Chapter 11
Ethics, Information Technology, and Public Health: Duties and Challenges in Computational Epidemiology

Kenneth W. Goodman and Eric M. Meslin

Abstract The use of powerful information technology tools in the practice of public health poses many interesting, difficult, and important ethical challenges. Under a modern, electronic standard of care, it can be as blameworthy not to apply such tools as it is to apply them inappropriately. Ethical guidelines can help public health scientists make sound decisions about what users and uses of IT are appropriate in public health. Even with these guidelines, however, there remain some gray areas, particularly with respect to maintaining the privacy and confidentiality of public health information.

The power of modern IT tools renders obsolete some previously sacrosanct guidelines about maintaining privacy and confidentiality. Indeed, it may blur these distinctions to the point of complete conflation. It is therefore necessary for public health practitioners to exercise "progressive caution" in applying information technology to the practice of public health. Developments such as bioinformatics pose acute challenges to maintaining privacy and confidentiality, as does the use of powerful computing technology as support for decisions about interventions.

Moreover, the completion of the map and sequence of the genome of humans (and other organisms) is a technological accelerant for public health ethics. New genetic technologies have spawned an emerging field – public health genomics—engaging the nature vs. nurture debate in new ways. Finally, the interests of ethics and sound public health practice collide in the application of such modern tools as meta-analysis and data mining to public health problems. Even the time-honored

K.W. Goodman, PhD (✉)
University of Miami Bioethics Program, 1400 NW 10th Ave., Suite 916 (M-825), Miami, FL 33136, USA
e-mail: kgoodman@med.miami.edu

E.M. Meslin, PhD
Indiana University Center for Bioethics, Indiana University School of Medicine, 410 W. 10th St., Suite 3100, Indianapolis, IN 46202, USA
e-mail: emeslin@iu.edu

J.A. Magnuson, P.C. Fu, Jr. (eds.), *Public Health Informatics and Information Systems*, Health Informatics, DOI 10.1007/978-1-4471-4237-9_11,

practice of using and publishing case studies in public health research presents challenges to maintaining confidentiality of information as the World Wide Web and other communication and education tools make it increasingly possible for readers to identify the individual(s) discussed in a case.

Keywords Decision support systems • Bioethics • Data synthesis • Privacy • Confidentiality • Security • Group confidentiality • Standard of care

Learning Objectives

1. Differentiate between appropriate and inappropriate uses and users of information technology in public health under an electronic "standard of care."
2. Explain why there is an ethical imperative to use appropriate IT tools under an electronic "standard of care" in public health, and why failure to use appropriate IT tools can be as blameworthy as inappropriately using such tools.
3. Explain the concept of "progressive caution" in the ethical application of information technology to public health.
4. Explain the ethical tension inherent in attempting to maintain confidentiality of individual information while using modern IT tools to store and use group data.
5. Explain why ethical considerations will not permit scientists to entrust decisions about public health interventions to computers alone.
6. Identify meta-analysis and data mining as tools in public health research, and explain why such tools can themselves pose ethical challenges for scientists in making public health decisions.

Overview

The use of powerful information technology tools in the practice of public health poses many interesting, difficult, and important ethical challenges. Under a modern, electronic standard of care, it can be as blameworthy not to apply such tools as it is to apply them inappropriately. Ethical guidelines can help public health scientists make sound decisions about what users and uses of IT are appropriate in public health. Even with these guidelines, however, there remain some gray areas, particularly with respect to maintaining the privacy and confidentiality of public health information.

The power of modern IT tools renders obsolete some previously sacrosanct guidelines about maintaining privacy and confidentiality. Indeed, it may blur these distinctions to the point of complete conflation. It is therefore necessary for public health practitioners to exercise "progressive caution" in applying

information technology to the practice of public health. Developments such as bioinformatics pose acute challenges to maintaining privacy and confidentiality, as does the use of powerful computing technology as support for decisions about interventions.

Moreover, the completion of the map and sequence of the genome of humans(and other organisms)is a technological accelerant for public health ethics. New genetic technologies have spawned an emerging field – public health genomics—engaging the nature vs. nurture debate in new ways. Finally, the interests of ethics and sound public health practice collide in the application of such modern tools as meta-analysis and data mining to public health problems. Even the time-honored practice of using and publishing case studies in public health research presents challenges to maintaining confidentiality of information as the World Wide Web and other communication and education tools make it increasingly possible for readers to identify the individual(s) discussed in a case.

Introduction

At least as much as any other domain in the health professions and sciences, epidemiology and public health are information-intensive. Public health is at ground, albeit not at heart, the collection, sharing, and analysis of data; precious little of this effort uses 3-by-5 cards. The ancient, or at least traditional, thrust of public health informatics is best appreciated by picturing Aristotle, Paracelsus, John Graunt, and others building databases, sending e-mail, and surfing the Web – perhaps even tweeting – in search of more and better information. We have digitized the Broad Street pump—along with its handle, its dirty water and, increasingly through social media, the very people who drink from it. With technological improvements have also come advances in speed, accuracy, storage capacity, and ease of dissemination. On balance, this is good news. But attention to the intersection of ethics and public health informatics requires us to look more closely and with greater precision at the ways information technology (IT) is used and the issues it raises. Some of these issues are not especially novel – there has long been an interest in the security of personal information. Among the developments we discuss below are the technological and policy changes that have transformed issues of personal privacy and confidentiality from matters of personal or immediate family concern to those affecting vast swaths of society.

To begin, it is noteworthy that we are dealing with three broad areas of human inquiry: ethics, computing, and public health. Previous related work has explored the marriage of (a) ethics in epidemiology and public health [1–3], (b) ethics, computing, and health care [4], and (c) ethics, genetics and public health [5–7]. So we have a number of tools (or at least predecessors) to guide us; this is good, given that the three-way intersection we are about to traverse is one formed by high stakes, the need for practical guidance, and the existence of principled disagreement.

Toward an Electronic Standard of Care

In science and in particular the health care professions, standards evolve or are stipulated for a number of reasons. These include: the need for a public (that is, an accountable and transparent) evaluation metric, a system of professionally-accepted goals and objectives, and a calculus for assigning blame. Failure to agree on which criterion is being used has led to disputes in public health research – such as occurred when a controversial placebo controlled trial for HIV prevention was critiqued for using a standard of care that was local rather than international [8, 9]. In this debate the term "standard" was used to describe a level to be reached as an aspirational ceiling. Yet at the same time, especially in law, ethical standards are also seen as a floor below which practitioners may not fall without being found negligent. When ethics functions in both of these ways – describing aspirational goals and minimal conditions of professionalism – it can lead to some confounding consequences. For example, contrary to what many people expect when ethics is given a seat at the policy assessment table, the result is not always nay-saying and handwringing; sometimes, perhaps often, ethics will *require* use of a new technology if it will promote or achieve independently scrutinized goals (e.g., better patient care, improved public health, etc.). Such situations imply that ethical commentary can serve to both set floors and propose ceilings. This was clear at the dawn of interest in the intersection of ethics and health informatics when it was noted that failure to use a computational tool might itself be blameworthy [10]. This argument has also been made recently, especially as large electronic databases are being used for health care evaluation [11] and research using biomaterials [12]. Common to both uses is the recognition that promotion or protection of the public's health is a social value of importance—so important that taking actions that promote the public's health may outweigh those actions that promote the health of an individual person. It is but a short step, ethically, to take the view (as we do) that if an institution (for example, a state health department) is committed to promoting the public's health, is authorized to exercise its legal authority to do so, and has the tools available to do so, then it would be acting *unethically* if it failed to take appropriate steps and use the legal and technical tools at its disposal. But we should be cautious about moving too quickly from "a commitment to promoting public health" to "it is unethical not to use available health information technology." The context, details, and ethical justification are jointly important.

The idea of a standard of care for public health informatics therefore consists of several considerations: (a) what constitutes a standard for public health practice, (b) what constitutes a standard for public health research, and (c) what constitutes a standard for the use of informatics technology? Such a standard will help make clear which uses and users of information systems are appropriate, why failure to use appropriate tools can be as blameworthy as inappropriate use, and why system evaluation is essential for an ethically optimized IT system. Throughout this tour, we will attend to a critical tension between the need for science to progress and the demands of a reasoned and robust ethics; we call this "progressive caution" [13].

Appropriate Uses and Users of IT in Public Health

It is reasonable to hypothesize that there is a broad social consensus that state health departments should maintain databases and use them to "promote the public's health." We further surmise that there would be substantial agreement (though perhaps not as much as with the prior statement) that a state health department should maintain a public database to track the incidence of illness caused by tainted food, including, say, infant formula [14], but it should not maintain a database to *market* or sell infant formula. What's the difference? What makes the one use appropriate and the other inappropriate? While we consider these questions in some detail in what follows, we can lay out here some general strategies for answering them.

First and perhaps most obviously, not all uses or users are equal. We can begin to sort them out by looking at intentions, consequences, and values. So, for example, a database created with public funds to improve public health and promote public welfare is, well, a *public* database. This means that such a database is available for use by authorized public representatives for public purposes. Indeed, very little generally needs to be added to the idea of "public welfare" in order to understand this first criterion. *Anything* that is funded with taxpayer money and used without prejudice to help all will qualify as an ethical use. We are (of course intentionally) forgoing a robust and formal discussion of what "to help all" means. For instance, free reproductive counseling may be seen by some as "helping the community," while others would strongly disagree. A potentially inappropriate use of the public database would therefore be for some sort of private gain or benefit. This is not a comment on or criticism of free enterprise or the free-market system. It is only to observe that public resources should not generally be used to benefit private interests. On the other hand, even private entities have moral obligations to the public: One would expect, perhaps even demand, of a company that makes infant formula that it inform the public about a tainted product; that is both good business and good business ethics. Indeed, such a company might be considered morally praiseworthy if it prospectively established a database to track tainted food.

In addition, even the most ardent libertarian would agree that if data are collected for proprietary purposes, the needs for transparency and the free flow of information require that the fact of the data collection be disclosed in advance, if for no other reason than to allow the sources of the data to negotiate for their share of the profit. But then, of course, if (i) a person were told that his or her personal information were to be stored for proprietary purposes, (ii) failed to reach an agreement over profit sharing, and (iii) that person then refused to allow the information to be used, then such a database would be less valuable, less useful, and less accurate as a *public health* resource. Not all databases are of equal utility.

So far, however, we have merely stipulated that when a database is publically constructed and funded, a good moral case can be made for its use in the service of promoting the public's health. More importantly and powerfully, we assign moral weight to the *intention* guiding the creation and maintenance of the database to benefit the public. Intentions matter in ethics because they can aim for good or ill.

In this case, the intention (creating a public database to reduce infant mortality) was a good one, and so hewing to it will constitute an appropriate use. We also assigned moral weight to the idea that the status of the organization (perhaps we can refer to it as its moral status) is a morally relevant consideration. This is why we emphatically did not say that proprietary uses are somehow inherently ill-intentioned – indeed they conceivably might be very well intentioned – only that the use of public health information for public health should be regarded as more praiseworthy by virtue of the greater benefits that will accrue. Indeed, a private company wishing to develop a database for marketing its infant formula would not be acting *unethically* if it made its intentions clear, and the public were aware of the purpose of the database. But it might be acting unethically if it misrepresented the database as principally meeting a public health need.

But suppose an evil database designer set about creating a computational resource for marketing untested home remedies, discriminating against minorities, or spreading panic? Surely this intention should not enjoy the same status as the other. Put differently, intentions (like information technology uses and users) are not created equal. They are distinguished by, among other things, the consequences of their realization and the value we attach to the intention (whether realized or not). In part because the evil database designer, if successful, will cause great harm, we judge her intentions to be morally inferior. Likewise, we value health over illness, stability over chaos, justice over discrimination.

Looking at matters in this way, we can also see why failure to use appropriate tools can be as blameworthy as inappropriate use – though this, of course, is true only when there is reason to believe the tools will have a positive or valued effect. Health IT tools require comprehensive and even systematic evaluation, and this evaluation must occur in the context of actual use. Indeed, it has been convincingly argued that there is an ethical imperative to conduct such evaluation [15]. We can here explicitly extend this insight to public health informatics, at least provisionally, as we sort out the idea of an "electronic standard of care." This is because system evaluation also helps us make sense of particular uses and users of public health IT systems, at least to the extent that we need to determine for individual uses and users their efficacy and thereby part of their propriety.

We can now look at particular uses and users and see if our intentions-consequences-values metric does any good. For the sake of discussion, let's identify registry maintenance and querying, decision support and data analysis as uses; government officials, students, and corporate investors as users. To be sure, there are many other actual and potential uses and users, and they might be combined in many ways. Indeed, with the lists just presented, we have nine possible scenarios (i.e., three potential uses multiplied by three potential users). We will not review them all; the idea is rather to give a sense of how the process might work. We can do this with two easy hypothetical cases (or one case with two variants):

Case 1. A tumor registry is funded by a federal appropriation from the U.S. Occupational Safety and Health Administration (OSHA). As part of a periodic monitoring program, a government scientist working for OSHA wants to query

the registry to identify the incidence and prevalence of a certain neoplasm in a particular population living near a toxic waste site. The registry was built with public funds, and patients with cancer had agreed to contribute to the bank. The scientist's *intention* is to obtain epidemiologic data that will be used to help determine whether there are empirical grounds for closing the site. One of the possible *consequences* of the query is closing the toxic waste site, thereby reducing correlated morbidity and mortality in future populations in that area. Assuming that we accept that the user was appropriate, the *intentions* were appropriate, the *consequences* of the actions were appropriate and – perhaps most importantly, the *value* we place on reduced morbidity and mortality was appropriate — then we have identified an appropriate use and user.

Case 2. Suppose now that the same registry is queried by a biopharmaceutical investor with the stated goal of identifying biomarkers in those same neoplasms that have especially unusual properties. While it is very likely that his instrumental *intention* is to identify markers that will be used to design better anti-cancer drugs (reducing morbidity and mortality from cancer), it is also clearly the case that he is immediately and directly keen to predict for the sake of financial gain which anti-cancer agents will enjoy the greatest markets in coming years. Let us assume his principal intent is commercial. Using the public database for private commercial gain has many *consequences*, not the least of which is eroded public confidence in database security. The *value* is entrepreneurship. The question of whether this was an appropriate use by an appropriate user should be easy to answer: this use (querying a public database for private gain) by this user (a private entrepreneur)is not ethically equivalent to the use in Case 1 (querying a public database for preventing mortality and morbidity) by the user (a government-supported epidemiologist).

Make no mistake: many or most cases are vastly more complex than these. Indeed, developments in translational science already suggest that the once-bright lines between public and private funding, and basic and applied research, are blurring (and that such blurring is being encouraged) [16]. Rarely are data – or intentions! – as unambiguous as implied in our examples. In Case 1, what about the problem of communicating health risks and the likelihood of engendering fear or even panic? What about people who lose their jobs if a factory is closed? In the revised version, is there nothing to be said about the virtues of data sharing? What would we think if the entrepreneurial investor's query led more quickly than expected to a medical breakthrough that actually reduced the impact of a devastating cancer?

As a general starting point, it makes sense to say that ethics can help guide thinking towards optimal solutions and away from sub-optimal ones. Of course, in the same way that it is simplistic to explain genetics using only a basic Mendelian example of two types of pea plants – smooth and wrinkly – so too is it simplistic to explain the ethics of database use using only virtuous government epidemiologists and profit-focused business people (indeed, one can imagine examples in which the moral attributes are reversed). Issues raised later in this chapter will give examples

of these nuanced differences. In fact, ethical issues related to the use of IT should be seen as a subset of the ethical issues that arise in several domains of human activity including epidemiology, public health, and health research, as well as national security, economic development, and social networking.

Such refinement, it is worth emphasizing, is precisely the task of applied ethics. The model is reasonably well evolved in clinical ethics (where patients, families and health care providers wrestle with difficult care decisions) and in research ethics (where researchers, research subjects and oversight bodies confront difficult choices). Applied ethics is a growing area of disciplinary expertise with rigorous peer-reviewed methods that must pass public and professional scrutiny. It is not, however, the mere rote application of existing rules and regulations. The growing interest in codes of ethics is positive and noteworthy—but codes, guidelines, and lists of best practices are no substitute for robust and ongoing ethics education and analysis.

"Progressive Caution"

Ethics thrives on new science and technology. This is no less true in epidemiology and public health than in any other science. In the health professions, where the stakes are consistently high, the role of ethics is complex. When it comes to new technology, what role do we want ethical analysis to have? Should we be stomping our feet, shaking our heads, and clucking our tongues at the new technology, Luddites at the gates of progress? Or should we prefer facile boosterism, cheering each new gadget independent of its utility or consequences, cheerleaders at the edge of the abyss? The answer, of course, is straightforward: Neither. We want thoughtful analyses and practical guidance. We want science to progress, but not at any cost. We want to minimize risk but not to the point of unreasonably restricting liberty. But we also emphasize that each of these paired goals is understood differently when the practice is about social institutions promoting the public health than when it is about physicians providing excellent patient care or researchers conducting meritorious experiments.

That is, we want a kind of "*progressive caution*" whereby we move forward, and that progress is tempered or leavened by attention to the kinds of details being scrutinized here. To be fair, we recognize that some nuance is at work here, but it is worth emphasizing: it is the difference between prohibiting an action but allowing certain exceptions, and enthusiastically encouraging an action but placing certain restrictions. The path that ethics has trod in health care and research is littered with such nuanced distinctions. More than 60 years ago the Nuremberg Code laid out the first modern set of ethical principles for medical research, strictly prohibiting all research involving humans *unless* they could give voluntary informed consent. Over time, this protectionist stance relaxed to the point where research on humans is widely permitted, even on children and those who cannot give fully informed consent themselves because of diminished capacity to consent, so long as certain

restrictions and procedures are followed. There has been, in other words, a progressive caution exercised about research involving human subjects. Indeed, this is seen in many areas of biotechnology assessment, from stem cell research and reproductive health to gene therapy.

In a slightly different context, the idea of progressive caution was introduced thus: "Medical informatics is, happily, here to stay, but users and society have extensive responsibilities to ensure that we use our tools appropriately. This might cause us to move more deliberately or slowly than some would like. Ethically speaking, that is just too bad" [13].

The idea of progressive caution is perhaps best or most productively put in the form of a question: How should we arrange things so that we enjoy the benefits of new technology while reducing, minimizing, or mitigating the (potential) harms? Given that both the use and the failure to use information technology raise ethical issues, the concept of progressive caution will help guide us as we consider the specific ethical issues that arise when information technology is used in epidemiology and public health.

Privacy, Confidentiality and Security

The technical issues associated with privacy, confidentiality, and security in health informatics are discussed in other chapters. Here, we will discuss privacy, confidentiality, and security with an emphasis on ethics.

The intersection of ethics and health informatics almost immediately brings to mind the challenges of privacy and confidentiality. These issues are indeed what most people, scientists and lay people included, worry about. We suspect that most people have a reasonably well-developed idea about what these topics concern and why they are important.

We begin by recalling the general difference between privacy and confidentiality. *Privacy* is best thought of as relating to *people* and their expectation, hope, goal, or right to be left alone and free of intrusion by others; you might, for instance, intrude in my private life by peering in my window to study my behavior. Privacy is intruded upon when someone gains access (especially physical access) to you without your permission. *Confidentiality* relates to the status of *information* about people, the "holy secrets" of Hippocrates; you might violate my confidentiality by looking at my medical chart, or by querying the database that contains some or all of that information, without my permission or knowledge. Indeed, one of the intriguing developments in bioethics has been the way privacy intrusion and confidentiality violation have traded places as the more worrisome ethical transgression: unauthorized access to a person (privacy intrusion) may have been worse than unauthorized access to information about a person when the harms of the former are seen as more damaging than the latter. Once medical charts became more widely available to more people with a "need to know," confidentiality may have become the more worrisome. Indeed, one of the landmark ethics reports which documented

the large number of health care providers in a hospital with access to a patient's medical chart referred to confidentiality as a "decrepit concept" [17]. And now that genome science has progressed to the point where tiny bits of DNA can identify individuals without ever having to physically interact with a person, it may be time to revisit the entire analysis.

So too will public health informatics require that we think about privacy and confidentiality in ways somewhat different than we might be accustomed to in clinical medicine, nursing, or psychology. The core problem with confidentiality and electronic health media is this: We want simultaneously to make information easily accessible to appropriate users and inaccessible to inappropriate users. This is a problem, because the means for accomplishing the one are often in conflict with the means for accomplishing the other. But this air of dilemma is resolvable in at least three ways [18, 19]:

- Technology, including security measures
- Institutional policies and procedures
- Education programs addressing the foundations and importance of confidentiality

These practical steps may be regarded as moral imperatives, measures to take as part of a comprehensive program to protect individuals' health information. But such protections cannot—and should not—be absolute. That is, there may be credible challenges to confidentiality, and many of the most interesting and important ones arise in public health.

Information, Consent, and Stigma

The most obvious way one might ethically set aside concerns about confidentiality breaches is with the consent of those about whom the information pertains. This is often the case in research contexts: Investigators need to have access to personal health information, and subjects/participants must agree to this access. Patients also routinely consent to release of information to third parties—e.g., insurers—for the sake of reimbursement of health professionals (though because they must provide such consent to be treated in the first place, one might plausibly wonder how voluntary such consent really is.) We also note the apparent ease with which individuals routinely "consent" to allow information to be used, collected, and shared to facilitate social networking, downloading of "apps" and website content. This gives rise to a new public health informatics reality arising from social media. For example, by relying on search queries alone, Google Flu Trends is able to measure influenza outbreaks faster and, some scientists argue, more accurately, than by relying on traditional health care system reports [20]. The key point is that if individuals voluntarily permit others to obtain and use information about them, then the information that has been shared is no longer confidential. It may have an impact if it is

shared, it may cause embarrassment, remorse, guilt, pain, or befuddlement, but the act of giving permission (and the assumption that one understood what one was giving permission for) renders the status of that information no longer confidential and thus outside the range of violation. This is why, for example, there is considerable interest in the world of biobanking to de-emphasize privacy and confidentiality protections to those being asked to donate samples and allow access to information, and to focus instead on providing clear information about possible uses.

Public health IT poses special challenges to the traditional clinical/research model, in part because there are many cases in which it would be logistically or practically impossible for epidemiologists or public health officials to obtain consent from all those whose information they want to collect or analyze. In other contexts, such as collecting information about transmission of various diseases, rates of vaccination, and so forth, society has set aside the notion of absolute confidentiality in exchange for the benefits of better health surveillance, monitoring, and analysis. Indeed, a great deal of personal health information is collected, stored, and processed by governments, universities, and other entities without any individual consent whatsoever. Institutional Review Boards (IRBs) oversee some of these efforts, but they do not oversee all public health surveillance, in part because some of these activities do not fall under standard definitions of research involving human subjects.

This is not as far-fetched as one might think. In environments where the public is confident that government officials will use previously collected health information in a trustworthy manner, consent is not always required [11, 21]. But that willingness is not to be presumed come what may: It is, we might surmise, a gift from citizens in open societies. They trust health authorities to make sound decisions and recommendations based on the best available evidence, and they trust those authorities to acquire the evidence in the least intrusive ways possible. One of the ways to accomplish this is to render the data anonymous in salient respects. For instance, many public health surveillance efforts do not require the collection or storage of unique identifiers such as name, address, or Social Security Number; all that is needed is case information, context, and so forth. Another way is to make explicit efforts to engage the community [22].

But the balance of the "special challenge" of public health IT is that health data achieve a distinctive synergy when they are stored in computers. For example, it might not matter that you do not know an individual's name if you know her disease, race, postal code, and sexual orientation [23, 24], or perhaps have a sample of her blood [25]. Either you will be able to identify this person – to pick her out of the crowd – anyway by virtue of these surrogate data ensembles, or your surveillance or research will come to associate her social, racial, ethnic, or other group with a malady or behavior in ways she would have objected to had she been given the opportunity to dissent.

Even in open societies, most people are ignorant of the ability of geographic information systems to characterize neighborhoods and draw inferences about ever-narrower social groups. Would people consent to these characterizations or inferences? Indeed, would they ever have agreed in the first place to allow their personal

information to be digitized if they knew the kinds of inferences that might be drawn? What we have come to call "group confidentiality," or the idea that population subgroups have privacy and confidentiality interests [25], has acquired increased currency, especially in genetics principally because genetic information is ultimately about the information that is shared by communities, be they families or persons who share a similar disease. In the case of families, knowing the genetic test results of a parent immediately conveys information about their biological children; testing an individual for the presence of a genetic mutation that is more prevalent in a racial or ethnic group will immediately convey information about that group.

The Case of Bioinformatics

Completion of the project to map and sequence the human genome is ushering in what many hope to be a golden age of molecular epidemiology. It is therefore important to provide a brief excursus on computational genomics or bioinformatics [26, 27]. For a variety of clinical and research purposes, including drug discovery, clinicians and scientists are increasingly able to digitize genetic information and store it in databases. Three key questions emerge from this effort, and they will continue to challenge our ability to get an ethical grip on this new technology:

1. Does it make any real sense to talk about confidentiality when computers processing genomic data (perhaps in conjunction with other information) provide a high-powered way of identifying individuals whose idea of confidentiality was a piece of paper in a locked desk?
2. Consent to acquire information increasingly needs to take into account the idea that people might—or might not—want to learn the results of aggregate genetic analysis. In other words, if I agree to let you store and analyze my genetic data, does that mean you will later let me know what you learn? Will you have an unanticipated duty to disclose risks and other incidental findings to people who might not want to hear of them?
3. What standards or assurances are available that error reduction is being addressed by the new technology? Complex databases and gene annotation protocols are ripe for both error and error-reduction strategies. With genomes as email attachments and digitized genetic information being included in very large databases, the job of valid consent will be as difficult as in any other aspect of biomedical research. There are several reasons for this. Some are independent of the role of information technology and some are greater because of computers.

As already noted, genetic information is not about one person; it is also information, in 1 degree or another, about others. These relatives might be identified in research (usually pedigree studies) without having consented to be subjects in the research. Genetic information is to some extent also about members of one's racial or ethnic group, increasing the risk of bias and stigma – even as we might make use of the information for standard epidemiologic purposes. And of course genetic

information is about people who share common genetic mutations that raise their level of risk of disease. Genetic information increases in scientific (and other) value over time. This is due to the fact that while we have sequenced the human genome, we are still mostly ignorant of the *function*s of most genes—what genes actually do when they make the proteins that form the parts of our human selves. As functional genomics progresses, we will acquire tomorrow the ability to conduct research that is not possible today. This increase in research potential is independent of the stored genetic information or tissue samples themselves. In other words, today's genetic database will increase in value tomorrow even if it is not changed or augmented.

Can valid consent rise to these challenges? There is every reason to believe it can, especially as we ensure that the concept of valid consent as a process and not an event does not collapse into platitude and cliché. Indeed, the idea that consent is a process which might, in fact, never end offers a way to ethically optimize the epidemiologic use of digitized genetic and, indeed, other information. Consider the potentially great value in special newsletters for subjects (and even communities) whose genetic information has been digitized and stored in an electronic database. Such newsletters can inform individuals, relatives, and communities of new and potential uses, including research, planned for the database. The database, if appropriately constructed, could provide the means for individual subjects to opt out of specific studies. For instance, suppose I am willing to consent to research in cancer genetics but not research on Alzheimer's disease. Once my genome is in your database you will be able to let me know of the contemplated use for Alzheimer's studies, and if I dissent you will be able to ensure that my genetic information is not included in your study.

Such a newsletter might also provide a much better way of including subjects in the broad sweep of the research in general by informing them of study results, related research, and even ethical issues raised by the research. Furthermore, imagine that not only would a newsletter or blog be available to patients, but that physicians received up-to-the-minute information about the relevance of these findings, with reminders, warnings, and special considerations, as they now often do when they write a prescription in a computerized physician order entry system. The positive potential for using genetics in the service of public health is only now starting to be explored. Among the most obvious targets for applying genomic science to population health is the focus on predictive, diagnostic, and therapeutic benefits for stratified populations and subpopulations rather than individuals. The benefits of population screening for familial hypercholesterolemia or inherited colorectal cancer are good examples of genetics helping public health. But many implementation and infrastructure challenges remain [28].

Decision Support

Our discussion of appropriate uses and users of IT systems will be of no small utility as we consider the issue of computational decision support in epidemiology and public health. In one sense, all computers used in epidemiology and public health are decision support systems—computers that help us navigate among the shoals of probabilistic data.

In clinical medicine and nursing, there are generally thought to be at least three kinds of decision support systems: *reminder* systems, *consultation* systems, and *educational* systems. Their functions are easily inferable from their names. Apart from seasonal reminders to "get your flu shot," it is not clear if decision support in epidemiology and public health runs parallel to these three uses—what constitutes a reminder in clinical medicine, for instance, has no ready analog in the public health sciences. We can however identify two functions of ethical interest in decision support in epidemiology and public health; they are (1) *interventions* and (2) *data synthesis*, including meta-analysis and data mining.

Interventions

A decision support system might be used to help decide whether and when to begin an intervention program and what kind of intervention would be best or most efficacious. Why is there an ethical issue here? To answer this question, let's turn to clinical medicine.

What has come to be called the "standard view" of decision support in diagnosis suggests that humans are better than machines at functions as complicated as diagnosis [29]. Humans *understand* data better than machines (even if computers might be able to *process* it better and faster. The answers to questions about whether to close a toxic waste site, commence an education program, or call for a quarantine are decisions that require more than digital firepower. They are decisions that require vast background knowledge, a scientific as well as an intuitive understanding of risk, and a more or less clear sense of how best to balance and trade off among competing goals. Computers cannot meet these criteria, and are unlikely to be able to for some time.

It follows that while we might have a duty to use computers to help in making tough calls, we must not let the computers make the tough calls. This stance is appropriate whether we are contemplating needle exchange programs or anthrax attack countermeasures, vaccination protocols or mutant flu quarantines. Another way of putting this is that public health decisions are rarely if ever exclusively scientific, statistical, or empirical. Public health scientists and officials are faced with a difficult array of decision points such that the correct or best answer will rarely be arrived at with more information or more computing power. Rather, scientists and officials need to analyze their intentions or the goals they hope to achieve, the consequences of various decisions they might make or actions they might take, and the values that guide them.

The question of whether to intervene and which intervention to commend is in part an ethical one precisely for these reasons. It is possible that a decision support system might one day be able to analyze human values as well as data sets—but it is very unlikely and, in any case, it will be quite a long time before that happens. The lesson in public health is the same as in clinical medicine and nursing: Computers should not be allowed to trump people [29].

Data Synthesis and Computer-Based Research

Ever-increasing demands for data and evidence to inform guidelines and best practices have made it clear that we need computers to help us sort out all our information. Indeed, we now turn with increasing frequency to various forms of research synthesis to make sense of the data. The computational tools of meta-analysis and data mining will give us our best examples; they provide ways of eliciting conclusions, answers, or even mere suggestions from the apparent mess of data. They provide us with many case studies about whether and when to use a computer in making scientific decisions. Debates over meta-analysis, which often turn on its methods and reliability, remain important for any discussion of ethics in epidemiology in general, and ethics-computing-and-epidemiology in particular [30].

Consider the important historic case of meta-analytic studies of the effects of environmental tobacco smoke. In 1993, the US Environmental Protection Agency, relying on a meta-analysis of 11 studies of smokers' spouses, classified environmental or "second-hand" tobacco smoke as a Group A carcinogen along with radon, asbestos, and benzene [31]. No problem so far—tobacco smoke is bad, people agree tobacco smoke is bad, a study shows that tobacco smoke is bad. The problem is that meta-analysis continues to engender intense debate about its accuracy and reliability. It might be, in other words and just for the sake of discussion, that we (in 1993) actually lacked adequate scientific warrant to rank environmental tobacco smoke as a Group A carcinogen. At any rate, the debate elicited the following remark: "Yes, it's rotten science, but it's in a worthy cause. It will help us to get rid of cigarettes and to become a smoke-free society" [32]. This quote is two-sided: on the one hand the self-righteous among us are prepared to accept a certain amount of scientific uncertainty so long as the public health policy goal is achieved – how sure do we have to be, scientifically, to recommend an anti-smoking policy for city restaurants? On the other hand, uncertain science is precisely the basis for the pushback by opponents of anti-smoking regulation. And what tobacco science was to the 1990s, climate change science is to the early part of the twenty-first century. How much certainty is required (and what counts as good data) that the planet is warming and that humans bear some responsibility before public health policy to restrict carbon emissions takes place?

The ethics-computing-public health tension has been described as follows:

> In one respect, the very idea is incoherent: If one believes the science to be flawed, then how can it support a worthy cause? How even can the cause become worthy in the absence of credible evidence? (If environmental smoke does not harm children, then there is no reason to protect them from it, and so protecting them cannot be worthy.) But granting for the sake of discussion that the cause is worthy, it is nevertheless a severe form of ethical shortsightedness to suggest that the credibility of scientists, government institutions, and policy makers is a fair trade for a victory on one policy issue. Even the most craven utilitarian would recognize this to be a bad bet [27].

Note that while the intention might be praiseworthy (to reduce environmental tobacco smoke) and the consequence a positive one (fewer people suffering the effects of second-hand smoke), the value we place on scientific method and credibility may sometimes outweigh the other considerations. It is also important to underscore that it can be very difficult to calculate future consequences – including future negative consequences.

Think of meta-analysis and data mining as secondary or *n*-ary uses of data. Such use matters, as it did with bioinformatics, because subjects or communities might have consented to the primary use but not necessarily the secondary or *n*-ary one. Now, this might matter little or not at all to research subjects, especially if the risks of such research are minimal or absent and if (as is usually the case with meta-analysis) individuals cannot be identified from or in the data. With data mining, also sometimes called "knowledge discovery" or "machine learning," we have the *n*-ary analysis of databases in search of patterns, trends, associations, and so on. Employed to great profit in science and business, data mining is emerging as a potentially valuable resource in health care.

Our concern is with valid consent in computational public health practice and research – specifically, the use of personal information for purposes other than originally intended (advocates for public health surveillance observe that if data are collected for public health, their use for public health is primary, not secondary). Data mining technology promises public health trend-spotting, quality assessment, and outcomes research of depth and breadth unimagined a few years ago. Since this information is *personal* information, we need to ask whether those people the information is about would agree to such use. We need to look at three key considerations:

1. Is the database analysis something that was disclosed and consented to when the information was obtained?
2. Is the purpose of the data mining scientific, commercial or both?
3. Are individuals identifiable in the database or as a result of the research?

The answer to question 1 is rarely "yes"; for question 2, the use might be commercial; the answer to question 3 will often be "generally" or "in principle." The feature of data mining that distinguishes it from more garden-variety forms of database research is the facility with which scientists (and others) can look through vast amounts of personal, identifiable information — again and again and again (it is, therefore, a question at least of degree and perhaps of kind). Each analysis is a further "experiment" for which we may generally presume that no consent has been obtained. Besides, tools such as newsletters are more useful for focused research programs in which the goals of the research can be itemized. In data mining, one might perform an analysis with all the effort and forethought that go into a PubMed search, for instance.

As with bioinformatics, more research is needed to clarify the ethical issues surrounding data mining. We include it here to give a sense of exciting new challenges to the standard model of valid consent (how best, for instance, might one describe data mining in lay language to prospective subjects?). For now, the best consent for data mining research is likely to be obtained in advance, for non-commercial research, and for studies where individual identifiers are either not available or can be readily hidden and secured.

Conclusion

The computational turn in epidemiology and public health offers extraordinarily powerful and intelligent tools to collect, analyze, and transmit the personal health information of millions of people. We have seen that it would be ethically irresponsible not to continue to develop and use these tools for the improvement of public health. As important, we have learned that the ordinary people who are the sources of that information have warrant to expect that its collectors, analyzers, and transmitters will safeguard it and ensure its appropriate use.

What counts as an appropriate use and who should be regarded as an appropriate user are questions whose answers will guide practitioners and policy makers as they balance the needs of public health and the rights of individuals. This balancing effort can be difficult and nuanced: the information at issue includes both the familiar and quotidian (on vaccinations and vital statistics) and the novel and complex (genetic data about individuals and groups).

Moreover, what is in some domains a comfortable demarcation between practice and research becomes fraught and controversial in epidemiology and public health. This is unavoidable, but it presents us with splendid opportunities to apply and evaluate the tools of applied ethics. This will be especially true as ever-grander computers and data networks link scientists and officials from around the world. We will judge them by how well they use the networks in the service of public health, and by how well they attend to the concerns of individuals who, in a flash (or a click), may find themselves and their genes and maladies and behaviors out there for all to see.

Review Questions

1. Many people think of ethics as prohibitive. What does it mean to say that use of a technology might be obligatory?
2. Explain why the concepts of "appropriate use" and "appropriate user" are given so much emphasis.
3. What is the point of "progressive caution" and why does it matters in public health informatics?
4. Differentiate among *privacy*, *confidentiality*, and *security*, as those terms relate to public health information.
5. Review the ways electronic health data might be made easily accessible to appropriate users and inaccessible to inappropriate users.
6. Say why "group confidentiality" is important in public health informatics.
7. Review some of the leading challenges that arise in bioinformatics.
8. In their discussion of decision support, the authors conclude that "Computers should not be allowed to trump people." Why do they say this? Do you agree? Why or why not?

Acknowledgement Part of Goodman's work on this chapter was supported by the University of Miami CTSI, funded by the National Center for Advancing Translational Sciences grant #1UL1TR000460.

References

1. Coughlin S, Beauchamp T, editors. Ethics and epidemiology. New York: Oxford University Press; 1996.
2. Coughlin S, Soskolne C, Goodman K. Case studies in public health ethics. Washington, DC: American Public Health Association; 1997.
3. Geissman K, Goodman KW, et al. Scientific ethics: an interactive, multimedia, computer-based training. Atlanta: Centers for Disease Control and Prevention and Agency for Toxic Substances and Disease Registry; 1998.
4. Goodman KW, editor. Ethics, computing and medicine: informatics and the transformation of health care. New York: Cambridge University Press; 1998.
5. Burke W, Burton H, Karmali M, Khoury MJ, Knoppers BM, Meslin EM, Stanley F, Wright CF, Zimmern RL, Hall AE. Extending the reach of public health genomics: what should be the agenda for public health in an era of genome-based and 'personalised' medicine? Genet Med. 2010;12(12):785–91.
6. Evans JP, Meslin EM, Marteau TM, Caulfield T. Deflating the genomic bubble. Science. 2011;331:861–2.
7. Meslin EM, Garba I. Biobanking and public health: is a human rights approach the tie that binds? Hum Genet. 2011;130(3):451–63.
8. Varmus H, Satcher D. Ethical complexities of conducting research in developing countries. New England Journal of Medicine. 1997;337:1003–5.
9. Lurie P, Wolfe SM. Unethical trials of intervention to reduce perinatal transmission of the human immunodeficiency virus. New England Journal of Medicine. 1997;337:853–5.
10. Miller RA, Schaffner KF, Meisel A. Ethical and legal issues related to the use of computer programs in clinical medicine. Ann Intern Med. 1985;102:529–36.
11. Stanley FJ, Meslin EM. Australia needs a better system for health care evaluation. Medical Journal of Australia. 2007;186:220–1.
12. Meslin EM, Goodman KG. An ethics and policy Agenda for Biobanks and Electronic Health Center for American Progress, Science Progress. http://www.scienceprogress.org/2010/02/bank-on-it/. Posted 25 Feb 2010.
13. Goodman KW. Bioethics and health informatics: an introduction. In: Goodman KW, editor. Ethics, computing and medicine: informatics and the transformation of health care. New York: Cambridge University Press; 1998. p. 1–31.
14. U.S. Food and Drug Administration. FDA, UC Davis, Agilent Technologies and CDC to create publicly available food pathogen genome database. News release, available at http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm311661.htm.
15. Anderson JG, Aydin CE. Evaluating medical information systems: social contexts and ethical challenges. In: Goodman KW, editor. Ethics, computing and medicine: informatics and the transformation of health care. New York: Cambridge University Press; 1998. p. 57–74.
16. Collins FS. Reengineering translational science: the time is right. Sci Transl Med. 2011;3:90cm17. doi:10.1126/scitranslmed.3002747.
17. Siegler M. Confidentiality in medicine—a decrepit concept. New England Journal of Medicine. 1982;307:1518–21.
18. National Research Council. For the record: protecting electronic health information. Washington, DC: National Academy Press; 1997.

19. Alpert SA. Health care information: access, confidentiality, and good practice. In: Goodman KW, editor. Ethics, computing and medicine: informatics and the transformation of health care. New York: Cambridge University Press; 1998. p. 75–101.
20. Pervaiz F, Pervaiz M, AbdurRehman N, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res. 2012;14(5):e125. doi:10.2196/jmir.2102.
21. Lee LM. The cornerstone of public health practice: public health surveillance, 1961–2011. Morb Mortal Wkly Rep. 2011;60:15–21.
22. Meslin EM. The value of using top-down and bottom-up approaches for building trust and transparency in biobanking. Public Health Genomics. 2010;13:207–14.
23. U.S. General Accounting Office. Record linkage and privacy: issues in creating new federal research and statistical information. Washington, DC: U.S. General Accounting Office (GAO-01-126SP); 2001.
24. Sweeney LA. Guaranteeing anonymity when sharing medical data: the Datafly system. In: Masys DR, editor. Proceedings of AMIA Annual Fall Symposium. Philadelphia: Hanley & Belfus; 1997. p. 51–5.
25. Alpert SA. Privacy and the analysis of stored tissues. In: Research involving human biological materials: ethical issues and policy guidance, Commissioned papers. IIth ed. Rockville: National Bioethics Advisory Commission; 2000.
26. Goodman KW. Ethics, genomics, and information retrieval. Comput Biol Med. 1996; 26:223–9.
27. Anderson JG, Goodman KW. Ethics and information technology: a case-based approach to a health care system in transition. New York: Springer; 2002 (esp. Chapter 5, "The Challenge of Bioinformatics", pp. 113–122).
28. Burke W, Burton H, Karmali M, Khoury MJ, Knoppers BM, Meslin EM, Stanley F, Wright CF, Zimmern RL, Hall AE. Extending the reach of public health genomics: what should be the agenda for public health in an era of genome-based and "personalised" medicine? Genet Med. 2010;12:785–91.
29. Miller RA. Why the standard view is standard: people, not machines, understand patients' problems. Journal of Medicine and Philosophy. 1990;15:581–91.
30. Goodman KW. Meta-analysis: conceptual, ethical and policy issues. In: Goodman KW, editor. Ethics, computing and medicine: informatics and the transformation of health care. New York: Cambridge University Press; 1998. p. 139–67.
31. Environmental Protection Agency. Respiratory health effects of passive smoking: lung cancer and other disorders. Washington, DC: Government Printing Office (EPA/600/6-90/006F; GPO: 0555-000-00407-2); 1993.
32. Feinstein AR. Critique of review article, environmental tobacco smoke: current assessment and future directions. Toxicol Pathol. 1992;20:303–5.