from two or more individuals that are more similar than would be expected by chance allows the inference of a putative transmission link. The associated privacy risk posed by inferences based on sampled individuals is mitigated by the possibility that one or more unsampled individuals could be present in between the sampled individuals along the transmission chain (ie, genetic linkage of two individuals does not prove direct transmission).[14] Although directionality cannot be inferred without additional epidemiological information, new analytical techniques can allow for informed guesses on the basis of other types of data[30] or knowledge of within-host viral evolution.[31] However, because of the criminal statutes regarding HIV transmission in 37 US states and numerous nations, including Canada, even the establishment of such putative linkages could have legal implications.[32,33]

Putative transmissions paired with sociodemographic data might lead to the unintended identification of individuals or groups with unique characteristics (eg, risk behaviour, travel history, etc), with the potential result of increased social stigma from inferences about both HIV status and these characteristics (eg, injection drug use). However, the widespread availability of such sequence data contributes numerous benefits to the clinical and research communities. For example, applications of transmission network analyses could improve the efficiency of public health prevention interventions through identification of individuals or groups at high risk for transmitting virus and selected groups with increased rates of epidemic growth. Such network analyses could also make it easier to observe and understand disease trends.

The absence of available methods to both quantify and preserve privacy, while making use of genetic sequence data for clinical and public health benefit, remains a stumbling block for HIV molecular epidemiology research. The simple solution would be to extend the protections noted in GINA to the community level and non-host genetic information. However, this approach would result in severe curtailment of aggregations of sequence data, resulting in substantial limitations to the burgeoning potential of molecular epidemiology and loss of the potential public health benefit. Such an approach would also not necessarily protect individuals participating in research studies, since even the results of the analyses of research sequence datasets have the potential to provide identifying information.[34] Indeed, these fears are currently limiting cooperation between research groups, pharmaceutical companies, and public health institutions to further realise the potential of molecular epidemiological analyses.

Privacy

In the USA, the present expectation of privacy protection in research involving human participants is described in the Health Insurance Portability and Accountability Act of 1996 (HIPAA). According to the HIPAA Privacy Rule, data are de-identified (ie, privacy protected) if at least one of two criteria are met: an expert, using accepted methods, determines that the risk of associating individual data items with any individual (ie, re-identification) is “very small” (so-called expert determination); or all of the 18 predefined information items or characteristics present in the data are removed (the so-called Safe Harbor rule; panel). Because of the challenges of satisfying the expert determination method,[35] investigators of most studies opt to implement the Safe Harbor rule, requiring the removal of the 18 items deemed linkable to other data sources.

However, in the context of genetic information, the Safe Harbor rule might not be safe enough.[5,6] Additionally, data not contained within the HIPAA’s list of items might still be interpreted in the context of a larger dataset or population, allowing inferences to be drawn that could compromise personal privacy. For example, when a person is newly diagnosed with syphilis, contact tracing is routinely done by public health agencies to identify sexual partners of the infected index case. Although contacted partners are provided with no information about the source of their exposure risk, if they have had only one sexual contact they will know the identity of the infected individual without having been provided with individually identifiable health information related to that person. In this case the released information, syphilis risk, would pass the Safe Harbor privacy test since none of the HIPAA’s 18 identifiers were provided to the sexual contact of the index individual, but the process of release allowed the individual to make an inference from the released data.

The alternative to the Safe Harbor method, the expert determination method of privacy preservation, requires experts to establish a threshold below which the risk of a privacy breach is deemed safe. This assessment requires that the risks associated with re-identification are quantifiable; however, empirical testing of a dataset to establish a threshold for safety is problematic. The fact that one particular researcher is not able to re-establish linkage to the study participant does not mean that another will be similarly unable.[5,36–38] Therefore, a non-empirical quantification of risk is needed.

The issue of privacy is particularly germane to the aggregation and analysis of data needed for large-scale HIV molecular epidemiological studies. Knowledge of sociodemographic information related to each HIV sequence is necessary to inform the models that can be used to predict transmission patterns and assess the effects of interventions. However, even if this socio-demographic information is stripped of the 18 HIPAA identifiers, having this information might still allow reasonable inferences to be made about the identity of individuals in the study. Thus, for molecular epidemiological analyses to preserve privacy, the ability to quantify privacy risks in a meaningful way will be crucial, and will represent the basis for risk–benefit analysis essential for sound policy making.

Unfortunately, increased efforts to protect individual privacy will require existing privacy-related practices to change, since there are limits to how far existing technological

safeguards can go.[39–41] For example, recent research suggests that quantifying the privacy properties of data (such as de-identified data as defined by HIPAA) is impossible.[42,43]

Panel: **Protected health information, as defined by the HIPAA Privacy Rule***

- Name
- Geographical subdivision smaller than US state, apart from the first three digits of a zip code (provided that the geographical area defined by those digits includes more than 20 000 individuals)
- All dates related to an individual (apart from years), and age if 89 years or older (eg, date of birth, date of hospital admission, etc)
- Telephone number
- Fax number
- Email address
- Social security number
- Medical record number
- Health plan beneficiary number
- Any account numbers
- Any certificate or licence numbers
- Vehicle identifiers and serial numbers
- Device identifiers and serial numbers (eg, for medical devices)
- Web URLs
- IP address numbers
- Biometric identifiers
- Facial photographs
- Any other unique identifier, characteristic, or code

List of 18 identifiers of personal health information, as defined by the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule.

URL=uniform resource locator. IP=internet protocol.

Differential privacy

A new and different approach shifts the focus of privacy from being regarded as a property of the data, as reflected in the HIPAA de-identification standard, to account for the process by which information is disclosed independently of the data. The emerging standard in privacy protection and disclosure control, termed differential privacy,[42,43] is based on this approach.

A benefit of moving from data-centred privacy to process-driven privacy is that this approach provides previously missing non-empirical quantitation of privacy. Because this measurement is based on the process of data disclosure, the process can be modified to reduce the risk of privacy loss below an agreed upon threshold. For example, differential privacy guarantees protection of privacy below a chosen threshold of data release through controlled randomness that is introduced in the disclosure process. This threshold could then be used to provide a quantitatively defined guarantee of strong privacy to research participants, or to set a reasonable level of privacy for public health analyses of the data. For example, in the case of syphilis contact tracing, rather than only identifying the sexual contacts of the index case, a social group of more than ten individuals including the contact of interest could be identified and tested for syphilis, thus reducing the risk of disclosure to the index individual.

An example of a differentially private approach to analysis could be envisioned for HIV sequence data. For example, a local HIV sequence database might be queried to determine the sociodemographic characteristics of the individuals who make up an expanding HIV injection-drug-use transmission cluster. Suppose that this particular cluster included individuals of generally homogeneous geography and sociodemographic characteristics, as well as a few individuals with uniquely distinctive characteristics (eg, race, vocation, etc). Release of these HIPAA-compliant data could still result in the unintentional disclosure of some members of the transmission cluster. A differentially private release of these data could output slightly perturbed counts of these distinctive characteristics within the cluster, where the variance of the perturbation is determined by the level of privacy required.

In addition to not always protecting privacy, the application of the HIPAA Safe Harbor method has been criticised for rendering data useless, particularly for epidemiology, because of the removal of too much information.[44,45] A process-oriented approach to privacy would allow several groups to aggregate their data into secure data warehouses for the purposes of queries or analysis on the combined dataset, without releasing all of the data into a public domain or even to the contributing groups.[46,47] This approach would allow useful information to be realised from the full dataset, but would preserve privacy risks to study participants in a way that can be quantified and specified for individual analyses.

Conclusions

The intersection of disease and genetics is growing rapidly, whereas privacy guidelines and policies have not kept pace with the technological advances that continue to expose new challenges to personal health-information privacy. Present methods of privacy protection in human research studies limit data sharing between researchers and constrain research opportunities. In the particular case of HIV molecular epidemiology, the risks to patient privacy are, at best, poorly quantified, and the recognised potential for loss of privacy limits research that could offer more effective methods to reduce HIV transmission.

The incorporation and assessment of differential privacy methods into collaborative data analyses could provide formal, data-independent guarantees that limit privacy risk while simultaneously increasing the yield from disparate datasets. For HIV particularly, the

adoption and further development by scientists and ethicists of modern formal definitions and methods of preserving privacy will enable rational policy development. Recognition and use of these privacy-preserving policies will pave the way for molecular epidemiology to maximally improve public health while providing strong protection of patient privacy.

Acknowledgments

We thank Davey Smith for his assistance in reviewing the report. We also thank our funders—the US National Institutes of Health (grants AI093163 [SRM], AI43638 [SJL], AI074621 [SJL], LM07273 [SAV], and HL108460 [SAV]) and Gilead Biosciences (SJL).

References

1. Birney E. The making of ENCODE: lessons for big-data projects. Nature. 2012; 489:49–51. [PubMed: 22955613]
2. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics—re-shaping scientific practice. Nat Rev Genet. 2009; 10:331–35. [PubMed: 19308065]
3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]
4. Wellcome Trust. Sharing data from large-scale biological research projects: a system of tripartite responsibility. Report of a meeting organized by the Wellcome Trust; 14–15 January 2003; Fort Lauderdale, USA. London: Wellcome Trust; 2003.
5. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013; 339:321–24. [PubMed: 23329047]
6. Rodriguez LL, Brooks LD, Greenberg JH, Green ED. The complexities of genomic identifiability. Science. 2013; 339:275–76. [PubMed: 23329035]
7. Horimoto T, Kawaoka Y. Designing vaccines for pandemic influenza. Curr Top Microbiol Immunol. 2009; 333:165–76. [PubMed: 19768405]
8. Gilbert JM, White DG, McDermott PF. The US national antimicrobial resistance monitoring system. Future Microbiol. 2007; 2:493–500. [PubMed: 17927472]
9. Hecht FM, Wolf LE, Lo B. Lessons from an HIV transmission pair. J Infect Dis. 2007; 195:1239–41. [PubMed: 17396989]
10. UNAIDS Reference Group on HIV and Human Rights. [accessed April 30, 2013] Statement on criminalization of HIV transmission and exposure. UNAIDS Reference Group on HIV and Human Rights. 2009. http://data.unaids.org/pub/Report/2009/20090303_hrrefgroupcrimexposure_en.pdf
11. Holmes EC. RNA virus genomics: a world of possibilities. J Clin Invest. 2009; 119:2488–95. [PubMed: 19729846]
12. Worobey M, Gemmel M, Teuwen DE, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 2008; 455:661–64. [PubMed: 18833279]
13. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT. for the UK HIV Drug Resistance Collaboration. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. J Infect Dis. 2011; 204:1463–69. [PubMed: 21921202]
14. Aldous JL, Pond SK, Poon A, et al. Characterizing HIV transmission networks across the United States. Clin Infect Dis. 2012; 55:1135–43. [PubMed: 22784872]
15. Dennis AM, Hué S, Hurt CB, et al. Phylogenetic insights into regional HIV transmission. AIDS. 2012; 26:1813–22. [PubMed: 22739398]
16. Smith DM, May SJ, Tweeten S, et al. A public health model for the molecular surveillance of HIV transmission in San Diego, California. AIDS. 2009; 23:225–32. [PubMed: 19098493]
17. Mehta SR, Wertheim JO, Delport W, et al. Using phylogeography to characterize the origins of the HIV-1 subtype F epidemic in Romania. Infect Genet Evol. 2011; 11:975–79. [PubMed: 21439403]
18. Wertheim JO, Kosakovsky Pond SL, Little SJ, De Gruttola V. Using HIV transmission networks to investigate community effects in HIV prevention trials. PLoS One. 2011; 6:e27775. [PubMed: 22114692]

19. Johnson VA, Brun-Vézinet F, Clotet B, et al. Update of the drug resistance mutations in HIV-1: December 2010. Top HIV Med. 2010; 18:156–63. [PubMed: 21245516]
20. Kuiken C, Korber B, Shafer RW. HIV sequence databases. AIDS Rev. 2003; 5:52–61. [PubMed: 12875108]
21. US Congress. [accessed Feb 11, 2014] HR 493 (110th): Genetic Information Nondiscrimination Act of 2008. https://www.govtrack.us/congress/bills/110/hr493/text
22. Gura T. Rare diseases: genomics, plain and simple. Nature. 2012; 483:20–22. [PubMed: 22382959]
23. Fernandez Vina MA, Hollenbach JA, Lyke KE, et al. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. Philos Trans R Soc Lond B Biol Sci. 2012; 367:820–29. [PubMed: 22312049]
24. Ragonnet-Cronin M, Hodcroft E, Hué S, et al. for the UK HIV Drug Resistance Database. Automated analysis of phylogenetic clusters. BMC Bioinformatics. 2013; 14:317. [PubMed: 24191891]
25. Yebra G, Holguín A, Pillay D, Hué S. Phylogenetic and demographic characterization of HIV-1 transmission in Madrid, Spain. Infect Genet Evol. 2013; 14:232–39. [PubMed: 23291408]
26. Balode D, Skar H, Mild M, et al. Phylogenetic analysis of the Latvian HIV-1 epidemic. AIDS Res Hum Retroviruses. 2012; 28:928–32. [PubMed: 22049908]
27. Karlsson A, Björkman P, Bratt G, et al. Low prevalence of transmitted drug resistance in patients newly diagnosed with HIV-1 infection in Sweden 2003–2010. PLoS One. 2012; 7:e33484. [PubMed: 22448246]
28. Brenner BG, Roger M, Stephens D, et al. and the Montreal PHI Cohort Study Group. Transmission clustering drives the onward spread of the HIV epidemic among men who have sex with men in Quebec. J Infect Dis. 2011; 204:1115–19. [PubMed: 21881127]
29. Wertheim JO, Leigh Brown AJ, Hepler NL, et al. The global transmission network of HIV-1. J Infect Dis. 2014; 209:304–13. [PubMed: 24151309]
30. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005; 22:1185–92. [PubMed: 15703244]
31. Kreimer, A.; Shaefer, MS.; Pfeifer, N. Defining HLA genotypes from bulk HIV sequences. 20th Conference on Retroviruses and Opportunistic Infections; Atlanta, GA, USA. March 3–6, 2013;
32. Metzker ML, Mindell DP, Liu XM, Ptak RG, Gibbs RA, Hillis DM. Molecular evidence of HIV-1 transmission in a criminal case. Proc Natl Acad Sci USA. 2002; 99:14292–97. [PubMed: 12388776]
33. Kaiser Family Foundation. [accessed Feb 11, 2014] Criminal statutes on HIV transmission. http://www.statehealthfacts.org/comparetable.jsp?ind=569&cat=11
34. Homer N, Szelinger S, Redman M, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 2008; 4:e1000167. [PubMed: 18769715]
35. US Department of Health and Human Services. [accessed April 1, 2013] Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html
36. Atreya RV, Smith JC, McCoy AB, Malin B, Miller RA. Reducing patient re-identification risk for laboratory results within research datasets. J Am Med Inform Assoc. 2013; 20:95–101. [PubMed: 22822040]
37. Kushida CA, Nichols DA, Jadrnicek R, Miller R, Walsh JK, Griffin K. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. Med Care. 2012; 50 (suppl):S82–101. [PubMed: 22692265]
38. McGraw D. Building public trust in uses of Health Insurance Portability and Accountability Act de-identified data. J Am Med Inform Assoc. 2013; 20:29–34. [PubMed: 22735615]
39. Chawla, S.; Dwork, C.; McSherry, F.; Smith, A.; Wee, H. Toward privacy in public databases. In: Killian, J., editor. Theory of cryptography—Second Theory of Cryptography Conference, TCC

2005 (Cambridge, MA, USA, February 10–12, 2005) proceedings (Lecture Notes in Computer Science 3378). Berlin, Heidelberg: Springer-Verlag; 2005. p. 363-85.

40. Dinur, I.; Nissim, K. Proceedings of the twenty-second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS). New York: Association for Computing Machinery; 2003. Revealing information while preserving privacy; p. 202-10.
41. Gutmann A. Data re-identification: prioritize privacy. Science. 2013; 339:1032. [PubMed: 23449576]
42. Dwork, C. In: Bugliesi, M.; Preneel, B.; Sassone, V.; Wegener, I., editors. Differential privacy; Automata, languages and programming—33rd International Colloquium, ICALP 2006; Venice, Italy. July 10—14, 2006; Berlin, Heidelberg: Springer-Verlag; 2006. p. 1-12.proceedings, part II (Lecture Notes in Computer Science 4052)
43. Dwork C, Pottenger R. Toward practicing privacy. J Am Med Inform Assoc. 2013; 20:102–08. [PubMed: 23243088]
44. Gostin LO, Nass S. Reforming the HIPAA Privacy Rule: safeguarding privacy and promoting research. JAMA. 2009; 301:1373–75. [PubMed: 19336713]
45. Tang PC. An AMIA perspective on proposed regulation of privacy of health information. J Am Med Inform Assoc. 2000; 7:205–07. [PubMed: 10730605]
46. Gardner J, Xiong L, Xiao Y, et al. SHARE: system design and case studies for statistical health information release. J Am Med Inform Assoc. 2013; 20:109–16. [PubMed: 23059729]
47. Mohammed N, Jiang X, Chen R, Fung BC, Ohno-Machado L. Privacy-preserving heterogeneous health data sharing. J Am Med Inform Assoc. 2013; 20:462–69. [PubMed: 23242630]