Title

Privacy Policy and Technology in Biomedical Data Science

Permalink

https://escholarship.org/uc/item/28p1f6xg

Journal

Annual Review of Biomedical Data Science, 1(1)

ISSN

2574-3414

Authors

Arellano, April Moreno
Dai, Wenrui
Wang, Shuang
et al.

Publication Date

2018-07-20

DOI

10.1146/annurev-biodatasci-080917-013416

Published in final edited form as:
Annu Rev Biomed Data Sci. 2018 July ; 1: 115–129. doi:10.1146/annurev-biodatasci-080917-013416.

Privacy Policy and Technology in Biomedical Data Science

April Moreno Arellano[#], **Wenrui Dai**[#], **Shuang Wang**, **Xiaoqian Jiang**, and **Lucila Ohno-Machado**
Department of Biomedical Informatics, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA; lohnomachado@ucsd.edu

[#] These authors contributed equally to this work.

Abstract

Privacyis an important consideration when sharing clinical data, which often contain sensitive information. Adequate protection to safeguard patient privacy and to increase public trust in biomedical research is paramount. This review covers topics in policy and technology in the context of clinical data sharing. We review policy articles related to (*a*) the Common Rule, HIPAA privacy and security rules, and governance; (*b*) patients' viewpoints and consent practices; and (*c*) research ethics. We identify key features of the revised Common Rule and the most notable changes since its previous version. We address data governance for research in addition to the increasing emphasis on ethical and social implications. Research ethics topics include data sharing best practices, use of data from populations of low socioeconomic status (SES), recent updates to institutional review board (IRB) processes to protect human subjects' data, and important concerns about the limitations of current policies to address data deidentification. In terms of technology, we focus on articles that have applicability in real world health care applications: deidentification methods that comply with HIPAA, data anonymization approaches to satisfy well-acknowledged issues in deidentified data, encryption methods to safeguard data analyses, and privacy-preserving predictive modeling. The first two technology topics are mostly relevant to methodologies that attempt to sanitize structured or unstructured data. The third topic includes analysis on encrypted data. The last topic includes various mechanisms to build statistical models without sharing raw data.

Keywords

health care privacy technology; health care policy; health care security technology; data sharing

1. INTRODUCTION

Research involving health information faces significant challenges, risks, and limitations related to patient privacy. There is increasing public awareness of the potential risks of sharing sensitive data and concern about patient privacy (1, 2). The widespread adoption of electronic health records (EHRs) and the desire for interoperable systems made these

DISCLOSURE STATEMENT
The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

concerns more salient, with patients becoming interested in controlling their own data (3). Recent work has suggested that many patients do not approve having all their health data shared with anyone, even when the main identifiers have been removed; instead, they want control over what should be accessed by whom (4). A patient's ability to granularly control their electronic health data is an important element in discussions about privacy (5). Existing policies and technologies have addressed some of these concerns from different perspectives.

Current health and technology policies include a focus on patient rights to privacy and human subjects' protections in research. The US Department of Health and Human Services Common Rule was designed to implement steps to improve human subjects' protection in research and to facilitate research (6). The Common Rule was updated for the first time since 1991, and main portions of the updated Common Rule may become effective in 2018. The updated Common Rule regulations provide more flexibility on the ethical conduct of human subjects in research, such as enabling more exemptions for low-risk studies and allowing broader patient consent practices to be implemented.

Additionally, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) rules are of particular importance. HIPAA contains privacy and security rules that help protect patient privacy.

The HIPAA privacy rule was designed to regulate the use of protected health information (PHI) held by so-called covered entities (e.g., health care organizations) within the United States. It requires appropriate privacy protection of health information to be applied (e.g., deidentification of PHI for secondary use of data in biomedical research). There are two deidentification mechanisms under the HIPAA privacy rule: expert determination and safe harbor. The former requires an expert to certify that there is no significant risk of the patient being identified from the data. The latter specifies a particular set of identifiers to be removed from the data.

The HIPAA security rule describes national standards for proper storage, use, and transmission of electronic personal health information handled by covered entities. It is designed to protect the confidentiality, integrity, and security of patient data (7).

As a complement to policy, many technology solutions were designed to preserve patient privacy beyond the requirements of HIPAA (8), and new technologies are continuously being developed to meet the emerging needs of researchers and data custodians. For example, algorithms have been developed to protect the process of building a model without sharing patient-level data (9), and various data anonymization techniques have been developed to support privacy-preserving dissemination of data (10, 11). Advanced security models can enable confidentiality during the entire data lifecycle, including computation on encrypted data (12). Technology solutions are not a replacement for policy, as they often focus on very specific tasks. Policies should describe when and how particular technologies should be employed.

In this article, we review the most relevant publications on clinical data privacy policy and technology that became available in MEDLINE, prioritizing those published in the past five years. Given the breath of technology and data types used in biomedicine, we are focusing mostly on EHRs since other data typically require an equal or lower level of protection. Because of the increasing integration of EHRs with genomic data, we also include some relevant techniques that emerged in recent years in the section titled Technology.

2. POLICIES TO PROTECT PATIENT PRIVACY

2.1. General Concerns About the Common Rule, HIPAA, and Governance

Revisions to the Common Rule include new requirements for patient consent, including a requirement for clinical trials consent forms to be publicly available online within 60 days. Further directives include different requirements for consent in order to conduct research with remnants of biospecimens and use of a single IRB for research involving multiple sites. Other features of the revised Common Rule include requirements to provide consent forms to patients in a more condensed and simplified format. Additionally, the definition of “vulnerable subjects” has been broadened to include individuals with limited cognitive abilities or that are educationally or economically disadvantaged (13). More leniency is now provided to researchers for obtaining broader consent to use the same data for possible future research on a not-yet-known topic. Additionally, two new exemptions include the use of identifiable (but publicly available) data for secondary use, as well as exemptions for long-term data and specimen storage if broad consent was given (14). A summary of key changes is shown in Table 1.

Protections of human subjects in IRB processes are currently not implemented uniformly, as the Common Rule can be subject to different interpretations (15). It is unclear whether the revised rules will lead to fewer variations.

HIPAA covers many different aspects of health care-related regulations. We focus on the privacy protection policies under HIPAA, which were designed to guide covered entities in handling PHI in a protected manner. The data deidentification methods defined under the HIPAA privacy rule include standard deidentification, expert determination and safe harbor methods. Standard deidentification involves methods where there is “no reasonable basis” for believing that the patient can be identified from the data. Covered entities must additionally meet one of the two requirements of data deidentification. Expert determination requires the use of statistical methods by trained experts, with the assurance and process documentation that there would be a “very small risk” of reidentification of patients from the data. However, there is no clear definition of “very small risk” to define or clarify the terms of these requirements. Safe harbor is another option that requires the omission of 18 types of patient identifiers from the data. The advantage of this method is that it is procedurally simple and can be automated. The main disadvantage is the decrease in data usefulness once all of the requirements are met (19). Benitez & Malin (20) evaluated privacy risks with respect to the HIPAA privacy rule and showed its vulnerability: The reidentification rate ranged from 0.01% to 0.25%. Figure 1 illustrates both mechanisms from a high level; interested readers can refer to Reference 19 for more details.

Despite policies and regulations, there are continuing concerns in the literature related to patient privacy. Liu et al. (21) discussed the HIPAA safe harbor policy for data sharing and its limitations, particularly for data from patients with similar or rare diseases and the risks of reidentification. Rodrigues et al. (22) discussed the necessary security requirements of sharing EHR data in the cloud using third-party cloud services. The efficiency of data deidentification policies and regulations is questionable. For example, some researchers have proposed a utility and reidentification space to efficiently quantify rule-based policies (23). By reviewing a wide range of approaches to reducing disclosure risk, O'Keefe & Rubin (24) discussed trade-offs between disclosure risk and data utility. Additional policy topics have surfaced. Glenn & Monteith (25) noted that various types of patient medical or health data are outside the scope of HIPAA. These data include mobile health and social media data owned by various business entities. Despite compliance with legal privacy protections in the health care and technology sectors, threats to patient privacy continue to pose serious concerns. A major question is: How can health care and clinical providers assure patients' trust in the privacy and security of their data (25, 26)? DeAngles (27) discussed the need, at the federal level, for EHR regulation to promote more widespread EHR use through the development of an interoperable national network of advanced EHRs: Federal and state laws need to be synchronized to minimize fragmented data waste and to reduce health care costs.

A clear data governance framework and easy-to-follow guidelines are key for policies and technology to meet privacy and security needs. Distributed research networks need governance guidelines to address the social and ethical concerns of patients, and some have included patients in the governance process (28). Holmes (29) discussed the changing landscape of data governance.

2.2. Patient Perspectives, Consent for Use of Data, and Research Ethics

Mamo et al. (28) discovered that patients were willing to share data to increase health care knowledge but also noted patient concerns with data security, as well as broader issues of social responsibility, commercialization of data, and the public benefit. Sources of concern included personal health data collected from a broad range of sources, such as from clinical, genome, social, behavioral, environmental and financial sources (28, 29).

Accessing data for health care improvement and biomedical research is critically important, but Luchenski et al. (30) highlighted the challenge of balancing these benefits with public concerns about security and privacy. Trachtenbarg et al. (31) discussed patients' interests in medical information privacy but also highlighted the costs of privacy and risks to patient health if important information is not shared. Their study reported that most patients were not willing to pay for increased data privacy and that they understood the importance of data sharing to improve health care. K.K. Kim et al. (32) addressed gaps in data sharing and privacy research in a study of California consumer views on privacy and security of EHR data for the purposes of research and improved health care. Bull et al. (33) discussed the increased support for data sharing for biomedical research in the literature, concluding that the support for data sharing by individuals was influenced by the existence of data sharing ethics and best practices policies. K.K. Kim et al. (34) identified factors required to increase patient trust, confidence, and willingness to share their electronic data: Consumer choice to share health data was affected by beliefs and attitudes towards EHRs, research benefits, and individual control of the data.

The consequences of selection bias are unclear when excluding data from patients who do not consent to data sharing for research. There are situations in which data sharing can be considered minimal risk and of high potential benefit to society (35), thus calling for a framework for balancing preferences, risks, and benefits.

In a discussion of data privacy for rare disease research, Mascalzoni et al. (36) argued for alternative approaches to data privacy and consent from the perspective of opportunity rather than barriers. They suggested IT-based approaches to provide a balance between patient data control and the opportunity to participate in research. Even in highly competitive situations such as in the context of therapeutic clinical trials, there are examples from the pharmaceutical industry in which researchers accessed clinical trial data and shared them with others while minimizing risks to privacy and confidentiality (37). Findings have shown that, in general, patients support the use of data for research, even though some insist on prior consent (1). An example of this development is iCONCUR (informed consent for clinical data and bio-sample use for research) (4), developed to provide patients with granular-level informed consent options for the use of their data and biospecimens for research purposes.

The transition from paper records to EHRs was identified as a challenge in working with longitudinal patient data, as both formats continue to exist at the present time. The current limitations of working with health record data in paper and electronic formats affect the ability of data to be used for research purposes. Challenges related to data privacy, as well as data presentation and interoperability, continue to exist (38).

3. TECHNOLOGY

We have identified various topics in health care technology, including HIPAA compliant deidentification, data anonymization, encryption methods, and privacy-preserving predictive modeling. At a high level, deidentification covers methods aiming at the HIPAA privacy rule; data anonymization provides an additional layer of protection to remove unique identifiability; encryption methods support secure data storage, retrieval, and sharing; and privacy-preserving predictive modeling uses a suite of combined technologies to fulfill data analysis needs while protecting privacy.

3.1. Deidentification Methods

Deidentification methods were introduced to protect patient privacy by removing sensitive information from health care data to comply with policy (i.e., the HIPAA privacy rule). Rule-and machine learning-based systems have been developed for automatic HIPAA-defined deidentification of EHRs. Rule-based systems leverage human expertise to define patterns to deidentify data and can be implemented in a straightforward manner for structured data (e.g., excluding the 18 identifiers defined by HIPAA safe harbor), but implementation is nontrivial for narrative text. Hanauer et al. (39) reported an experiment to quantify the human annotation efforts needed to build a deidentification system for narrative patient records using the MITRE Identification Scrubber Toolkit for statistical deidentification of history and physical notes, as well as social worker notes. The results show a good F_1 score (0.95) after 21 rounds of iterative annotation (this is quite laborious when compared with pure machine learning systems). Meystre et al. (40) discussed automated deidentifying clinical information text based on the 2010 i2b2 (Informatics for Integrating Biology and the Bedside) natural language processing challenge corpus and a corpus of clinical notes from the Veterans Health Administration (VHA), which still resulted in clinical information with 0.81% exact overlap and 1.78% partial overlap with PHI. Gardner & Xiong (41) developed a framework named HIDE (Health Information for Deidentification) to deidentify unstructured medical data using a simple Bayesian classifier, which showed an overall accuracy of 0.75 (identifying medical record number, account number, age, date, name, etc.) on 100 textual pathology reports from a cancer center. Ferrández et al. (42) presented BoB, a hybrid clinical system automating text deidentification that combined rule- and machine learning-based models. The proposed system was evaluated for generalizability and portability based on a manually annotated VHA corpus, as well as the i2b2 challenge corpus of 2006. BoB demonstrated recall of 92.6% and had reasonable precision (83.6%). Dernoncourt et al. (43) developed another deidentification system based on artificial neural networks. In comparison with existing systems, their model did not require specifically designed features or rules and achieved better deidentification performance on the publicly available i2b2 2014 deidentification challenge data set (F_1 score = 97.85) and the MIMIC (F_1 score = 99.23) data set. To date there is no single method that is considered sufficient or required for narrative text deidentification.

3.2. Data Anonymization Methods

The goal of patient data anonymization is to hide information from an individual in a cohort of others so that no one is uniquely identifiable. This cannot be guaranteed using deidentification methods alone. Approaches proposed in this category aim at removing the unique identifiability, a major vulnerability that puts individual participants at risk of reidentification. It is important to note why unique identifiability is important to protect privacy and how it goes beyond HIPAA deidentification protections. For example, while it would be difficult to determine the identity of an individual from a database containing biometrics and diagnoses alone, if the goal is to determine the diagnosis of an individual whose biometrics (e.g., fingerprints, genome sequence) we obtained by legal or illegal means, it would be trivial to match the biometrics in a disclosed database and look up the diagnosis associated with it. That is, deidentification does not take into account information external to the database that can be linked to the data being disclosed.

Various techniques have been developed to anonymize data while maintaining data utility as much as possible. One of the pioneer anonymization definitions is called k-anonymity (44), which aims at guaranteeing that each individual patient's information cannot be distinguished from the information of at least $k - 1$ other patients in a single database. This type of method and its various derived methods do not address the problem of data being linkable to external databases. For example, l-diversity and t-closedness principles extended k-anonymization. Several studies were conducted to anonymize data based on these definitions using suppression and generalization operations. For example, Aristodimou et al. (45) proposed a pattern-based multidimensional suppression technique (kPB-MS) for privacy-preserving data publishing to minimize the information loss through k-anonymity in which feature selection was incorporated to reduce data dimensionality and combine attributes for record suppression. Yoo et al. (46) described a generalization method satisfying both *k*-anonymity and *l*-diversity that used conditional entropy to measure the loss of information as well as mutual information for sensitive attributes.

Generalization methods have also been adopted for anonymizing structured terminologies in EHRs [i.e., ICD-9 (International Classification of Diseases, Ninth Revision), ICD-10, and SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms)]. Tamersoy et al. (47) presented a method to support secure sharing of patient-specific longitudinal data for biomedical research. Sequence aligning and clustering methods were adopted to aggregate temporal and diagnostic information while preserving data utility. Martínez et al. (48) developed a generalized semantic framework to support statistical disclosure control on non-numerical attributes in EHRs. Structured medical knowledge represented in SNOMED-CT was incorporated to build the semantic operators for comparison, aggregation, and sorting of these non-numerical attributes. S. Kim et al. (49) designed privacy-preserving data cubes with an anonymization process for preserving the privacy of EHR data sets. Three variations of data cubes were constructed corresponding to the *k*-anonymity and *l*-diversity rules based on global generalization, local generalization, and bucketization.

Loukides & Gkoulalas-Divanis (50) proposed a method to anonymize diagnosis codes with generalization and suppression, taking into consideration that a patient's identity could be linked with genome sequences using diagnosis codes. Hughes et al. (37) developed an online system to anonymize patient-level clinical trial data using replacement and suppression with an objective to maximize the utility for research. Heatherly et al. (51) employed a *k*-anonymization algorithm to anonymize clinical profiles for a patient hypothyroidism study at three medical centers, and utility was assessed at different population levels. The study demonstrated that anonymization on the entire EHR reduced the amount of record generalization when compared to the anonymization approach that focuses on a specific cohort, and most generalized codes (~70%) do not lose too much granularity. Poulis et al. (52) presented another approach to anonymize the demographics and diagnosis codes for protection from reidentification and to minimize information loss while ensuring the usefulness of anonymized data for targeted analyses (utility constraints were specified by data owners to limit the amount of generalization).

Dwork (53) developed an elegant framework for controlling the degree of reidentification in a manner that is independent of knowledge of potential linkable databases. Differential privacy methods ensure the specific degree of reidentification risk that is associated with the disclosure of a database or statistics. Controlled noise is added to the data or statistic to make reidentification hard, creating a trade-off between data utility and reidentification risk. Several authors have explored the use of this framework in biomedical research, including EHRs (54–56). To reduce the noise that needs to be introduced to support differential privacy, Ji et al. (57) proposed a differential privately distributed model of logistic regression integrating public and private data sets. H. Li et al. (58) presented a hybrid support vector machine based on both public and private data sets in which an RBF (radial basis function) kernel was leveraged to handle nonlinearly separable cases. Differentially private solutions by Simmons et al. (59, 60) demonstrated the practicability of returning meaningful GWAS (genome-wide association study) results while protecting privacy with appropriate privacy budgets. Their results on a rheumatoid arthritis data set outperform those from previous studies based on differential privacy (61–63) with high accuracy. Barriers for the implementation of differential privacy in health sciences include, but are not limited to, the lack of an objective approach for determining the right amount of noise to achieve an acceptable balance between privacy protection and utility of the disclosed data, as well as the lack of corresponding policies.

3.3. Security Methods for Safeguarding Medical Data Privacy in Storage, Retrieval, and Sharing

Several encryption methods have been studied to ensure data confidentiality in storage, retrieval, and sharing, including secure data exchange, searchable encryption, and attribute-based encryption. A line of research has been devoted to encryption-based secure data exchange. Thilakanathan et al. (64) proposed a secure protocol for users to control the encryption keys for their personal health data, independent of cloud services, which enables patients and health care providers to securely share sensitive medical data. Similar work was proposed by C.-L. Chen et al. (65), who utilized the features of mobile devices to support secure medical data exchange in a cloud environment. Bredfeldt et al. (66) suggested secure messaging for higher quality diabetes care to prevent the disclosure of PHI. Their proposed method includes a flagging mechanism to prevent the unintentional sharing of PHI when transferring health data for research in distributed network environments. Public key encryption-based searchable encryption is another active area of study to support secure retrieval of encrypted medical data from multiple sources. When users provide an encrypted keyword trapdoor (generated by a keyword selected by the user and her secret key), the server will return matching data in an encrypted format to ensure confidentiality during the data retrieval process. Y.-C. Chen et al. (67) identified an index for accessing EHRs by patients and clinicians while still preserving patient privacy using a keyword search over encrypted data. This method is susceptible to keyword guessing attack, in which case an attacker can learn the keyword used to generate the trapdoor. Wu et al. (68) developed a secure channel-free encryption method against keyword guessing attacks for EHRs. Similarly, Yuan et al. (69) implemented a privacy-preserving cohort discovery service for distributed clinical research networks using elliptic curve cryptography to provide a strong data privacy guarantee. The software was highly parallelized and tested over an encrypted database of 7.1 million records from the Healthcare Cost and Utilization Project, specifically including three kinds of cohorts: elderly cervical cancer patients with radical hysterectomy, oropharyngeal and tongue cancer patients with robotic transoral surgery, and female breast cancer patients with mastectomy. Another area of research is attribute-based encryption, which prevents collusion by making the decryption possible only if the attribute of the ciphertext matches that of the user key. Eom et al. (70) enhanced security for EHR systems using improved attribute-based encryption to enhance patient control over their data while protecting individual privacy. Zhang et al. (71) proposed another strategy based on anonymous attribute-based encryption to enforce the security of patient data while providing fine-grained access.

3.4. Privacy-Preserving Predictive Modeling

Federated data analysis has been widely studied for privacy-preserving predictive modeling with applications for both regression and classification. It facilitates secure collaboration among multiple organizations without revealing sensitive patient-level information. A common goal of federated data analysis is to produce the number of patients with a particular characteristic (i.e., number of patients in a cohort) across multiple health care institutions. Wyatt et al. (72) implemented the Federated Aggregate Cohort Estimator for cohort discovery through a collaboration between the University of Alabama at Birmingham, The Ohio State University, the University of Massachusetts Medical School, and the Denver Health and Hospital Authority. Such examples are also common in PCORnet (National Patient-Centered Clinical Research Network) (73) and other clinical data research networks. Requesting counts, averages, proportions, and other statistics are just some simple ways of consulting federated databases, but each already carries reidentification risks. Multivariate analyses, deep learning, and other statistical and machine learning (i.e., artificial intelligence) methods can also be executed over distributed databases and have been subjects of research and pilot implementations. Motivated by the idea of building shared models without sharing data (74) on horizontally partitioned data (i.e., data from different patients located in different databases containing the same variables), Y. Li et al. (75) introduced a dual optimization method to solve logistic regression problems with vertically partitioned data (i.e., data from a single patient hosted at different organizations such as hospitals and sequencing centers). Wu et al. (76) proposed a grid-based response with multicategory ordinal and multinomial logistic regressions on horizontally distributed data. Both approaches provide an accurate global solution based on data from distributed sources. Two distributed privacy-preserving ensemble strategies were developed based on boosting and parallel boosting algorithms for EHRs horizontally distributed across multiple agents (77). These methods can be constructed without sharing patient-level information and reduce the risk of deidentification during information exchange.

Another area of research is to keep the entire data set encrypted during analysis. Brumen et al. (78) demonstrated the feasibility of an encrypted decision tree approach using standard encryption algorithms (e.g., advanced encryption standard) and technology solutions to outsource medical data analysis to overcome legal, privacy, and confidentiality issues. Advanced predictive modeling requires homomorphic encryption, a technique that is only starting to be explored in biomedicine and is aimed at supporting direct computation over encrypted data. Liu et al. (79) leveraged a naive Bayesian classifier to develop a patient-centered privacy-preserving clinical decision support system using Paillier homomophic encryption. Rahulamathavan et al. (80) presented a protocol for clinical decision support based on a Gaussian kernel-based classification of encrypted medical data from the public database at the University of California, Irvine. Bos et al. (12) adopted homomorphic encryption for the privacy of predictive analysis tasks with encrypted clinical data including logistic regression and Cox proportional hazards. Graepel et al. (81) showed the potential of applying a machine learning model to encrypted data, and a later work by Dowlin et al. (82) demonstrated the feasibility of using artificial neural networks on homomorphically encrypted health care data. The recent iDASH (Integrating Data for Analysis, Anonymization, and Sharing) genomic data privacy competitions have also promoted the progress of encryption-based methods in protecting sensitive data to support specific data analyses (83, 84).

4. DISCUSSION AND CONCLUSION

Privacy and its preservation are complex topics that can only partially be addressed by policy and technology. It is encouraging that regulations are being revised and that public opinion is shaping the discussion. It is also increasingly apparent that data governance should include the perspective of those whose data are being shared and that patients are interested in control of their data.

Some aspects of key changes to the Common Rule reveal components that were supported or unwelcome depending on the perspective of the individual. In general, aspects of increased protections for human subjects were seen as more favorable, such as more concise clarification of informed consent (13). Some disagreed with the increased lenience of previous Common Rule aspects such as the more relaxed permissions of broad consent across research institutions, as they raise accountability concerns (17). Most of the disapproval for the revised Common Rule concerned the new requirements for unidentified partial biospecimens to be reviewed by the IRBs (13), since many noted that specimens could not ever be truly unidentifiable (17, 18). The HIPAA privacy rule cannot completely ensure patient privacy. As a result, it is necessary for researchers and policymakers to consider new approaches to this issue more proactively through technological advancements coupled with policies and to help data stewards navigate the complex human subjects data-sharing environment.

We have observed impressive progress in privacy technologies; for example, machine learning-based approaches to deidentification of clinic narratives are pushing the limits of accuracy and can be more efficient than rule-based methods devised by humans. However, machine learning methods are not perfect, and much work is still needed to increase their usefulness in practice, as even the weak HIPAA safe harbor rule has shown an equivalent of a 0.013% reidentification rate (85) with structured EHRs, which is one order of magnitude lower than state-of-the-art deidentification techniques for clinical narratives (43). We also observed that, in terms of data anonymization, many recent articles still refer to the traditional *k*-anonymization standard to sanitize data, despite its identified limitations [e.g., attribute disclosure and inference attacks (86, 87)]. The alternative approach of differential privacy (53) has a provable deidentification guarantee and has attracted some recent algorithm development (88–91); however, it has not yet been intensively used in real applications. One prominent challenge for differential privacy is to define an appropriate privacy budget and to determine how to renew it when the predefined budget is depleted. This is a policy challenge rather than a technological limitation, but decision makers need a clear understanding of the differential privacy concept. Reformulating privacy and security policies in light of the rapidly evolving technology and adoption is a complex problem.

In this article, we reviewed several security technologies in the context of data storage, retrieval, and sharing that are intended to provide data safety guarantees but may have important impact on protecting privacy, especially in commercial cloud environments. To enable privacy-protecting predictive analytics, we often need to combine deidentification, anonymization, and encryption techniques to maximize data utility and minimize risks to privacy. The choice of appropriate techniques to strike the right balance between utility and privacy requires deep understanding of the capabilities of individual methods and of the analytical task at hand.

Based on our survey of privacy concerns, technology, and current policies impacting research involving patient data, privacy concerns will continue to grow as access to patient data becomes more ubiquitous and detailed. Researchers will need to remain mindful of policy changes concerning patient data while also remaining mindful of associated emerging technologies. To answer the question, How can health care and clinical providers assure patients' trust in the privacy and security of their data?, we need continued vigilance through proactive policies and technological developments.

ACKNOWLEDGMENTS

A.M.A. was supported in part by the National Institutes of Health (NIH) grant T15LM011271. W.D. was supported in part by NIH grant U01EB023685. S.W. was supported in part by NIH awards R00HG008175, U01EB023685 and R01GM124111. X.J. was supported in part by the NIH awards R01GM118609, R01GM118574, and U01EB023685. L.O.M. was supported in part by NIH awards U54HL108460 and 0T30D025462, by Veterans Administration (VA) grant IIR 12–068, and by the Patient-Centered Outcomes Research Institute (PCORI), contract CDRN-1306–04819. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the VA.

LITERATURE CITED

1. Page SA, Manhas KP, Muruve DA 2016 A survey of patient perspectives on the research use of health information and biospecimens. BMC Med. Ethics 17(1):48 [PubMed: 27527514]

2. Menachemi N, Collum TH 2011 Benefits and drawbacks of electronic health record systems. RiskManag. Healthc. Policy 4:47–55

3. Meingast M, Roosta T, Sastry S 2006 Security and privacy issues with health care information technology. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 28th, New York, N.Y., 30Aug.–3 Sept., pp. 5453–58. New York: IEEE

4. Kim H, Bell E, Kim J, Sitapati A, Ramsdell J, et al. 2016 iCONCUR: informed consent for clinical data and bio-sample use for research. J. Am. Med. Inform. Assoc. 24(2):380–87

5. Caine K, Hanania R 2013 Patients want granular privacy control over health information in electronic medical records. J. Am. Med. Inform. Assoc. 20(1):7–15 [PubMed: 23184192]

6. Off. Hum. Res. Prot. 2017 Revised Common Rule. Regul. Guid, updated Jan. 19 https://www.hhs.gov/ohrp/regulations-and-policy/regulations/finalized-revisions-common-rule/index.html

7. Off. Civ. Rights. 2017 The Security Rule. Regul. Guid., updated May 12 https://www.hhs.gov/hipaa/for-professionals/security/index.html

8. Jiang X, Sarwate AD, Ohno-Machado L 2013 Privacy technology to support data sharing for comparative effectiveness research: a systematic review. Med. Care. 51:S58–65 [PubMed: 23774511]

9. Ohno-Machado L, Agha Z, Bell DS, Dahm L, Day ME, et al. 2014 pSCANNER: patient-centered Scalable National Network for Effectiveness Research. J. Am. Med. Inform. Assoc. 21(4):621–26 [PubMed: 24780722]

10. Gardner J, Xiong L, Xiao Y, Gao J, Post AR, et al. 2013 SHARE: system design and case studies for statistical health information release. J. Am. Med. Inform. Assoc. 20(1):109–16 [PubMed: 23059729]

11. Li H, Xiong L, Jiang X 2015 Differentially private histogram and synthetic data publication In Medical Data Privacy Handbook, ed. Gkoulalas-Divanis A, Loukides G, pp. 35–58. Cham, Switz.: Springer Int.

12. Bos JW, Lauter K, Naehrig M 2014 Private predictive analysis on encrypted medical data. J. Biomed. Inform. 50:234–43 [PubMed: 24835616]
13. Menikoff J, Kaneshiro J, Pritchard I 2017 The Common Rule, updated. N. Engl.J. Med. 376(7): 613–15 [PubMed: 28103146]
14. Wanerman RE, Armstrong MS, Davidsen BS 2017 Six key changes to the common rule. Health Care and Life Sciences Client Alert, Epstein Becker & Green, P.C http://www.ebglaw.com/content/uploads/2017/02/HCLS-Client-Alert-Six-Key-Changes-to-The-Common-Rule-13Feb17.pdf
15. Lidz CW, Appelbaum PS, Arnold R, Candilis P, Gardner W, et al. 2012 How closely do institutional review boards follow the common rule? Acad. Med. 87(7):969–74 [PubMed: 22622205]
16. Kennedy S 2015 The Common Rule (1991). IMARC Blog, 9 24 http://www.imarcresearch.com/blog/the-common-rule-1991
17. Hudson KL, Collins FS 2015 Bringing the Common Rule into the 21st century. N. Engl. J. Med. 373(24):2293–96 [PubMed: 26509903]
18. Rivera SM, Nichols L, Brako L, Croft G, Russo T, Tran T 2017 CTSA institution responses to proposed Common Rule changes: Did they get what they wanted? J. Empir. Res. Hum. Res. Ethics 12(2):79–86 [PubMed: 28421883]
19. Off. Civ. Rights. 2015 Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Regul. Guid, updated Nov. 6 https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html
20. Benitez K, Malin B 2010 Evaluating re-identification risks with respect to the HIPAA privacy rule. J. Am. Med. Inform. Assoc. 17(2):169–77 [PubMed: 20190059]
21. Liu X, Li X-B, Motiwalla L, Li W, Zheng H, Franklin PD 2016 Preserving patient privacy when sharing same-disease data. ACMJ. Data Inf. Qual. 7(4):17
22. Rodrigues JJPC, de la Torre I, Fernandez G, Lopez-Coronado M 2013 Analysis of the security and privacy requirements of cloud-based electronic health records systems. J. Med. Internet Res. 15(8):e186 [PubMed: 23965254]
23. Xia W, Heatherly R, Ding X, Li J, Malin BA 2015 R-U policy frontiers for health data de-identification. J. Am. Med. Inform. Assoc. 22(5):1029–41 [PubMed: 25911674]
24. O'Keefe CM, Rubin DB 2015 Individual privacy versus public good: protecting confidentiality in health research. Stat. Med. 34(23):3081–103 [PubMed: 26045214]
25. Glenn T, Monteith S 2014 Privacy in the digital world: medical and health data outside of HIPAA protections. Curr. Psychiatry Rep. 16(11):494 [PubMed: 25218603]
26. Shenoy A, Appel JM 2017 Safeguarding confidentiality in electronic health records. Camb. Q. Healthc. Ethics 26(2):337–41 [PubMed: 28361730]
27. DeAngles M 2015 National electronic health record network regulation and synchronization of national and state privacy laws needed to increase efficiency and reduce costs in healthcare. J. Leg. Med. 36(3–4):413–19 [PubMed: 28256941]
28. Mamo LA, Browe DK, Logan HC, Kim KK 2013 Patient informed governance of distributed research networks: results and discussion from six patient focus groups. AMIA Annu. Symp. Proc. 2013:920–29 [PubMed: 24551383]
29. Holmes JH 2016 Privacy, security, and patient engagement: the changing health data governance landscape. eGEMs 4(2):1261 [PubMed: 27141525]
30. Luchenski S, Balasanthiran A, Marston C, Sasaki K, Majeed A, et al. 2012 Survey of patient and public perceptions of electronic health records for healthcare, policy and research: study protocol. BMC Med. Inform. Decis. Mak. 12:40 [PubMed: 22621621]
31. Trachtenbarg DE, Asche C, Ramsahai S, Duling J, Ren J 2017 The benefits, risks and costs of privacy: patient preferences and willingness to pay. Curr. Med. Res. Opin. 33(5):845–51 [PubMed: 28166481]
32. Kim KK, Joseph JG, Ohno-Machado L 2015 Comparison of consumers' views on electronic data sharing for healthcare and research. J. Am. Med. Inform. Assoc. 22(4):821–30 [PubMed: 25829461]

33. Bull S, Roberts N, Parker M 2015 Views of ethical best practices in sharing individual-level data from medical and public health research: a systematic scoping review. J. Empir. Res. Hum. Res. Ethics 10(3):225–38 [PubMed: 26297745]

34. Kim KK, Sankar P, Wilson MD, Haynes SC 2017 Factors affecting willingness to share electronic health data among California consumers. BMC Med. Ethics 18(1):25 [PubMed: 28376801]

35. Mann SP, Savulescu J, Sahakian BJ 2016 Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue. Philos. Trans. R. Soc. A 374(2083):20160130

36. Mascalzoni D, Paradiso A, Hansson M 2014 Rare disease research: breaking the privacy barrier. Appl. Transl. Genom. 3(2):23–29 [PubMed: 27275410]

37. Hughes S, Wells K, McSorley P, Freeman A 2014 Preparing individual patient data from clinical trials for sharing: the GlaxoSmithKline approach. Pharm. Stat. 13(3):179–83 [PubMed: 24668938]

38. Samuels JG, McGrath RJ, Fetzer SJ, Mittal P, Bourgoine D 2015 Using the electronic health record in nursing research: challenges and opportunities. West. J. Nurs. Res. 37(10):1284–94 [PubMed: 25819698]

39. Hanauer D, Aberdeen J, Bayer S, Wellner B, Clark C, et al. 2013 Bootstrapping a de-identification system for narrative patient records: cost-performance tradeoffs. Int.J. Med. Inform. 82(9):821–31 [PubMed: 23643147]

40. Meystre SM, Ferrandez O, Friedlin FJ, South BR, Shen S, Samore MH 2014 Text de-identification for privacy protection: a study ofits impact on clinical text information content. J. Biomed. Inform. 50:142–50 [PubMed: 24502938]

41. Gardner J, Xiong L 2009 An integrated framework for de-identifying unstructured medical data. Data Knowl. Eng. 68(12):1441–51

42. Ferrandez O, South BR, Shen S, Friedlin FJ, Samore MH, Meystre SM 2012 Generalizability and comparison of automatic clinical text de-identification methods and resources. AMIA Annu. Symp. Proc. 2012:199–208 [PubMed: 23304289]

43. Dernoncourt F, Lee JY, Uzuner O, Szolovits P 2017 De-identification of patient notes with recurrent neural networks. J. Am. Med. Inform. Assoc. 24(3):596–606 [PubMed: 28040687]

44. Sweeney L 2002 k-anonymity: a model for protecting privacy. Internat. J. Uncertain. Fuzziness Knowl.-Based Syst. 10(05):557–70

45. Aristodimou A, Antoniades A, Pattichis CS 2016 Privacy preserving data publishing of categorical data through k-anonymity and feature selection. Healthc. Technol Lett. 3(1):16–21 [PubMed: 27222728]

46. Yoo S, Shin M, Lee D 2012 An approach to reducing information loss and achieving diversity of sensitive attributes in k-anonymity methods. Interact. J. Med. Res. 1(2):e14 [PubMed: 23612074]

47. Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B 2012 Anonymization of longitudinal electronic medical records. IEEE Trans. Inf. Technol. Biomed. 16(3):413–23 [PubMed: 22287248]

48. Martinez S, Sanchez D, Valls A 2013 A semantic framework to protect the privacy of electronic health records with non-numerical attributes. J. Biomed. Inform. 46(2):294–303 [PubMed: 23228807]

49. Kim S, Lee H, Chung YD 2017 Privacy-preserving data cube for electronic medical records: an experimental evaluation. Int. J. Med. Inform. 97:33–42 [PubMed: 27919391]

50. Loukides G, Gkoulalas-Divanis A 2013 Utility-aware anonymization of diagnosis codes. IEEEJ. Biomed. Health Inform. 17(1):60–70

51. Heatherly R, Rasmussen LV, Peissig PL, Pacheco JA, Harris P, et al. 2016 A multi-institution evaluation of clinical profile anonymization. J. Am. Med. Inform. Assoc. 23:e131–37 [PubMed: 26567325]

52. Poulis G, Loukides G, Skiadopoulos S, Gkoulalas-Divanis A 2017 Anonymizing datasets with demographics and diagnosis codes in the presence of utility constraints. J. Biomed. Inform. 65:76–96 [PubMed: 27832965]

53. Dwork C. Bugliesi M. Preneel B. Sassone V. Wegener I. Differential privacy; Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006; Venice, Italy. July 10–14, 2006; Berlin: Springer-Verlag; 2006. 1–12. Proceedings, Part II

54. Vinterbo SA, Sarwate AD, Boxwala AA 2012 Protecting count queries in study design. J. Am. Med. Inform. Assoc. 19(5):750–57 [PubMed: 22511018]

55. Gkoulalas-Divanis A, Loukides G, Sun J 2014 Publishing data from electronic health records while preserving privacy: a survey ofalgorithms. J. Biomed. Inform. 50:4–19 [PubMed: 24936746]

56. Dankar FK, El Emam K 2012 The application of differential privacy to health data. Proc. 2012 Jt. EDBT/ICDT Workshops Berlin, Ger., 30Mar., pp. 158–66. New York: Assoc. Comput. Mach.

57. Ji Z, Jiang X, Wang S, Xiong L, Ohno-Machado L 2014 Differentially private distributed logistic regression using private and public data. BMC Med. Genom. 7:S14

58. Li H, Xiong L, Ohno-Machado L, Jiang X 2014 Privacy preserving RBF kernel support vector machine. Biomed. Res. Int. 2014:827371

59. Simmons S, Sahinalp C, Berger B 2016 Enabling privacy-preserving GWASs in heterogeneous human populations. Cell Syst. 3(1):54–61 [PubMed: 27453444]

60. Simmons S, Berger B 2016 Realizing privacy preserving genome-wide association studies. Bioinformatics 32(9):1293–300 [PubMed: 26769317]

61. Johnson A, Shmatikov V 2013 Privacy-preserving data exploration in genome-wide association studies. Proc. Int. Conf. Knowl. Discov. Data Min., 19th, Chicago, 1ll., 11–14 Aug., ed. Ghani R, Senator TE, Bradley P, Parek R, He J, pp. 1079–87. New York: Assoc. Comput. Mach.

62. Yu F, Fienberg SE, Slavkovic AB, Uhler C 2014 Scalable privacy-preserving data sharing methodology for genome-wide association studies. J. Biomed. Inform. 50:133–41 [PubMed: 24509073]

63. Yu F, Ji Z 2014 Scalable privacy-preserving data sharing methodology for genome-wide association studies: an application to iDASH healthcare privacy protection challenge. BMC Med. Inform. Decis. Mak. 14:S3 [PubMed: 25521367]

64. Thilakanathan D, Calvo RA, Chen S, Nepal S, Glozier N 2016 Facilitating secure sharing of personal health data in the cloud. JMIR Med. Inform. 4(2):e15 [PubMed: 27234691]

65. Chen C-L, Yang T-T, Shih T-F 2014 A secure medical data exchange protocol based on cloud environment. J. Med. Syst. 38(9):112 [PubMed: 25037716]

66. Bredfeldt CE, Compton-Phillips AL, Snyder MH 2011 Effects ofbetween visit physician-patient communication on Diabetes Recognition Program scores. Int.J. Qual. Health Care 23(6):664–73 [PubMed: 21937586]

67. Chen Y-C, Horng G, Lin Y-J, Chen K-C2013 Privacy preserving index for encrypted electronic medical records. J. Med. Syst. 37(6):9992 [PubMed: 24158427]

68. Wu Y, Lu X, Su J, Chen P 2016 An efficient searchable encryption against keyword guessing attacks for sharable electronic medical records in cloud-based system. J. Med. Syst. 40(12):1–9 [PubMed: 26573639]

69. Yuan J,Malin B,Modave F, Guo Y, Hogan WR, et al.2017Towards a privacy preserving cohort discovery framework for clinical research networks. J. Biomed. Inform. 66:42–51 [PubMed: 28007583]

70. Eom J, Lee DH, Lee K 2016 Patient-controlled attribute-based encryption for secure electronic health records system. J. Med. Syst. 40(12):253 [PubMed: 27714562]

71. Zhang L, Wu Q, Mu Y, Zhang J 2016 Privacy-preserving and secure sharing of PHR in the cloud. J. Med. Syst. 40(12):267 [PubMed: 27730393]

72. Wyatt MC, Hendrickson RC, Ames M, Bondy J, Ranauro P, et al. 2014 Federated Aggregate Cohort Estimator (FACE): an easy to deploy, vendor neutral, multi-institutional cohort query architecture. J. Biomed. Inform. 52:65–71 [PubMed: 24316052]

73. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS 2014 Launching PCORnet, a national patient-centered clinical research network. J. Am. Med. Inform. Assoc. 21(4):578–82 [PubMed: 24821743]

74. Wu Y, Jiang X, Kim J, Ohno-Machado L 2012 Grid binary logistic regression (GLORE): building shared models without sharing data. J. Am. Med. Inform. Assoc. 19(5):758–64 [PubMed: 22511014]

75. Li Y, Jiang X, Wang S, Xiong H, Ohno-Machado L2015 Vertical grid logistic regression (VERTIGO). J. Am. Med. Inform. Assoc. 23(3):570–79 [PubMed: 26554428]

76. Wu Y, Jiang X, Wang S, Jiang W, Li P, Ohno-Machado L 2015 Grid multi-category response logistic models. BMC Med. Inform. Decis. Mak. 15:758–64

77. Li Y, Bai C, Reddy CK 2016 A distributed ensemble approach for mining healthcare data under privacy constraints. Inf. Sci. 330:245–59

78. Brumen B, Hericko M, Sevcnikar A, Zavrsnik J, Holbl M 2013 Outsourcing medical data analyses: Can technology overcome legal, privacy, and confidentiality issues? J. Med. Internet Res. 15(12):e283 [PubMed: 24342053]

79. Liu X, Lu R, Ma J, Chen L, Qin B 2016 Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification. IEEEJ. Biomed. Health Inform. 20(2):655–68

80. Rahulamathavan Y, Veluru S, Phan RC-W, Chambers JA, Rajarajan M 2014 Privacy-preserving clinical decision support system using Gaussian kernel-based classification. IEEE J. Biomed. Health Inform. 18(1):56–66 [PubMed: 24403404]

81. Graepel T, Lauter K, Naehrig M 2012 ML Confidential: machine learning on encrypted data In Information Security and Cryptology, ed. Kwon T, Lee MK, Kwon D, pp. 1–21. Berlin: Springer-Verlag

82. Dowlin N, Gilad-Bachrach R, Laine K, Lauter K, Naehrig M, Wernsing J 2016 CryptoNets: applying neural networks to encrypted data with high throughput and accuracy. J. Mach. Learn. Res. 48:201–10

83. Jiang X, Ohno-Machado L, Malin B, Tang H, Wang S, et al. 2014 A community assessment of data perturbation techniques on privacy protection for human genome data. BMC Med. Inform. Decis. Mak. 14(1):S1 [PubMed: 25521230]

84. Tang H, Jiang X, Wang X, Wang S, Sofia H, et al. 2016 Protecting genomic data analytics in the cloud: state of the art and opportunities. BMC Med. Genom. 9(1):63

85. El Emam K, Jonker E, Arbuckle L, Malin B 2011 A systematic review of re-identification attacks on health data. PLOS ONE 6(12):e28071 [PubMed: 22164229]

86. Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M 2006 l-diversity: privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data 1(1):3

87. Li N, Li T, Venkatasubramanian S 2007 t-closeness: privacy beyond k-anonymity and l-diversity. Proc. IEEE Int. Conf Data Eng., 23rd, Istanb., Turk., 15–20Apr., pp. 106–15. New York: IEEE

88. Li H, Xiong L, Ji Z, Jiang X 2017 Partitioning-based mechanisms under personalized differential privacy. Proc. Adv. Know/edge Discov. Data Mining, 21st, Jeju, S. Korea, 23–26 May, pp. 615–27. Cham, Switz.: Springer Int.

89. Xu S, Su S, Xiong L, Cheng X, Xiao K 2016 Differentially private frequent subgraph mining. Proc. Int. Conf. Data Eng., 32nd, Hels., Finl., 16–20 May, pp. 229–40. New York: IEEE

90. Li H, Xiong L, Jiang X, Liu J 2015 Differentially private histogram publication for dynamic datasets: an adaptive sampling approach. Proc. ACM Int. Conf. Inf. Knowl. Manag., 24th, Melb., Aust., 18–23 Oct., pp. 1001–10. New York: Assoc. Comput. Mach.

91. Mohammed N, Chen R, Fung BC, Yu PS, Philip SY 2011 Differentially private data release for data mining. Proc. Int. Conf. Know/. Discov. Data Mining, 17th, San Diego, Calif., 21–24 Aug., pp. 493–501. New York: Assoc. Comput. Mach.

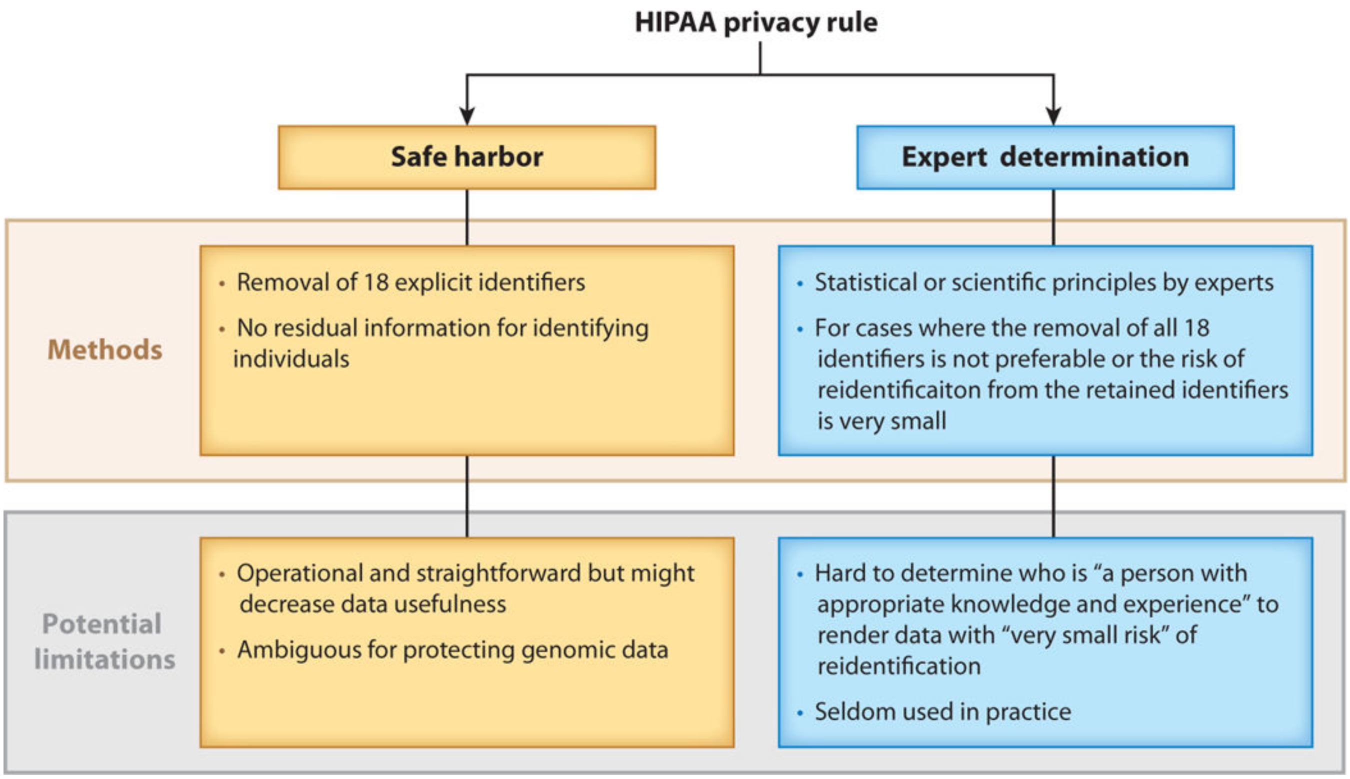
HIPAA privacy rule
Safe harbor
Expert determination
Methods
• Removal of 18 explicit identifiers
• No residual information for identifying individuals
• Statistical or scientific principles by experts
• For cases where the removal of all 18 identifiers is not preferable or the risk of reidentificaiton from the retained identifiers is very small
Potential limitations
• Operational and straightforward but might decrease data usefulness
• Ambiguous for protecting genomic data
• Hard to determine who is "a person with appropriate knowledge and experience" to render data with "very small risk" of reidentification
• Seldom used in practice

Figure 1.

Summary and limitations of two deidentification mechanisms, safe harbor and expert determination, under the HIPAA (Health Insurance Portability and Accountability Act) privacy rule.

Table 1

Comparison of key changes in the updated Common Rule

Topics	Previous Common Rule features	Revised Common Rule features
Informed consent requirements	There are transparency requirements to improve subject understanding of the consent details (16). A single posting on a site location is required (16).	There must be a clear and concise description of the research (13, 14). A consent form for clinical trials must be available online for 60 days (14).
Consent for researching individual data on biospecimens or other identifiable data	Only in “very rare cases” can researchers work with biospecimens without consent (16).	Research with broad consent is permitted for both current and future research with the data/specimens (13, 14).
Exempt research criteria	There are six categories of exemptions, including research using educational methods, educational exams, and “benign interventions” with adults (14, 17).	There are eight categories of exemptions, including stricter requirements for protecting educational class time or performance, protections from identification, protection from risk of individual harm, or further IRB approval (14). The two new categories of exceptions involve the use of publicly available data for secondary use and the storage and maintenance of publicly available data (14).
IRB authorization for interorganizational research	Each organization involved in collaborative research must submit separate IRB applications (14).	A single authorization must be requested for interorganizational cooperative research; the compliance requirement is delayed until Jan. 20, 2020(14, 18).
IRB approval criteria	There are requirements for minimal risk to participants, analysis of benefit-risk factors, and equity in selecting research subjects (14).	The definition of “vulnerable populations” is changed to include the socioeconomically disadvantaged, children, prisoners, and subjects with limited cognitive ability to make decisions (14).
Details on the ongoing approval of IRB	There is continued IRB review of studies involving minimal risk (13, 14).	Criteria not requiring continued IRB review are expanded. There is continued review for cases where exemptions are based on IRB limited review (13, 14).
Human subjects term expansion for biospecimens that cannot be identified	Biospecimens with nonidentifiable features do not require IRB approval (18).	Expanded protections require IRB approval for the use of deidentified partial biospecimens (13, 18).
Adverse outcomes reporting	There are exemptions for research involving “benign interventions” (17).	Researchers are required to report adverse outcomes on a federal website (18).

Abbreviation: IRB, institutional review board.