Abstract

IMPORTANCE Digital technology is part of everyday life. Digital interactions generate large amounts of data that can reveal information about the health of individual consumers (the digital health footprint).

OBJECTIVE To describe health privacy challenges associated with digital technology.

DESIGN, SETTING, AND PARTICIPANTS For this qualitative study, in-depth, semistructured, qualitative interviews were conducted with 26 key experts from diverse fields in the US between January 1 and July 31, 2018. Open-ended questions and hypothetical scenarios were used to identify sources of digital information that contribute to consumers’ health-relevant digital footprints and challenges for health privacy. Participants also completed a survey instrument on which they rated the health relatedness of digital data sources.

MAIN OUTCOMES AND MEASURES Health policy challenges associated with digital technology based on qualitative responses to expert interviews.

RESULTS Although experts’ ratings of digital data sources suggested a possible distinction between health and nonhealth data, qualitative interviews uniformly indicated that all data can be health data, particularly when aggregated across sources and time. Five key characteristics of the digital health footprint were associated with health privacy policy challenges: invisibility (people are unaware of how their data are tracked), inaccuracy (data in the digital health footprint can be inaccurate), immortality (data have no expiration date and are aggregated over time), marketability (data have immense commercial value and are frequently bought and sold), and identifiability (individuals can be readily reidentified and anonymity is nearly impossible to achieve). There are virtually no regulatory structures in the US to protect health privacy in the context of the digital health footprint.

CONCLUSIONS AND RELEVANCE The findings suggest that a sector-specific approach to digital technology privacy in the US may be associated with inadequate health privacy protections.

Key Points

Question What challenges for health privacy are associated with digital technology?

Findings In this qualitative study, 5 key challenges for health privacy were associated with digital technology: invisibility (people unaware of how they are tracked), inaccuracy (flawed data), immortality (data never expire), marketability (data are frequently bought and sold), and identifiability (individuals can be readily reidentified).

Meaning The findings suggest that a sector-specific approach to digital technology privacy in the US may be associated with inadequate health privacy protections.

Introduction

By June 2019, there were more than 4.4 billion internet users, an 83% increase in 5 years.1 More than half of the global population uses email.2 Seventy-three percent of Americans frequently access their bank accounts online.3 Every minute on Facebook, 510,000 comments are posted, 293,000 statuses are updated, and 136,000 photographs are uploaded.4 Digital interactions are obligatory, with central roles at home and work, resulting in a recorded stream of personal information.

Digital interactions, using mobile applications, searching the Internet, wearing connected devices, or conversing on social media, often generate health-relevant information. Smart watches and smartphone applications are in widespread use for tracking physical activity, fertility, and blood

Open Access. This is an open access article distributed under the terms of the CC-BY License.
glucose levels. One step removed, data scientists have been able to identify chronic disease risk or depressed mood based on Internet searches and social media posts. Cars now record whether their drivers have gained or lost weight or strayed from their lanes. We refer to the sum of these health-relevant data as an individual’s digital health footprint.

Although the European Union implemented broad new consumer digital privacy regulations in 2018, the US has not adopted a comprehensive regulatory approach. Instead, the US has taken a sector-specific approach, with differential protections conferred on health care encounter data through the Health Insurance Portability and Accountability Act (HIPAA). In addition, genetic information, thought to be particularly sensitive, receives protections under the Genetic Information Nondiscrimination Act (GINA). These regulations leave wide swaths of digital consumer privacy unregulated.

Some privacy risks are attributable to illegal hacking. Those risks are managed by security systems and law enforcement. However, many of the challenges to health privacy are associated with data practices that are currently legal. We explored the privacy challenges associated with the digital health footprint through interviews with multidisciplinary experts. Those interviews informed a framework for considering the genesis, transformation, and application of the digital health footprint as well as challenging characteristics of the digital health footprint that may require policy attention.

Methods

Participants

For this qualitative study, we conducted interviews between January 1 and July 31, 2018, using purposive and convenience sampling to recruit 26 participants with diverse expertise in emerging digital technology and applications to health care and research. Specific areas of expertise included applications of digital technology to health (n = 12), data analytics and data mining (n = 12), health care innovation and business (n = 9), consumer behavior and preferences (n = 9), marketing (n = 7), health policy (n = 7), computer science (n = 7), privacy law (n = 4), ethics (n = 3), data security (n = 3), consumer advocacy (n = 3), and machine learning (n = 3) (categories not mutually exclusive). Experts were drawn from a range of sources, including national and international privacy committees and commissions and related research publications, and through convenience strategies beginning with our project advisory committee. Interview participants were compensated with $200. This study was reviewed and declared exempt by the institutional review board at the University of Pennsylvania. We were granted a waiver of written informed consent but obtained verbal informed consent from participants before their interview. All data were deidentified. This study followed the Standards for Reporting Qualitative Research (SRQR) reporting guideline.

Design

Interviews were conducted using in-depth, semistructured, qualitative methods. The interview guide (eAppendix in the Supplement) was informed by a consequential ethics framework in which the presence or absence of a substantial risk of harm associated with a loss of privacy determines the need for protections. Through open-ended questions and hypothetical scenarios, we asked experts to identify current and emerging sources of digital information from outside health care that contribute to consumers’ health-relevant digital footprints. We also asked them to describe current and potential future applications of that information, anticipate potential harms and benefits, and consider approaches to addressing privacy concerns. The 20- to 60-minute interviews were audio-recorded and conducted over the telephone or in person by a trained research coordinator (A.L.). After each interview, a web-based follow-up questionnaire was sent to the participant. The questionnaire included a list of data sources, and experts were asked to rate the health relatedness, potential harm (ie, if disclosed), and potential benefit (ie, to individuals or society) on a scale of 0 to 100, with 0 being the least and 100 being the most. Interviews were recorded, transcribed by a
professional transcription service, and deidentified before being uploaded to NVivo, version 12 (QSR International) for analysis.\textsuperscript{18}

**Statistical Analysis**

The study team developed a codebook through line-by-line, iterative reading and notation of transcripts, which produced 12 key categories.\textsuperscript{19} Two research coordinators (X.L.M., A.L.) trained in qualitative data analysis used the codebook to complete coding. To establish agreement, 14 interview transcripts were double coded. Interrater reliability was measured using percent agreement (96.5\%) and the Cohen $\kappa$ (0.68). After agreement was established, the researchers individually coded the remaining transcripts, which were then summarized in memos that were reviewed and discussed by the study team to identify patterns and synthesize cross-cutting themes. The results are reported thematically, with supporting quotes, to distill the most salient challenges for health policy.

**Results**

A total of 26 experts were interviewed. The interviews informed a conceptual framework for understanding the digital health footprint and identifying potential leverage points for regulation or policy action. They also revealed 3 key themes: (1) the digital ecosystem offers no clear distinction between health and nonhealth information, (2) key characteristics of the digital footprint merit policy attention, and (3) few regulatory structures currently protect consumer privacy.

**Conceptual Framework**

Consumers’ everyday activities generate the digital health footprint, which is routinely aggregated, transferred (or commodified), and applied in a range of settings, including health care, business, and research. Figure 1 synthesizes the dynamic formation and transformation of the digital health footprint, as described by experts in this study.

![Figure 1. Findings From Interviews With Key Experts](https://jamanetwork.com/)

This figure shows the steps leading to the creation of the digital health footprint (including data generation and aggregation) and subsequent applications of digital data by a range of users across different sectors. This entire continuum is affected by overarching environmental factors that are associated with the collection and use of digital health data and the absence and/or presence of oversight and regulation.
Within the digital ecosystem, social, economic, and governmental norms, practices, and policies are the main contributors to the increasing reliance on digital technologies for core life tasks, including pragmatic economic transactions (eg, banking), governmental functions (eg, tax filing), and social exchanges (eg, texting). Together, these data-generating activities contribute to the digital health footprint, which is a person-specific, dynamically evolving collection of digital information that can be used to infer current health states or estimate future health states. Consumer data are aggregated or linked across platforms, allowing for more nuanced inferences about health than can be derived from any single data source.

Experts projected that, with the potential for improved estimation of health states, the digital health footprint will become more commercially valuable (eg, as health care systems increasingly rely on predictive analytics to manage patient care). Digital health footprints are being transferred from their original custodians to new commercial (and other) entities for applications unrelated to the original purpose of the data collected. Experts emphasized that the regulatory landscape attends almost exclusively to the ethical and legal concerns arising from electronic health records and genetic testing. However, wide-ranging information originating beyond the protected domains of health care and genetic testing contributes to the digital health footprint, with limited regulatory oversight or agreement regarding best practices.

**Distinction Between Health and Nonhealth Information**

Experts rated distinct information streams that contribute to the digital health footprint (Figure 2) on a scale from 0 (not at all health related) to 100 (highly health related). They assigned highly variable scores to different information streams, with high scores for the electronic health record (median score, 100; IQR, 85-100), followed by fitness trackers (median score, 72.5; IQR, 52.5-80.0). Lower health-relatedness scores were assigned to commercial genetic profiles (median score, 60; IQR, 35-80).

## Figure 2. Responses of Experts Regarding the Health Relatedness of Digital Data Sources

Experts were asked to rate the health relatedness of sources of digital data on a scale of 0 to 100, with 0 being the least and 100 being the most. These results are presented with quotes that demonstrate that interviewed experts were not able to draw distinctions between health and nonhealth data. A1C indicates glycated hemoglobin; GPS, Global Positioning System.
50-75), toll-tracking devices (median score, 10; IQR, 5-20), and frequent flyer accounts (median score, 7.50; IQR, 2.75-10.00).

The experts uniformly indicated that there are no clear distinctions between health-related and non-health-related data (Figure 2) and that all data can become health data. They noted that data are routinely aggregated across domains and over time, allowing for additional predictive analytics and increasingly precise characterization of health or risks. As one expert noted, “We’re moving from a time where health was measured directly using clinical measures to a new era where health is measured indirectly using...all the available information we leak on a daily basis.”

Experts summarized that “the line between just general digital data and health data is going to become so blurred...and the regulations aren’t going to catch up,” potentially introducing the risk of “discrimination based off of just 1 or 2 streams of information.”

Key Characteristics of the Digital Health Footprint

The experts identified 5 characteristics of the digital health footprint that may be associated with threats to consumer privacy (Table). The first characteristic was invisibility. An expert noted, “It would be very, very odd if someone followed you around...making notes of everywhere you went...and how much money you spent and what you saw and who you interacted with...We would call it stalking, right? But in the digital world, that’s just common behavior.”

A refrain was that consumers are largely unaware of how and where their data are being tracked, used, and sold. In addition, consumers are fundamentally denied the opportunity to opt out of passive data collection (eg, surveillance cameras and facial recognition). The invisibility of the digital health footprint may contribute to low levels of consumer vigilance, especially in the context of unwieldy privacy policies.

Another identified characteristic was inaccuracy. An expert stated, “If you don’t take that information seriously and you think, ‘Oh, it’s just some quiz,’ and maybe you just randomly answer some [joking] response, it might still stick with you.”

The experts cautioned that the digital health footprint can generate inaccurate inferences because machines are literal in their data interpretation. Thus, the digital health footprint may contain ambiguous information about health behaviors or the social determinants of health. For example, a location tracker could note a visit to a clinic with abortion services, which may incorrectly signal that the person had an abortion. A subset of inaccuracy is the inadvertent bystander effect in which data from neighbors, friends, and social network members may be used to infer a person’s own behaviors, which may or may not be concordant. Experts raised concern that consumers have limited control to correct inaccuracies in their digital records.

The third characteristic was immortality of data. An expert noted, “Say I build a wellness app and I ask you to fill out extensive surveys about yourself...It gets mildly popular and I get an acquisition off of it...the data is going to go into a data broker and get endlessly resold and segmented.”

Experts were concerned that an infinite lifetime for health-relevant digital data presents a high risk of potential misuse, exposure, or other breaches. A small risk, sustained during a long period, translates into a high risk of an adverse event. Experts cautioned that consumers should (but usually do not) have opportunities to review and destroy their own personal data, including data they perceive to be potentially damaging.

Marketability was a fourth characteristic identified through the expert interviews. One participant mentioned, “There’s a lot of questions about whether it’s right that companies are selling people’s individual consumer data and then the buyer of data turns it into profitable products and the consumer never benefits from that in any way.”

Experts underscored that consumers’ digital information holds potential for scientific and clinical advances as well as for commercial gain, with a low likelihood of compensation to the people whose digital data are being traded and sold. Experts highlighted that there are few, if any, safeguards against exploitation and no established mechanisms for compensating the individuals...
who contribute to advances derived from digital health footprints. Moreover, commercialization opportunities may further motivate data collection and development of new applications.

The last characteristic was identifiability. An expert noted, "Eighty-five percent of people can be re-identified based on 3 GPS points; my home, my office, my children's school, narrow me down to a really small number of people that that could uniquely be."

Experts indicated that individuals can easily be re-identified through the merging of data streams, thus undermining promises of confidentiality. Identifiability may be used as a tool for screening and identification of risk. For example, aggregated data and improved algorithms may identify problematic or dangerous behavior, such as suicidality, before a consumer (or their health care practitioner) is aware. However, identifiability may also allow for unwanted targeting, for example, efforts to shape consumer opinions and behavior (eg, Cambridge Analytica's purchase of

| Theme       | Quotes                                                                 | No. (%) of respondents (N = 26) |
|-------------|------------------------------------------------------------------------|----------------------------------|
| Invisibility| "One of the problems has been the use of these kinds of data in conjunction with other data sets. And so, one of the precautions when a consumer agrees to share their particular data, not having—having the ability to control that the data doesn't get combined with other information about them. So taken by itself, the idea that I might buy chocolate isn't such a bad data point. Combining that with other things may point a particular kind of risk factors that I don't want other people knowing or understanding about me."
|             | "So that's kind of like when people get a little bit of the tip of the iceberg and it freaks them out. And behind that little pair of shoes following you around is this huge infrastructure, really, like this kind of hidden web that's bigger than the web you see, in a sense, and that is where all that information is getting collected..."
|             | "Well, part of it is that it's very opaque, right? Companies who own this data are not necessarily revealing how they're using it."
|             | 22 (85)                                                                |
| Inaccuracy  | "A close friend of mine had a type of cancer, and so I began doing some online searches that were focused on finding a physician to treat this person...about two weeks later, I'm on a national newspaper site, and I get an ad for a physician for treating this particular type of cancer pop up on the right-hand side of my page on the newspaper. It's like ick, like ick, because they have targeted that based on my searches."
|             | "We know that actually the data collected through these kind of devices is notoriously patchy in terms of its—well, certainly in terms of its accuracy and also in terms of its completeness, so I think findings taken from these kinds of data is—always needs to be taken with a pinch of salt."
|             | 7 (27)                                                                 |
| Immortality | "I think the way the systems are set up now, they allow for infinite selling and reuse and storage of this data for the targeted advertising."
|             | "So how long does this stuff get kept? And the answer is usually forever because hard drives that they store it on are really cheap. So we store it forever. I kinda think whatever problems that we may have with these types of information being collected, you kind of exponentially get worse and they exponentially accrue. So if there's data leaks or something, let's say, some corporation has a list of everybody's medical interests."
|             | 8 (31)                                                                 |
| Marketability| "I think the bigger impact is around the kind of political economy of these devices and the data which is generated, in that we have become creators of potentially valuable data for corporations in most cases."
|             | "How many times do you and I use Google all day long? A lot—10, 20, 30, 40 times. And it's free because you and I are the product. We are what's being sold... And I think most people don't really kind of exactly get that."
|             | 14 (54)                                                                |
| Identifiability| "Once data leaves that first party and starts to get matched up with other datasets where you're highly identifiable, then, now, you get—people often build these databases that are really, really, really powerful. So they know not only like, okay, here's all your ads, here's are all your social behavior, but here's your actual financial information or here's your medical information. And that becomes very tricky because then people build full models of who you are."
|             | "Well, it turned out [Strava] exposes the location of secret military bases in Afghanistan because it turns out military people like to do quantified self-stuff and didn't realize that their data was gonna reveal the precise patterns of their patrols. So they can literally be seen from space in the data. So those sorts of things are starting to happen and they are going to crest and accelerate as people realize that by connecting these things to the internet and that it becomes possible to assemble this data in ways that no one ever expected."
|             | "Lots of people have done quite credible research using public data sets and identifying individuals in these, and I think that's the problem here too. Even if you say, 'oh we use all this data anonymously,' by definition the power of combining all these data sets means you can identify people quite easily."
|             | 9 (35)                                                                 |
Facebook data to shape political opinions (e.g., hiring decisions or insurance pricing).

**Few Regulatory Structures for Consumer Privacy Protection**

One expert commented, “We have a lot of work to do in the US. We don’t currently have an omnibus privacy protection law...we’ve also got a technology environment that has allowed for a relatively Wild West approach to the use and sharing and re-use of personal data.” The experts consistently described current regulatory protections as limited and sector specific. They additionally noted a reliance on corporate and other entities to self-monitor and protect consumers’ interests.

**Discussion**

We identified 3 key findings. First, there are no clear distinctions between data that are and are not health related. Second, the digital health footprint is associated with enduring health privacy challenges that transcend specific technologies or applications. Third, the digital health footprint is largely unregulated. These findings may have implications for health privacy and policy.

Data scientists draw inferences about health from wide-ranging, routinely collected data. Facebook has assessed linguistic nuance to identify mental health problems, smart mattresses monitor sleep habits, and location tracking can identify individuals who visit abortion clinics. Beyond these focused applications, data brokers are now commodifying aggregated digital data to fuel predictive analytics that can be applied in different settings (e.g., health risk scores). The enactment of HIPAA and GINA reflected regulators’ intent to confer special status on health information and therefore heightened consumer protections. A key finding from this study—that all data are health data—suggests that the privacy protections of HIPAA and GINA are inadequate and obsolete. The views of the experts who we interviewed are in line with an increasing body of academic and lay literature on the relevance of digital information to health. In the current digital landscape, a multisectoral regulatory approach is necessary to protect consumers’ health privacy.

Several challenges of the digital health footprint may transcend evolving technologies, posing persistent questions for health policy. Policy lessons can be drawn from long-standing debates about genetic privacy because genetic data share these qualities. Ethicists have addressed the case of Henrietta Lacks, whose cervical cancer cells were culled at Johns Hopkins Hospital and transformed into an immortal cell line still used today to advance scientific research. Like Henrietta Lacks’ cells (HeLa cells), the digital health footprint is immortal, having no set expiration date and no clear way to destroy data, especially when the chain of custody is long and digital copies exist across multiple platforms. In addition, policy makers must contend with fundamental questions of data ownership, consent, and just compensation to the people whose personal data (genetic or digital) are being repurposed or sold.

As with genetic information, the digital health footprint is highly person specific. Data scientists have demonstrated that, even without direct identifiers (e.g., name and address), individuals can be identified readily in population databases (e.g., location tracking) using a relatively small number of data points. Recognizing that HIPAA only protects the fraction of health information generated in health care encounters, new standards must be developed for the deidentification, use, and protection of information contained in the digital health footprint. In advancing new regulatory approaches, a key challenge is to balance consumer privacy and the limiting of potential harms with the potential for health care advances that may be derived from the digital health footprint.

Breaches of health information privacy can lead to social stigma, embarrassment, and economic harm (e.g., insurance discrimination). The digital health footprint expands the scope and scale of potential data breaches and raises additional concerns. For example, machine learning and algorithms have been shown to perpetuate rather than eliminate racial biases and discrimination. Benjamin has described this phenomenon as the New Jim Code, referring to the embedded,
invisible, and potentially injurious bias in automated systems. In a recent study,49 a health system algorithm relied on health care utilization as a proxy for illness severity. At similar levels of illness complexity, black patients had used fewer health services than white patients; thus, the algorithm did not see how sick the black patients actually were. Therefore, the code systematically underidentified black patients in need of supportive health interventions.49,50

Even without a specific social or economic harm, ethicists and policy makers have made the case that if a consumer is forced to live in a society in which nothing is private, consumers will lose their sense of individual dignity,16,17 potentially eroding social trust or generating adverse psychological sequelae.51 Without trust, it is difficult for individuals to have meaningful relationships in personal and professional spheres of their lives with long-term social consequences.52

The US is notable because of its sector-specific regulatory approach, which shaped HIPAA as the approach to privacy in the health sector. Although specific policy solutions were beyond the scope of our interviews with experts, many pointed to the model of the European Union General Data Protection Regulation. The General Data Protection Regulation establishes individual privacy rights across sectors, codifying expanded transparency of data collection and use, increasing consumer control and data access, and limiting the immortality of data through provisions, such as the right to be forgotten.53

In the US, the California Consumer Privacy Act became law as of January 1, 2020, allowing residents to access personal information collected about them digitally and to opt out of the commercialization of their personal data.54 Extensive efforts are needed to understand how these protections will be interpreted, adopted, enforced, and replicated or expanded elsewhere in the US.

Limitations
This study has limitations. First, this study is qualitative and therefore intended to identify the breadth of issues to consider around digital health privacy as opposed to quantifying the prevalence of views. The relatively small sample size for this qualitative approach means that our experts may not be representative of the broader population of experts in the respective fields we sampled. In addition, despite efforts to achieve a diverse sample, some perspectives (eg, employees of companies with proprietary interests) may be underrepresented. Second, the digital privacy landscape is rapidly evolving and the subject of intensive media focus.55 Results are reflective of the period (2018) during which the study was conducted. Third, our experts were not sampled to identify the full breadth of views regarding privacy law and potential policy solutions, and our interview guide did not seek to arrive at policy solutions. Fourth, social desirability also may have been operative.

Conclusions
This study suggest that there is no distinction between health and nonhealth data. Far-reaching sources of data contribute to the digital health footprint, with implications for most, if not all, US individuals. The findings also suggest that the US should reconsider definitions of health privacy and develop appropriate safeguards as digital technology permeates nearly all aspects of everyday life.
Author Affiliations: Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia (Grande, Merchant, Asch, Cannuscio); Perelman School of Medicine, Division of General Internal Medicine, University of Pennsylvania, Philadelphia (Grande, Asch); Center for Public Health Initiatives, University of Pennsylvania, Philadelphia (Luna Marti, Feuerstein-Simon, Cannuscio); Perelman School of Medicine, Department of Emergency Medicine, University of Pennsylvania, Philadelphia (Merchant); Penn Medicine Center for Health Care Innovation, Philadelphia, Pennsylvania (Asch); Department of Psychology, Indiana University-Purdue University Indianapolis, Indianapolis (Lewson); Perelman School of Medicine, Department of Family Medicine and Community Health, University of Pennsylvania, Philadelphia (Cannuscio).

Author Contributions: Dr Grande had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Grande, Merchant, Asch, Cannuscio.

Acquisition, analysis, or interpretation of data: Grande, Luna Marti, Feuerstein-Simon, Merchant, Lewson, Cannuscio.

Drafting of the manuscript: Grande, Luna Marti, Feuerstein-Simon, Merchant, Lewson, Cannuscio.

Critical revision of the manuscript for important intellectual content: Grande, Luna Marti, Merchant, Asch, Cannuscio.

Obtained funding: Grande.

Administrative, technical, or material support: Grande, Luna Marti, Feuerstein-Simon, Merchant, Lewson.

Supervision: Grande, Cannuscio.

Conflict of Interest Disclosures: Dr Grande and Drs Merchant reported receiving grants from the National Human Genome Research Institute, National Institutes of Health (NIH) during the conduct of the study. Dr Merchant reported receiving grants from the National Heart, Lung, and Blood Institute, NIH during the conduct of the study. Dr Asch reported receiving grants from the NIH during the conduct of the study; receiving personal fees from GSK, Meeting Designs, Capital Consulting, and the National Alliance of Health Care Purchaser Coalitions; and receiving personal fees and nonfinancial support from Cosmetic Boot Camp, the Health Care Financial Management Association, the Alliance for Continuing Education in the Health Professions, Deloitte, the American Association for Physician Leadership, and the North American Center for Continuing Medical Education outside the submitted work. No other disclosures were reported.

Funding/Support: This research was supported by grant R01HG009655-03 from the National Human Genome Research Institute.

Role of the Funder/Sponsor: The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES
1. Internet World Stats. World internet users statistics and 2019 world population stats. Published November 6, 2019. Accessed December 19, 2019. https://www.internetworldstats.com/stats.htm
2. Tschabitscher H. How many people use email worldwide? Lifewire. Published June 24, 2019. Accessed December 20, 2019. https://www.lifewire.com/how-many-email-users-are-there-1171213
3. American Bankers Association. Survey: bank customers preference for digital channels continues to grow. Published November 5, 2019. Accessed December 20, 2019. https://www.aba.com/about-us/press-room/press-releases/survey-bank-customers-preference-for-digital-channels-continues-to-grow
4. Schultz J. How much data is created on the internet each day? Published August 6, 2019. Accessed November 22, 2019. https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/
5. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med. 2016;13(2):e1001953. doi:10.1371/journal.pmed.1001953
6. Nguyen T, Tran T, Luo W, et al. Web search activity data accurately predict population chronic disease risk in the USA. J Epidemiol Community Health. 2015;69(7):693-699. doi:10.1136/jech-2014-204523
7. Park S, Lee SW, Kwak J, Cha M, Jeong B. Activities on Facebook reveal the depressive state of users. J Med Internet Res. 2013;15(10):e217. doi:10.2196/jmir.2718
8. Prieto VM, Matos S, Álvarez M, CACHEDA F, Oliveira JLG. Twitter: a good place to detect health conditions. PLoS One. 2014;9(1):e86191. doi:10.1371/journal.pone.0086191
9. Merchant RM, Asch DA, Crutchley P, et al. Evaluating the predictability of medical conditions from social media posts. PLoS One. 2019;14(6):e0215476. doi:10.1371/journal.pone.0215476
Eichstaedt JC, Smith RJ, Merchant RM, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci USA. 2018;115(44):11203-11208. doi: 10.1073/pnas.1802331115

Hanvey B. Your car knows when you gain weight. The New York Times. Published May 20, 2019. Accessed December 20, 2019. https://www.nytimes.com/2019/05/20/opinion/car-repair-data-privacy.html

Kerry CF. Why protecting privacy is a losing game today and how to change the game. Published July 12, 2018. Accessed January 9, 2020. https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/

Bari L, O’Neill DP. Rethinking patient data privacy in the era of digital health. Published December 12, 2019. Accessed December 12, 2019. https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/

Glenn T, Monteith S. Privacy in the digital world: medical and health data outside of HIPAA protections. Curr Psychiatry Rep. 2014;16(11):494. doi: 10.1007/s11920-014-0494-4

Hudson KL, Holohan MK, Collins FS. Keeping pace with the times—the Genetic Information Nondiscrimination Act of 2008. N Engl J Med. 2008;358(25):2661-2663. doi: 10.1056/NEJMp0803964

Gostin LO. Health information privacy. Cornell Law Rev. 1995;80(3):451-528.

Donaldson MS, Lohr KN. Health Data in the Information Age: Use, Disclosure, and Privacy. National Academies; 1994.

NVivo Qualitative Data Analysis Software. Version 12. QSR International Pty Ltd; 2018. Accessed June 22, 2018. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77-101. doi: 10.1191/1478088706qp063oa

Cadwalladr C, Graham-Harrison E. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. Published March 17, 2018. Accessed May 17, 2020. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Na L, Yang C, Lo CC, Zhao F, Fukuoaka 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. doi: 10.1001/jamanetworkopen.2018.6040

Sobhani M, Saxon L. All our data is health data. Medium. Published August 14, 2019. Accessed January 15, 2020. https://medium.com/@usccbc/all-our-data-is-health-data-57d3c0f0336d

Mamlin BW, Tierney WM. The promise of information and communication technology in healthcare: extracting value from the chaos. Am J Med Sci. 2016;351(1):59-68. doi: 10.1016/j.ajms.2015.10.015

Card C. How Facebook AI helps suicide prevention. Published September 10, 2019. Accessed January 9, 2020. https://about.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/

Appleby J. Your wake-up call on data-collecting smart beds and sleep apps. Kaiser Health News. Published May 30, 2019. Accessed May 30, 2019. https://khn.org/news/a-wake-up-call-on-data-collecting-smart-beds-and-sleep-apps/

Pressman A. Anti-abortion groups sending ads to women in Planned Parenthood clinics. Fortune. Published May 26, 2016. Accessed January 9, 2020. https://fortune.com/2016/05/26/anti-abortion-groups-planned-parenthood/

Allen M. Health insurers are vacuuming up details about you—and it could raise your rates. Published July 17, 2018. Accessed April 8, 2019. https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates

Millenson ML. Big data on social determinants: improved health and unaddressed privacy concerns. NEJM Catalyst. June 5, 2018.

Blumenthal D, McGraw D. Keeping personal health information safe: the importance of good data hygiene. JAMA. 2015;313(14):1424. doi: 10.1001/jama.2015.2746

Choi YB, Capitan KE, Krause JS, Streeper MM. Challenges associated with privacy in health care industry: implementation of HIPAA and the security rules. J Med Syst. 2006;30(1):57-64. doi: 10.1007/s10916-006-7405-0

Korobkin R, Rajkumar R. The Genetic Information Nondiscrimination Act: a half-step toward risk sharing. N Engl J Med. 2008;359(4):335-337. doi: 10.1056/NEJMp0804352

Rothstein MA. Putting the genetic information nondiscrimination act in context. Genet Med. 2008;10(9):655-656. doi: 10.1097/GIM.0b013e31818337bd

Warzel C. All your data is health data. New York Times. Published August 13, 2019. Accessed August 12, 2019. https://www.nytimes.com/2019/08/12/opinion/health-data.html

Raine L, Anderson J. The Future of Privacy. Pew Research Center; 2014.
35. Clayton EW, Evans BJ, Hazel JW, Rothstein MA. The law of genetic privacy: applications, implications, and limitations. *J Law Biosci*. 2019;6(1):1-36. doi:10.1093/jlb/lsz007

36. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. *Science*. 2013;339(6117):321-324. doi:10.1126/science.1229566

37. Burgess MM. Beyond consent: ethical and social issues in genetic testing. *Nat Rev Genet*. 2001;2(2):147-151. doi:10.1038/35052579

38. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. *Nat Rev Genet*. 2008;9(5):406-411. doi:10.1038/nrg2360

39. Skloot R. *The Immortal Life of Henrietta Lacks*. Broadway Books; 2017.

40. Hudson KL. Genomics, health care, and society. *N Engl J Med*. 2011;365(11):1033-1041. doi:10.1056/NEJMra1010517

41. de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD. Unique in the crowd: the privacy bounds of human mobility. *Sci Rep*. 2013;3:1376. doi:10.1038/srep01376

42. Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun*. 2019;10(1):3069. doi:10.1038/s41467-019-10933-3

43. Lubarsky B. Re-identification of “anonymized data”. *UCLA Low Rev*. 2010;1701:1754.

44. Barth-Jones D. The debate over ‘re-identification’ of health information: what do we risk? *Health Affairs* blog. Published August 10, 2012. Accessed January 9, 2020. https://www.healthaffairs.org/do/10.1377/hblog20120810.021952/full/

45. Rothstein MA. Predictive health information and employment discrimination under the ADA and GINA. *J Med Ethics*. 2020;48(2):1-13. doi:10.2139/ssrn.3544331

46. Cohen IG, Mello MM. HIPAA and protecting health information in the 21st century. *JAMA*. 2018;320(3):231-232. doi:10.1001/jama.2018.5630

47. Ives M. Data breaches dent Singapore’s image as a tech innovator. *New York Times*. Published January 29, 2019. Accessed January 10, 2020. https://www.nytimes.com/2019/01/29/world/asia/singapore-data-breach-hiv.html

48. Benjamin R. *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons; 2019.

49. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/science.aaax2342

50. Benjamin R. Assessing risk, automating racism. *Science*. 2019;366(6464):421-422. doi:10.1126/science.aax3873

51. Putnam RD. *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster; 2000.

52. Fukuyama F. *Trust: The Social Virtues and the Creation of Prosperity*. Vol 99. Free Press; 1995.

53. Voigt P, Von dem Bussche A. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer International Publishing; 2017. doi:10.1007/978-3-319-57959-7

54. Harding EL, Vanto JJ, Clark R, Hannah Ji L, Ainsworth SC. Understanding the scope and impact of the California consumer privacy act of 2018. *J Data Protection Privacy*. 2019;2(3):234-253.

55. Opinion: The Privacy Project. *New York Times*. Accessed December 8, 2019. https://www.nytimes.com/series/new-york-times-privacy-project

**SUPPLEMENT.**

**eAppendix.** Digital Health Privacy Expert Interview Guide