Chapter 5
Enhancing Surveillance: New Data, New Technologies, and New Actors

5.1 Introduction

In many areas of the globe, the capacities of public health to gather information about health are rudimentary at best. Health care, transportation, power, and communication infrastructures may be very limited, if not entirely absent. In some of these areas, new technologies are filling gaps. To take perhaps the most impressive example, cell phone and even smart phone access is increasing rapidly using local solar power and serving areas without hard infrastructure. Data collected through these mechanisms can be critical sources of information about users, their health, their locations, and their activities.

In wealthier nations, too, surveillance now uses sources that go far beyond surveys conducted by public health agencies and data collected through the reporting requirements of state and national laws. Electronic medical records, patient registries, research data bases, direct to consumer testing, smart phone apps, internet searches, social media sites, webcams, and smart devices making up the “internet of things” all contribute highly useful information about health. Much of this is used by public health, and much more is used by private entities to survey the health of the public.

These data sources are far richer in the information they contain and may also be more representative of the population than are sources available to traditional public health. Combining data from these sources may also reveal unexpected patterns and correlations that can be especially helpful for syndromic surveillance, a recently developed form of surveillance that uses artificial intelligence to identify patterns that may signal disease clusters or unusual events. Devices such as smart phones may give location information in addition to content that can enable detection of hot spots for exposures. The information may be available in real time, thus enabling quick responses to emergencies. The use of existing data sources also can avoid the redundancy and expense of efforts by public health to collect its own data.
But there are disadvantages and risks to these innovations (e.g. Kostkova 2018). Data initially collected for purposes other than public health may not meet the standards of public health. Information may be collected in different formats that are hard to link together and that risk reduplication or misidentification of individuals. It may contain unexpected inaccuracies or selection biases. The data collected for another purpose may be partially or imperfectly relevant for public health. People may be surprised to find that data have been collected and used, even when the use is for public health purposes of which they would otherwise have approved. Data that have public health uses may also be put to other purposes that people regard as less acceptable, such as marketing, development of expensive for-profit drugs, or criminal investigations. People may be outed, doxed, shamed, harassed, bullied, arrested, deported, or subjected to other risks when data are drawn from other sources to be used by public health. The power of these new technologies allows far more to be inferred or known about individuals, their movements to be traced, and quite complete portraits of them to be created—all, perhaps, without their knowledge or participation in any way. Anger, protests, and even efforts to have the data destroyed may be the result.

Technology is evolving so rapidly that it would be foolhardy to assume current uses will continue to predominate or that likely new data sources are predictable. Consequently, the discussion in this chapter uses important illustrations drawn from what is happening with data today rather than claiming to be comprehensive. In what follows, we describe these novel data sources, many of which are in private hands or use techniques of artificial intelligence for analysis:

– Interoperable electronic health records
– Retained blood spots from newborn screening
– Biobanks and other genetic databases
– Patient registries
– Information gained in research
– Direct to consumer testing, including genetic testing
– Smartphones and smartphone apps
– Wearables (e.g. fitbit, AppleWatch) and biosensors
– Robots and smart devices

Data from some or all of these sources are increasingly being linked, creating vast reservoirs of data.

Novel actors are also entering the surveillance space. These include non-state partners with WHO and health care providers. They also include internet search engines such as Google and social media sites such as Facebook. While these may also be both sources and users of data for many purposes, the information they collect and some of the uses they make of that information may be beneficial for public health. These business entities also have commercial interests that may be at odds with the goals of public health.

Cutting across some of these new data and new actors are features that are particularly important to the core considerations we outlined at the beginning of this volume and that we think are critical to sustain trust in surveillance. Many involve
“big data”: vast data reservoirs that are huge in volume, velocity and variety. These data are rich resources for the development of analytic and predictive algorithms for the health of populations and subpopulations, communities and groups, and families and individuals. Although the data often are initially de-identified, they may be combined in ways that significantly increase the risks that individuals may be re-identified or that inferences may be drawn about individuals. Inclusion in these data reservoirs may be without notice or consent at all, with opt-out or opt-in methods of agreement, or by “broad” consent that may stretch for years or even lifetimes.

5.2 “Big” Health Data and AI

Technological abilities to produce, collect, store, and analyze data are growing exponentially. Discussions of the ethics of big data are also growing apace (e.g. Mittelstadt and Floridi 2016). In this section and the next two, we focus on three of these overarching ethical issues with novel data sources of particular relevance to public health surveillance: artificial intelligence analytics and bias, identification of or inferences about individuals, and notice and consent.

What makes data “big” is itself subject to debate. Mittelstadt and Floridi, in a meta-analysis of publications on the ethics of big data, note that definitions include the numbers of individuals in the data base, the sheer amount of data, the computing technology needed to analyze the data, and the types of analyses that can be conducted. “Big” data does not mean data that are complete, accurate, or without bias, however. To take just one example, critics note that big data assembled from electronic medical records may include errors in data entry and selection bias in what is entered (Hoffman and Podgurski 2013).

Predictive analytics may be applied to these data sets. In what is termed a “learning health care system,” for example, algorithms may be developed that predict patient outcomes from the data. As more and more data are accumulated, the algorithms may be refined for greater and greater predictive accuracy, enabling the health care system to refine care delivery. One of the issues with the use of these predictive algorithms is that the patterns that they identify in the data may not be predictable in advance. Thus with techniques such as syndromic surveillance it may only be possible to tell people after the fact what was discovered. While some goals can be outlined—improving health or detecting drug side effects, for example—any precision about what might be linked to these goals may be unknown. Thus models of informed consent that require disclosures of possible risks may not be possible (Francis et al. 2009).

These predictive algorithms are used not only in health care but also in areas that may have differential impacts on the health of members of a population: for example, predicting whether someone is a good credit risk (and thus can qualify for a loan or a mortgage), a reliable tenant (and thus can be rented an apartment in a competitive market), or a likely re-offender (and thus should not be placed on
parole). Facebook has used algorithms to predict suicide risk and google to predict influenza outbreaks—with, as we describe below, success that has been uneven.

Bias can be introduced at many levels in this process. The data itself are an initial problem. Critics of the use of predictive algorithms in policing point out that if police are deployed primarily in minority neighborhoods, and drug arrests are higher in those neighborhoods, this does not mean that drug use is actually higher in those neighborhoods. It only means that data have been gathered from those neighborhoods (e.g. Ferguson 2017). Similarly, if biobanks include primarily Caucasians of Northern European descent, algorithms will not “learn” from data about persons with origins in which other genotypes are more common.

Algorithms used in predictive analytics are often not transparent; in the words of Frank Pasquale (2016), their use has created a “black box” society. Commercial entities developing these techniques argue that they should be protected as trade secrets (Meyers 2019). The ability to scrutinize bias in their development and application is limited by these protections.

Even if these algorithms were public, there is controversy about what it would mean for an algorithm to be “fair.” Friedler and colleagues (2016) argue that “fairness” has been understood in fundamentally different ways. Algorithms link decisions (for example, what treatment to recommend for a particular patient) to constructs (for example, a score about likelihood of survival) based on observed proxies for the construct (for example, heart ejection fraction or kidney function). Bias can come in at any of these points, including the observations collected, what are determined to be proxies for the construct, and whether the construct is appropriately linked to the decision space. With recidivism, to take an example from criminal law, prior arrest record might be an observation used as a proxy for the person’s “propensity” to commit a crime, which is then used to predict the likelihood that someone will reoffend and therefore should not be let out on parole. There may be structural bias in whether the construct space is appropriately represented by the observed space as a proxy. This may be skewed by groups such as race, sex, or age as well (Zou and Schiebinger 2018).

A further justice problem presented by AI is the “divide” between data subjects and those with the resources to analyze the data (Benkler 2019). Data may be scraped from all kinds of sources used by ordinary people, from grocery store discounts to medical records. However, big data collection, maintenance, and analysis take massive resources. If data are repurposed and used for gains of others—as they might be for decisions such as whether to grant credit, whether to rent housing, how to treat patients who can access care, or where to locate a primary care clinic—distrust might be the predictable result (Francis and Francis 2014).

5.3 The Debate About Re-identification

A standard distinction in data protection law lies between data that can be used to identify individuals and data that have been de-identified. The GDPR (2020), the U.S. rules governing research with human subjects, and the U.S. rules governing the
privacy and security of health information possessed by covered health care entities and health insurers all apply only to identifiable information. Once information has been appropriately de-identified—and it must be emphasized that these standards are stringent—the only further regulatory issue is that it not be re-identified. Data use agreements may prohibit re-identification efforts but ironically by so doing may also rule out external assessments of whether the data are at risk of re-identification (Sweeney et al. 2018). Privacy policies and terms and conditions from websites and social media sites typically draw this line as well. But the line has become increasingly controversial. Possibilities of re-identification pose one set of problems. Possibilities of harms to groups and harmful inferences about individuals pose another.

As data sets are increasingly combined, information becomes richer and richer. The more data are enriched, the greater the likelihood that properties will combine so that individuals ultimately stand out. The Harvard computer scientist Latanya Sweeney has conducted a number of well-known re-identification “attacks” (e.g. Yoo et al. 2018; Sweeney et al. 2018; Southern Illinoisan 2006). While some of these attacks required sufficient sophistication for Illinois courts to decide that the release of information collected by the state about cancer patients did not put individual privacy at risk, others have become increasingly easy. One study also indicates that reidentification of law students may be more likely if the students are Black or Hispanic, thus introducing the possibility of disparate impact discrimination in reidentification risk (Sweeney et al. 2018).

Identity-masking techniques have been developed to help to disguise individual identities. Some forms of identity masking involve taking out sufficient amounts of information so that individuals cannot be mapped to cells containing very small numbers. An example is the U.S. HIPAA privacy rule which presumes that data are deidentified if a list of eighteen kinds of information are removed. Other forms introduce noise to the extent that differential privacy is achieved in the sense that it is not possible to tell whether any given individual was included in the data set. The choice of anonymization techniques has costs, however. Removing information or introducing noise may reduce the utility of the data set, especially if this is done to the extent needed to reduce reidentification risks in light of current reidentification technologies. Xu and Zhang (2020) argue in addition that the choice of deidentification techniques may affect the extent to which disparities are observed from the data that could support claims of disparate impact discrimination. If noise insertion is the anonymization technique, and disparities are understood as the separation between groups, it is less likely that disparities will be identified because noise insertion may tend to increase the standard deviation of outcomes within population subgroups.

Even when de-identification is achieved successfully, further ethical concerns remain. When individuals can be mapped successfully to groups that are sufficiently small to allow inferences to be drawn about them with high probability, individuals may suffer stigmatization, shaming, or other deleterious consequences. Group-level harms may also occur, as when the conclusions about tribal origins were drawn about the Havasupai from the research described later in this chapter. Moreover, the economic incentives may be much greater to develop understanding about the
characteristics of groups rather than individuals: “high-payoff” targets for commercial actors are not single individuals but individuals as members of groups (Floridi 2014).

5.4 The Absence of Real-Time Notice or Consent

Further ethical issues are posed by whether individuals have notice of or can in any way be said to consent to data collection or use. If de-identified data are considered no longer to be about the individual, as is standard in legal regimes governing data protection, the conclusion is drawn that individuals need not be given notice. Standard boilerplate in the terms and conditions for website users, for example, is that information that does not reveal identities may be used or transferred. The Google Privacy Policy is an illustration: “We may share non-personally identifiable information publicly and with our partners—like publishers, advertisers, developers, or rights holders. For example, we share information publicly to show trends about the general use of our services. We also allow specific partners to collect information from your browser or device for advertising and measurement purposes using their own cookies or similar technologies” (Google 2020). The result may be surprise on the part of individuals about how information is being used or transferred, especially if the information is combined with other information that can be used to identify them or draw inferences about them. Individuals may accept this from Google—the service is free and people are used to the advertising they receive as a result—but be distressed if it is used in ways that might affect their health or the health of the public.

Some data users employ “opt in” or “opt out” mechanisms to secure individual agreement to data collection and use. Examples are the ubiquitous “I agree” hot buttons when users seek services over the internet. “Opt in” mechanisms are far less likely to result in user participation because they require positive action, for example signing up to receive content from a newsletter. As discussed below for newborn screening, opt in requirements have greatly reduced participation in letting samples be retained for further public health uses. On the other hand, the U.K. collection of patient data from the National Health has chosen to use an opt out model to increase the extent to which the data are representative of population health (Understanding Patient Data 2018). The concern about this choice is that people may remain largely unaware about collection and use of information drawn from their health care.

“Broad consent” is a controversial strategy for agreement to use of information on an ongoing basis. It is consent, but open-ended. For example, consent to allow medical records to be used in “future health-related research” until revoked is broad consent. This form of consent has been used not only for biobanks but also for registries and many other uses of identifiable information. It has the advantage that individuals do not need to be re-contacted each time a new use of information about
them is proposed. Instead, protection of the individuals would rest in any other required approval of the use. For research, the approval might include whether it is health-related and thus within the scope of the consent, whether it appropriately protects individual confidentiality, and whether its benefits outweigh risks. But consent may be far broader. For example, patientslikeme, a site that allows individuals to share their health information, has a trademarked “digitalme ignite” pilot study through which individuals can contribute blood samples, health updates, regular examinations, and other biospecimens. Information is kept indefinitely unless permission is revoked and may be used by researchers or for-profit companies. (patientslikeme 2020).

On the one hand, it seems that individuals should be able to say that they are willing to participate in an ongoing project collecting health information for health-related research or other activities, and to be included in any further research projects as long as their identities are protected. This could include public health research as related to health. Re-contacting individuals is inconvenient and costly and might be regarded as an annoying reminder of health issues.

On the other hand, memories fade. If people are not sent updates about what has been learned from the ongoing data use to which they have consented, they may be surprised and concerned if they hear about it. Commentators have suggested that broad consent is too empty to constitute genuine understanding, that the idea of “health-related” is pliable and may change, and that there should be some kind of expiration, requirement for re-consent, or the opportunity to opt out of continuing at regular intervals. One particularly important moment for re-consent might be adulthood. Some broad consent is given by parents for their children to enter into disease registries as infants and individuals thus included might be very surprised and distressed to discover that information about them has been updated ever since (Francis 2014).

Ploug and Holm (2016) propose “meta consent” as a solution to problems of broad consent. This idea is that people should be asked at the point of data collection to state their preferences about future consent. These choices might be never to be asked, always to be asked, to be asked in some cases or not others, and so on. They argue that this idea far better reflects differences in values: people might want to be asked about some data uses but be fully willing to agree to many other data uses without further contact. They argue that this approach is better able to respond to individual vulnerabilities and sources of mistrust as well. One suggestion that they do not make, but that could also enhance trust, is that when possible any kind of agreement to ongoing data use should be coupled with requirements to provide updates on a regular basis. Such updates might be easily arranged through electronic methods of communication. For example, if information is being collected from electronic health records, as described in the next section, people could be sent updates just as they are sent other communications such as reminders from their health care providers.
5.5 Interoperable Electronic Health Records (EHRs)

Records of individual health care are potentially a very rich source of information for public health. Interoperable EHRs, now widely used in health care, have great advantages for both individual health care and public health. They may be accessible in real time and across long distances, as long as the Internet itself is working and connections can be established. They will likely contain much of the relevant information about the health of individuals who regularly see medical providers. Depending on their design, they are readily searchable and thus have the capacity for public health to follow up and to gather new information as needed. They may be used for syndromic surveillance to identify unusual patterns that might warrant further investigation. They may provide information about rare drug side effects or patient outcomes with experimental medicines; analysis of data in patient medical records is the basis of the sentinel system in the U.S. for monitoring drug safety, for example (Coyle 2017). They may be configured with metadata that record information about the time and source of data collection, thus enabling judgments about reliability or relevance. They likely contain contact information about patients, yet they also may be programmed to de-identify information or to mask particular data fields considered irrelevant or sensitive. They are interactive, allowing alerts or reminders to be delivered to physicians, patients, or others. Use of EHRs may be more efficient for providers to transmit data to public health quickly and accurately than reports given over the telephone or typed into a website. There thus may be considerable pressure to rely on EHR data to supplement or supplant separate collection of data for public health, especially if funding for public health is restricted.

On the other hand, use of EHR data for public health—or at least, reliance on this data to a significant extent in the public health enterprise—presents major challenges even in well-off nations where automated data systems are ubiquitous and information technology is robust. EHR design may not capture data that are important to public health or organize data in a way that is readily accessible for public health purposes. When groups within the population do not have reliable access to health care, EHRs will be unrepresentative of the population as a whole and likely to miss disadvantaged groups. Use of these records for purposes other than patient care may be objectionable to some patients—so much so that they may lose trust in health care and avoid interaction with health care providers as much as they can. Suspicion may intensify with any suggestion that public health information may also be used for law enforcement or security.

Arguably, public health is at least health and many of its activities are devoted to improving health care and population health broadly construed. Nonetheless, access to these records presents serious privacy risks, especially if they can be entered, copied, and downloaded at the flick of a switch without patient agreement or even knowledge. Public health law scholar Wendy Mariner argues that public health use of medical records without individual consent is “mission creep” that threatens individual rights (Mariner 2007). These risks are magnified by the richness of the
information within EHRs; even de-identified EHR information may be re-identifiable when combined with other data sources.

### 5.5.1 EHRs in the United States

In the U. S. today, the vast majority of medical records are electronic. These records are protected by a patchwork of legal standards. These standards include the federal constitution and federal statutory and regulatory protections for covered health information. They also include a variety of state laws.

At the level of the federal constitution, a decision in 1977 held that a state requirement to report controlled substance prescriptions to a state data base did not violate privacy rights (*Whalen* 1977). The reasoning relied on the state’s compelling interest in preventing drug abuse and the structuring of the data base to prevent unauthorized disclosures of the information. This is the only Supreme Court decision dealing with data privacy, and it was handed down over forty years ago. The extent to which there is federal constitutional protection for informational privacy remains clouded; one federal appellate court held in 2020 that sexual assault victims could not sue local officials for releasing information that might identify them because the constitutional right to informational privacy was not clearly established (*Dillard* 2020).

The Health Insurance Portability and Accountability Act (HIPAA) is the federal statutory structure that governs these EHRs when held by covered entities, which are health care providers, insurers, and information exchanges. The HIPAA regulations set out standards for security and privacy (HIPAA 2020). HIPAA distinguishes between individually identifiable information and information that has been de-identified to prescribed standards, in which case regulatory requirements no longer apply except to assure that the information will not be re-identified. Under HIPAA, many uses and disclosures of identifiable information require individual agreement, called “authorization,” including uses for research or commercial purposes. Other uses and disclosures do not require any individual notice or authorization, such as uses or disclosures for treatment, payment, and health care operations, and uses or disclosures for public health as provided by law.

Thus HIPAA permits state or federal public health authorities to collect or use EHR information for public health without notice to individuals or their authorization. Health care providers may analyze data from their EHRs to assess community health needs or to decide which services to expand or close, also without individual notice or authorization, as this is considered part of health care operations. Public health researchers may not use EHR data that identifies patients without their authorization, however, unless the research qualifies for a waiver as explained below.

HIPAA gives individuals a right to request an electronic copy of most of the treatment-related information in their EHRs (called the “designated record set”) or to request that a copy be transferred. Once EHR information has been transferred out by the covered entity, it is no longer within the protections of HIPAA. Likewise, the vast array of actors collecting health information in electronic form that are not
HIPAA-covered entities also are not subject to the HIPAA constraints. Instead, these entities are far more lightly regulated by the Federal Trade Commission Act’s prohibition of unfair or deceptive trade practices. These entities collecting health information directly from individuals—including everything from Facebook to medical information sites such as WebMD to direct to consumer genetic testing companies—may have far greater scope of action, even when the information they receive originated from EHRs.

Within U.S., state laws may vary in the extent to which they protect information that has come into the hands of public health. These variations may create uncertainty on the part of the public about whether information transferred from EHRs to public health will be fully protected. One state law case has dealt with this issue. Journalists sought data in the state cancer registry obtained from patient records to investigate whether toxic exposures might have been related to cancer clusters of neuroblastoma in their region (Menderski ccc). They requested the information by zip code, date of diagnosis, and cancer type, but did not request any other identifying information. The health department refused to reveal the information, citing an analysis by Latanya Sweeney that patients could be re-identified through linkages with other publicly available data sets. The Illinois courts held that the Illinois Freedom of Information Act (FOIA) law permitted disclosure (Southern Illinoisan 2006). The trial court reasoned that the health department had not shown that the information would reasonably tend to reveal the identity of persons in the cancer registry—the statutory standard—because of the complex data skills required for the reidentification. The Illinois appellate courts upheld the ruling as correctly applying the state’s statutory standard regarding information in the cancer registry. We discuss registries more fully below; suffice it to say here that without sufficient transparency about public health data collection from EHRs, people may be surprised to know what data the government possesses and how it may be used.

To take another example, in Utah information from a state data base of controlled substance prescriptions was released without a warrant as part of an investigation of drug diversion from ambulances. Two firefighters were erroneously accused of drug fraud based on information from the data base. The result was that the legislature passed a statute protecting the data base against warrantless searches by state and local law enforcement, but the federal Drug Enforcement Agency still has the ability to access the data base without a warrant (Whitehurst 2017). State auditors have also warned that cybersecurity protection for the data base is inadequate (Carlisle 2019).

Electronic health records (EHRs) contain vast amounts of patient data that might be used to supplement or substitute for at least some public health data collection needs. Although EHRs are now ubiquitous in U.S. health care, their use in public health has fallen far short of what many believe it could be (Friedman et al. 2013). The HIPAA privacy and security rules were something of an afterthought when concerns were raised that electronic transfers for insurance billing might threaten data privacy or security. EHRs are limited in the information they contain in part due to this history. Data are structured with fields that are relevant for billing and may contain other important information in natural language text that is more difficult to search (Hoffman 2018). They will only contain information as gathered
through medical care encounters, including visits and information entered through portals where patients can enter their own information remotely. Information in EHRs will be incomplete if patients do not follow up after an initial visit or if they are seen by providers who use different EHR systems. There are also concerns about inaccuracies in the data entered into EHRs, although as physicians become more accustomed to technology the error rate may be decreasing (Hoffman 2018).

The fragmentation of health care delivery in the U.S. is another part of why EHRs are not as useful for public health as they might otherwise be. Different EHRs are used by different health care systems and the data in them cannot readily be analyzed together. Health information exchanges may be helpful in addressing this problem, as they process all data from providers and patients who participate (Saks et al. 2020). All payer claims data bases (ACPDs) are other efforts by state health departments to gather data about particular patients from all care for which payment claims were submitted. ACPDs have been very useful in developing fuller pictures of costs of care and of care utilization. For example, they can trace whether reduced outpatient costs of care for diabetic patients are correlated with more frequent hospitalization, which would raise costs and signal inadequate blood sugar control, or whether the reduced costs may be due to improved patient management. But ACPD data contain the information needed for reimbursement rather than a fuller picture of the patient’s condition. Moreover, a Supreme Court decision in 2016 held that the federal Employee Retirement Income Security Act (ERISA) preempts state requirements for employer health insurance plans to submit their claims data (Gobeille 2016). This decision may eventually mean that fewer plans submit data to ACPDs and likely has discouraged the development of ACPDs in more than the sixteen states that currently have established them (Curfman 2017).

Public health itself is fragmented, too; responsibilities for public health lie primarily at the state level and may be devolved to local health departments which lack extensive information technology infrastructure (Williams and Shah 2016). EHR utilization in the US was heavily incentivized as part of the economic stimulus response to the 2008 recession. Providers who met standards for “meaningful use” of these records received extra payments and EHR use was eventually required for Medicare reimbursement. The actual public health benefits of the meaningful use incentives were limited because some public health systems were unable to receive data transfers from EHRs. Meaningful use for public health as it was implemented after the 2008 recession did include the ability of EHRs to interface with immunization and cancer registries, the ability to report cases to other specified registries, and the capacity to retrieve and submit data for syndromic surveillance, primarily as optional functionalities. It also included the capacity to retrieve and submit certain reportable clinical laboratory results such as positive tests for hepatitis. These reporting capabilities are only a small subset of the legal requirements imposed on health care providers to report patient information to public health, however. Moreover, unless EHRs are standardized in how they record aspects of disease and patient function, comparisons will continue to be elusive and EHRs will not contribute critical information about population health trends (Kruse et al. 2018).
Information equity is another significant problem for EHRs. Minority populations in the U.S. receive health care less frequently and less adequately than other population subgroups. EHR data are likely to be better in quality and more complete for patients with consistent insurance coverage (Saks et al. 2020). Notably in regard to information equity, non-white and ethnic minority populations express privacy concerns about the use of EHRs at far higher levels than do U.S. whites (Clayton et al. 2018). The result may be that these populations are less likely to agree to use of their data, including uses that may benefit population health or subgroups of which they are members (Saks et al. 2020).

5.5.2 EHRs in the UK

Because with the National Health Service (NHS) health care in the UK is far less fragmented than in the U.S., the UK presents greater opportunities for use of patient care data for public health. Challenges remain, however. Among them are that in practice there is fragmentation of information technology systems across the UK and data sharing agreements are complex (digitalhealth 2016). Public concerns about privacy risks and the importance of ensuring patient trust in data use have come increasingly to the fore. A series of reports under the leadership of Dame Fiona Caldicott has addressed these tensions in data use through the development of what have been named “Caldicott principles.”

The original Caldicott principles were:

1. every use or transfer of confidential personal data should be clearly defined, scrutinized, and documented; continuing reviews should be regularly reviewed; these should all be conducted by an appropriate data guardian.
2. Identifiable information should be only used when essential for the specified purpose.
3. Identifiable data used should be the minimum necessary.
4. Access to personal confidential data should be on a strict need to know basis, and then only for the minimally necessary data.
5. All those handling personal confidential data should be fully aware of their obligations to protect confidentiality.
6. All use of personal confidential data should be in accord with law.

A seventh Caldicott principle, adopted in 2013, is that the duty to share information can be as important as the duty to protect patient confidentiality.

The NHS Constitution stipulates that patients have the right to request that their confidential information not be used beyond their own care and treatment, to have their objections to additional uses considered, and to be given a legal explanation where their wishes cannot be honored (Caldicott 3, 2016, at 5). This constitution applies to health care but not to social care such as home help with activities of daily living. Controversy about NHS data use emerged in 2014 as the NHS sought to make greater use of technology to improve patient care. Patients were told by the NHS that records from their general practitioner would not be shared in identifiable form with the Health and Social Care Information Centre (HSCIC) if they so requested. They were also told that they may request that the HSCIC not further
share their identifiable information. The NHS pamphlet with this information was confusing and resulted in the request for Caldicott 3, which produced a third version of the Caldicott principles.

Caldicott 3 relied importantly on the UK National Data Guardian (NDG) review of data protection to safeguard patients (Caldicott 3, 2016). It proposed a system under which patients could opt out of use of their records, rather than having to opt in to record use. This approach significantly increases the likelihood of record availability, since many patients are unlikely to take the steps needed to opt out. Patients were to be given two opt outs; declining to allow their records to be used to improve local services and running the NHS, and declining to allow their records to be used for research and improved treatment—although this last proposal was to be put out to further study. The basic value underlying Caldicott 3 was the need to enhance trust despite allowing record use. The Report begins: “This is a report about trust.” (Caldicott 3, 2016, at 3) Important aspects of Caldicott 3 are recommendations for clear descriptions to patients about when information might be used, the requirement that there should be no surprise that an appropriate professional has access to information, and patient confidence that only the minimum necessary information is shared. Other aspects singled out as critical to trust included adequate data security, full public consultation, and dialogue with the public. The NDG could allow anonymized data to be used without consent, but subject to enhanced sanctions for failures to protect this anonymity.

The British standard for privacy rests in “reasonable expectations” in the sense that there should be no surprises to an individual about how data are used (Office for the National Data Guardian & Connected Health Cities 2018). To understand the idea of reasonable expectations (and whether it differs from the idea that there are circumstances in which consent might be implied), the NDG convened a three-day citizens’ jury in 2018 (NDG 2016). The jury was given scenarios about a hypothetical patient and asked to discuss them in light of the legal standard for reasonable expectations: what an “average person of normal sensibilities” would think about whether the information sharing would be surprising and unacceptable. Jury answers thought information sharing among providers to improve individual patient care would be expected. They were less certain about the need to share information with administrators and other non-clinical personnel beyond information that would be minimally necessary to contact patients, schedule appointments, and bill for services. Expected sharing included discussions with experts at other facilities about difficult care questions. Jury members were not as sure, however, that patients would reasonably expect information to be shared to help experts diagnose their own patients. A majority of the jury did not expect information to be shared with patients’ relatives; they thought that without consent this sharing would not improve care and might lessen trust. Extended discussion of reasons for and against data sharing increased the likelihood that jurors would want to share the data.

In 2018, the National Data Guardian (NDG) was established in the UK as a statutory office, enabling the NDG to issue official guidance about processing of health and adult social care data. Dame Fiona Caldicott was appointed as the initial holder of the position in early 2019. Commentators have argued that with this arrangement
Britain could be a model for balancing use of data to improve health in an aging population with protecting privacy (Chan et al. 2016).

5.5.3 **EHRs in the European Union: The General Data Protection Regulation and Public Health**

The European Union’s General Data Protection Regulation (GDPR) went into effect in May 2018. It sets out a common set of regulations that all EU members must adopt for data protection. EU member states are in the ongoing process of statutory revisions to implement the GDPR (Molnár-Gábor 2018, on Germany). According to Molnár-Gábor, European privacy regulation is most strongly influenced by French and German law; the fundamental anchor of French law is personal integrity and the fundamental anchor of German law is human dignity and the right to informational self-determination.

The advent of the GDPR caused considerable consternation about the availability of sensitive health information for research or public health, although there are exceptions to the requirement of individual consent for both. Genetic and biometric data are specifically mentioned as sensitive. Explicit consent of the data subject is required but there is an exemption for research when there are suitable safeguards and the processing is carried out for reasons of substantial public interest. Suitable safeguards may involve an oversight body such as an ethics review committee (see description in Shabani and Borry 2018). Pseudonymized data is treated as within the exemption for research. (This is data where a key code, kept separately, is substituted for identifying information. Records with the same key code can be linked, but only those with access to the key will know who the individuals are).

For public health, Recital 54 of the GDPR provides that the processing of special categories of personal data may be necessary for reasons of public interest in the area of public health without consent of the data subject. When this is done, the processing must be subject to suitable, specific measures to protect the rights and freedoms of natural persons. Public health is interpreted as in Regulation (EC) no 1338/2008 to involve all elements related to health, including health status, mortality and morbidity, provision of health care, access to health care, health needs and resources, health care expenditures and financing, and causes of mortality. Recital 54 stipulates that data processing for public health should not result in processing by third parties such as employers, insurers, or banking companies for other purposes (EU 2019).

Researchers and public health advocates voiced concerns that the GDPR would discourage EHR use for these purposes despite the stated exceptions. COVID-19 and the need for rapid collaboration about disease spread brought these concerns to the fore (McLennan et al. 2020). Digital patient records, if available across systems and countries, are a critical source of information about the disease incidence, course, and possible treatments. Many of the early research about COVID-19 relied
instead on small case series of patients in a single institution and thus did not pro-
vide evidence regarded as high quality. Because of differences between the U.S. and
the EU regarding data protection, proposals for combining data sets across the
Atlantic may be legally difficult. McLennan and colleagues, however, argue that
solidarity requires pursuing this option and that the GDPR does permit data use for
this purpose with appropriate technical and organizational safeguards.

The EU has embarked on an initiative on the Digital Transformation of Health
and Care (DigiCare). There is a proposal to create an innovation cloud within the
DigiCare environment that would be consistent with GDPR protections (Aerestrup
et al. 2020). One of the goals of this technology is to be able to utilize shared com-
putational and storage resources that can enable processing of huge data sets needed
to detect weak signals about possible correlations. Another is to enable use of
pseudonymized data that cannot be downloaded so that individual identities can be
protected. As these and other initiatives continue to develop, transparency, over-
sight, and purpose specification will continue to be important to prevent mistrust. It
seems likely that people will be far more willing to accept the use of data sets such
as these to provide early warning signs of public health emergencies or to gain
urgently needed knowledge such as for COVID-19 than for purposes that are not as
immediately urgent.

5.6 Bloodspots Retained from Newborn Screening

Since the 1960s, newborns with access to good health care have been screened for
metabolic conditions such as phenylketonuria (PKU) that require early treatment to
prevent severe disability or even death. Screening programs represent a population-
based intervention with a very strong public health aim: identifying individual
health conditions that can be addressed through timely interventions. Although the
conditions now screened for may vary with the jurisdiction, they typically include
sickle cell, cystic fibrosis, PKU, congenital hypothyroidism, and certain other inher-
ited metabolic conditions. The screening is conducted by a heel stick; blood spots
are collected on a card. If they are retained, the blood spot cards can be an enor-
mously rich source of material not only for research about genetics but also for
public health issues such as infection seroprevalence, toxic exposures, epigenetics,
and other information available from a blood sample. The samples have the poten-
tial to form a population-wide data bank of genetic information and environmental
information available from blood that could be linked to clinical outcomes on an
ongoing basis (Lewis et al. 2012). The information in the samples that is genetic
can, with sufficient other information, be linked back to individuals and thus is not
fully de-identifiable. It could in principle provide important warnings to individuals
about carrier status or risks of adult onset conditions.

The information obtained from retained newborn screening samples has several
features that are especially important ethically. It is collected from individuals (new-
borns), who have no knowledge of the collection and who are unlikely to ever
become aware of it unless they are told about it, or encounter consequences from it, in later life. In many jurisdictions, the screening is mandatory, although retention of the spots beyond the short time periods needed for confirmation and perhaps also program improvement may require separate consent. If consent is required, it is given by parents on behalf of their children; such consent, however, comes at a time of great excitement when attention is likely to be elsewhere. Parents may remember the heel stick if the baby cries, but memories that bloodspots may be saved for future research or public health are likely to fade very quickly. Because newborn screening applies to everyone, retained samples perhaps also represent the most comprehensive snapshot of the population at any given point in time; these spots are the available data least likely to omit vulnerable population sub-groups, at least in jurisdictions where screening is mandatory. They also will contain complete information from age cohort to age cohort, so they present the opportunity to make comparisons among individuals born in different places and times. Just knowledge of the date and place of birth about an individual, however, may convey critical, powerful, and potentially stigmatizing information about them, depending on what has been learned from these samples. Finally, the genetic information in them can be linked to individuals, although the samples may also be useful for studies such as levels of environmental exposures at a given point in time that do not require individually identifying information.

A comprehensive study of the use of retained newborn screening samples in Denmark illustrates the potential uses of this resource (Nordfalk and Ekstrøm 2019). Newborn screening and bloodspot retention for research are routine in Denmark. Although the legal possibility to opt out of the screening exists, no informed consent is required for sample storage and use in research. No records are kept of the frequency with which samples are actually used in research. Nordfalk and Ekstrøm used a survey of publications to identify secondary uses of screening samples and their frequency. Their survey strategy was the best available but did not capture any uses of the samples that were not published. The researchers estimated that just under 40% of the samples had been used for purposes other than newborn screening. Most but not all of the studies had research ethics approval. The most-studied disease topic was mental illness (about one-fifth of the uses), followed by metabolic disorders and diabetes. It seems possible to hypothesize that individuals would be less likely to expect newborn screening samples to be used in studies of mental illness than in studies of diseases of a type more closely resembling the conditions represented in the screening panel itself, which are largely metabolic disorders. Nordfalk and Ekstrøm argue that information about actual uses of samples is important for both discussion of the permissible range of uses and for public transparency and trust.

For many years, blood spots obtained in screening were routinely retained by programs in many U.S. states and other jurisdictions. Families—and of course the newborns themselves—were utterly unaware that the samples could be made available for a variety of research, public health, and other uses. When awareness of the existence and use of the samples emerged, it sparked protests and litigation by privacy advocates in Texas, in Minnesota, in Ireland, and elsewhere. The result of
litigation in Texas and Minnesota was destruction of millions of samples. Some have judged the destruction a disaster for public health. At best, the destruction is a clear example of what can happen when transparency in data use is lacking and people lose trust in data retention (IOM 2010).

Concerns about blood spot retention have adversely affected newborn screening—and thus health, too. While in most jurisdictions screening is considered routine, some jurisdictions permit parents to opt out. Some even require explicit parental consent for the procedure. Although by far the majority of parents are supportive of the method for identifying the low risk of serious disease, some object to the pain of the needle stick for their child, some have religious objections, and some find methods of communicating positive results unsatisfactory (Etchegary et al. 2016). A great concern of public health officials is that repurposing of screening samples will scare people away from allowing their children to be screened—and that even the knowledge that such uses may be possible will be sufficient disincentive for screening (Editorial 2011). To avoid risks that parents will opt out, the U.S. National Society of Genetic Counselors supports mandatory screening and uniform, transparent policies to govern any use of residual samples (Blount et al. 2014).

Jurisdictions such as Michigan now have moved to separate blood spot retention from initial screening by adopting opt-in consent mechanisms to govern retention of blood spots for research or public health. The Michigan BioTrust retains samples for up to 100 years. Use of the samples for research requires parents to opt into this possibility shortly after delivery. The Michigan BioTrust also contains legacy samples obtained between 1984 and 2010, before initiation of the opt-in process; these samples may be used in research unless individuals request their destruction (Michigan DHHS 2019). Parental opt-in rates to research use are low, possibly because parents are never asked to opt into this use during the period of excitement and confusion surrounding birth of a newborn. Pilot programs using social media to enhance awareness have not been particularly effective, although further development of these mechanisms might enhance transparency and understanding about ongoing collection and retention practices (Platt et al. 2013). Michigan’s experience suggests that without more discussion and changes in practice, opt-in consent is unlikely to yield blood spot retention rates that are useful for public health.

In general, commentators remain concerned that participants and the public are largely unaware of the possibility that blood spots might be stored and re-used. Some uses may be judged unacceptable by many (e.g. Thiel et al. 2014). A study of public preferences in Canada also indicates that people prefer that permission be asked and that discussion that increases understanding of the important research uses of the information for health may increase willingness to allow samples to be used without permission (Hayeems et al. 2016). A study in Illinois revealed similar preferences for permission, along with demographic differences in support. Willingness to support research use was more likely to be found among those with higher levels of education and less likely to be found among Blacks (Hart et al. 2015). These findings suggest the importance of reasonable expectations and the concerns that might arise with surprises about data use. They suggest that if uses of
retained samples are not transparent and people are surprised about data collection and use, protests like those that occurred in Minnesota and Texas may recur. Finally, they suggest continuing concerns about discrimination in research and health care in the United States.

Uses of retained blood spots to increase understanding of disease and improve population health may also result in information that can be useful for the health of individuals. For example, it may indicate that individuals are at highly elevated risks of developing certain cancers, Huntington’s disease, or dementias. Whether, when, and how, any such results should be returned to individuals raises additional vexing questions. On the one hand, information return may diminish trust, especially if the information is troubling, raises questions about identity, carries risks of discrimination, or creates anxiety and uncertainty. Imagine learning suddenly about risks for a serious disease, without forewarning and without any previous knowledge that a governmental entity had information about the risk. On the other hand, the fact that some have potentially life- or health-altering information about others, but do not reveal it when it could be beneficial or actionable in some way, is also troubling. It may be especially troubling if the institutions holding the information are public health authorities—authorities that have improving health as their mission. Moreover, if such failures to reveal critical health information come to light, they may cast added suspicion on the function of newborn screening programs. A further complexity is that if the secondary use of the samples occurred without awareness or any kind of agreement, no information will be available about individual preferences for return of results. These preferences may be especially important for adult-onset conditions and for conditions that are not clinically actionable, such as Huntington’s disease, but that nonetheless may be of great importance to non-health related choices such as whether to have children or save for retirement (Lewis and Goldenberg 2015).

Bloodspot retention presents one of the clearest illustrations of the need for communication and transparency. Yet the timing of sample collection remains challenging: samples are collected at birth, any consent is not obtained from the individual from whom the sample is taken but from parents, and sample utility may not become apparent until far in the future. A sense of surprise when sample use came to light generated concern, protest, destruction, and regulation. Not only was public health the loser, but in some jurisdictions, newborns also were put at risk when their parents refused the testing.

5.7 Biobanks

Recognition of complex relationships between genotypes and phenotypes and of the importance of genetic information in tailoring treatments to individuals has led to extensive efforts to collect databases of genetic information and banks of biological samples, especially when these can be linked to clinical records on a continuing basis. Some of these biobanks are established by entities administering health care to individuals. For example, Intermountain Healthcare has been collecting a
biobank of tissue samples and patient outcomes that can be used for precision medical care for patients with cancer and other conditions (Intermountain 2019).

Other biobanks are society wide. For example, UK Biobank is a charity that recruited 500,000 people from across the UK beginning in 2006; participants gave blood samples and agreed to have their medical information collected on a continuing basis. On entering the UK Biobank, participants gave what is called “broad consent”: consent to the ongoing collection and use of the data for “health-related” research (UK Biobank 2018). Patients who opt out of use of their NHS records for research pursuant to Caldicott 3 do not thereby opt out of UK Biobank; that would require separate notice to the biobank. All data made available for research by UK Biobank are de-identified; the biobank maintains identifying information only to properly establish linkages.

Denmark has one of the most extensive biobanking systems. The Danish Biobank Register lists the Danish National Biobank, which includes newborn blood spots since 1982; the Danish twin registry with data about 86,000 pairs; the Danish Cancer Biobank; and information from many other biobanks in Denmark. All together, these banks contain information about 5.7 million people, nearly the entire population of Denmark; they can be linked to other data such as location of residence that may be especially important for health (Staten Serum Institut 2020). Commentators have raised concerns that members of the Danish public are largely unaware of the numbers of studies that have used the Danish National Biobank (Nordfalk and Ekstrøm 2019). All studies using samples in the Biobank must be approved by the Danish Data Protection Agency, which can waive consent if there are no health-related risks and the use does not burden the participants. Some commentators have questioned whether the GDPR exceptions for research provide adequate protection in the case of biobanks, and have urged that national laws implement oversight and transparency standards (Staunton et al. 2019).

5.8 Registries

Registries collect information from or about people who share a common procedure or condition. The information they contain may be gleaned from medical records, from research, or directly from patients or their families. Some registries are maintained by public health agencies, some by practitioners or medical societies such as cardiology, some by researchers, some by charitable organizations devoted to particular diseases, and some even by pharmaceutical companies (Francis and Squires 2018). Registries may be international in scope, especially those devoted to rare diseases. Some registries only contain information as minimal as a list of contact information for people with a given condition who may be interested in participating in research about the condition. Others follow people throughout their lives, assessing disease course, treatment outcomes, and quality of care. People may be entered into registries as infants or children when they are first diagnosed, but in the United States will be asked to re-consent to participation when they reach adulthood (Francis 2014).
These registries may serve as valuable surveillance methods for assessing potential causes of increases or decreases in disease occurrence. For example, in the U.S. under the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute, every state maintains a tumor registry (NCI 2021). These registries collect information about all cases of cancer diagnosed in the U.S., including identifying patient information, tumor site and morphology, tumor stage, and initial treatment. Deidentified information from these registries is released to the SEER program to compile population statistics about cancer in the U.S. It was supposedly deidentified information that was released from the Illinois cancer registry that was the subject of the controversy described earlier in this chapter.

According to Fairchild and colleagues (2007), when the U.S. adopted the Privacy Act protecting information held by agencies of the federal government, it did not include a proposal for people to be notified when information about them was collected by the federal government. U.S. law governing the confidentiality of medical records does not require that patients be notified when information in medical records is shared with public health (HIPAA 2020). Patients are not made aware that data drawn from their medical records has been shared with the SEER cancer surveillance program, for example. Law professor Wendy Mariner (2007) has argued that these means for sharing information from patients’ health records with public health without consent are violations of the constitutional right to privacy. Others argue that the statutes are sufficiently narrowly tailored to address important state interests in protecting health (McLaughlin et al. 2010). We think it might even be argued that people who are receiving care from shared resources should reasonably expect that data from them may be used on an ongoing basis to improve health—but only if the considerations we outlined at the beginning of this volume such as transparency, protection from harm, and equity are appropriately implemented.

In the U.S., the non-profit Cystic Fibrosis Foundation maintains a registry of people with CF who agree and are treated in centers accredited by the Foundation. The registry supports research about CF care, uses data to improve care quality, and follows the natural history of the disease and care outcomes (Schechter et al. 2014). Similar registries exist elsewhere as well, such as the UK CF Registry. Drug companies also sponsor or support registries, sometimes in cooperation with disease foundations, in the effort to develop treatments that may prove effective.

There are also registries organized on an international basis to collect information from individuals with rare diseases. For example, the Orphan Disease Center has built registries that seek information directly from patients, including information from their medical records (ODC 2020). Although these international efforts may be critical for the understanding of rare conditions, their efforts to collect data also raise complex questions about meeting the legal and ethical requirements of different jurisdictions for data protection. Given the proliferation and variety of registries, it is not surprising that people might be confused about the entities that possess and use health information for many different purposes.
5.9 Information Gained in Medical Research

Information gained for or in the conduct of medical research might also be a useful source of information for public health. The many research biobanks are a good illustration of information that can be used to link individual genotypes and outcomes. The U.S. “all of us” precision medicine initiative to construct a biobank for a million lives is designed to be one of the largest and most representative of these, but there are many others. For example, the U.S. National Library of Medicine maintains the database of Genotypes and Phenotypes (dbGaP) to collect the data and results of genetics studies in humans funded by the federal government (NCBI 2019).

When patients agree to participate in medical research, they give explicit informed consent to the project. This informed consent typically includes information about who may have access to the information gained in the research, how the information will be protected, and how the information will be used. The consent may be very broad—for example, to any further use in research studies—but it is consent nonetheless. Thus research participants may have particular grounds for surprise if data gained in the research are used for new purposes. This kind of surprise is envisioned by the Fair Information Practice principle that individual consent should be sought for secondary uses of data. Even when people are willing to allow data originally collected as part of a research project to be shared, one study indicates, they still want to be asked (Ludman et al. 2010).

These features of medical research suggest that the case against repurposing of research data for public health is stronger than the case against repurposing data from health care. Reactions may be particularly intense when consent to the research was for a topic of particular health interest to the participants, when the public health purpose is far different, or when the use appears inequitable or stigmatizing. A further concern is that consent was sought initially and may not have included any further data use. The highly publicized case of genetic research involving the Havasupai Indian tribe illustrates. Researchers from Arizona State University conducted research on diabetes using blood samples from tribal members. Members of the tribe were encouraged to participate in the trial because of the particularly heavy health burden of diabetes for them. When researchers later used the samples that had been stripped of identifiers to study schizophrenia, migration patterns, and tribal inbreeding, the tribe were outraged. The tribe sued, and the University ultimately settled for compensation and destruction of the samples (Sterling 2011).

This of course was a second research study, not a study by public health. Notably, it also used data that were considered deidentified. But it is not difficult to imagine similar outrage if the state public health department had used the data to assess whether tribes were experiencing high burdens of schizophrenia or inbreeding. Finally, given the specific inclusion criteria for many research studies and the widespread availability of other data, it seems unlikely that medical research data will be uniquely important as a source for public health. In most circumstances, therefore, the balance of considerations would seem to weigh against repurposing data from
medical research for public health. At a very minimum, however, if it does occur, it
should be subject to the same constraints as the use of information from health care.

5.10 Direct to Consumer Testing, Including Genetic Testing

Direct to consumer testing is another potential resource for public health. Consumers
now submit samples to commercial companies such as 23andMe or Ancestry.com
for personal enjoyment and in some cases health testing. These companies offer
consumers information about their ancestral origins and possible relatives. They
also may provide lower-cost alternatives for identifying deleterious variants of
genes such as BRCA1 or BRCA2. For example, 23andMe now offers an approved
test for three of the most common variants of these genes that are associated with
high risks of breast cancer (23andMe 2019).

When individuals sign up for these direct to consumer products, they are given
terms of service and asked to consent to uses of the data. Companies vary widely in
the protections they provide to consumers (e.g. Bunnik et al. 2014). Jurisdictions
also vary widely in the legal protections that they give consumers. For example, the
U.S. the primary protection at the federal level is the Federal Trade Commission Act
(15 U.S.C. § 45 (2019)) prohibition of unfair or deceptive trade practices. This pro-
hibits companies from misrepresenting their data protection practices but does not
prohibit companies from explicitly declaring that they will be quite open about shar-
ing information they have gained. States also have their own protections of unfair
trade practices that may mirror the federal.

Controversy has erupted in the U.S. about whether these data bases may be used
for law enforcement purposes. For example, GEDmatch is an online service that
allows individuals to upload data that they have obtained from other sources such as
the direct to consumer testing companies. Individuals who upload their data may
search for other relatives who have submitted their DNA. Law enforcement accessed
GEDmatch to find relatives who might be linked to a killer who had been respon-
sible for multiple murders, rapes, and burglaries in California between 1974 and
1986, and whose identity had frustrated law enforcement. This successful search for
the “Golden State killer” highlighted the potential utility of these data bases to solve
crimes but also the privacy risks to people who had uploaded data without imagin-
ing its crime-fighting possibilities (Kaiser 2018). It is not hard to see that public
health might have similar interests in this information, for example if genetic mate-
rial is available about an unknown individual who has been the source of a deadly
infection.

Preliminary studies indicate public support for access to these databases to solve
violent crimes, crimes against children, or sexual assaults, but far less support than
when the access is to solve nonviolent crimes (Guerrini et al. 2018). This would
suggest that there might be support for public health to consult these data bases in
the unlikely event of the immediate need to identify someone who is at imminent
risk of transmitting deadly infections, but not otherwise. If public health were
permitted to do this, moreover, it should be subject to the ethical concerns we raised about contact tracing in Chapter 2. With these constraints, public health access would at least minimize additional risks to persons thus identified.

5.11 Smartphones and Smartphone Apps

According to estimates by researchers at the Pew Research Center, over 5 billion people used mobile phones in 2019; this represents just over 70% of the world’s population. About half of these devices are smartphones. These percentages are higher in advanced economies and more variable but growing rapidly in emerging economies. Smartphone ownership rates are higher among people who are better off, better educated and younger in age. Even in some advanced economies, younger adults are far more likely to own smart phones than older adults; in Japan, for example, nearly all adults under 35 (96%) own smart phones while far fewer adults over 50 have these devices (44%). To illustrate with one notable gap by education, in Nigeria 51% of people with a secondary education or more use smart phones, whereas only 6% with less education do so. In most emerging economies, men and women use smartphones at roughly comparable rates; the exception is India, where men (at 34%) are over twice as likely than women (at 15%) to own smartphones. Given these differences, data collected from smartphones will have notable gaps, particularly among the elderly, the less well educated, and the poor. These gaps are closing in more advanced economies but at least for now are growing in less advanced areas as smartphone adoption is rising vary rapidly among the young and better off (Taylor and Silver 2019).

Despite the gaps, smartphones are judged to be a significant advance in their potential for disease surveillance. Mobile phones allow information to be transmitted quickly, cheaply, and from locations where communication may otherwise be unavailable. Smartphones add information, too, especially about location in real time (e.g. Lee et al. 2016). Efforts to address malaria transmission have taken particular advantage of cellphones. The World Health Organization used smartphone technology to assist in the efforts to eliminate malaria in Bhutan. The technology allowed essential real-time transmission of geolocation information (WHO 2019). A group of researchers have used travel surveys and cellphone data to provide a more complete map of transmission of malaria in Bangladesh, including travel from urban residents to remote rural locations (Chang et al. 2019). Without the analysis of the mobile phone data, it was widely believed that travel to the countryside was infrequent and did not play a significant role in transmission; it is now understood that people travel home fairly often. Still, gaps in the data remain, as in rural areas cellphone towers are less frequent. An earlier study by some of the same researchers used mobile phone call data records to assess the travel patterns of 15 million cell phone owners in Kenya, allowing them to create detailed risk maps of the movement of malaria parasites (Wesolowski et al. 2012). A different group of researchers studying public health surveillance using mobile phones in sub-Saharan Africa
concluded that the primary use of cellphones in surveillance was enabling public health workers to transmit timely information about infectious diseases such as malaria or influenza (Brinkel et al. 2014). The transmissions often were done by public health workers using their own phones, raising questions about the source of repayment for the phone use. A major advantage of cell phone use, however, was that it allowed for two-way communication. This study concluded that cell phone use for surveillance in sub-Saharan Africa remained fragmented and small in scale. Cellphone data also were used during the Ebola epidemic to assess the need to control movement and the efficacy of travel restrictions (Peak et al. 2018).

Twitter, one of the most frequently used smartphone apps, also provides geolocation information. One group of researchers used information from Twitter to understand the triggers of influenza epidemics. These researchers found that the strongest predictors of influenza incidence were the host populations’ socio- and ethno-demographic properties, specific weather traditions, viral antigenic drift, and the host population’s land travel habits, among others. To reach these conclusions, they used multiple data sources including Twitter: insurance records capturing the dynamics of influenza-like illness, weather measurements, air travel and geographic proximity, Twitter data revealing long and short distance human movement patterns, and census data (Chattopadhyay et al. 2018).

In addition to the enhanced communication they provide, geolocation data in real time and the ability to track movement over time are at present the primary additional contributions of smartphone data to surveillance. As the research we have described indicates, these contributions may be highly informative additions to the data that can be accessed by public health from other sources. However, these capabilities enhance risks to individuals because they enabled them to be located and tracked in a way that information in static databases does not. These risks increase the need to ensure that, if these methods are used, they are coupled with the protections that we discussed in Chapter 2 for advanced COVID-19 surveillance.

In addition, geolocation information from smartphones enables individuals to be tracked over extended periods of time. In the words of the U.S. Supreme Court, “A cell phone faithfully follows its owner beyond public thoroughfares and into private residences, doctor’s offices, political headquarters, and other potentially revealing locales ….Moreover, the retrospective quality of the data here gives police access to a category of information otherwise unknowable” (Carpenter 2018, p. 13). If cell-phone information is routinely used to monitor individuals’ movements across time and space, individuals may come to fear that a shadowy Big Brother really is watching over them. Many commentators have raised the alarm that continual surveillance will chill freedom of thought and expression, especially when people are tracked over time by what they perceive to be agents of the government (Gullo 2016). These concerns indicate the importance of oversight, transparency, and limited use by public health of devices that can track movement over time.
Another potential new data frontier for public health is the presence of devices that can observe the individual directly. Robots may watch an individual’s every movement, recording everything the individual says and does. Wearables track exercise, sleep patterns, blood pressure, and blood sugar levels. Biosensors may identify whether individuals have taken medication, how that medication has been absorbed, and even features of the individual microbiome.

Domestic robots are common in many households in wealthier countries. “Companion” robots are advertised as a way to make it easier and safer for older people to live alone and remain in their own homes. “Buddy,” headquartered in France, advertises that it can provide companionable social interaction, medication reminders, easy access to the internet, and warnings about falls for seniors who would otherwise be isolated (Blue Frog Robotics 2019). But Buddy could just as easily send messages to public health about the health status of the elderly in their community. Care robots are increasingly used in care facilities to aid in feeding, hygiene, and monitoring; designed to look like pets, they can also provide comfort and measure distress. These robots are cost effective and can extend personnel when there are labor shortages. But they are also ethically controversial. A recent systematic review of ethical arguments about care robots concludes that respect for privacy, behavioral control, and deception were major ethical concerns (Vandemeulebroucke et al. 2018). This review concludes that transparency and stakeholder engagement are critical to determining whether robot use is ethically permissible in particular circumstances. Like in-home robots, these robots could conceivably transmit a great deal of information to public health—from information about the moods and possibly quality of life of people in their care, to information about abuse or neglect in care centers.

Wearables also collect a great deal of health information, often through smartphone apps. Fitbit measures steps and sleep patterns, multiple apps are on the market for measuring blood pressure, and the iPhone will measure breathing patterns and heart rate to help users with mindful meditation. Many people use these mechanisms for personal enhancement, but they also can be recommended by health care providers and tethered to electronic medical records. For example, the British National Institute for Health and Clinical Excellence recommends wearables for blood glucose monitoring in certain circumstances. The devices can also be used to track periods of fertility for people who want to conceive—or to avoid conception. Whether these constant monitors do more harm than good was the subject of the medical ethics debate between Oxford University and Cambridge University in 2018. Although Oxford won the debate for the side in favor of their use, considerable concerns were raised about the risks of third-party access to data on cellphones left casually around, data security protection, and access to data by cellphone providers and app designers for their own purposes. Another major issue was access by the government to data stored on cellphones. Such constant monitoring of bodily
functions may also medicalize everyday life, worrying the well and demanding hyper-vigilance on the part of the sick (Gilmartin et al. 2018).

Biosensors may be even more intrusive than wearables. In 2017, the U.S. Food and Drug Administration gave marketing approval for the antipsychotic Abilify MyCite in a form that allows tracking of whether the drug has been ingested (FDA 2017). It is not difficult to imagine applying the same technology to controlled substances such as opioids to assess whether they have been ingested by the patient for whom they were prescribed or been diverted. Such mechanisms could also be applied to whether women have taken their contraceptives, data of possible interest to boyfriends concerned about deception or Medicaid programs concerned about whether women are staying on prescribed birth control. Biosensors can measure dietary intake, alcohol consumption, or use of illegal drugs. The day is not too far off in which it will be possible to continuously monitor the entirety of the gut microbiome.

The potential for public health of these monitoring devices is not at all farfetched. There are reports of their use in China to ensure that people who have tested positive for COVID-19 remain in self-isolation. Moreover, they will yield information that cannot be acquired as readily in any other way. While we have presented these devices in a provocative manner, we have done so to reflect what may likely be public reactions to collection of data from them by public health. Any hint of such collection is likely to be met with revulsion and resistance. Reactions are likely to be especially strong if the collection comes from within the body. At present, we think, public health would be well advised to stay away from such biosensor measurements, except in very carefully defined circumstances in which robustly informed consent is obtained from willing participants.

5.13 Public Health Surveillance by Actors in the Private Sector

Many non-state actors are now engaging in what might be regarded as public health surveillance—that is, surveillance of public health even though not surveillance by public health. Many of these actors are using some of the novel forms of data that we have just described. These actors range from formally established partners with the World Health Organization, to health care providers, to internet search engines such as Google and social media sites such as Facebook.

Reliance on non-state actors for public health surveillance is not new. Amy Fairchild and colleagues describe how from colonial times innkeepers, ship operators, and even family members were obligated to report deadly contagion (Fairchild et al. 2007, p. 2). They locate the professionalization of public health—and the reservation of public health functions to public health—with the development of bacteriology in the late nineteenth century. We might also note that the late nineteenth century was a time of more general professionalization, not only of medicine
but also of law, although medicine was a particularly powerful instance of the trend. Kristin Luker’s study (1984) of the history of reproductive policy, for example, reveals how the efforts of physicians to take control from midwives changed policies around abortion and childbirth. That non-state actors might have resources and skills beyond what are available in the public sector is also not new. What might be newer today is the extent to which entities in the private sector are interested in surveilling for their own purposes along with public health and the scale on which they are so doing. And what is also newer are the kinds of data to which they have access and the power of the analytical tools they use.

Public health governmental agencies, moreover, are not the only entities engaging in activities that might be regarded as public health surveillance with new data forms. These private sector actors may not be scrutinized or regulated to the extent that government agencies are subject to oversight. In the U.S., in order to claim exemptions from federal income taxes, non-profit health care providers operating hospitals must perform community health needs assessments every three years and adopt implementation strategies to address the needs thus identified (IRS 2018). Data from electronic medical records are a frequent new source of data that may be used for public health purposes by non-profit health systems as well as by public health agencies. Direct-to-consumer genetic testing contains a wealth of health-related information about individuals. Credit card expenditures reveal health information from purchases and purchasing patterns; grocery store loyalty cards show dietary changes and indiscretions. Massive amounts of health information are collected, analyzed, and transmitted by internet search engines and through social media sites. This includes misinformation, too. Smart devices from robots to refrigerators are highly revealing about day to day activities.

In addition to the kinds of questions that raised by novel and large-scale data, additional issues of oversight are implicated when non-state actors engage in surveillance. In some jurisdictions, notably in the U.S., these actors may be subject to limited regulatory restrictions on what they do. At the U.S. federal level, for example the primary statute regulating non-state commercial actors who are not governed by HIPAA is the Federal Trade Commission Act prohibition on unfair and deceptive trade practices, 42 U.S.C. § 45. Non-state actors may have their own interests, consume resources in ways that duplicate or even perhaps undermine the efforts of public actors, present conflicts of interest, or perhaps even act in ways that are judged by others to be misconduct. To illustrate with a conflict before the advent of big data between a well-intentioned charity and public health, Fairchild and her colleagues (2007, p. 152) describe efforts of the National Foundation for Infantile Paralysis during the 1940s to get direct reports of polio cases. The Foundation argued that they needed the data to get proper services and treatment to patients; public health professionals were concerned about unnecessary duplication and lack of coordination and argued that the Foundation’s efforts should be supportive only. The tension reportedly faded with the development of polio vaccines.
5.13.1 WHO and Non-state Actors

Established over seventy years ago, the World Health Organization has become an international body that can identify and respond to public health emergencies and that can provide information and support for global public health. WHO was created at a time when there were few other international organizations with a public health mission. But WHO has not been uniformly successful. The new International Health Regulations that entered into force in 2007 (2005) vastly expanded expectations for surveillance, allowing WHO to receive reports from non-state actors and to investigate potential health emergencies of international concern beyond the limited list of contagious diseases that had previously been within its purview. Acting under these regulations, WHO was criticized for overreacting to the outbreak of influenza in Mexico in 2009 and for reacting far too slowly to the Ebola outbreak in West Africa in 2014. Criticism of the handling of the Ebola epidemic led to extensive debate over the effectiveness of the organization in addressing epidemics and in coordinating strategies to address their containment and reduction.

The debate over the role and effectiveness of WHO had not diminished even by 2019: “The planet’s premier health agency has announced drastic reforms. Critics say they aren’t drastic enough” (Kupferschmidt 2019). Problems cited by critics included persistent shortfalls in funding and difficulties in coordination between WHO headquarters in Geneva and regional offices. Eigil Sørensen, a former WHO official who is now on the public health faculty at Thammasat University in Thailand, writes that national public health agencies, once dependent on WHO for technical assistance, have gained a great deal in competence although they continue to rely on WHO for its publications on global health statistics (Sørensen 2018). Over the years, Sørensen observes, organizations such as the Global Fund to Fight AIDS, Tuberculosis and Malaria, The Vaccine Alliance, and Unitaid have arisen in part out of frustration with what they judged to be the inadequacy of WHO efforts. But these non-state organizations may have their own priorities—priorities that may not always align with the priorities of WHO or with the priorities of other non-state actors.

The challenge for WHO is what role it may be able to play as new actors emerge in surveillance and other efforts to address public health. It may be able to play a mobilizing role in forming coalitions to respond to new pandemics or epidemics. Or, WHO may try to find new income streams that allow it to set budgetary priorities. Or, it might focus its role on serving as a forum to build consensus sustaining public health values, perhaps becoming an honest broker in disputes among member states. If non-state actors play increased roles in public health emergencies, as they are doing around the globe, unless WHO has adequate funding it may be in a position of stepping back into roles of advising and encouraging rather than serving as an active participant.

Under the International Health Regulations, the WHO is now permitted to take reports of potential health events of concern from non-state actors and to investigate these reports according to established science (WHO 2016, Art. 9). In 2016, WHO
adopted a framework of engagement with non-state actors that recognized their importance in the promotion of global public health activities but also aimed to protect “WHO’s integrity, reputation, and public health mandate” (WHO 2016, ¶ 4). This framework also included separate provisions for NGOs, commercial entities, philanthropic foundations, and academic institutions. In explaining the need for the framework, WHO singles out concerns for transparency, undue influence, conflicts of interests, and potential risks to its credibility raised by partnerships with non-state actors. WHO is also concerned that through the support they are given non-state actors may acquire competitive advantages, pursue their own interests, and whitewash their images. The WHO Director-General now reports annually on the status of engagement with non-state actors. The 2019 Report of the WHO director-general indicated both success with and ongoing challenges of establishing partnerships with non-State entities under the framework; challenges included operationalizing the requirement that non-State actors confirm that they do not further the interests of the tobacco industry and cosponsoring global health events with private sector entities with potential commercial interests (WHO director general 2019). It remains to be seen how the role of WHO in conjunction with non-state actors will continue to evolve.

If the US had followed through on withdrawing from the WHO, it would have been even more dependent on non-state actors for surveillance because of the funding loss.

5.13.2 U.S. Non-profit Hospitals and Community-Based Needs

In the U.S., non-profit hospitals are expected to engage in surveillance-like activities in order to qualify for exemptions from federal income taxes. These “community-based needs assessments” were introduced by the Affordable Care Act (ACA) and require the hospital to identify health needs of their communities and develop plans to address these needs as a condition to maintaining their tax exemptions. If the entire ACA is struck down by the Supreme Court as the Trump administration urged (Federal Respondents 2020), this statutory requirement would no longer stand, although hospitals would still need to demonstrate that they meet the standards of the Internal Revenue Service charitable tax exemption.

The needs assessment requirement is an opportunity for hospitals to coordinate with local health departments but also raised the concern that there may be inadequate consultation with public health in developing priorities that engage the public and are equitable (ASTHO 2012). A 2019 study of hospitals in North Carolina suggests there may indeed be reason for concern; the bulk of hospital expenditures were for financial assistance to individual patients rather than for activities aimed to improve community health or address social determinants of health (Fos et al. 2019). Another study suggested the irony that hospitals in areas of very high health needs may face demands to devote more resources to meeting patient needs that reduce their ability to contribute to the overall health of their communities (Singh et al. 2018).
5.13.3 **Internet Search Engines: Google**

When Google launched “Google Flu Trends” in 2008 and claimed to have predicted influenza activity in advance of public health, it caught public attention. The algorithm later required adjustment, however, because of changes in search queries (Cook et al. 2011), and was discontinued. But interest remains in the improved use of internet search engine data with machine learning to detect health trends such as influenza outbreaks. For example, a collaboration between researchers in Mexico and researchers in Boston has recently developed models for predicting influenza in several Latin American countries where disease burdens are high and prediction therefore quite useful (Clemente et al. 2019).

Google itself continues to be interested in health surveillance. DeepMind Health is the medical unit of an AI company acquired by Google in 2014; it was brought directly into Google in 2018. DeepMind has contracts with the British National Health Service, raising concerns among observers that Google will have unprecedented access to patient health data (Stokel-Walker 2018). An earlier agreement between DeepMind and the NHS was determined by the Information Commissioner’s Office to have violated patient data protection regulations (ICO 2017a). Under that agreement, Royal Free Hospital in London had agreed to share identifiable data about 1.6 million patients to test the clinical safety of an application, Streams, used to identify risks for acute kidney injury. Royal Free had believed that the use of the data qualified as processing for direct patient care and had not required that the data be anonymized. Royal Free also had not been transparent with patients about the use of the data. The Information Commissioner found that this was a use of personal data that patients would not reasonably expect and that the data use was not what was minimally necessary to achieve the purpose of the safety testing (ICO 2017b). The data use was thus determined to have violated a number of the Caldicott principles discussed earlier in this chapter.

5.13.4 **Social Media: The Facebook Example**

Many observers now regard the ability to understand what people have in common as a signal achievement of social media. Social media sites such as Facebook have developed algorithms that create pathways to link people together. These algorithms can be adjusted by machine learning techniques and refined in specific areas in considerable detail. But they have of course also brought much criticism of social media. Facebook is at the center of these controversies, as it is the largest social media site with over two and a half billion users in 2020. Facebook also owns Instagram and WhatsApp, among many other companies.

Facebook uses algorithms to prioritize timelines, send news feeds, and serve content from advertisers. In 2018, in response to criticism that they were circulating propaganda or fake news that may have distorted electoral judgments, and that they
were prioritizing brand advertising over meaningful interactions with friends, Facebook adopted a new algorithm for structuring news distributions to users. The algorithm itself is protected intellectual property and not open to scrutiny, but according to Montells (2019) appears to prioritize factors such as how frequently a friend has visited another friend’s page (affinity), the quality of the post for generating clicks (clickbait), the interactions with the post by others in the network, and the recency of the post.

Through algorithm mediated interactions, Facebook can realize—or undermine—public health goals in several ways. Health information is frequently shared through news feeds. Facebook has announced that it would try to stop misinformation about vaccinations from circulating through its news feeds; it will no longer recommend anti-vaccination groups or allow advertising that contains false information about vaccines (Thebault 2019). It is taking similar steps against groups that post misinformation about supposed cancer cures such as baking soda, apple cider vinegar, or frankincense, and supposed cures for autism such as swallowing bleach (Ohlheiser 2019). However, these steps only downrank misinformation and groups sharing it. Facebook continues to host many such groups and simple google searches will reveal them to interested users (e.g. VRM 2019).

For COVID-19, Facebook has become somewhat more aggressive against misinformation. In March 2020, the site launched an “Information Center” which it features at the top of the news feed to convey accurate information about the pandemic. A month later, it extended fact-checking, attached warnings to debunked stories, reduced the distribution of these stories, removed misinformation that could cause imminent physical harm (such as that drinking bleach could cure COVID-19), and began sending news feed messages to people who had interacted with such harmful misinformation (Rosen 2020). Removing misinformation and sending warnings to people who might have accepted it represents new steps for Facebook.

On Facebook, friends may also share news about their own health or the health of others they know. News feeds will report happy news such as births or a child’s first steps. They will also reveal less happy news such as daily reports of a grueling course of chemotherapy or shock at a recent diagnosis. People can also join groups devoted to specific health issues. Some of these groups are open membership while others are “private” groups requiring approval for membership. Groups can also be hidden, so that they can only be found by their members.

All of this information—what people say, who people friend, what groups people join, and what information attracts attention—can be used for surveillance. The information in social media sites such as Facebook may be particularly useful because it may contain not only information about individual health conditions and health trends, but because it will also contain information about attitudes towards health and health care. Through Facebook, public health may not only learn whether someone has influenza, but also whether somebody—or many somebodies—is taking the possibility of an influenza outbreak seriously. Research has also suggested that data from social media sites such as Twitter, in combination with other data digital data streams such as reports from smart thermometers, may facilitate providing early warning of emerging outbreaks of COVID-19 (Kogan et al. 2020).
Facebook’s algorithm may also be put to service in making inferences about the health situation of particular individuals. Facebook now engages in suicide prevention efforts that have caught national attention. The company will call emergency services if it believes a user may be at very high risk because of the user’s interactions on its platform. It uses an algorithm to assess risk; although it has offered some details on the algorithm, the full algorithm has not been made public. Ian Barnett and John Torous (2019) have recently criticized this limited transparency for medical research, pointing out that the credentials of Facebook’s Community Operations team who reviews these data and the outcomes from approximately 3500 notifications to local emergency services to date are not fully clear. Barnett and Torous also ask whether a practice such as this should more properly be for public health:

Facebook’s suicide prevention efforts leads to the question of whether this falls under the scope of public health. Considering the amount of personal medical and mental health information Facebook accumulates in determining whether a person is at risk for suicide, the public health system it activates through calling emergency services, and the need to ensure equal access and efficacy if the system does actually work as hoped, the scope seems more fitting for public health departments than a publicly traded company whose mandate is to return value to shareholders. What happens when Google offers such a service based on search history, Amazon on purchase history, and Microsoft on browsing history? In an era where integrated mental health care is the goal, how do we prevent fragmentation by uncoordinated innovation? (565)

Youth suicide is a major public health problem in the U.S. and Facebook’s efforts may be beneficial. But it is also easy to imagine that they might prove quite intrusive. If the predictive algorithms are wrong, especially if they lead to false positives in a low incidence population, the results may be very disturbing. And it is not hard to imagine algorithms such as these being put to many other uses: trying to detect whether someone is at risk of committing violent acts against others, coming down with a dangerous infection, experiencing an opioid overdose, or even experiencing early signs of an undetected cancer.

All of these uses of social media have potential benefits for individuals or for those they might harm. But they are deeply troubling as well. The inequities of predictive algorithms are well known; algorithms can yield skewed judgments if the data themselves are skewed in some way. For example, many have argued that predictive policing or detention algorithms in use in the U.S. are skewed by race and that algorithms predicting the likelihood of addiction may be skewed by poverty (Eubanks 2018). There is little or no transparency about how these algorithms work because they are protected intellectual property.

Finally, in the U.S. especially there is limited oversight of how social media may be developing public health functions. Social media sites are commercial entities with interests that may not always align with the public interest. Their support through advertising in particular may influence how they are structured, how they appeal to users, and how they harvest and use data shared through their sites.
Many of these kinds of data—patient medical records, bloodspots retained from newborn screening, and some tissue samples in biobanks, some information in registries, and some information from research—were initially produced in the course of individual health care. Whether individuals are likely to expect information from their health care to be assembled and repurposed for public health is not clear. Individuals may expect their records to be used for their health care and perhaps for the improvement of health care by the providers from which they have received treatment. They are less likely to expect, however, that information gleaned in their care might be used to serve the public more generally, even when the use is in the service of health overall. This is especially true when the information concerns children or others who may be unaware that they even had the medical care that was the source of the information. The calls for destruction of bloodspots retained in newborn screening are a cautionary tale for what may be the result of what people regard as unfair surprise.

We would not necessarily draw the conclusion, however, that in most cases individual consent should be required for public health use of data gained in medical care. Rather, these are uses that must be considered in light of the considerations we raised at the outset of this volume. Openness and transparency are particularly important to avoid surprises such as those with newborn screening. As we have suggested, electronic means of communication available for use with patient records may make it far easier to communicate with patients about public health uses of medical records and their benefits. In addition, public health uses of information from medical care must not increase risks to patients. When public health uses information drawn from medical care, it must be subject to the same standards that are applied to these medical records for data security and disclosures. Public health must not access identifiable patient information without performing adequate security risk assessments and implementing appropriate protections against security threats. Further, public health must not use information in ways that significant numbers of people would find objectionable. For example, the British Wellcome Trust commissioned a study of patient attitudes towards commercial access to health records. The study found that participants were more likely to accept the access when it was for clear public benefit rather than for private benefit and that uses such as for medical research in the public interest could be acceptable while uses for insurance underwriting or for marketing would not be (Ipsos MORI 2016). The study also found that data security was essential to the public. Public health should also attend to equity issues in the use of data it has gained from medical care. There should also be oversight of public health attention to these concerns, as with the Caldecott guardian structure in the UK.

New data, new users, and new analytic methods will continue to present ethical challenges for public health. In weighing the ethics of these uses, public health will need to consider what these uses and users may contribute to the enterprise of
improving overall health. Enhanced risks and ways of mitigating them will need to be identified. Public reactions of surprise are likely if data are repurposed for goals they do not accept or if the data collection or use seems unduly intrusive or threatening. Equity is an ongoing challenge. In general, public health will need to do all that it can to ensure that the ethical considerations we put forth in the beginning of this volume are addressed in ways that can enhance trust and thus sustain surveillance.

References

23andMe. 2019. Do You Speak BRCA? https://www.23andme.com/brca/. Accessed 24 July 2020.
Aarestrup, F.M., A. Albeayatti, W.J. Armitage, et al. 2020. Towards a European Health Research and Innovation Cloud (HRIC). *Genome Medicine* 12: 18.
Association of State and Territorial Health Officials (ASTHO). 2012. Maximizing the Community Health Impact of Community Health Needs Assessments Conducted by Tax-exempt Hospitals. http://www.astho.org/Programs/Access/Community-Health-Needs-Assessment/Consensus-Statement/. Accessed 29 July 2020.
Barnett, Ian, and John Torous. 2019. Ethics, Transparency, and Public Health at the Intersection of Innovation and Facebook’s Suicide Prevention Efforts. *Annals of Internal Medicine* 170 (8): 565–566.
Benkler, Yochai. 2019. Don’t Let Industry Write the Rules for AI. *Nature* 569 (7755): 161.
Blount, Carrie, Cate Walsh Vockley, Amy Gaviglio, Michelle Fox, Brook Croke, Lori Williamson Dean, and The Newborn Screening Task Force on behalf of the NSGC Public Policy Committee. 2014. Newborn Screening: Education, Consent, and the Residual Blood Spot., The Position of the National Society of Genetic Counselors. *Journal of Genetic Counseling* 23 (1): 16–19.
Blue Frog Robotics. 2019. Buddy for Senior Citizen(s) Who Live at Home. http://www.bluefrogrobotics.com/en/buddy-elderly-senior-robot-alone/. Accessed 29 July 2020.
Brief for the Federal Respondents. 2020. *California v. Texas*, Nos. 19-840 and 19-1019. United States Supreme Court. https://www.supremecourt.gov/DocketPDF/19/19-840/146406/2020062520555069_19-840bsUnitedStates.pdf. Accessed 24 July 2020.
Brinkel, Johanna, Alexander Krämer, Ralf Krumkamp, Jürgen May, and Julius Fobil. 2014. Mobile PhoneBased mHealth Approaches for Public Health Surveillance in Sub-Saharan Africa: A Systematic Review. *International Journal of Environmental Research and Public Health* 11 (11): 11559–11582.
Bunnik, Eline M., A. Cecile, J.W. Janssens, and Maarje H.N. Schermer. 2014. Informed Consent in Direct-to-Consumer Personal Genome Testing: The Outline of a Model Between Specific and Generic Consent. *Bioethics* 28 (7): 343–351.
Carlisle, Nate. 2019. Utah’s Controlled Substance Database susceptible to hacking, auditors warn. The Salt Lake Tribune (February 13). https://www.sltrib.com/news/politics/2019/02/13/utahs-controlled/. Accessed Feb 20, 2021.
*Carpenter v. United States*. 2018. https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf. Accessed July 29, 2020.
Chang, Hsiao-Han, Amy Wesolowski, Ipsita Sinha, Christopher G. Jacob, Ayesha Mahmud, Didar Uddin, Sazid Ibna Zaman, Md Amir Hossain, M. Abul Faiz, Aniruddha Ghose, Abdullah Abu Sayeed, M. Ridwanur Rahman, Akramul Islam, Mohammad Jahirul Karim, M. Kamar Rezwan, Abul Khair Mohammad Shamsuzzaman, Sanya Tahmina Jhora, M.M. Akhtaruzzaman, Eleanor Drury, Sonia Gonçalves, Mihir Kekre, Mehul Dhorda, Ranitha Vongpromek, Olivo Miotto, Kenth Engø-Monsen, Dominic Kwiatkowski, Richard J. Maude, and Caroline Buckee. 2019. Mapping Imported Malaria in Bangladesh Using Parasite Genetic and Human Mobility Data. *Epidemiology and Global Health*. eLife 8: e43481. https://doi.org/10.7554/eLife.43481. Accessed 29 July 2020.
References

Chattopadhyay, Ishanu, Emre Kiciman, Joshua W. Elliott, Jeffrey L. Shaman, and Andrey Rzhetsky. 2018. Conjunction of Factors Triggering Waves of Seasonal Influenza. Epidemiology and Global Health. eLife 7: 330756. https://doi.org/10.7554/eLife.30756. Accessed 29 July 2020.

Chan, Tom, Concetta Tania Di Iorio, Simon de Lusignan, Daniel Lo Russo, Craig Kuziemsky, and Siaw-Teng Liaw. 2016. The UK National Data Guardian for Health Care’s Review of Data Security, Consent and Opt-outs: Leadership in Balancing Public Health with Rights to Privacy? Journal of Innovation in Health Informatics 23 (3): 627–632.

Clayton, Ellen W., Colin M. Halverson, Nila A. Sathe, and Bradley A. Malin. 2018. A Systematic Literature Review of Individuals’ Perspectives on Privacy and Genetic Information in the United States. PLoS One 13 (10): e0204417.

Clemente, Leonardo, Fred Lu, and Mauricio Santillana. 2019. Improved Real-Time Influenza Surveillance: Using Internet Search Data to Find Eight Latin American Countries. JMIR Public Health Surveillance 5 (2): e12214.

Cook, Samantha, Corrie Conrad, Ashley L. Fowlkes, and Matthew H. Mohebbi. 2011. Assessing Google Flu Trends Performance in the United States During the 2009 Influenza Virus A (H1N1) Pandemic. PLoS One. https://doi.org/10.1371/journal.pone.0023610. Accessed 29 July 2020.

Coyle, D. Tyler. 2017. Sentinel System Overview. https://www.fda.gov/media/110795/download. Accessed 29 July 2020.

Curfman, Gregory D. 2017. All-Payer Claims Databases After Gobille. Health Affairs Blog (Mar 3). https://www.healthaffairs.org/do/10.1377/hblog20170303.058995/full/. Accessed 22 July 2020.

digitalhealth. 2016. Caldicott 3: Easy to Say, Hard to Do? (Oct 6). https://www.digitalhealth.net/2016/10/caldicott-3-easy-to-say-hard-to-do/. Accessed 23 July 2020.

Dillard v. O’Kelley. 961 F.3d 1048 (8th Cir. 2020).

Editorial. 2011. There Will Be Blood. Nature 475 (7355): 139.

Etchegary, Holly, Stuart G. Nicholls, Laure Tessier, Charlene Simmonds, Beth K. Potter, Jamie C. Brehaut, Daryl Pullman, Robyn Hayeems, Sari Zelenietz, Monica Lamoureux, Jennifer Milburn, Lesley Turney, Pranesh Chakraborty, and Brenda Wilson. 2016. Consent for Newborn Screening: Parents’ and Healthcare Professionals’ Experiences of Consent in Practice. European Journal of Human Genetics 24 (11): 1530–1534.

Eubanks, Virginia. 2018. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: St. Martin’s Press.

European Union (EU). 2019. Recital 54: Processing of Sensitive Data in Public Health Sector. https://gdpr-info.eu/recitals/no-54/. Accessed 29 July 2020.

Fairchild, Amy L., Daniel Wolfe, James Keith Colgrove, and Ronald Bayer. 2007. Searching Eyes: Privacy, the State, and Disease Surveillance in America. Berkeley: University of California Press.

Ferguson, Andrew Guthrie. 2017. The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. New York: NYU Press.

Floridi, Luciano. 2014. Open Data, Data Protection, and Group Privacy. Philosophy & Technology 27: 1–3.

Food and Drug Administration (FDA). 2017. FDA Approves Pill with Sensor That Digitally Tracks If Patients Have Ingested Their Medication. https://www.fda.gov/news-events/press-announcements/fda-approves-pill-sensor-digitally-tracks-if-patients-have-ingested-their-medication. Accessed July 29, 2020.

Fos, Elmer, Michael Thompson, Christine Elnitsky, and Elena Platonova. 2019. Community Benefit Spending Among North Carolina’s Tax-Exempt Hospitals After Performing Community Health Needs Assessments. Journal of Public Health Management and Practice 25 (4): E1–E8.

Francis, Leslie P. 2014. Adult Consent to Continued Participation in Patient Registries. St Louis University Journal of Health Law & Policy 7 (2): 389–405.

Francis, John G., and Leslie P. Francis. 2014. Privacy, Confidentiality, and Justice: Using Large-Scale Sets of Health Data to Improve Public Health. Journal of Social Philosophy 45 (3): 408–431.
Francis, Leslie P., and Michael Squires. 2018. Patient Registries and Their Governance: A Pilot Study and Recommendations. Indiana Health Law Review 16: 43–65.

Francis, Leslie P., Margaret P. Battin, Jay Jacobson, and Charles B. Smith. 2009. Syndromic Surveillance and Patients as Victims and Vectors. Journal of Bioethical Inquiry 6 (2): 187–195.

Friedler, Sorelle A., Carlos Scheidegger, and Suresh Venkatasubramanian. 2016. On the (im)possibility of Fairness. https://arxiv.org/pdf/1609.07236.pdf?source=post_page%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D%2D. Accessed 27 July 2020.

Friedman, Daniel J., R. Gibson Parrish, and David A. Ross. 2013. Electronic Health Records and US Public Health: Current Realities and Future Promise. American Journal of Public Health 103 (9): 1560–1567.

General Data Protection Regulation (GDPR). 2020. https://gdpr-info.eu/. Accessed 21 July 2020.

Gilmartin, Chris, Edward H. Arbe-Barnes, Michael Diamond, Sasha Fretwell, Euan McGivern, Myrto Vlazaki, and Liment Zhu. 2018. Varsity Medical Ethics Debate 2018: Constant Health Monitoring—The Advance of Technology into Healthcare. Philosophy, Ethics, and Humanities in Medicine 13: 12. https://doi.org/10.1186/s13010-018-0065-0. Accessed 29 July 2020.

Gobeille v. Liberty Mutual Insurance Company, 577 U.S. ___, 136 S. Ct. 936 (2016).

Google. 2020. Privacy Policy (updated July 1). https://policies.google.com/privacy?hl=en-US. Accessed 29 July 2020.

Guerrini, Christi J., Jill O. Robinson, Devan Petersen, and Amy L. McGuire. 2018. Should Police Have Access to Genetic Genealogy Databases? Capturing the Golden State Killer and Other Criminals Using a Controversial New Forensic Technique. PLOS Biology (Oct 2). https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2006906. Accessed 29 July 2020.

Gullo, Karen. 2016. Surveillance Chills Speech—As New Studies Show—And Free Association Suffers. https://www.eff.org/deeplinks/2016/05/when-surveillance-chills-speech-new-studies-show-our-rights-free-association. Accessed 29 July 2020.

Hart, Alexa, Michael Petros, Joel Charrow, Claudia Nash, and Catherine Wicklund. 2015. Storage and Use of Newborn Screening Blood Specimens for Research: Assessing Public Opinion in Illinois. Journal of Genetic Counseling 24 (3): 482–490.

Hayeems, Robin Z., Fiona A. Miller, Carolyn J. Barg, Yvonne Bombard, Celine Cressman, Michael Painter-Main, Brenda Wilson, Julian Little, Judith Allanson, Denise Avard, Yves Giguerre, Pranesh Chakraborty, and June C. Carroll. 2016. Using Newborn Screening Bloodspots for Research: Public Preferences for Policy Options. Pediatrics 137 (6): e20154143.

Health Insurance Portability and Accountability Act Rules (HIPAA). 2020. 45 C.F.R. Parts 160, 162, and 164. https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/combined/hipaa-simplification-201303.pdf. Accessed 23 July 2020.

Hoffman, Sharona. 2018. Big Data Analytics: What Can Go Wrong. Indiana Health Law Review 15: 227–246.

Hoffman, Sharona, and Andy Podgurski. 2013. Big Bad Data: Law, Public Health, and Biomedical Databases. Journal of Law, Medicine and Ethics 41 (Suppl. 1): 56–60.

Information Commissioner’s Office (ICO). 2017a. Royal Free – Google DeeMind Trial Failed to Comply with Data Protection Law (July 3). https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/. Accessed 29 July 2020.

———. 2017b. Letter from Elizabeth Denham, ICO Commissioner, to Sir David Sloman (July 3). https://ico.org.uk/media/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf. Accessed 29 July 2020.

Institute of Medicine (IOM). 2010. Challenges and Opportunities in Using Residual Newborn Screening Samples for Translational Research. Washington, DC: National Academies Press.

Intermountain Healthcare. 2019. Intermountain Biorepository. https://intermountainhealthcare.org/about/transforming-healthcare/innovation/intermountain-biorepository/about/. Accessed 29 July 2020.
References

Internal Revenue Service (IRS). 2018. Community Health Needs Assessment for Charitable Hospital Organizations—Section 501(r)(3) (Nov 18). https://www.irs.gov/charities-non-profits/community-health-needs-assessment-for-charitable-hospital-organizations-section-501r3. Accessed 29 July 2020.

Ipsos MORI. 2016. The One-Way Mirror: Public Attitudes to Commercial Access to Health Data. https://www.ipsos.com/sites/default/files/publication/5200-03/sri-wellcome-trust-commercial-access-to-health-data.pdf. Accessed 29 July 2020.

Kaiser, Jocelyn. 2018. We Will Find You: DNA Search Used to Nab Golden State Killer Can Home in on About 60% of White Americans. Science (Oct 11) [online]. https://www.sciencemag.org/news/2018/10/we-will-find-you-dna-search-used-nab-golden-state-killer-can-home-about-60-white. Accessed 29 July 2020.

Kogan, Nicole E., Leonardo Clemente, Parker Liautaud, Justin Kaashoek, Nicholas B. Link, Andre T. Nguyen, Fred S. Lu, Peter Huybers, Bernd Resch, Clemens Havas, Andreas Petutschng, Jessica Davis, Matteo Chinazzi, Backtosch Mustafa, William P Hanage, Alessandro Vespignani, and Mauricio Santillana. 2020. An Early Warning Approach to Monitor COVID-19 Activity with Multiple Digital Traced in Near Real-Time. (July 7). https://arxiv.org/pdf/2007.00756.pdf. Accessed 24 July 2020.

Kostkova, Patty. 2018. Disease Surveillance Data Sharing for Public Health: The Next Ethical Frontiers. Life Sciences, Society and Policy 14: 16–21.

Kruse, Clemens Scott, Anna Stein, Heather Thomas, and Harmander Kaur. 2018. The us of Electronic Health Records to Support Population Health: A Systematic Review of the Literature. Journal of Medical Systems 42 (11): 214.

Kupferschmidt, Kai. 2019. The Planet’s Premier Health Agency Has Announced Drastic Reforms. Critics Say They Aren’t Drastic Enough. Science (Mar 12) [online]. https://www.sciencemag.org/news/2019/03/planet-s-premier-health-agency-has-announced-drastic-reforms-critics-say-they-arent. Accessed 29 July 2020.

Lee, Elizabeth C., Jason M. Asher, Sandra Goldlust, John D. Kraemer, Andrew B. Lawson, and Shweta Bansal. 2016. Mind the Scales: Harnessing Spatial Big Data for Infectious Disease Surveillance and Inference. The Journal of Infectious Diseases 214 (Suppl. 4): S409–S413.

Lewis, Michelle Huckaby, and Aaron J. Goldenberg. 2015. Return of Results from Research Using Newborn Screening Dried Blood Samples. The Journal of Law, Medicine, & Ethics 43 (3): 559–568.

Lewis, Michelle Huckaby, Michael E. Scheurer, Robert C. Green, and Amy L. McGuire. 2012. Research Results: Preserving Newborn Blood Samples. Science Translational Medicine 4 (159): 159cm12.

Ludman, Evette J., Stephanie M. Fullerton, Leslie Spangler, Susan Brown Trinidad, Monica M. Fujii, Gail P. Jarvik, Eric B. Larson, and Wylie Burke. 2010. Glad You Asked: Participants’ Opinions of Re-Consent for dbGaP Data Submission. Journal of Empirical Research on Human Research Ethics 5 (3): 9–16.

Mariner, Wendy K. 2007. Mission Creep: Public Health Surveillance and Medical Privacy. Boston University Law Review 87: 347–396.

McLaughlin, Robert H., Christina A. Clarke, LaVera M. Crawley, and Sally L. Glaser. 2010. Are cancer registries unconstitutional? Social Science & Medicine 70 (9): 1295–1300.

McLennan, Stuart, Leo Anthony Celi, and Alena Buyx. 2020. COVID-19: Putting the General Data Protection Regulation to the Test. JMIR Public Health Surveillance 6 (2): e19279.

Menderski, Maggie. 2014. Superfund Legacy Remains in Taylorsville and Beyond. The State Journal Register (Sept 13) [online]. https://www.sj-r.com/article/20140913/NEWS/140919731. Accessed 29 July 2020.

Meyers, Jessica M. 2019. Artificial Intelligence and Trade Secrets. American Bar Association. Landslide II (3). https://www.americanbar.org/groups/intellectual_property_law/publications/landslide/2018-19/january-february/artificial-intelligence-trade-secrets-webinar/. Accessed 27 July 2020.
Singh, Simone R., Geri R. Cramer, and Gary J. Young. 2018. The Magnitude of a Community’s Health Needs and Nonprofit Hospitals’ Progress in Meeting Those Needs: Are We Faced With a Paradox? *Public Health Reports* 133 (1): 75–84.

Sørensen, Eigil. 2018. Challenges for the World Health Organization. *Tidsskr Nor Legeforen* 138 (1). [https://doi.org/10.4045/tidsskr.17.0412](https://doi.org/10.4045/tidsskr.17.0412). Accessed 5 Aug 2020.

*Southern Illinoisan v. Illinois Department of Public Health*, 844 N.E.2d 1 (Ill. 2006).

Statens Serum Institut. 2020. The Danish Biobank Register. [https://www.danishnationalbiobank.com/danish-biobank-register](https://www.danishnationalbiobank.com/danish-biobank-register). Accessed 28 July 2020.

Staunton, Ciara, Santa Slokenberga, and Deborah Mascalzoni. 2019. The GDPR and the Research Exemption: Considerations on the Necessary Safeguards for Research Biobanks. *European Journal of Human Genetics* 27: 1159–1167.

Sterling, Robyn L. 2011. Genetic Research Among the Havasupai: A Cautionary Tale. *AMA Journal of Ethics (Virtual Mentor)* 13 (2): 113–117.

Stokel-Walker, Chris. 2018. Why Google Consuming DeepMind Health Is Scaring Privacy Experts. *Wired* (Nov 18) [online]. [https://www.wired.co.uk/article/google-deepmind-nhs-health-data](https://www.wired.co.uk/article/google-deepmind-nhs-health-data). Accessed 29 July 2020.

Sweeney, Latanya, Michael von Loewenfeldt, and Melissa Perry. 2018. Saying It’s Anonymous Doesn’t Make It So: Re-identifications of “Anonymized” Law School Data. *JOTS Technology Science* (Nov 13) [online]. [https://techscience.org/a/2018111301/](https://techscience.org/a/2018111301/). Accessed 27 July 2020.

Taylor, Kyle, and Laura Silver. 2019. Smartphone Ownership Is Growing Rapidly Around the World, But Not Always Equally. Pew Research Center (Feb 5). [https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/02/Pew-Research-Center_Global-Technology-Use-2018_2019-02-05.pdf](https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/02/Pew-Research-Center_Global-Technology-Use-2018_2019-02-05.pdf). Accessed 20 Feb 2021.

Thebault, Reis. 2019. Facebook Says It Will Take Action Against Anti-vaccine Content. Here’s How It Plans to Do It. *The Washington Post* (Mar 7) [online]. [https://www.washingtonpost.com/business/2019/03/07/facebook-says-it-will-take-action-against-anti-vax-content-heres-how-they-plan-do-it/?noredirect=on&utm_term=.063f6b2ead69](https://www.washingtonpost.com/business/2019/03/07/facebook-says-it-will-take-action-against-anti-vax-content-heres-how-they-plan-do-it/?noredirect=on&utm_term=.063f6b2ead69). Accessed 29 July 2020.

Thiel, Daniel B., Tovah Platt, Jody Platt, Susan B. King, and Sharon L.R. Kardia. 2014. Community Perspectives on Public Health Biobanking: An Analysis of Community Meetings on the Michigan BioTrust for Health. *Journal of Community Genetics* 5 (2): 125–138.

UK Biobank. 2018. Information About the New General Data Protection Regulation (GDPR). [https://www.ukbiobank.ac.uk/wp-content/uploads/2018/10/GDPR.pdf](https://www.ukbiobank.ac.uk/wp-content/uploads/2018/10/GDPR.pdf). Accessed 24 July 2020.

Understanding Patient Data. 2018. Why an Opt-out Rather than an Opt-in or Consent? (May 25). [https://understandingpatientdata.org.uk/news/why-an-opt-out](https://understandingpatientdata.org.uk/news/why-an-opt-out). Accessed 29 July 2020.

Vaccine Resistance Movement (VRM). 2019. Updates and News. [https://www.facebook.com/groups/VaccineResistanceMovement](https://www.facebook.com/groups/VaccineResistanceMovement). Accessed 29 July 2020.

Vandemeulebroucke, Tijs, Bernadette Dierckx de Casterlé, and Chris Gastmans. 2018. The Use of Care Robots in Aged Care: A Systematic Review of Argument-Based Ethics Literature. *Archives of Gerontology and Geriatrics* 74: 15–25.

Wesolowski, Amy, Nathan Eagle, Andrew J. Tatem, David L. Smith, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2012. Quantifying the Impact of Human Mobility on Malaria. *Science* 338 (6104): 276–270.

Whitelhurst, Lindsay. 2017. APNewsBreak: Utah to obey order for DEA drug database search (August 22). [https://apnews.com/article/a0985c75b8ce43c8a834131a42b7fdbc](https://apnews.com/article/a0985c75b8ce43c8a834131a42b7fdbc). Accessed Feb 20, 2021.

Whalen *v. Roe*, 429 U.S. 587 (1977).

Williams, Karmen S., and Gulzar H. Shah. 2016. Electronic Health Records and Meaningful Use in Local Health Departments: Updates From the 2015 NACCHO Informatics Assessment Survey. *Journal of Public Health Management and Practice* 22 (Suppl. 6): S27–S33.

World Health Organization (WHO). 2016. Framework of Engagement with Non-State Actors (May 28). [http://apps.who.int/gb/ebwha/pdf_files/wha69/a69_r10-en.pdf](http://apps.who.int/gb/ebwha/pdf_files/wha69/a69_r10-en.pdf). Accessed 29 July 2020.
Xu, Heng, and Nan Zhang. 2020. Implications of Data Anonymization on the Statistical Evidence of Disparity. Working Paper presented at Privacy Law Scholars Conference (PLSC 2020). SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3662612. Accessed 28 July 2020.

Yoo, Ji Su, Alexandra Thaler, Latanya Sweeney, and Jinyan Zang. 2018. Risks to Patient Privacy: A Re-identification of Patients in Maine and Vermont Statewide Hospital Data. JOTS Technology Science (Oct 9) [online]. https://techscience.org/a/2018100901/. Accessed 27 July 2020.

Zou, James, and Londa Schiebinger. 2018. AI Can Be Sexist and Racist – It’s Time to Make It Fair. Nature 559 (7714): 324–326.