

POLICY ANALYSES

Respect, justice and learning are limited when patients are deidentified data subjects

Marielle S. Gross^{1,2}  | Amelia J. Hood²  | Joshua C. Rubin³ | Robert C. Miller Jr.⁴

¹University of Pittsburgh Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh Center for Bioethics and Health Law, Johns Hopkins Berman Institute of Bioethics, Pittsburgh, Pennsylvania, USA

²Johns Hopkins Berman Institute of Bioethics, Baltimore, Maryland, USA

³Learning Health Systems Initiative, University of Michigan Medical School, Ann Arbor, Michigan, USA

⁴Flashbots, New York, New York, USA

Correspondence

Marielle S. Gross, MD, MBE, University of Pittsburgh Department of Obstetrics, Gynecology and Reproductive Sciences University of Pittsburgh Center for Bioethics and Health Law, Johns Hopkins Berman Institute of Bioethics 300 Halket Street, Pittsburgh, PA 15213, USA.
Email: grossms@upmc.edu, mgross23@jhmi.edu

Abstract

Introduction: Critical for advancing a Learning Health System (LHS) in the U.S., a regulatory safe harbor for deidentified data reduces barriers to learning from care at scale while minimizing privacy risks. We examine deidentified data policy as a mechanism for synthesizing the ethical obligations underlying clinical care and human subjects research for an LHS which conceptually and practically integrates care and research, blurring the roles of patient and subject.

Methods: First, we discuss respect for persons vis-a-vis the systemic secondary use of data and tissue collected in the fiduciary context of clinical care. We argue that, without traditional informed consent or duty to benefit the individual, deidentification may allow secondary use to supersede the primary purpose of care. Next, we consider the effectiveness of deidentification for minimizing harms via privacy protection and maximizing benefits via promoting learning and translational care. We find that deidentification is unable to fully protect privacy given the vastness of health data and current technology, yet it imposes limitations to learning and barriers for efficient translation. After that, we evaluate the impact of deidentification on distributive justice within an LHS ethical framework in which patients are obligated to contribute to learning and the system has a duty to translate knowledge into better care. Such a system may permit exacerbation of health disparities as it accelerates learning without mechanisms to ensure that individuals' contributions and benefits are fair and balanced.

Results: We find that, despite its established advantages, system-wide use of deidentification may be suboptimal for signaling respect, protecting privacy or promoting learning, and satisfying requirements of justice for patients and subjects.

Conclusions: Finally, we highlight ethical, socioeconomic, technological and legal challenges and next steps, including a critical appreciation for novel approaches to realize an LHS that maximizes efficient, effective learning and just translation without the compromises of deidentification.

KEYWORDS

common rule, data use policy, deidentification, deidentified data, ethics, federal regulations, HIPAA, human subjects research, learning health systems, structural justice

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Learning Health Systems* published by Wiley Periodicals LLC on behalf of University of Michigan.

1 | BACKGROUND

In the U.S., HIPAA¹ does not restrict sharing health data when 18 classes of identifiers are removed and human subjects protections do not apply to secondary research on deidentified data. This regulatory “safe harbor” for deidentified data connects the clinical context of Hippocratic duties and HIPAA with biomedical research and the Common Rule.² Strengthening this bridge, the HITECH Act (2009)³ mandated digitization of medical records and the Cures Act (2016)⁴ broadened data sharing and use of “real world data” for the advancement of health technology. Relaxed regulatory requirements for use of deidentified patients’ data and tissue has been instrumental for accelerating learning from care, transforming medicine for the digital age.

The concept of Learning Health Systems (LHS) captures this modernization of health care by envisioning continuous and seamless incorporation of research into practice, blurring historic ethical and practical distinctions between patients and subjects.⁵ Faden et al⁶ propose a novel framework for upholding the principles of respect for persons, beneficence and justice in LHS that combines clinical and research ethics. A radical departure from tradition, the framework asserts that patients are obligated to contribute to learning and that the system is obligated to translate knowledge into improved care. Safe harbor for deidentified data facilitates the use of patient data to promote the implementation of an ethical LHS. We examine whether the privacy protections afforded by deidentification are consistent with the optimal synthesis of moral obligations to individuals as both patients and subjects when diverse clinical, pre-clinical and translational research are systematically embedded in care.

First, we reflect on deidentification’s function of permitting data created in the fiduciary context of clinical care to be used without informed consent or duty to benefit the individual. Second, we consider the effectiveness of deidentification for privacy preservation, impact on scientific advancement, and efficiency for an integrated model of care and research. Third, we examine whether systemic use of deidentified data may exacerbate documented structural injustice⁷ in health care by enforcing patients’ obligation to contribute to learning without ensuring equitable distribution of benefits. We conclude with a survey of ethical, legal, socioeconomic and technological implications and topics for further investigation.

1.1 | Respect for patients as subjects

The Belmont Report⁸ bases respect for persons on dual imperatives to treat individuals as autonomous agents, and to protect those with diminished autonomy. Informed consent is a primary mechanism for respecting autonomy in both research and clinical contexts. Classically, informed consent demands a triad of information, comprehension and voluntariness. Informed consent requirements are more rigorous in research contexts because subjects do not benefit from physicians’ fiduciary duty of beneficence. In LHS, research participation is embedded in care, and de facto all patients serve as research

subjects. Traditional, in-person, prospective informed consent, with its significant time and human resource consumption, is neither practical nor desirable for the speed, volume, scale and risk profile of data-driven research.⁹ Transparency and engagement are essential for ensuring respect for persons in this setting.¹⁰ However, best practices for transparency and engagement of patients when research is omnipresent and evolving are yet undefined.

Once patient data has been deidentified, studies using that data are not considered human subjects research. Hence, investigators are incentivized to use deidentified data because informed consent is not required. Recent updates to the Common Rule (2017)² added broad consent as an alternative to reduce the burden of research on patient data and tissue.¹¹ In practice, broad consent may be combined with deidentification to maximize ease of research operations while minimizing patient and institutional risks. Explicit, but generic disclosure of ongoing research on clinical data may be necessary for LHS.¹⁰ Today, patients may not be informed of deidentified data use, or disclosure may be limited to broad consent forms that incorporate unspecified research.

Patients’ consent to physician examinations and procedures to generate clinical data and tissue specimens is predicated on trust. Deidentification demonstrates respect for patients’ privacy apropos the extent of patients’ exposure to the medical gaze in their most vulnerable moments. Figure 1 references The Scar Project, a photo series depicting young breast cancer patients, widely recognized for unprecensored and “shockingly raw” exposure of the “visible world few have seen:” the “true reality of cancer” (www.thescarproject.org). Here, the unaltered portraits are shown alongside the same images in which the subjects’ eyes are covered, mimicking the privacy-preserving technique typical of medical texts. This figure may help illuminate how deidentification may or may not maximize respect for persons when sensitive health data are shared.

The Scar Project photographs symbolize the intimate and personal nature of clinical data and tissue, offering a powerful metaphor for patients’ contributions to research through care. The identified images in Figure 1 evoke transparency and trust, as the subjects are consenting and engaged, aware of their vulnerability and exposure, and contributing their data to learning that serves their own interests and those of other patients. By contrast, the subjects’ consent and engagement are not readily apparent in the deidentified images. When their eyes are covered, the viewer’s attention is drawn to the scar, away from the identifiable face of the human subject. Likewise, patients’ expectations of medical privacy may be discordant with the reality of distribution and use of deidentified data under safe harbor policy. Systemic, nonvoluntary secondary use as an ongoing background process may not be consistent with transparency absent defined strategies for ensuring comprehension of data use practices.

When prospective informed consent is not a viable or desirable option, the return of results to patients could support transparency regarding secondary data use. However, deidentification renders the return of results difficult or impossible. In some cases, reidentification would be technically straightforward, however, U.S. policy explicitly



FIGURE 1 references The Scar Project, (www.thescarproject.org), courtesy of the photographer, a photo documentary of young breast cancer patients. We provide this image as a heuristic to refer to with regard to our arguments about respect for persons, beneficence, justice and how identification may or may not serve individuals with regard to these fundamental principles. Identified images are juxtaposed with one in which the subjects' eyes are covered, a traditional privacy-preserving method widely used in medical textbooks

prohibits reidentification and contact of deidentified subjects, regardless of clinical actionability of findings. Therefore, deidentification does not appear to support transparency or direct engagement of patients as stakeholders in learning.

Deidentification facilitates the treatment of all clinical care as a learning opportunity by reducing burdens traditionally associated with human subjects research. Without informed consent, there is no opportunity to opt out of secondary research. This is justified insofar as risks and burdens of participation are minimized and no further actions by patients are required.^{5,9} Deidentification enforces patients' obligation to contribute to learning, consistent with the LHS ethical framework.¹² In the Belmont Report, persons with diminished autonomy are entitled to protection, where "the extent of protection afforded should depend on the risk of harm and likelihood of benefit" (Part B, Section 1).⁸ Respect for persons under this condition of decreased autonomy requires protection of individuals' best interests regarding potential benefits and corresponding risks of harm. Translational, patient-centered and next-generation precision medicine research seek to produce real-time benefits relevant to the individuals studied. Potential benefits of returning research findings may increasingly outweigh the risk of harm from reidentification in these contexts.

LHS patients have a secondary role as research subjects. Blocking return of results, which may be timely and actionable, under the auspices of privacy protection may not optimally balance risks and benefits for individuals. Without ensuring the return of relevant results, secondary benefits of learning from data and tissue collected for clinical purposes may appear to trump the fiduciary duty to benefit an individual with the products of their own care. Deidentification could be considered dehumanizing if the advancement of generalizable knowledge from a patient's care takes precedence over that individual's welfare.

Current deidentification policy and practice may not be consistent with transparency, engagement or translation of benefits required to demonstrate respect for persons in LHS. Exemption of studies on deidentified data from human subjects research regulations may express a lack of appreciation for the individuals whose care is capitalized upon for the benefit of others. Without adequate transparency and engagement, commercial use of deidentified data may appear exploitative, particularly for individuals whose data or tissue have an outsized value.^{13,14} Respect for patients in clinical settings requires a dynamic and ongoing informed consent process sensitive to invasiveness, timeliness, risks and prospective benefits. Respect for patients as subjects in an LHS may call for a similar approach.¹⁰

1.2 | Promoting beneficence in learning health systems

Learning from clinical data carries a risk of harm that may occur should third parties gain unauthorized access to protected health information (PHI), for which there is an expectation of privacy. Removal of specified identifiers reduces the risk of secondary data use to the small, but non-zero, “possibility that de-identified data could be linked back to the identity of the patient” (US Department of Health and Human Services Office for Civil Rights [US HHS OCR] 2010).¹⁵ Safe harbor for deidentified data has been widely utilized to decrease barriers to learning. Research in an LHS is a public work to which everyone must contribute and from which everyone should benefit.¹⁶ Knowledge has advanced substantially, though there is no established metric for assessing effectiveness or efficiency for maximizing benefits for contributing patients. We consider the impact of deidentification on privacy preservation, learning, and translation as core elements of risk/benefit calculus in LHS.

Massive amounts of identifiable health data now exist both within and outside of traditional health care contexts. Removal of 18 identifiers conceals a data subject’s identity to the human eye of a third-party investigator, though the dataset remains detailed enough to be unique to a specific person.¹⁷ For clinician-investigators who may already know patients’ identities, remaining details in deidentified records may be sufficient for reidentification. Meanwhile, the vastness of data and the power of AI present challenges for genuine deidentification of digital biospecimens for any prospective data user. The possibility of third-party reidentification increases further via triangulation with many publicly available datasets.^{18,19}

Limited oversight for deidentified data use may increase the risk of reidentification and potential misuse by third-parties. Researchers may themselves to produce algorithms or other knowledge products with unintended effects absent adequate ethical oversight. For example, specific individuals may be identified and harmed, for example, via denial of insurance benefits or other algorithmic biases, even if their identities are never revealed to third parties. Deidentification of digital biospecimens using legacy techniques may neither preserve the privacy of identity nor prevent downstream harms, in the era of big, multimodal data and ever-advancing technologies.

Likewise, the Common Rule dictates that biospecimens “can be used to generate information unique to individuals and therefore cannot be truly deidentified”,^{11,20} thus informed consent is required for secondary use. This statute does not apply retroactively to previously deidentified banked samples, in which case identity protection relies upon data users refraining from reidentification. Pursuant to the acknowledged privacy risks, health systems and researchers increasingly provide broad notifications of deidentified data use practices or seek broad consent for use of deidentified data and tissue. Yet, while informed consent may mitigate liability should patient identity be compromised, privacy itself is not better protected. Broadly opting out of secondary data or tissue use may not be feasible in our current systems or desirable for long-term LHS goals. Practically speaking, most users of deidentified datasets may be solely interested in what

they can learn from the data, rather than the application of that knowledge to respective individuals, and may therefore have no incentive to reidentify subjects.

Nevertheless, concerns about reidentification have spurred efforts to improve privacy preservation via more advanced techniques for concealing data subject identities. Anonymization is a technically challenging option, which produces a dataset that is much less likely to permit reidentification. Privacy risks are minimized, but there are recognized compromises in data quality and value that limit prospective benefits.^{21,22} Alternate methods of protecting privacy, such as encryption, pseudonymization or use of “honest brokers,”²³ could enable learning without compromising data quality. Despite advantages for privacy, the increased cost, complexity and compromises of these techniques disincentivize implementation absent a concrete obligation to exceed existing regulatory requirements. Such efforts may be onerous for investigators, and unlikely to be widely utilized given the context of a regulatory safe harbor intended to ease barriers to secondary data use. By comparison, traditional deidentification, in which pre-set identifiers are removed from datasets before distribution and use, is quick and easy.

Today, secondary use of deidentified data and tissue is broadly accepted: a normative background process occurring in parallel with clinical care. Methods that rely on removing or obscuring underlying identifying information functionally impoverish datasets.^{24,25} Such practices may hinder continuous, global assessment of the data landscape, potentially reinforcing knowledge silos, preventing longitudinal engagement and thereby limiting scientific progress.²⁶ Researchers utilize deidentified data for ease of access and minimization of oversight, however, removal of key identifiers makes datasets incomplete, and compromises their value for learning. Preventing reidentification and contact may also frustrate the seamless integration of new knowledge into practice. A veil of deidentification between research and care may be suboptimal from the perspective of an LHS which seeks to maximize both learning and translational value.

In addition to challenges for identity protection and compromises for learning, patients have indicated how deidentification only partially protects the privacy of sensitive data and tissue.²⁷ Bodily privacy and dignity may be compromised when these data and tissue are shared, even if identity is never exposed. Consider how the patients in Figure 1 may feel if their deidentified images are distributed without their explicit knowledge or consent. Though their identities are concealed, they are still recognizable as individuals, similar to a deidentified medical record. This thought experiment illustrates the potential for violation of bodily privacy that is unmitigated by maintaining the privacy of identity or broad consent for use of deidentified data.

Protecting patients in research and iteration of clinical practice requires oversight and peer review: planned inefficiencies that ensure ethical treatment and optimal outcomes.¹⁵ In reality, inefficiency of translation far exceeds beneficent intentions.²⁸ Deidentification promotes learning by reducing friction for data use but induces a compromise in the quality of learning and efficiency of translation. LHS seeks to realize precision medicine as the standard of care. This requires

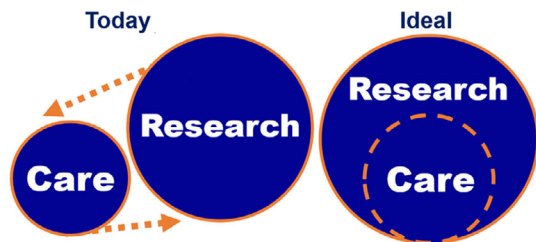


FIGURE 2 Efficiency & effectiveness of learning health system design

harmonization of research and clinical activities which may not be optimized by deidentification, particularly as actionable findings are increasingly produced (Figure 2). Forgoing IRB oversight and obligations to benefit patients from research on their own deidentified data and tissue may not maximize benefits of LHS for individuals or society, especially given the availability of alternative privacy-preserving technologies that may render deidentification obsolete.²⁹

1.3 | Distributive justice in LHS design

In the Belmont Report,⁸ justice is concerned with identifying a fair distribution of research benefits and burdens. The Report gives special weight to both potential benefits and risks for immediate research subjects. Accordingly, “an injustice occurs when some benefit to which a person is entitled is denied without good reason or when some burden is imposed unduly.”⁸ Thus, the potential for an individual to benefit from their research participation factors into considerations of justice, although the Report does not require patients to benefit directly from their research contributions given expected delays between learning and translation.³⁰ Traditionally, the vision entailed learning from current patients for the benefit of future patients: a process that is beneficial and fair to patients overall.

Distributive justice in an LHS is founded on a compact, which asserts that patients are obligated to contribute to learning as a condition of receiving care, and the health system is responsible for making continuous improvements.² LHS implementation begins with all patients inheriting an immediate and inescapable duty to contribute their data and tissue to learning for the sake of future patient-facing benefits. The health system simultaneously adopts an obligation to translate resulting knowledge for the benefit of patients overall, though the fulfillment of the duty is neither immediate nor enforceable. There are no predetermined metrics to ascertain the efficiency or effectiveness of the learning-translation cycle, and the potential for specific patients' care to improve as a result of their research contributions is not addressed through this mechanism.

Safe harbor aims to unlock the societal benefits of learning from data by reducing both risks for individuals and burdens for researchers. Deidentification ensures that patients fulfill their duty to contribute to learning as no one is exempt from secondary use of their deidentified data. The minimal risks and passive nature of participation in data-driven research reduce the concerns regarding burdens

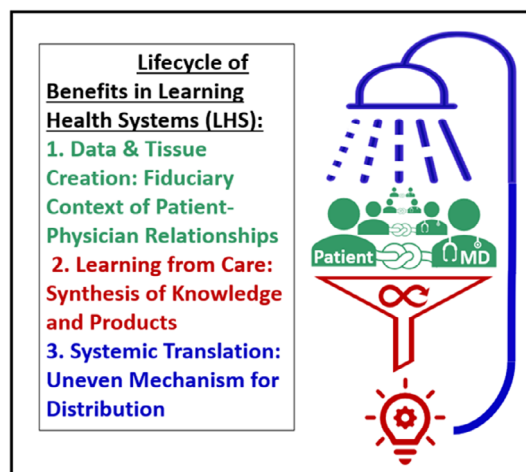


FIGURE 3 Lifecycle of benefit in learning health system (LHS)

for individual subjects. Universal inclusion of patients in data-driven research may therefore alleviate justice concerns regarding equitable subject selection, though it does not ensure equitable distribution of research benefits. The core ethical challenge for LHS emerges as patients are obligated to contribute, though there are no assurances that the health system will justly distribute the benefits of learning, either with respect to individuals or society.

The Belmont Report stipulates that, for therapeutic devices and procedures developed via government-funded research, “justice demands both that these not provide advantages only to those who can afford them and that such research should not unduly involve persons from groups unlikely to be among the beneficiaries of subsequent applications of the research.”⁸ Given the lack of universal health care and disparities in access to cutting-edge treatment for U.S. populations,^{7,28} underserved individuals may experience a less favorable ratio of benefits to burdens for research on deidentified data. Safe harbor may contribute to disparities by producing generalizable knowledge from care without mechanisms to ensure equal access to resulting benefits. Further, more advanced disease creates greater learning opportunities, suggesting that patients whose presentations are delayed due to limited access to care may contribute disproportionately to knowledge advancement. The historical pattern in which “the burdens of serving as research subjects fell largely upon poor ward patients, while the benefits of improved medical care flowed primarily to private patients⁴” may be replicated in an LHS that obligates all patients to contribute to learning but does not maintain the provenance of duty to benefit individuals from their specific contributions.

Studies that produce results, which are most likely to be clinically actionable for deidentified subjects³¹ may raise the most distributive justice concerns should deidentification present legal or practical barriers for translation.^{32,33} Figure 3 illustrates the flow of value in an LHS which systematically leverages clinical data and tissue as byproducts of care with an infrastructure, which may allow a disproportionate flow of benefits to advantaged institutions and individuals. When all patients are subjects, justice may require further attention to ensuring universal access to the fruits of the research enterprise. When deidentified

patient data and tissue are developed through collaboration with private corporations, the absence of obligations to ensure universal access to commercial products may permit a stark disparity between underserved individuals' contributions to and benefits from the LHS.

Safe harbor for deidentified data presumes that patients as a whole benefit from advancements in generalizable knowledge and health technology. Yet, since the mass digitization of U.S. health records and ensuing secondary use of deidentified data, health inequities and mortality have increased.^{7,30} Routine deidentification of patient data and tissue for administrative convenience during secondary research may contribute to this discordance.

First, safe harbor as a structural mechanism for learning from care does not address disparities in its risk/benefit calculus. Second, by accelerating learning processes in the absence of a mechanism for ensuring a just distribution of benefits, asymmetric translation may be exaggerated. Third, deidentification could hinder our ability to recognize existing or new disparities affecting LHS patients by shielding researchers from identifiers that are essential for characterizing marginalized populations. Further, universal inclusion of deidentified subjects and lack of IRB oversight may allow researchers to focus on populations of interest or convenience at their own discretion, potentially enabling an unfair distribution of resources and attention to health issues affecting privileged groups.

Barriers to direct translation of research results may manifest in knowledge asymmetries across the LHS, particularly as precision medicine and other next-generation research findings are relevant to individuals. Prohibiting reidentification and contact in the event of clinically actionable, potentially timely results may be disproportionately burdensome for underserved individuals. The most appropriate scheme for justly distributing benefits to LHS patients remains undefined and may vary based on context. However, both what is owed to specific individuals and how our LHS design helps mitigate preexisting structural injustices must be priorities for a system that aspires to maximize efficiency, effectiveness and justice in its integration of research and care. The unprecedented trend in decreasing life expectancy for the U.S. population that has precipitated in tandem with advancing platform-based learning may reflect an underlying structural injustice in our LHS.³⁴ Despite its administrative convenience, safe harbor for deidentified data may not produce a ratio of benefits and burdens that optimally serves patients as it decouples learning from any requirement to share what is learned for the benefit of those who rendered it possible.

2 | LIMITATIONS AND FUTURE DIRECTIONS

Patients are subjects by design in an LHS. The corresponding ethical framework seeks to synthesize the duties of research and care. Safe harbor for deidentified data is a prominent strategy for integrating these adjacent domains. We have highlighted potential challenges for using deidentification to facilitate learning from care vis-a-vis respect, beneficence and justice. Yet, deidentification practices embedded in current ethical, socioeconomic, technological and legal norms may present

seemingly insurmountable barriers to reimagining the status quo. Significant normative and empirical considerations must be addressed if an alternative to deidentification is sought for the future of LHS.

Deidentification is a particular manner of reconciling clinical and research ethics. However, the optimal approach to respect patients as subjects in an LHS remains undetermined, and potentially varies substantially across settings. Contemporary data-driven research and the vast amount of health data created in clinical and non-clinical contexts have complicated our reliance on informed consent as the primary means of respecting patients as subjects. Traditional notions of consent and associated voluntariness of research participation do not comport with the LHS obligation for patients to contribute their data and tissue to systemic learning. Instead, transparency and engagement become key for centering patients as persons and protecting their best interests in these instances of diminished autonomy regarding research participation. We propose an exploration of a "reasonable patient" standard that may be useful for guiding intuitions regarding ideal transparency and engagement at various points throughout the LHS lifecycle.

New models and technologies may be necessary for implementing LHS values within the digitally integrated care-research setting. Challenging deidentification for an LHS may prompt reconsideration of how we conceptually and practically integrate various learning activities into care. Whereas current LHS models may focus on benefits to society and risks to individuals, an approach grounded in patient care may be necessary to balance benefits for individuals and risks for society. More work is needed to develop the conceptual basis of novel roles of patient-qua-subject and physician-qua-researcher, and how to interpret the fiduciary duties thereby attached. For example, the standard view that publication is researchers' only responsibility for ensuring translation of their research into practice needs to be interrogated within this LHS framework.¹⁰ Development of pathways to enable timely disclosure of clinically actionable findings to deidentified patients and their physicians must be prioritized, including incorporation into funding mechanisms.

Socializing LHS values is critical for implementation. Normalization of secondary use of deidentified data and tissue for general learning without specific duties to benefit respective individuals supports safe harbor as a fixture of the LHS infrastructure. The true extent of safe harbor's social acceptability is unknown, however, as secondary uses of clinical data and tissue may not be fully appreciated by most patients. Broad but discrete disclosures regarding ongoing secondary use of deidentified data or tissue may be insufficient for transparency and engagement, especially in light of normative expectations of medical privacy and the primacy of clinical concerns in the patient experience. Centering patients and physicians as stakeholders at the nexus of both learning and translation will be essential for the evolution of the culture surrounding simultaneous learning and care. Further investigations should explore how to best deliver a transparent, engaged model of learning that works with, not merely about, patients, apropos current technology and pace of learning within the broader social context in which citizens are data subjects.

Elsewhere, we have suggested that innovations in information technology, like blockchain and privacy-preserving computation, could

inform and may be central to ethical LHS design.^{35,36} Holistically, these innovations have the potential to minimize tensions between data privacy and utility, and they may provide enhanced trust, transparency and security of health data supply chains. Substantial investment will be required to optimize and scale these technologies for LHS. As with other technical solutions (eg, encryption for privacy preservation,³⁷), a new standard of care and compliance enforcement may be required for adoption. Development of automated processes for embedding the return of results into the LHS architecture may be essential for feasibility. Non-trivial challenges remain for engineering and deployment of a system that optimizes privacy, learning and translation.

By obligating patients to contribute to learning as a condition of care, the ethical framework of LHS implies that research participation is a civic duty rather than a voluntary act. Legal protections of patients and subjects may need to be updated accordingly to ensure a just distribution of benefits and burdens. Safety regulations and liability reform should also address new risks related to the rapid iteration of clinical practice and potential vulnerabilities of a new era in which learning at scale aims to inform precision insights rather than generalizable knowledge. Safe harbor for deidentified data traverses several major US statutes and corresponding regulations, however, its elimination for the sake of advancing an ethical LHS may come at too high a cost. Increased IRB oversight for deidentified data may be necessary, and treating all research on patient data as human subjects research may be an appropriate next step. Legal alternatives to safe harbor may include an extension of covered entity status with enhanced security requirements for identified datasets or updated criteria for deidentification under HIPAA and refinement of broad consent or enhanced benefit distribution requirements for deidentified subjects under the Common Rule.

We focus on U.S. policy, however, the ethics of learning from deidentified patients must be considered for LHS worldwide. For example, countries with universal health care and other regulatory and cultural norms may alter the calculus of risks and benefits. Examination of E.U. experiences regarding implementation of informed consent and data protection mandated by GDPR may inform U.S. policy development.³⁸ The advancement of international ethical standards for health data is a priority, particularly as the ultimate vision of LHS is a global phenomenon and the potential exploitation of patients-quasi-subjects in poorer nations may be of particular concern.³⁹

One important function of deidentification is to facilitate the use of clinical care to accelerate the development of health technologies.⁴⁰ Commercial uses of deidentified and broadly consented data or tissue under the 21st Century Cures Act⁴ may be bolstered by comprehensive legislation to ensure that these publicly funded, crowd-sourced advances are available to patients regardless of payer status. Legal standards surrounding data ownership and intellectual property are priorities for further study. The next steps may include advancement of mechanisms to ascertain the effectiveness and distributive justice regarding the systemic use of data and tissue to advance collective interests. Further work may also prioritize engagement of diverse audiences on the view of data and tissue produced during clinical care as an asset to be invested, rather than a byproduct to be donated.

3 | CONCLUSIONS

Safe harbor for deidentified data facilitates learning, but it is not without tradeoffs. Fiduciary duties to benefit patients are the premise of clinical data and tissue collection.

By separating systemic secondary use of these assets from their primary purpose, deidentification may not adequately respect patients as subjects. Further, deidentification may not sufficiently protect privacy or maximize the efficiency and effectiveness of learning from care at scale. Policies restricting reidentification and contact appear to hinder the direct implementation of individually relevant insights. LHS architecture relying on deidentification may not optimally distribute benefits and burdens, potentially exaggerating existing disparities. A structurally just LHS has a duty to ensure equitable translation, and learning infrastructure must address the potential for the asymmetry between individuals' contributions and benefits. Despite its established advantages, deidentification may not optimize respect, beneficence or justice for patients and subjects-- and it may no longer be necessary. Novel technological approaches could improve privacy protection as well as the utility and value of learning from care. Ethical and legal alternatives to current safe harbor policies may be essential for maximizing efficiency, effectiveness and justice for patients as contributors to and beneficiaries of learning in LHS of the future.

ACKNOWLEDGEMENTS

We would like to acknowledge the vital contributions of The Scar Project, including photographer David Jay, along with Shay S, Sara M, and Eliza H, whose portraits (see Figure 1) powerfully illustrate how identification, not deidentification, maximizes respect, justice and utility of fundamentally non-fungible clinical data and tissue.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Marielle S. Gross  <https://orcid.org/0000-0002-3009-4082>

Amelia J. Hood  <https://orcid.org/0000-0002-9702-7693>

REFERENCES

1. U.S. Dept. of Labor, *Employee Benefits Security Administration*. The Health Insurance Portability and Accountability Act (HIPAA). 1996.
2. Dept of Health and Human Services. 2017. Federal Policy for the Protection of Human Subjects. 6 CFR § 46. Accessed August 20, 2021. <https://www.govinfo.gov/content/pkg/FR-2017-01-19/pdf/2017-01058.pdf>
3. Health Information Technology for Economic and Clinical Health (HITECH) Act, Title XIII of Division A and Title IV of Division B of the American Recovery and Reinvestment Act of 2009 (ARRA). 2009. Accessed January 21, 2022. <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/hitechact.pdf>
4. 21st Century Cures Act, Pub. L. No. 114-255. 2016. Accessed January 21, 2022. <https://www.govinfo.gov/app/details/PLAW-114publ255>
5. Kass NE, Faden RR, Goodman SN, Pronovost P, Tunis S, Beauchamp TL. The research-treatment distinction: a problematic approach for determining which activities should have ethical oversight. *Hastings Center Rep*. 2013;43(s1):S4-S15. doi:10.1002/hast.133

6. Faden, Ruth R., Nancy E. Kass, Steven N. Goodman, Peter Pronovost, Sean Tunis, and Tom L. Beauchamp. 2013. "An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics." *Hast Center Special Rep Ethic Oversight Learn Health Care Syst*, 43 (1): S16–S27. [10.1002/hast.134](https://doi.org/10.1002/hast.134)
7. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *New England J Med*. 2003; 348(26):2635-2645. doi:[10.1056/NEJMs022615](https://doi.org/10.1056/NEJMs022615)
8. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. "The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research." US Dept of Health & Human Services, Office of Human Research Protections. Accessed August 20, 2021. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>.
9. Bernabe RDLC, van Thiel GJM, Raaijmakers JAM, van Delden JJM. The fiduciary obligation of the physician-researcher in phase IV trials. *BMC Med Ethics*. 2014;15(1):11. doi:[10.1186/1472-6939-15-11](https://doi.org/10.1186/1472-6939-15-11)
10. Kass NE, Faden RR. Ethics and learning health care: the essential roles of engagement, transparency, and accountability. *Learn Health Syst*. 2018;2(4):e10066. doi:[10.1002/lrh2.10066](https://doi.org/10.1002/lrh2.10066)
11. Hudson KL, Collins FS. Bringing the common rule into the 21st century. *N-Perspective Massachusetts Medical Society*. World. December 9, 2015;373:2293-2296. [10.1056/NEJMp1512205](https://doi.org/10.1056/NEJMp1512205)
12. Sterling RL. Genetic research among the Havasupai: a cautionary tale. *AMA J Ethics*. 2011;13(2):113-117. doi:[10.1001/virtualmentor.2011.13.2.hlav1-1102](https://doi.org/10.1001/virtualmentor.2011.13.2.hlav1-1102)
13. Jaggi R, Griffith KA, Sabolch A, et al. Perspectives of patients with cancer on the ethics of rapid-learning health systems. *J Clin Oncol*. 2017;35(20):2315-2323. doi:[10.1200/JCO.2016.72.0284](https://doi.org/10.1200/JCO.2016.72.0284)
14. Jones RD, Krenz C, Griffith KA, et al. Governance of a learning health care system for oncology: patient recommendations. *JCO Oncol Pract*. 2021;17(4):e479–e489. doi:[10.1200/OP.20.00454](https://doi.org/10.1200/OP.20.00454)
15. US Department of Health and Human Services Office for Civil Rights. 2010. "1.3 De-Identification and Its Rationale." In Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Accessed August 20, 2021. <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#top>.
16. Faden RR, Beauchamp TL, Kass NE. Learning health care systems and justice. *Hastings Center Rep*. 2011;41(4):3-3.
17. Miller K. De-Identifying Medical Patient Data Doesn't Protect Our Privacy. Stanford HAI Published July 19, 2021. Accessed August 20, 2021. <https://hai.stanford.edu/news/de-identifying-medical-patient-data-doesnt-protect-our-privacy>
18. Na L, Yang C, Lo C-C, Zhao F, Fukuoka Y, Aswani A. 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 Netw Open*. 2018;1(8):e186040–e186040. doi:[10.1001/jamanetworkopen.2018.6040](https://doi.org/10.1001/jamanetworkopen.2018.6040)
19. Wakabayashi D. Google and the University of Chicago are sued over data sharing. *The New York Times*; 2019 June 26, 2019, sec. Dent Tech. Accessed August 20, 2021. <https://www.nytimes.com/2019/06/26/technology/google-university-chicago-data-sharing-lawsuit.html>
20. Angrist M. Eyes wide open: the personal genome project, citizen science and veracity in informed consent. *Pers Med*. 2009;6(6):691-699. doi:[10.2217/pme.09.48](https://doi.org/10.2217/pme.09.48)
21. Emam E, Khaled SR, Malin B. Anonymising and sharing individual patient data. *The BMJ*. 2015;350:h1139. doi:[10.1136/bmj.h1139](https://doi.org/10.1136/bmj.h1139)
22. Ploem MC, Essink-Bot ML, Stronks K. Proposed EU data protection regulation is a threat to medical research. *BMJ*. 2013;346:f3534. doi:[10.1136/bmj.f3534](https://doi.org/10.1136/bmj.f3534)
23. Dhir R, Patel AA, Winters S, et al. A multidisciplinary approach to honest broker services for tissue banks and clinical data. *Cancer*. 2008; 113(7):1705-1715. doi:[10.1002/cncr.23768](https://doi.org/10.1002/cncr.23768)
24. Mathews DJH, Jamal L. Revisiting respect for persons in genomic research. *Genes*. 2014;5(1):1-12. doi:[10.3390/genes5010001](https://doi.org/10.3390/genes5010001)
25. SACHRP Committee. 2016. "Summary Table of Recommendations on the HIPAA Privacy Rule." US Dept of Health & Human Services, Office of Human Research Protections, Secretary's Advisory Committee on Human Research Protections. Accessed August 20, 2021. <https://www.hhs.gov/ohrp/sachrp-committee/recommendations/2004-september-27-letter-summary/index.html>.
26. Lund H, Brunnhuber K, Juhl C, et al. Towards evidence based research. *BMJ*. 2016;355:i5440. doi:[10.1136/bmj.i5440](https://doi.org/10.1136/bmj.i5440)
27. Mathews DJH, Rabin JT, Quain K, et al. Secondary use of patient tissue in cancer biobanks. *Oncologist*. 2019;24(12):1577-1583. doi:[10.1634/theoncologist.2018-0376](https://doi.org/10.1634/theoncologist.2018-0376)
28. Fernandez-Moure JS. Lost in translation: the gap in scientific advancements and clinical application. *Front Bioeng Biotechnol*. 2016; 4:43. doi:[10.3389/fbioe.2016.00043](https://doi.org/10.3389/fbioe.2016.00043)
29. Kocabaş Ö, Soyata T. Medical data analytics in the cloud using homomorphic encryption. *Handbook of Research on Cloud Infrastructures for Big Data Analytics*. IGI Global; 2014 Accessed April 24, 2021. <https://www.igi-global.com/gateway/chapter/103226>
30. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med*. 2011;104(12):510-520.
31. "Our Programs." 2016. Patient-Centered Outcomes Research Institute. June 28, 2016. Accessed April 24, 2021. <https://www.pcori.org/about-us/our-programs>.
32. Strickland, Eliza. 2015. Big Data Beats Cancer. *IEEE Spectrum: Technology, Engineering, and Science News*. Accessed April 24, 2021. <https://spectrum.ieee.org/biomedical/diagnostics/big-data-beats-cancer>
33. How REMAP Is Being Used in COVID-19 Fight. UPMC HealthBeat 2020. Accessed April 24, 2021. <https://share.upmc.com/2020/12/understanding-upmcs-remap-project-and-the-future-of-clinical-trials/>
34. Murphy SL, Xu J, Kochanek KD, Arias E. Mortality in the United States, 2017. National Center for Health Statistics; 2018. Accessed August 20, 2021. <https://www.cdc.gov/nchs/products/databriefs/db328.htm>
35. Gross M, Miller RC. Ethical implementation of the learning healthcare system with Blockchain technology. *BHTY*. 2020;2:2. doi:[10.30953/bhty.v2.113](https://doi.org/10.30953/bhty.v2.113)
36. Gross MS, Hood AJ, Miller Jr, RC. Nonfungible Tokens as a Blockchain Solution to Ethical Challenges for the Secondary Use of Biospecimens: Viewpoint. *JMIR Bioinformatics and Biotechnology*. 2021;2(1):e29905. doi:[10.2196/29905](https://doi.org/10.2196/29905)
37. Fast Healthcare Interoperability Resources® (FHIR), eCQI Resource Center. eCQI Resource Center. n.d. Accessed April 24, 2021. <https://ecqi.healthit.gov/fhir>
38. Hoofnagle CJ, van der Sloot B, Borgesius FZ. The European Union general data protection regulation: what it is and what it means. *Information & Communications Technology Law*. 2019;28(1):65-98. doi:[10.1080/13600834.2019.1573501](https://doi.org/10.1080/13600834.2019.1573501)
39. Fernando B, King M, Sumathipala A. Advancing good governance in data sharing and biobanking - international aspects. *Wellcome Open Res*. 2019;4:184. doi:[10.12688/wellcomeopenres.15540.1](https://doi.org/10.12688/wellcomeopenres.15540.1)
40. FDA. 21st Century Cures Act [Internet]. FDA Office of the Commissioner. FDA; 2020 [cited 2021 Jul 9]. Available from: <https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/21st-century-cures-act>

How to cite this article: Gross MS, Hood AJ, Rubin JC, Miller RC Jr.. Respect, justice and learning are limited when patients are deidentified data subjects. *Learn Health Sys*. 2022; 6(3):e10303. doi:[10.1002/lrh2.10303](https://doi.org/10.1002/lrh2.10303)