REPORT 26 OF THE BOARD OF TRUSTEES (A-19)
Research Handling of De-Identified Patient Information
(Reference Committee on Amendments to Constitution and Bylaws)

EXECUTIVE SUMMARY

At the 2018 Annual Meeting, Policy D-315.975, "Research Handling of De-Identified Patient Information," was adopted by the House of Delegates. This policy directs the American Medical Association (AMA) to study the handling of de-identified patient data and report the findings and recommendations to the House of Delegates at the 2019 Annual Meeting. This report outlines appropriate and inappropriate use of de-identified patient data, perspectives from stakeholders in organized medicine, potential ethical concerns of the commercial use of such data, regulatory implications, and recommendations for the future use of de-identified patient data

Protected health information (PHI) includes many common identifiers (e.g., name, address, birth date, Social Security Number) when they can be associated with patient health information. The HIPAA Privacy Rule sets limits and conditions on the uses and disclosures that may be made of such information without patient authorization. Security of PHI safeguards patients from the risk of their data being released or used in manners that could result in discrimination, stigmatization, or embarrassment. However, the use, sale, or distribution of de-identified patient data is not prohibited under HIPAA, since once PHI is de-identified in accordance with the HIPAA Privacy Rule, it is no longer considered PHI and, thus, may be used and disclosed by a covered entity or health information organization (HIO) for any purpose.

REPORT OF THE BOARD OF TRUSTEES

B of T Report 26-A-19

Subject:	Research Handling of De-Identified Patient Information
Presented by:	Jack Resneck, Jr., MD, Chair
Referred to:	Reference Committee on Amendments to Constitution and Bylaws (William Reha, MD, Chair)

INTRODUCTION

At the 2018 Annual Meeting, Policy D-315.975, "Research Handling of De-Identified Patient Information," was adopted by the House of Delegates. This policy directs the American Medical Association (AMA) to study the handling of de-identified patient data and report the findings and recommendations to the House of Delegates at the 2019 Annual Meeting. This report outlines appropriate and inappropriate use of de-identified patient data, perspectives from stakeholders in organized medicine, potential ethical concerns of the commercial use of such data, regulatory implications, and recommendations for the future use of de-identified patient data.

BACKGROUND

Health-related information collected during the course of clinical care has always been of great interest for a number of secondary use cases, including scientific research in the academic and commercial settings, marketing for pharmaceutical and medical device companies, and a wide variety of other uses. More recently, a new and substantial interest has been raised from technology companies who seek to use patient data to build new clinical tools using machine learning and "big data." Clinical data is the topic of significant ethical guidance and regulation at both the state and federal levels, focused primarily on the appropriate use and handling of identifiable patient information. Little guidance exists, however, on the use of de-identified patient data.

A variety of entities, including provider organizations, clinical laboratories, and commercial entities such as personal genomics companies, may collect patient data intended for clinical use or to deliver genetics information, and then resell de-identified data to other entities for other purposes. For example, 23andMe, a personal genomics and biotech service, sells de-identified user data to pharmaceutical companies that use it to conduct research on various diseases. Concerns arise in that when the data is de-identified, it is no longer considered PHI and therefore patient authorization or consent for use is not required and therefore not solicited—meaning that patients are not always aware how their data is being used.[1] For example, research using de-identified data such as biologic specimens may result in scientific knowledge that has commercial value. Proper consent for use and/or disclosure of commercial interest in this research is ideal but not always documented, sometimes resulting in legal action against physicians or researchers.[2]

In addition, there is a perceived lack of transparency and regulation in how patients' data is being sold, distributed, or used outside of their direct health care. Risk of re-identification, which some studies have demonstrated to be possible through matching data to other publicly available data sources, is another issue related to the use of de-identified data. There are also concerns about

access to such information that is sought for marketing purposes on behalf of commercial entities that have financial interests in physicians' treatment and/or prescribing behavior. In addition, the sale of de-identified data by clinicians and provider organizations may create a real or perceived conflict of interest, which could lead to a loss of patient confidence.

What is Protected Health Information

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) provides extensive protections for patient data that is considered protected health information (PHI).[3] PHI is information, including demographic information, which relates to an individual's past, present, or future physical or mental health or condition; the provision of health care to the individual; or the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual.[4] PHI includes many common identifiers (e.g., name, address, birth date, Social Security Number) when they can be associated with the health information listed above. The HIPAA Privacy Rule sets limits and conditions on the uses and disclosures that may be made of such information without patient authorization.[5] Security of PHI safeguards patients from the risk of their data being released or used in manners that could result in discrimination, stigmatization, or embarrassment.[6, 7] Section 164.514(a) of the HIPAA Privacy Rule establishes standards for de-identifying PHI so individuals can no longer be identified by any portion of the data. The use, sale, or distribution of de-identified patient data is not prohibited under HIPAA, since once PHI is de-identified in accordance with the HIPAA Privacy Rule, it is no longer considered PHI and, thus, may be used and disclosed by a covered entity or health information organization (HIO) for any purpose.[8]

In addition to regulation at the federal level, state lawmakers have exhibited a general trend toward establishing stricter guards on the use of patient data and the requirement for patient consent, some of which reflect standards set forth in the European Union's recent General Data Protection Regulation (GDPR).[9] Some states are considering and passing laws to protect consumer privacy as it relates to the use of their personal information. For example, California in June 2018 passed the California Consumer Privacy Act of 2018 (effective January 1, 2020), which protects consumers' right to: (1) know what personal information a for-profit business has collected about them, where it was sourced from, what it is being used for, whether it is being disclosed or sold, and to whom it is being disclosed or sold; (2) "opt out" of allowing a business to sell their personal information to third parties; (3) have a business delete their personal information, with some exceptions; and (4) receive equal service and pricing from a business, even if they exercise their privacy rights under the Act.[10] California's law does not apply to information covered by HIPAA, de-identified personal data, or aggregate consumer data, however, as long as the de-identification measures meet the Act's strict standards.[11]

What is de-identified patient data?

De-identified patient data is information about a patient or user of a health-related service that has been stripped of individually identifiable health information. Removing identifiers from PHI mitigates privacy risks to individuals and thereby supports the secondary use of data for comparative effectiveness studies, policy assessment, life sciences research, and other endeavors.[4] Information can be de-identified by either of two means: (1) a formal determination by a qualified expert (expert determination); or (2) the removal of specified individual identifiers and an absence of actual knowledge by the covered entity that residual information could be used to identify the individual (safe harbor).
The identifiers removed from PHI in the safe harbor method include:[4]

- Names
- All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:
    - The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and
    - The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000
- All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
- Telephone numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Fax numbers
- Device identifiers and serial numbers
- Email addresses
- Web URLs
- Social security numbers
- Internet Protocol addresses
- Medical record numbers
- Biometric identifiers, including finger and voice prints
- Health plan beneficiary numbers
- Full-face photographs and any comparable images
- Account numbers
- Any other unique identifying number, characteristic, or code, except as permitted
- Certificate/license numbers

How is de-identified data used?

De-identified data is used for research to derive information and knowledge about treatment and outcomes, as well as other patient care-related purposes. Outside of health care organizations and researchers, de-identified patient data is used by a variety of organizations and industries for various purposes, including many not related to patient care. De-identified data is sourced, collected, and used by a variety of organizations, including health care provider organizations such as hospitals or academic medical centers, and commercial enterprises such as personal genomics and biotechnology companies. Pharmaceutical manufacturers and retail pharmacies may also find use in de-identified health data to target their advertising. Health care providers use this data typically in research or the direct care of patient populations. The data can also be used to help reduce costs of care, improve treatment options, and support public health initiatives.

Machine learning is a family of methods used by some health care and data solution organizations to help predict certain outcomes and better prepare for and treat patients identified to be at risk. Machine learning models establish predictive rules using vast amounts of computing power. The more data a machine learning model has, the more complex the rules and the more accurate the predictions.[12] However, machine learning models are vulnerable to biases induced by data that does not adequately represent the patient population, such as data collected from only one institution or one geographic region. In order to develop clinical decision support tools that can be effectively used to treat the diverse patient populations in the United States, large amounts of data are

required, and often data from many different providers across the country are required to avoid bias. This data is often sourced from de-identified or anonymized patient records. Allscripts, for example, used 50 million de-identified patient records, and the application of an advanced machine learning algorithm, to "train" its systems and further improve its clinical decision support tools.[13] Organizations like Orion Health and Precision Driven Health are using datasets like these to generate machine learning aimed at improving health care decisions, and driving operational and cost efficiencies.[12, 14] By combining multiple datasets, such as behavioral data, device use data, patient claim data and socioeconomic and geographic data, these organizations are developing advanced predictive analytics to further improve precision health care.[14] The data used for the purposes of data mining and honing machine learning algorithms are either sourced and used at the organizational level, or de-identified or anonymized when used for external research, such as the analysis done by Allscripts. Data may be sourced via publicly available de-identified datasets, databases established through collaborative research agreements, or via the purchase of bulk de-identified data, on an exclusive or non-exclusive basis. Since this technology is relatively new in the health care space its implications for patient data are not well-studied. As artificial intelligence and advanced machine learning proliferate in the health care space, the value and number of potential uses of patient health data will inevitably increase. Stakeholders should be prepared for increasing concerns about related patient privacy and data security.

Commercial entities, such as personal genomics companies, may collect data to deliver genetics information to subscribers and then subsequently sell the de-identified data to another entity for another purpose. For example, 23andMe, a genomics and biotech service, sells de-identified user data to pharmaceutical companies that use it to conduct research on various diseases. Concerns arise in that when the data is de-identified, it is no longer considered PHI and therefore patient authorization or consent for use is not required and therefore not solicited—meaning that patients are not always aware how their data is being used.[1] For example, research using de-identified data such as biologic specimens may result in scientific knowledge that has commercial value. Proper consent for use and/or disclosure of commercial interest in this research is ideal but not always documented, sometimes resulting in legal action against physicians or researchers.[2]

In addition, there is a perceived lack of transparency and regulation in how patients' data is being sold, distributed, or used outside of their direct health care. Risk of re-identification, which some studies have demonstrated to be possible through matching data to other publicly available data sources, is another issue related to the use of de-identified data. There are also concerns about access to such information that is sought for marketing purposes on behalf of commercial entities that have financial interests in physicians' treatment and/or prescribing behavior.

AMA POLICY

The AMA has multiple policies expressing its recognition of the importance of data privacy and protection of PHI, as well as policies expressing commitment to ensuring safe and appropriate use of de-identified data.

Board of Trustees Report 21-A-18, "Ownership of Patient Data," outlines federal and state laws that establish who owns a patient's medical records. The report also highlights the importance of ensuring patients have appropriate access to their data and physicians have the tools and controls they need to be good stewards of their patients' information while at the same time maintaining the ability to share information to seamlessly coordinate the best care. In support of these initiatives, the AMA has actively engaged with the U.S. Department of Health and Human Services (HHS), the Office of Inspector General, the Office of Civil Rights, and the Office of the National

Coordinator for Health Information Technology (ONC), and has broad policy in place covering all aspects of patient record maintenance, access and control.

AMA Policy H-315.978, "Privacy and Confidentiality," states that where possible, informed consent should be obtained before personally identifiable health information is used for any purpose. However, in those situations where specific informed consent is not practical or possible, either (1) the information should have identifying information stripped from it or (2) an objective, publicly accountable entity must determine that patient consent is not required after weighing the risks and benefits of the proposed use. Re-identification of personal health information should only occur with patient consent or with the approval of an objective, publicly accountable entity.

AMA Policy H-315.974, "Guiding Principles, Collection and Warehousing of Electronic Medical Record Information," expresses the AMA's commitment to advocating that physicians, as trusted stewards of PHI, should be the owners of all patient claims data and de-identified aggregate data that is established and maintained by the physician practice, specifically including data stored in the electronic health record or practice management system. The policy establishes principles around the use of these data that include compliance with HIPAA, requires physician consent for analysis of the data, and requires data to remain accessible to authorized users for purposes of treatment, public health, patient safety, quality improvement, medical liability defense, and research.

AMA Policy H-315.983, "Patient Privacy and Confidentiality," states that whenever possible, medical records should be de-identified for purposes of use for utilization review, panel credentialing, quality assurance, and peer review. This policy also states our AMA will guard against the imposition of unduly restrictive barriers to patient records that would impede or prevent access to data needed for medical or public health research or quality improvement and accreditation activities, and that whenever possible, de-identified data should be used for these purposes. Policy H-315-983 posits that in the event of a sale or discontinuation of a medical practice, only de-identified and/or aggregate data should be used for "business decisions," including sales, mergers, and similar business transactions when ownership or control of medical records changes hands. This policy includes extensive language emphasizing the AMA's commitment to protecting PHI, and that it will continue its advocacy for privacy and confidentiality regulations, including: (a) The establishment of rules allocating liability for disclosure of identifiable patient medical information between physicians and the health plans of which they are a part, and securing appropriate physician control over the disposition of information from their patients' medical records; (b) The establishment of rules to prevent disclosure of identifiable patient medical information for commercial and marketing purposes; and (c) The establishment of penalties for negligent or deliberate breach of confidentiality or violation of patient privacy rights.

In Policy H-315.975, "Police, Payer, and Government Access to Patient Health Information," the AMA commits to advocating for narrow and clearly defined bounds for the appropriate use of patient information by law enforcement, payers and government entities, for operations that cannot be reasonably undertaken with de-identified data. AMA Policy H-315.987, "Limiting Access to Medical Records," further defines who should and should not have access to this information.

The AMA's Code of Medical Ethics includes an opinion on "Access to Medical Records by Data Collection Companies." Opinion E-3.2.4 asserts that disclosing information to third parties for commercial purposes without consent undermines trust, violates principles of informed consent and confidentiality, and may harm the integrity of the patient-physician relationship. The opinion further expresses that physicians who wish to permit third-party access to *specific patient information* for commercial purposes should: (a) only provide data that has been de-identified, and

(b) fully inform each patient whose record would be involved about the purpose(s) for which access would be granted. This opinion, with respect to requests for permission to allow access to or disclose a *full medical record*, prohibits disclosing identifiable information for commercial purposes *without obtaining consent* from the patient to do so.

The authors of Resolution 3-A-18, which established policy D-315.975 and is the subject of this report, expressed particular concern that this Code of Medical Ethics Opinion may contradict itself in its emphasis on informing the patient of how their de-identified data will be used and the subsequent emphasis on the importance of obtaining consent. The key difference between the two elements of the opinion lies in the description of the patient information being requested (specific, de-identified patient information vs. full medical record), thus our AMA does not agree that these statements are contradictory.

The authors also expressed that this Opinion may be in disharmony with the rules set forth in the HIPAA Privacy Rule, specifically stating that authorization, rather than consent, is sometimes mandated for the release of PHI when being requested for purposes not related to treatment, payment, or health care operations (TPO). HIPAA defines three such uses or disclosures for which written authorization of the patient is required: (1) use and disclosure of psychotherapy notes; (2) use and disclosure of PHI for marketing; and (3) any sale of PHI.

Ethical Opinion E-3.2.4 was originally issued in 1994 and updated in 1998, prior to the enactment of the HIPAA Privacy Rule, yet provides an even higher standard than the Rule with respect to requirements for consent to disclose patient data, including data that has been de-identified. With respect to authorization requirements, Opinion E-3.2.4 does not include a statement about when authorization, rather than consent, is appropriate and/or required. Guidance provided in the Code of Ethics is provided by standards of conduct that define the essentials of honorable behavior for the physician. They cover broad ethical principles and are not intended to align with law or specific regulations that may be legally enforceable. During a comprehensive eight-year modernization process that ended in 2017, the AMA *Code of Medical Ethics* was reviewed for relevance/timeliness of guidance, clarity, and consistency of guidance. Opinion E-3.2.4 was reorganized in this process, taking the HIPAA provisions into consideration during the process. Care was taken to ensure the Council on Ethical and Judicial Affairs was conservative in suggesting substantive change, doing so only where needed to ensure that guidance remains relevant in the face of changes in biomedical science and conditions of medical practice. No contradictions or points of discord with HIPAA were identified in that review.

DISCUSSION

Oversight of patient information

The use of de-identified patient data is not heavily regulated. The HIPAA Privacy Rule does not restrict the use or disclosure of de-identified health information, since it is not considered PHI.[2, 5] HIPAA permits secondary uses of de-identified data for purposes such as public health initiatives, research, law enforcement, and other public interest endeavors.[5, 15] In addition, commercial entities that sell or use de-identified data, such as biotech and pharmaceutical companies, are not considered covered entities under HIPAA. Through their interactions with pharmacy benefit managers, pharmacies, payers, physicians and patients, however, they are indirectly impacted by privacy rules and must structure their transactions, projects, and internal data programs such that their partners that are covered entities or business associates thereof meet data privacy requirements under HIPAA and any other applicable standards.

Studies that use de-identified data are exempt from regulations that govern human subject research.[2, 16] Entities that collect and use consumer data, such as pharmaceutical companies or academic institutions conducting research, should employ privacy protections into their practices, such as data security, reasonable collection limits, sound retention and disposal practices, and data accuracy to protect privacy, as guided in recommendations from the Federal Trade Commission (FTC).[17] For example, Harvard University, like many academic institutions receiving federal grants, implements strict policy to govern the collection, storage and use of research data, including PHI.[18] In addition to the enforcement of strict policy, all human subjects research is subject to approval by the institution's Institutional Review Board (IRB). It is the responsibility of IRBs to specify the security level for research projects they review and approve, obtain confirmation that the relevant security controls are being implemented and decide if the human subject must give consent or in the case of de-identified information, approve the research under an exempt status from obtaining the consent.

Human subject research conducted or supported by certain federal departments or agencies is governed by the Federal Policy for the Protection of Human Subjects ("Common Rule"). Revisions to the Common Rule in 2017 were adopted in response to shifts in science, technology, public engagement, and public expectations that have raised concerns about the limitations of the existing ethical framework in research.[19] The rapid pace of change in the availability, utility, and value of patient data, including PHI and de-identified data, will continue to necessitate regular reconsideration of the ethical oversight of patient data and how it is protected by researchers and other entities.

Risks and ethical concerns

There are ethical concerns about the disclosure and use of de-identified health data that are rooted in the risk of re-identification. Studies have shown that certain elements of patient records, although not exclusive or unique to individual patients, increase the risk of re-identification if not removed from individual-level data.[20, 21] Elements such as gender, date of service, date of birth or zip code can potentially be linked back to other sources of data, such as voter registration lists, and could put the data at risk of re-identification.[21, 22] Organizations that collect, store, transfer and distribute de-identified data should take steps to reduce this risk, such as replacing a specific date of birth or date of service with a year.

Studies have been undertaken to assess the risk of re-identification after steps have been taken to de-identify the data, and have found gaps that can put de-identified patient health data at risk of being re-identified.[20, 23, 24] While these findings are significant and should not be ignored, one review of some of these studies concluded that many of them were small and did not use data that was de-identified according to existing standards (those set forth in the HIPAA Privacy Rule), so caution should be taken when making generalizations based on the few cases identified in the studies.[25]

In addition to risk of re-identification, there are general ethical concerns with the availability and use of patient health data, even if it's de-identified, without explicit authorization from patients. For example, pharmaceutical companies may use de-identified data to target marketing or advertising efforts to specific physicians, therefore influencing treatment plans for patient populations with specific diseases or conditions. Accountable Care Organizations (ACOs), as business associates of the ACO participants or a covered entity, may use de-identified data to analyze quality measures, population risk scores and patient behaviors.[26] Other for-profit entities may use de-identified data for the development of new technology or clinical innovations. These sales of patient records for profit by provider organizations may raise concerns from the public that providers have an ulterior

motive for collecting their data during clinical encounters. In addition, patient record licensing contracts with exclusive rights may raise questions about the appropriate stewardship of patient data, as such exclusive contracts may be seen to benefit specific licensees at the expense of others, rather than enabling research and product development across the entire marketplace.

Consent and authorization

Issues that arise in the potential risks of patient data use can be mitigated by proactively obtaining appropriate authorization or consent from patients for the use of their data. These issues primarily apply to PHI covered under HIPAA, however, and not de-identified data. The HIPAA Privacy Rule permits, but does not require, a covered entity voluntarily to obtain patient consent for uses and disclosures of PHI for TPO. Covered entities that decide to obtain consent have complete discretion to design a process that best suits their needs. By contrast, an authorization is required by the Privacy Rule for most uses and disclosures of PHI not otherwise allowed by the Rule. Where the Privacy Rule requires patient authorization, voluntary consent is not sufficient to permit a use or disclosure of PHI. An authorization is a detailed document that gives covered entities permission to use PHI for specified purposes (e.g., sale or marketing of PHI) or to disclose PHI to a third party specified by the individual. An authorization must include a number of elements, including a description of the PHI to be used and disclosed, the person authorized to make the use or disclosure, the person to whom the covered entity may make the disclosure, an expiration date, and, in some cases, the purpose for which the information may be used or disclosed.[27]

PHI may be used and disclosed for research without an authorization in limited circumstances: (1) Under a waiver of the authorization requirement; (2) as a limited data set with a data use agreement; (3) preparatory to research; and (4) for research on decedents' information. Limited data sets exclude 16 categories of direct identifiers, rather than the 18 identifiers removed in de-identified data. The information in a limited data set is considered PHI and its use or disclosure requires a data use agreement between the covered entity and the entity that will receive or use the data.

Non-covered entities that use de-identified health data for purposes such as genomics services or research are not regulated under HIPAA, but most have policies and procedures in place to protect the privacy of their subscribers or participants, and to ensure transparency in the use of the data. 23andMe, for example, obtains personal information from its subscribers and through its service identifies genetic information that is stored within its databases. According to its Privacy Policy, 23andMe "implements physical, technical, and administrative measures to prevent unauthorized access to or disclosure of your information, to maintain data accuracy, to ensure the appropriate use of information, and otherwise safeguard your Personal Information."[28] Subscribers can voluntarily consent to allow their information to be used in research, and can also choose what level of de-identified data they consent for use. 23andMe stores and allows access to both aggregate and individual level data to third-party service providers such as marketing and analytics organizations and targeted advertising service providers that contribute to the service provided by 23andMe. It also sells de-identified user data to pharmaceutical companies for the purposes of research.

Other entities may use anonymous, de-identified data in manners that are legal but may be perceived as ethically questionable since they may not have obtained patient consent for the use of the data. For example, a startup artificial intelligence business, funded by executives at a cancer center, has received exclusive access to the cancer center's database of millions of tissue slides.[29] The cancer center holds an equity stake in the organization along with two of its top leaders, and other board members are initial investors in the new venture. The company's leadership indicated that some patients had provided consent for the use of their data, others did not and their data was

subsequently stripped of its identifying factors. Still, pathologists at the cancer center, and their patients, have expressed concern about the potential conflict of interest in the cancer center leadership's relationship with the startup, as well as the use of patient data for a profit-driven venture. In this case, it was reported that the enterprise had been reviewed and approved by an IRB.[29]

Standards and guidance

ONC publishes the "Guide to Privacy and Security of Electronic Health Information" to help physicians, other health care providers and practices work to comply with federal requirements in collecting, storing and using patients' data.[30]

In addition to the policy set by the AMA and the guidance provided in the AMA *Code of Medical Ethics*, other physician and health care organizations provide guidelines and standards on the use of de-identified patient data. For example, the American Academy of Family Physicians published a "Data Stewardship" policy that facilitates the appropriate collection, storage, transmission, analysis, and reporting of de-identified patient data.[31] This policy includes guidance on establishing and maintaining a proper patient and physician consent process, as well as the appropriate use of data by third parties and policies that establish requirements for third party use.

The American College of Physicians (ACP) policy encourages clinical entities and physicians to publish electronically their policies and procedures for sharing patient data and ensuring privacy. ACP's policy also states that in keeping with HIPAA, patients should know what information exists about them, its purpose, who can access and use it, and where it resides. While ACP supports the use of appropriately de-identified patient data for socially important activities, such as population health efforts and retrospective research, it does recommend tighter controls on the risks of re-identification of de-identified data.[32]

CONCLUSION

Access to de-identified patient data is important for the future of health care. Its benefits to the field of research have significant implications for our ability to make progress in refining the practice of medicine, reducing health care costs, reducing and preventing chronic disease, identifying cures for deadly conditions, and much more. In practice-level interventions, de-identified data can help practice administrators recognize patterns and gaps in processes and treatment plans across clinicians. In the genomics and biotechnology fields the study of patient data, stripped of identifying factors, can contribute to global innovation in medical technology and pharmaceutical solutions. There are numerous ways in which the use of de-identified patient data contributes to the continuum of improvement that is much needed across health care.

Its use does not come without risks, however. In 1951, the development of the HeLa cell line led to many significant research accomplishments in medicine. However, the lack of patient consent in the development of the cell line raises serious ethical concerns, which were further compounded by the commercial use of the cell line for profit, which was not shared with the patient or her family. Though in recent times, substantial effort has been made to correct this historical wrong by the National Institutes of Health and other organizations, much of the harm done to patients who's clinically obtained samples were used without consent can never be undone. Today, a new revolution in health science powered by big data is in process, and there is little doubt that the research accomplishments derived from this data will transform the practice of medicine. However, all stakeholders involved now have an opportunity to ensure that there is not a repeat of the ethical mistakes of the past. Risk mitigation is the responsibility of all stakeholders, from the individual

clinician and patient to the administrators and third-party data users. The privacy and security of the patient data must be protected at every point, and its use needs to be ethically conducted with the appropriate level of consent or authorization required. The HIPAA provisions, regulations enacted at the state level, and organizational policies and procedures, ensure compliance with standards developed to protect the patient. If followed appropriately, these measures can effectively protect patient data from misuse.

RECOMMENDATIONS

The Board of Trustees recommends that the following be adopted and the remainder of this report be filed:

1. That our American Medical Association (AMA) reaffirm Policies H-315.974, "Guiding Principles, Collection and Warehousing of Electronic Medical Record Information," H-315.983, "Patient Privacy and Confidentiality," H-315.975, "Police, Payer, and Government Access to Patient Health Information," H-315.978, "Privacy and Confidentiality," and H-315.987, "Limiting Access to Medical Records." (Reaffirm HOD Policy)

2. That our AMA support state-based efforts to protect patient privacy including the patient's right to know whether information is being disclosed or sold and to whom and the right to opt out of the sale of their data. (New HOD Policy)

3. That our Council on Ethical and Judicial Affairs consider re-examining existing guidance relevant to the confidentiality of patient information in light of new practices regarding de-identified patient data, including the use of exclusive de-identified data licensing agreements in healthcare. (Directive to Take Action)

4. That Policy D-315.975, "Research Handling of De-Identified Patient Information," be rescinded, as having been fulfilled by this report. (Rescind HOD Policy)

Fiscal note: Minimal – Less than $500

REFERENCES

1. U.S. Department of Health and Human Services and National Institutes of Health. *Clinical Research and the HIPAA Privacy Rule*. 2004 06/22/04 [cited 2018 August 24]; Available from: https://privacyruleandresearch.nih.gov/clin_research.asp.
2. Rothstein, M.A., *Is Deidentification Sufficient to Protect Health Privacy in Research?* The American journal of bioethics : AJOB, 2010. 10(9): p. 3-11.
3. United States, *The Health Insurance Portability and Accountability Act of 1996*. 1996, U.S. Department of Labor Employee Benefits Security Administration: Washington, D.C.
4. U.S. Department of Health and Human Services. *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the HIPAA Privacy Rule*. 2015 11/06/15 [cited 2018 August 24]; Available from: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.
5. U.S. Department of Health and Human Services, *Summary of the HIPAA Privacy Rule*. 2013.
6. Rothstein, M.A., *Currents in contemporary bioethics. Access to sensitive information in segmented electronic health records.* J Law Med Ethics, 2012. 40(2): p. 394-400.
7. Gammon, A. and D.W. Neklason, *Confidentiality & the Risk of Genetic Discrimination: What Surgeons Need to Know.* Surgical oncology clinics of North America, 2015. 24(4): p. 667-681.
8. U.S. Department of Health and Human Services, *HIPAA FAQs: May a health information organization (HIO), acting as a business associate of a HIPAA covered entity, de-identify information and then use it for its own purposes?* 2008.
9. Klein, D. *Comparing the California Consumer Privacy Act (CCPA) and the EU's General Data Protection Regulation (GDPR)* 2018.
10. State of California, *California Consumer Privacy Act of 2018*, in *1.81.5*. 2018.
11. Mathews, K. and C. Bowman, *The California Consumer Privacy Act of 2018*, in *Proskauer Privacy Law Blog*, Kristen J. Mathews, Editor. 2018: California.
12. Orion Health, *Introduction to Machine Learning in Healthcare*. 2016, Orion Health.
13. Fatima Paruk, *How to increase adoption of machine learning in healthcare*, in *Vital Signs Blog*. 2018, Modern Healthcare.
14. Orion Health, *Orion Health Unveils New Predictive Intelligence Using Machine Learning to Help Save Billions in Healthcare Costs*. 2018, PRNewswire.
15. NCVHS Ad Hoc Workgroup for Secondary Uses of Health Data, *HIPAA Privacy Rule and Secondary Uses of Health Information*. 2007, United States Department of Health and Human Services.
16. U.S. Department of Health and Human Services, *Code of Federal Regulations Title 45 Public Welfare Part 46 Protection of Human Subjects*, in *45*. 2009.
17. Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change*. 2012.
18. Harvard University, *Harvard Research Data Security Policy*, Office of the Vice Provost for Research, Editor. 2014.
19. Federal Register, Federal Policy for the Protection of Human Subjects. Vol. 82, No. 12. January 19, 2017. Available from: https://www.govinfo.gov/content/pkg/FR-2017-01-19/pdf/2017-01058.pdf
20. Benitez, K. and B. Malin, *Evaluating re-identification risks with respect to the HIPAA privacy rule.* J Am Med Inform Assoc, 2010. 17(2): p. 169-77.
21. Sweeney, L., A. Abu, and J. Winn, *Identifying Participants in the Personal Genome Project by Name (A Re-identification Experiment)*. 2013.
22. Sweeney, L., *Weaving Technology and Policy Together to Maintain Confidentiality.* J Law Med Ethics, 1997. 25(2-3): p. 98-110.
23. Na, L., et al., *Feasibility of reidentifying individuals in large national physical activity data sets from which protected health information has been removed with use of machine learning.* JAMA Network Open, 2018. 1(8): p. e186040.

24. Gymrek, M., et al., *Identifying personal genomes by surname inference.* Science, 2013. 339(6117): p. 321-4.
25. El Emam, K., et al., *A Systematic Review of Re-Identification Attacks on Health Data.* PLoS ONE, 2011. 6(12): p. e28071.
26. Sakowitz-Klein, J., *Proposed ACO Rule Implicates HIPAA*, in *Health Reform Resource Center*. 2011, Akin Gump Strauss Hauer & Feld, LLP.
27. U.S. Department of Health and Human Services, *HIPAA FAQs: What is the difference between "consent" and "authorization" under the HIPAA Privacy Rule?* 2013.
28. 23andMe.com. *Privacy Policy*. 2018 July 17, 2018 September 17, 2018; Available from: https://www.23andme.com/about/privacy/.
29. Advisory Board *Memorial Sloan Kettering insiders face controversy over AI startup*. 2018.
30. Office of the National Coordinator for Health Information Technology, *Guide to Privacy and Security of Electronic Health Information*. 2015, U.S. Department of Health and Human Services.
31. American Academy of Family Physicians. *Data Stewardship*. 2014 [cited 2018 October 3]; Available from: https://www.aafp.org/about/policies/all/data.html.
32. American College of Physicians, *Policy Compendum*, D.o.G.A.a.P. Policy, Editor. 2016: Washington, DC.