AMA *Code of Medical Ethics*

3.3.4 Research Handling of De-Identified Patient Data

Within health care systems, identifiable private health information, initially derived from and used in the care and treatment of individual patients, has led to the creation of massive de-identified datasets. As aggregate datasets, clinical data takes on a secondary promising use as a means for quality improvement and innovation that can be used for the benefit of future patients and patient populations. While de-identification of data is meant to protect the privacy of patients, there remains a risk of re-identification, so while patient anonymity can be safeguarded it cannot be guaranteed. In handling patient data, individual physicians thus strive to balance supporting and respecting patient privacy while also upholding ethical obligations to the betterment of public health.

When clinical data are de-identified and aggregated, their potential use for societal benefits through research and development is an emergent, secondary use of electronic health records that goes beyond individual benefit. Such data, due to their potential to benefit public health, should thus be treated as a form of public good, and the ethical standards and values of health care should follow the data and be upheld and maintained even if the data are sold to entities outside of health care. The medical profession's responsibility to protect patient privacy as well as to society to improve future health care should be recognized as inherently tied to these datasets, such that all entities granted access to the data become data stewards with a duty to uphold the ethical values of health care in which the data were produced.

As individuals or members of health care institutions, physicians should:

(a) Follow existing and emerging regulatory safety measures to protect patient privacy.

(b) Practice good data intake, including collecting patient data equitably to reduce bias in datasets.

(c) Answer any patient questions about data use in an honest and transparent manner to the best of their ability in accordance with current federal and state legal standards.

Health care entities, in interacting with patients, should adopt policies and practices that provide patients with transparent information regarding:

(d) The high value that health care institutions place on protecting patient data.

(e) The reality that no data can be guaranteed to be permanently anonymized, and that risk of re-identification does exist.

(f) How patient data may be used.

(g) The importance of de-identified aggregated data for improving the care of future patients.

Health care entities managing de-identified datasets, as health data stewards, should:

(h) Ensure appropriate data collection methods and practices that meet industry standards to support the creation of high-quality datasets.

(i) Ensure proper oversight of patient data is in place, including Data Use/Data Sharing Agreements for the use of de-identified datasets that may be shared, sold, or resold.

(j) Develop models for the ethical use of de-identified datasets when such provisions do not exist, such as establishing and contractually requiring independent data ethics review boards free of conflicts of interest and verifiable data audits, to evaluate the use, sale, and potential resale of clinically derived datasets.

(k) Take appropriate cyber security measures to seek to ensure the highest level of protection is provided to patients and patient data.

(l) Develop proactive post-compromise planning strategies for use in the event of a data breach to minimize additional harm to patients.

(m) Advocate that health- and non-health entities using any health data adopt the strongest protections and seek to uphold the ethical values of the medical profession.

There is an inherent tension between the potential benefits and burdens of de-identified datasets as both sources for quality improvement to care as well as risks to patient privacy. Re-identification of data may be permissible, or even obligatory, in rare circumstances when done in the interest of the health of individual patients. Re-identification of aggregated patient data for other purposes without obtaining patients' express consent, by anyone outside or inside of health care, is impermissible.

AMA Principles of Medical Ethics: IV

Background report(s):

CEJA Report 2-A-24, Research Handling of De-Identified Patient Data

REPORT 2 OF THE COUNCIL ON ETHICAL AND JUDICIAL AFFAIRS (A-24)
Research Handling of De-Identified Patient Data
(D-315.969)

EXECUTIVE SUMMARY

In adopting policy D-315.969, "Research Handling of De-Identified Patient Data," the House of Delegates directed the Council on Ethical and Judicial Affairs (CEJA) to examine guidance related to the use of de-identified patient data and the risks of re-identification.

This report articulates a series of recommendations on how best to respond to the increasing collection, sale, and use of de-identified patient data and the associated risks. The report outlines how health data exist within digital information ecosystems, how such complex ecosystems pose challenges to data privacy, how de-identified data functions as a public good for clinical research, and how de-identified data derived within the context of health care institutions lead to certain ethical standards for and protections of that data.

Because CEJA recognizes both the promise of de-identified datasets for advancing health and the concerns surrounding the use of de-identified patient data including the risks of re-identification that extend from the level of individual physicians collecting clinical data to hospitals and other health care institutions as repositories and stewards of data, this report proposes a new Code of Medical Ethics opinion be adopted in conjunction with amendments to four existing opinions to provide ethics guidance in this rapidly evolving digital health ecosystem.

REPORT OF THE COUNCIL ON ETHICAL AND JUDICIAL AFFAIRS*

CEJA Report 2-A-24

Subject:	Research Handling of De-Identified Patient Data (D-315.969)
Presented by:	David A. Fleming, MD, Chair
Referred to:	Reference Committee on Amendments to Constitution and Bylaws

Policy D-315.969, "Research Handling of De-Identified Patient Data," adopted by the American Medical Association (AMA) House of Delegates in November 2021, asked the Council on Ethical and Judicial Affairs (CEJA) to examine guidance related to the use of de-identified patient data and the risks of re-identification.

In its informational report on de-identified data [CEJA 6-A-23], CEJA examined a range of challenges that health care professionals and institutions are now confronted with as technological innovations rapidly evolve both within and outside of health care, blurring the boundary distinctions between these spheres. CEJA's exploration suggested that in this dynamic environment, foundational ethical concepts of privacy and consent likely need to be revisited to better reflect that personal health information today exists in digital environments where responsibilities are distributed among multiple stakeholders.

This report expands on the previous work to articulate a series of recommendations on how best to respond to the increasing collection, sale, and use of de-identified patient data and the associated risks. The report outlines how health data exist within digital information ecosystems, how such ecosystems pose challenges to data privacy, what the *Code* says about data privacy and informed consent, how de-identified data functions as a public good for clinical research, how privacy scholars are reconceptualizing privacy as contextual integrity, and how de-identified data derived within the context of health care institutions lead to certain ethical standards for and protections of that data.

Because CEJA recognizes both the promise of de-identified datasets for advancing health and the concerns surrounding the use of de-identified patient data including the risks of re-identification that extend from the level of individual physicians collecting clinical data to hospitals and other health care institutions as repositories and stewards of data, this report proposes a new ethics opinion in conjunction with amendments to four existing opinions to provide ethics guidance in this rapidly evolving digital health ecosystem.

* Reports of the Council on Ethical and Judicial Affairs are assigned to the Reference Committee on Amendments to Constitution and Bylaws. They may be adopted, not adopted, or referred. A report may not be amended, except to clarify the meaning of the report and only with the concurrence of the Council.

HEALTH DATA & DIGITAL ECOSYSTEMS

De-identified patient data are a subset of health data that exists within larger digital health information ecosystems [1]. Such ecosystems are highly dynamic and distributed, with health information often being combined from multiple datasets and distributed among multiple stakeholders [1]. Traditionally, health data has referred to patient health information produced from patient–physician interactions and stored by health care organizations [2]. This type of data is typically recorded as identifiable patient data and entered into the patient's electronic medical record (EMR); from there, it can be de-identified and bundled together with other patent data to form an aggregated dataset. In the age of Big Data, however, where large datasets can reveal complex patterns and trends, diverse sets of information are increasingly brought together. Health data now extends to all health-relevant data, including data collected anywhere from individuals both passively and actively that can reveal information about health and health care use [2].

Within digital health ecosystems, health-related data can be generated by health care systems (e.g., EMRs, prescriptions, laboratory data, radiology), the consumer health and wellness industry (e.g., wearable fitness tracking devices, wearable medical devices such as insulin pumps, home DNA tests), digital exhaust from daily digital activities (e.g., social media posts, internet search histories, location and proximity data), as well as non-health sources of data (e.g., non-medical records of race, gender, education level, residential zip code, credit history) [2]. The ethical challenges raised by such widely distributed data ecosystems, with their vast array of data types and multiple stakeholders, require a holistic approach to the moral issues caused by digital innovation. Digital ethics has arisen as a theoretical framework to analyze these recent challenges and examine such ethical concerns from multiple levels of abstraction. The digital ethics framework takes into account the general environment in which ethical concerns arise and examines ethical dilemmas as they relate to information and data, algorithms, practices and infrastructure, and their impact on the digital world [3].

CHALLENGES TO DATA PRIVACY

In the U.S., the Health Insurance Portability and Accountability Act (HIPAA) imposes constraints on the sharing of "protected health information," including individually identifiable health information contained in the EMR, by "covered entities," including physicians, hospitals, pharmacies, and third-party payers. HIPAA's scope is narrow and does not cover other health-relevant data, such as data generated voluntarily by patients themselves, for example, through the use of commercial health-related apps or devices, or identifiable data individuals provide to municipal authorities, utilities, retailers, or on social media. Furthermore, information that began in the medical record can take on a new, independent life when linked with personal information widely available through datasets generated outside of health care. As McGraw and Mandl explain, "since HIPAA's coverage is about 'who' holds the data, but not what type of data, much of the health-relevant data collected today are collected by entities outside of HIPAA's coverage bubble and thus resides outside of HIPAA's protections" [2]. HIPAA is thus limited in its ability to protect patient data within digital health information ecosystems.

Complicating the matter is the fact that once patient health data has been de-identified, it is no longer protected by HIPAA, and can be freely bought, sold, and combined with other datasets. Hospitals now frequently sell de-identified datasets to researchers and industry. Recent developments in AI and its use within health care have similarly created new difficulties.

Patients, and patient privacy advocates, are often concerned about who has access to their data. As data ecosystems have grown larger and more distributed, this has become increasingly more

difficult to ascertain. In the age of Big Data, the global sale of data has become a multibillion-dollar industry, with individuals' data viewed by industry as "new oil" [1]. The global health care data monetization market alone was valued at just over $0.4 billion in 2022 and is expected to grow to $1.3 billion by 2030 [4]. Industry often purchases hospital datasets to improve marketing and sales, predict consumer behaviors, and to resell to other entities. Within health care and research settings, the massive datasets collected from clinical data—used initially in the care and treatment of individual patients—have created the potential for secondary use as a means for quality improvement and innovation that can be used for the benefit of future patients and patient populations [5].

The dynamic and distributed nature of today's digital health information ecosystems challenges the prevailing procedural model for protecting patient privacy: informed consent and de-identification. In a world where the secondary use of patient data within large datasets can easily enter into a global marketplace, the intended use is almost impossible to discern. Patients cannot be honestly and accurately informed about the specific terms of interactions between their collected data and the data collector and any potential risks that may emerge [1,6]. Therefore, patients are unable to truly give informed consent. Furthermore, whether de-identifying datasets truly prevents individual data subjects from being re-identified has been increasingly called into question. Removing the 18 identifiers specified in HIPAA does not ensure that the data subject cannot be re-identified by triangulation with identifying information from other readily available datasets [7]. Machine learning and AI technologies have advanced to the point that virtually all de-identified datasets risk re-identification, such that "even when individuals are not 'identifiable', they may still be 'reachable'" [6].

A final avenue to consider with respect to private health information and patient privacy is the risk of health care data breaches. Raghupathi et al note, "[h]ealthcare is a lucrative target for hackers. As a result, the healthcare industry is suffering from massive data breaches" [8]. The number of health care data breaches continues to increase every year, exposing the private health information of millions of Americans. Despite being heavily targeted by cybercriminals, health care providing institutions are widely considered by cybersecurity experts to lack sufficient security safeguards [8]. Raghupathi et al note, "healthcare entities gathering and storing individual health data have a fiduciary and regulatory duty to protect such data and, therefore, need to be proactive in understanding the nature and dimensions of health data breaches" [8].

CLINICAL DATA AND PRIVACY

Within the *Code,* Opinion 3.1.1, "Privacy in Health Care," distinguishes four aspects of privacy:

> personal space (physical privacy), personal data (informational privacy), personal choices including cultural and religious affiliations (decisional privacy), and personal relationships with family members and other intimates (associational privacy).

The *Code* does not explicitly examine whether personal medical or health information are ethically distinct from other kinds of personal information (e.g., financial records) or in what way. Current guidance treats the importance of protecting privacy in all its forms as self-evident, holding that respecting privacy in all its aspects is of fundamental importance, "an expression of respect for autonomy and a prerequisite for trust" [Opinion 3.1.1]. However, Opinion 3.3.3, "Breach of Security in Electronic Medical Records," directly acknowledges that data security breaches create potential "physical, emotional, and dignity harms" to patients. Similarly, Opinion 7.3.7, "Safeguards in the Use of DNA Databanks," states that breaches of confidential patient information

"may result in discrimination or stigmatization and may carry implications for important personal choices."

Violations of privacy can result in both harm—tangible negative consequences, such as discrimination in insurance or employment or identity theft—and in wrongs that occur from the fact of personal information being known without the subject's awareness, even if the subject suffers no tangible harm [7]. Price and Cohen note that privacy issues can arise not only when data are known, but when data mining enables others to "generate knowledge about individuals through the process of inference rather than direct observation or access" [7].

CLINICAL DATA AND INFORMED CONSENT

With respect to Opinion 2.1.1, "Informed Consent," in the *Code,* successful communication is seen as essential to fostering trust that is fundamental to the patient–physician relationship and to supporting shared decision making. Opinion 2.1.1 states: "[t]he process of informed consent occurs when communication between a patient and physician results in the patient's authorization or agreement to undergo a specific medical intervention." In seeking a patient's informed consent, physicians are directed to include information about "the burdens, risks, and expected benefits of all options, including forgoing treatment" [Opinion 2.1.1]. It should be noted, however, that no direct mention of patient data is discussed in the opinion, other than that documentation of consent should be recorded in the patient's medical record.

CLINICAL DATA, DATASETS, AND THE PUBLIC GOOD

Because aggregated clinical data has the potential for secondary use that can benefit all of society, it has been argued that such data should be treated as a form of public good [5]. When clinical data are de-identified and aggregated, the potential use for societal benefits through research and development is an emergent, secondary side effect of electronic health records that goes beyond individual benefit. Larson et al argue that not only does the public possess an interest in safeguarding and promoting clinical data for societal benefits, but all those who participate in health care systems have an ethical responsibility to treat such data as a form of public good [5]. They propose:

> all individuals and entities with access to clinical data inherently take on the same fiduciary obligations as those of medical professionals, including for-profit entities. For example, those who are granted access to the data must accept responsibility for safeguarding protected health information [5].

This entails that any entity that purchases private health information, whether or not it has been de-identified, has an ethical obligation to adhere to the ethical standards of health care where such data were produced. Hospitals thus have an ethical responsibility to ensure that their contracts of sale for datasets insist that all entities that gain access to the data adhere to the ethical standards and values of the health care industry.

This is particularly important when we recall that the wide distribution of digital health information ecosystems increasingly includes non-health-related parties from industry that may have market interests that conflict with the ethical obligations that follow health data. Within this framework, the fiduciary duty to protect patient privacy as well as to society to improve future health care follows the data and thus applies to all entities that use that data, such that all entities granted access to the data become data stewards, including for-profit parties [5]. This also includes patients, such that they bear a responsibility to allow their data to be used for the future improvement of

health care for society, especially when we recognize that current health care has already benefited from past data collection [5].

While the re-identification of aggregated patient data should generally be prohibited, there are rare exceptions. There may be occasions when researchers wish to re-identify a dataset, such as sometimes occurs in the study of rare diseases that rely on international registries; in such situations, all individuals must be re-contacted, and their consent obtained in order to re-identify their data since this would represent a significant change to the initial research protocols and respective risks [9]. Re-identification of datasets for research is uncommon, however, because obtaining re-consent can be difficult and can lead to flawed research if data is lost because patients do not re-consent. The other situation in which it may be permissible, or even obligatory, to re-identify aggregated patient data is when doing so would be in the interest of the health of individual patients, such as might occur in the study of a rare genetic disorder. Even within these exceptions, the risks associated with re-identification remain and re-identified data should thus never be published. Re-identification of de-identified patient data for any other purposes, by anyone inside or outside of health care, must be avoided.

AN ALTERNATIVE APPROACH: PRIVACY AS CONTEXTUAL INTEGRITY

Within today's digital health information ecosystems, physicians and hospitals face several challenges to protecting patient privacy. Barocas and Nissenbaum contend that "even if [prevailing forms of consent and anonymization] were achievable, they would be ineffective against the novel threats to privacy posed by big data" [6]. A more effective option, Nissenbaum has argued, would understand privacy protection as a function of "contextual integrity," i.e., that in a given social domain, information flows conform to the context-specific informational norms of that domain. Whether a transmission of information is appropriate depends on "the type of information in question, about whom it is, by whom and to whom it is transmitted, and conditions or constraints under which this transmission takes place" [10]. The view of privacy as contextual integrity—that our conception of privacy is contextual and governed by various norms of information flow—recognizes that there exist different norms regarding privacy within different spheres of any distributed digital ecosystem [7,11]. The challenge within health care, as we have seen, is how to balance these various norms when they conflict and how to ensure that health care's ethical standards and values are maintained throughout the distributed use of de-identified private health information.

THE CONTEXTUAL INTEGRITY OF DE-IDENTIFIED HEALTH DATA

In handling patient data, individual physicians strive to balance supporting and respecting patient privacy while also upholding ethical obligations to the betterment of public health. Through their own actions, as well as through their membership organizations and through their health care organizations, physicians should: (1) ensure that data entered into electronic records are accurate and reliable to the best of their ability; (2) be transparent with patients regarding the limited extent to which their data can be safely protected, how their data may be used, and why the use of such data is crucial for improving health care outcomes within society; and (3) ensure that proper oversight and protections of data are in place, including contractual provisions that any data sold or shared with outside entities stay in alignment with the ethical standards of the medical profession, and that meaningful sanctions or penalties are in place and enforced against any actors that violate those ethical standards. It is critical to recognize, as is outlined in the *Code,* that the patient–physician relationship is built on trust, and that this trust relies heavily on transparency.

It is important for both patient care and research that clinical data entered into the EMR be as accurate and complete as possible. Some data capture practices, such as copying-and-pasting daily progress notes from previous encounters, which may contribute to efficiency, can lead to documentation errors [12]. One avenue for improving EMR accuracy is that, under HIPAA, patients have the right to access their data and request any perceived errors be amended. While there is no one solution to improving accuracy of EMR data, further study into how to improve EMR accuracy is important. One challenge to both EMR accuracy and completeness is the limited interoperability of different EMR systems. Matching digital health records for the same patient across and within health care facilities can be a challenge, further contributing to the potential for EMR errors. Standardization of recording data elements, such as capturing patient address and last name in a consistent format, may improve matching of patient records and thus improve the accuracy of the EMR [13].

Another challenge to EMR data quality is the risk of bias, primarily due to implicit bias in EMR design and underrepresentation of patients from historically marginalized groups, low socioeconomic status, and rural areas [14,15]. Critically important for research involving data collected from EMRs, available EMR data only reflects those with access to health care in the first place. While certain study designs and tools have been developed to reduce these biases in research, physicians and health care institutions should be looking into ways to reduce bias within EMRs, such as features to optimize effective EMR use and to consistently capture patient data, especially data on race/ethnicity and social determinants of health that are often inconsistently and inaccurately captured in EMR systems [14,15,16].

Patients have a right to know how and why their data are being used. While physicians should be able to answer questions regarding patient data as they relate to HIPAA protections, it is the responsibility of health care institutions to provide more detailed information regarding expectations of data privacy, how patient data may be used, and why such use is important to improve the future of health care. Health care systems may consider fulfilling this ethical obligation by creating a patient notification of data use built into the patient registration process (using language similar to the National Institutes of Health's (NIH) Introduction-Description component, meant to provide prospective research participants with an introduction to and description of the planned storage and sharing of data and biospecimens [17]).

As stewards of health data, health care institutions have an ethical responsibility to protect data privacy. This fiduciary duty to patient data should be seen as following the data even after they are de-identified and leave the institution where they were initially captured [5,8]. While hospitals and health care organizations increasingly come under cyberattack, they consistently lag behind other industries in cybersecurity [18]. With regards to protecting the data they maintain, health care institutions have a responsibility to make more significant investments in cybersecurity.

In order to ensure that the ethical standards of health care are maintained even after data leaves health care institutions, McGraw and Mandl propose that companies collecting or using health-relevant data could be required to establish independent data ethics review boards [2]. They write that such boards could be similar to Institutional Review Boards but should focus more on privacy than on participant risk, evaluating proposed data projects for legal and ethical implications as well as their potential to improve health and/or the health care system [2]. In practice, ethics review boards involved with industry face challenges to both independence and efficacy. Independence can be compromised by influences such as conflicts of interest, while efficacy can be compromised by the absence of authority, procedures, and systems to enact recommendations made by these review bodies. To be effective, data ethics review boards must be independent and free of conflicts of interest from the company or organization whose data research proposal(s) they are evaluating

and have systems in place for both transparency and implementation of feedback for remediations of privacy and other quality and ethics concerns. Though not a comprehensive solution, independent data ethics review boards could be an effective safeguard against industry conflicts of interest and should be considered as a required part of contracts of sale of health data, with contracts stipulating that any future resale of the data also undergo review by a data ethics review board.

An additional safeguard is the implementation of regular data audits to assess the quality and use of shared data [19]. These regulatory measures could be implemented as requirements outlined in Data Use Agreements or Data Sharing Agreements (DSAs). Such agreements have the potential to establish data governance policies and practices within health care institutions regarding "what data can be shared, with whom, under what conditions, and for what purposes." In developing DSAs, hospital administrators should engage all relevant stakeholders, require a neutral entity be designated as an independent custodian of shared data, limit the types and/or characteristics of shared data to certain purposes, and apply additional safeguards to protect the data [20].

The need for more transparent disclosure to patients regarding their data use as well as the importance of building the values of medical ethics into the contracts of sale of aggregate datasets created by hospitals highlights the fact that the ethical responsibilities to respond to the risks of de-identified data should not be borne by physicians alone. Respecting patient privacy and their informed consent are responsibilities that physician member organizations and health care institutions must take on because the risks to these rights that patients face within digital health ecosystems radiate far beyond the patient–physician relationship to areas where individual physicians have little influence.

RECOMMENDATIONS

In light of the challenges considered with regard to constructing a framework for holding stakeholders accountable within digital health information ecosystems, the Council on Ethical and Judicial Affairs recommends:

1. That the following be adopted:

    Within health care systems, identifiable private health information, initially derived from and used in the care and treatment of individual patients, has led to the creation of massive de-identified datasets. As aggregate datasets, clinical data takes on a secondary promising use as a means for quality improvement and innovation that can be used for the benefit of future patients and patient populations. While de-identification of data is meant to protect the privacy of patients, there remains a risk of re-identification, so while patient anonymity can be safeguarded it cannot be guaranteed. In handling patient data, individual physicians thus strive to balance supporting and respecting patient privacy while also upholding ethical obligations to the betterment of public health.

    When clinical data are de-identified and aggregated, their potential use for societal benefits through research and development is an emergent, secondary use of electronic health records that goes beyond individual benefit. Such data, due to their potential to benefit public health, should thus be treated as a form of public good, and the ethical standards and values of health care should follow the data and be upheld and maintained even if the data are sold to entities outside of health care. The medical profession's responsibility to protect patient privacy as well as to society to improve future health care should be recognized as inherently tied to these

datasets, such that all entities granted access to the data become data stewards with a duty to uphold the ethical values of health care in which the data were produced.

As individuals or members of health care institutions, physicians should:

(a) Follow existing and emerging regulatory safety measures to protect patient privacy;

(b) Practice good data intake, including collecting patient data equitably to reduce bias in datasets;

(c) Answer any patient questions about data use in an honest and transparent manner to the best of their ability in accordance with current federal and state legal standards.

Health care entities, in interacting with patients, should adopt policies and practices that provide patients with transparent information regarding:

(d) The high value that health care institutions place on protecting patient data;

(e) The reality that no data can be guaranteed to be permanently anonymized, and that risk of re-identification does exist;

(f) How patient data may be used;

(g) The importance of de-identified aggregated data for improving the care of future patients.

Health care entities managing de-identified datasets, as health data stewards, should:

(h) Ensure appropriate data collection methods and practices that meet industry standards to support the creation of high-quality datasets;

(i) Ensure proper oversight of patient data is in place, including Data Use/Data Sharing Agreements for the use of de-identified datasets that may be shared, sold, or resold;

(j) Develop models for the ethical use of de-identified datasets when such provisions do not exist, such as establishing and contractually requiring independent data ethics review boards free of conflicts of interest and verifiable data audits, to evaluate the use, sale, and potential resale of clinically-derived datasets;

(k) Take appropriate cyber security measures to seek to ensure the highest level of protection is provided to patients and patient data;

(l) Develop proactive post-compromise planning strategies for use in the event of a data breach to minimize additional harm to patients;

(m) Advocate that health- and non-health entities using any health data adopt the strongest protections and seek to uphold the ethical values of the medical profession.

There is an inherent tension between the potential benefits and burdens of de-identified datasets as both sources for quality improvement to care as well as risks to patient privacy. Re-identification of data may be permissible, or even obligatory, in rare circumstances when done in the interest of the health of individual patients. Re-identification of aggregated patient data

for other purposes without obtaining patients' express consent, by anyone outside or inside of health care, is impermissible. (New HOD/CEJA Policy); and

2. That Opinion 2.1.1, "Informed Consent"; Opinion 3.1.1, "Privacy in Health Care"; Opinion 3.2.4, "Access to Medical Records by Data Collection Companies"; and Opinion 3.3.2, "Confidentiality and Electronic Medical Records" be amended by addition as follows:

a. Opinion 2.1.1, Informed Consent

Informed consent to medical treatment is fundamental in both ethics and law. Patients have the right to receive information and ask questions about recommended treatments so that they can make well-considered decisions about care. Successful communication in the patient-physician relationship fosters trust and supports shared decision making. Transparency with patients regarding all medically appropriate options of treatment is critical to fostering trust and should extend to any discussions regarding who has access to patients' health data and how data may be used.

The process of informed consent occurs when communication between a patient and physician results in the patient's authorization or agreement to undergo a specific medical intervention. In seeking a patient's informed consent (or the consent of the patient's surrogate if the patient lacks decision-making capacity or declines to participate in making decisions), physicians should:

(a) Assess the patient's ability to understand relevant medical information and the implications of treatment alternatives and to make an independent, voluntary decision.

(b) Present relevant information accurately and sensitively, in keeping with the patient's preferences for receiving medical information. The physician should include information about:

(i) the diagnosis (when known);

(ii) the nature and purpose of recommended interventions;

(iii) the burdens, risks, and expected benefits of all options, including forgoing treatment.

(c) Document the informed consent conversation and the patient's (or surrogate's) decision in the medical record in some manner. When the patient/surrogate has provided specific written consent, the consent form should be included in the record.

In emergencies, when a decision must be made urgently, the patient is not able to participate in decision making, and the patient's surrogate is not available, physicians may initiate treatment without prior informed consent. In such situations, the physician should inform the patient/surrogate at the earliest opportunity and obtain consent for ongoing treatment in keeping with these guidelines. (Modify HOD/CEJA Policy)

b. Opinion 3.1.1, Privacy in Health Care

Protecting information gathered in association with the care of the patient is a core value in health care. However, respecting patient privacy in other forms is also fundamental, as an expression of respect for patient autonomy and a prerequisite for trust.

Patient privacy encompasses a number of aspects, including personal space (physical privacy), personal data (informational privacy), personal choices including cultural and religious affiliations (decisional privacy), and personal relationships with family members and other intimates (associational privacy).

Physicians must seek to protect patient privacy in all settings to the greatest extent possible and should:

(a) Minimize intrusion on privacy when the patient's privacy must be balanced against other factors.

(b) Inform the patient when there has been a significant infringement on privacy of which the patient would otherwise not be aware.

(c) Be mindful that individual patients may have special concerns about privacy in any or all of these areas.

(d) Be transparent with any inquiry about existing privacy safeguards for patient data but acknowledge that anonymity cannot be guaranteed and that breaches can occur notwithstanding best data safety practices. (Modify HOD/CEJA Policy)

c. Opinion 3.2.4, Access to Medical Records by Data Collection Companies

Information contained in patients' medical records about physicians' prescribing practices or other treatment decisions can serve many valuable purposes, such as improving quality of care. However, ethical concerns arise when access to such information is sought for marketing purposes on behalf of commercial entities that have financial interests in physicians' treatment recommendations, such as pharmaceutical or medical device companies.

Information gathered and recorded in association with the care of a patient is confidential. Patients are entitled to expect that the sensitive personal information they divulge will be used solely to enable their physician to most effectively provide needed services. 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.

Physicians who propose to permit third-party access to specific patient information for commercial purposes should:

(a) Only provide data that has been de-identified.

(b) Fully inform each patient whose record would be involved (or the patient's authorized surrogate when the individual lacks decision-making capacity) about the purpose(s) for which access would be granted.

Physicians who propose to permit third parties to access the patient's full medical record should:

(c) Obtain the consent of the patient (or authorized surrogate) to permit access to the patient's medical record.

(d) Prohibit access to or decline to provide information from individual medical records for which consent has not been given.

(e) Decline incentives that constitute ethically inappropriate gifts, in keeping with ethics guidance.

Because de-identified datasets are derived from patient data as a secondary source of data for the public good, health care professionals and/or institutions who propose to permit third-party access to such information have a responsibility to establish that any use of data derived from health care adhere to the ethical standards of the medical profession. (Modify HOD/CEJA Policy)

d. Opinion 3.3.2, Confidentiality and Electronic Medical Records

Information gathered and recorded in association with the care of a patient is confidential, regardless of the form in which it is collected or stored.

Physicians who collect or store patient information electronically, whether on stand-alone systems in their own practice or through contracts with service providers, must:

(a) Choose a system that conforms to acceptable industry practices and standards with respect to:

(i) restriction of data entry and access to authorized personnel;

(ii) capacity to routinely monitor/audit access to records;

(iii) measures to ensure data security and integrity; and

(iv) policies and practices to address record retrieval, data sharing, third-party access and release of information, and disposition of records (when outdated or on termination of the service relationship) in keeping with ethics guidance.

(b) Describe how the confidentiality and integrity of information is protected if the patient requests.

(c) Release patient information only in keeping with ethics guidance for confidentiality and privacy. (Modify HOD/CEJA Policy); and

3. That the remainder of this report be filed.

Fiscal Note: Less than $500

REFERENCES

1. Ruotsalainen P, Blobel B. Health information systems in the digital health ecosystem—problems and solutions for ethics, trust and privacy. Int J of Envir Res and Pub health 2020;17(9):3006.
2. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. NPJ dig med 2021;4(1):2.
3. Floridi L, Cath C, Taddeo M. Digital ethics: its nature and scope. *The 2018 yearbook of the digital ethics lab* 2019:9-17.
4. Data monetization in healthcare market size report, 2030. In: Data Monetization In Healthcare Market Size Report, 2030. https://www.grandviewresearch.com/industry-analysis/data-monetization-healthcare-market-report.
5. Larson DB, Magnus DC, Lungren MP, Shah NH, Langlotz CP. Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. Radiology 2020;295(3):675-82.
6. Barocas S, Nissenbaum H. Big data's end run around anonymity and consent. *Privacy, big data, and the public good: Frameworks for engagement* 2014;1:44-75.
7. Price WN II, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25:37–43.
8. Raghupathi W, Raghupathi V, Saharia A. Analyzing Health Data Breaches: A Visual Analytics Approach. AppliedMath 2023;3(1):175-99.
9. Hansson MG, Lochmüller H, Riess O, Schaefer F, Orth M, Rubinstein Y, Molster C, Dawkins H, Taruscio D, Posada M, Woods S. The risk of re-identification versus the need to identify individuals in rare disease research. Eur J Hum Genet 2016;24(11):1553-1558. doi:10.1038/ejhg.2016.52.
10. Nissenbaum, H. Respecting context to protect privacy: why meaning matters. Sci Eng Ethics 2018;24:831–52.
11. Nissenbaum H. Privacy as contextual integrity. Wash. L. Rev 2004;79:119-158.
12. Weng CY. Data Accuracy in Electronic Medical Record Documentation. JAMA Ophthalmol 2017;135(3):232–233. doi:10.1001/jamaophthalmol.2016.5562
13. Pew. Enhancing patient matching is critical to achieving full promise of digital health records: Accurately linking individuals with their health records essential to improving care. Pew Trust Report October 2, 2018. https://www.pewtrusts.org/en/research-and-analysis/reports/2018/10/02/enhanced-patient-matching-critical-to-achieving-full-promise-of-digital-health-records
14. Boyd AD, Gonzalez-Guarda R, Lawrence K, Patil CL, Ezenwa MO, O'Brien EC, Paek H, Braciszewski JM, Adeyemi O, Cuthel AM, Darby JE. Equity and bias in electronic health records data. Contemp Clinic Trials 2023;130:107238.
15. Boyd AD, Gonzalez-Guarda R, Lawrence K, Patil CL, Ezenwa MO, O'Brien EC, Paek H, Braciszewski JM, Adeyemi O, Cuthel AM, Darby JE. Potential bias and lack of generalizability in electronic health record data: reflections on health equity from the National Institutes of Health Pragmatic Trials Collaboratory. J of the Am Medl Informat Ass 2023:ocad115.
16. Khurshid S, Reeder C, Harrington LX, Singh P, Sarma G, Friedman SF, Di Achille P, Diamant N, Cunningham JW, Turner AC, Lau ES. Cohort design and natural language processing to reduce bias in electronic health records research. NPJ Digital Med 2022;5(1):47.
17. NIH. Informed consent for secondary research with data and biospecimens: Points to consider and sample language for future use and/or sharing. NIH Office of Science Policy Office of Extramural Research May 2020. https://osp.od.nih.gov/wp-content/uploads/Informed-Consent-Resource-for-Secondary-Research-with-Data-and-Biospecimens.pdf
18. Skahill E, West DM. Why hospitals and healthcare organizations need to take cybersecurity more seriously. Brookings August 9, 2021. https://www.brookings.edu/articles/why-hospitals-and-healthcare-organizations-need-to-take-cybersecurity-more-seriously/

19. Winter JS, Davidson E. Big Data Governance of personal health information and challenges to Contextual Integrity. The Information Society 35:36–51. doi: 10.1080/01972243.2018.1542648
20. Allen C, Des Jardins TR, Heider A, et al. Data Governance and data sharing agreements for community-wide health information exchange: Lessons from the Beacon Communities. eGEMs (Generating Evidence & Methods to improve patient outcomes) 2:5. doi: 10.13063/2327-9214.1057