

The Report Committee for Lucius Lomax
Certifies that this is the approved version of the following report:

THE NIGHTINGALE PROJECT

SUPERVISING COMMITTEE:

Ying Ding, Co-Supervisor
Lance Hayden, Co-Supervisor

THE NIGHTINGALE PROJECT

By

Lucius Lomax

Master's Report

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Information Studies

The University of Texas at Austin

May 2022

ABSTRACT

Revealed by the *Wall Street Journal* in 2019, the Nightingale Project was designed as an agreement between Google Health and the Ascension Healthcare Family to transfer patient data, without patient permission, from Ascension's 2600 hospitals and clinics to Google for further data wrangling and to inform Google's own significant healthcare inquiries. First viewed by independent advocates as a privacy nightmare and a further encroachment by Big Tech on individual liberties, it is the contention of this paper that, in light of the COVID-19 pandemic, Nightingale and similar projects may *not go far enough* in data-sharing, and should include cooperation across international borders. One possible method to ensure that exchange may be the creation of markets for healthcare data from across the globe.

Table of Contents

Introduction	5
Nightingale Takes Off	9
What will be transferred from Ascension to Google?	10
What will be done with the data?	12
What is the direct clinical importance of Big Data in health care?	14
What is Personalized Medicine?	17
The Business of Health Care	19
What does Google really want?	21
What does the pandemic mean for the Nightingale Project?	24
What are the downsides?	26
The Ethical Conundrum	27
Which special populations are most at risk?	29
Privacy and the Market for Healthcare Data	31
The China Solution	33
Conclusion	36
Bibliography	39

INTRODUCTION

Pat Ireland Nixon is one of the most famous physicians ever to practice in Texas. He was born in 1883 in what was then rural Guadalupe County, outside San Antonio, and graduated from the University of Texas in 1905 and from the Johns Hopkins University School of Medicine four years later. He returned to Texas, to San Antonio, and over the next half-century Dr. Nixon worked as a general practitioner and advocate for improved public health in the state. What he is best known for, however, is documenting his life as a physician in the first half of the 20th century. According to the authoritative *Handbook of Texas* (Burns, 1995), a publication of the Texas State Historical Association of which the good doctor was also once president, “Nixon was the first physician to write extensively on the history of medicine in Texas, a field to which he contributed several articles and three books.” His posthumously published autobiography *Pat Nixon of Texas* (Nixon, 1976) is of interest today because in it the author reveals the minutiae of learning the profession of medicine and practicing it, not as a world-famous specialist, but as a well-educated yet everyday working doctor in America during the first half of the 20th century.

Since the foundations of medicine in pre-history, as an art more than a science, to the publication of the first medical text—apparently written on papyrus in Egypt—through the work of Hippocrates, the so-called “father of medicine,” in 5th century BC Greece, the tools and the practices of the profession have been recorded as they evolved. Pat Nixon offered a snapshot of his studies and work during a period of time which most healthcare historians

would consider to be part of the age of what we now call modern medicine, at least when compared with the shamans and bloodletting of prior eras. But a kind of medicine quickly disappearing into the rear-view mirror of what we call health care today, that is heavily data-driven.

As a student at Johns Hopkins, young Dr. Nixon recorded the barbarities of the age—including different standards of care for minorities and for the poor—and also the barbarities of an early 20th century medical education, including having to memorize vast swathes of human anatomy, long lists of symptoms and differential diagnoses and having to “stand and deliver,” that is, being called upon to recite from memory all of the above in front of one’s classmates. As a working physician post-World War One, blood tests and advanced diagnostics were unheard of. An excellent example of what Dr. Nixon did *not* have to help him in patient treatment is X-ray technology. The first patient radiography had actually been taken in 1886, barely a decade after Pat Nixon’s birth (Howell, 2016). “However, it was only a decade later,” according to a 2016 study “Early Clinical Use of the X-Ray,” in *Transactions of the American Clinical and Climatological Association*, “after the underlying structure of the hospital changed *due to importation of technologies from business* [italics added] that x-ray images became seen as part of routine patient care.” Nowhere in his autobiography—Dr. Nixon died in 1965—was there mention of DNA, phenotypes or generalized data. Illnesses and conditions occurred and were cured, or not, but what the treatment of one patient could mean for treatment of the next patient was *limited* to common anatomy, or known infectious pathology, as in sexually transmitted diseases, or how to approach surgical procedures, the location of organs in one body being the same in the next. Most of the time. Patient histories and findings from physical

exams—using a stethoscope and reflex hammer—were paramount in making individualized diagnoses.

Pat Nixon was, nonetheless, a very well-educated and well-prepared practitioner of the healing arts. In Baltimore at Johns Hopkins he crossed paths with the so-called “father of modern medicine,” the most exalted physician of the modern era of health care, the Canadian William Osler who was one of the four founders of Johns Hopkins and who was leaving Hopkins for Oxford University just as Pat Nixon was arriving. Later, after graduation, and despite not liking surgery, the young Dr. Nixon ventured to Rochester, Minnesota, to observe a family of physicians who were devoted to new practices in the surgical arts—the Mayos—founders of the Mayo Clinic. By the standards of the time Dr. Nixon may have been an outlier, at the positive end, in terms of preparation. But standards change. Today, many professions can require a whole new skillset every few years. Indeed, entire professions (telegraph operator, for example) may come and go in a few years. But in perhaps no other profession is change in standards and in tools as true as in medicine, where society has a special and vested interest in seeing that practitioners are up to date in knowledge and in use of new science and new technology. Today the Harvard Medical School website recommends that applicants to their program have taken biostatistics, genetics and biochemistry. Among the most common majors of students seeking entry to the Harvard medical program (Harvard, 2022) are engineering and computer science, and medical school curriculums themselves may include engineering, in order to teach doctors, it is said, “to think like engineers.” Genetic profiles are now used everywhere in medicine—for machine learning and artificial intelligence development—but these advanced modalities require big datasets, in other words a lot of patient information. The

outcomes can be exceptional, however, with some A.I. programs now outperforming specialists in selected cancer diagnoses.

It is my thesis here that so-called Big Data, together with data-sharing and changes in the business end of medicine, such as the changes in hospitals that led to acceptance of radiography, mentioned above, are the keys to a new age of health care. In this view, post-pandemic, with gigabytes of super-important data related to COVID-19 alone, privacy and national security concerns will have to be reformulated in order to assure maximum transfer of *de-identified* patient data, one hopes, to all researchers, wherever they may be, in order to ensure a robust response by healthcare authorities when other public health crises arrive. One way to do that is through markets for healthcare data, much like the markets for other tangibles (such as wheat and gold) and intangibles (cryptocurrencies or currency futures), to encourage use of valuable patient data, while also protecting patient privacy.

NIGHTINGALE TAKES OFF

Little more than two years ago the *Wall Street Journal* reported (Copeland, 2019) that Ascension-Seton Healthcare network, which has 2600 health facilities in the United States, including half a dozen hospitals in the Austin area, had quietly reached a data-sharing agreement with Google Health that will allow the multinational technology company access to tens of millions of Ascension-Seton patient records, in order to feed machine learning and artificial intelligence and give “Dr. Google,” as the tech giant’s healthcare search engine is informally called, an edge. The agreement is The Nightingale Project.

At one event during a podcast just prior to the onset of the pandemic (South By Southwest, 2019) Clay Johnston, then-dean of the University of Texas Dell Medical School, greeted Dr. David Feinberg—then-head of Google Health, which has since been disbanded as a separate unit of the company but whose functions continue under the Google mantle. The two physicians began a discussion of future trends in health care. Dr. Feinberg, speaking of Google, which he claimed gets one billion hits daily, said of the people who use Google Health, “We know everything about you,” a claim that did little to reduce any potential concerns about privacy. His were prophetic words nonetheless.

Preliminary reaction to the Nightingale revelation was almost uniformly skeptical (Copeland & Needleman, 2019) with most objections focusing on privacy issues—not on the traditional concerns of data scientists, which usually regard inferences drawn from statistics, in other words what the stats really mean. With Nightingale the question became, should the

stats be used at all? The Office of Civil Rights of the U.S. Department of Health and Human Services, prompted by both Republican and Democratic members of Congress, began a formal inquiry related to Ascension-Seton's release of even de-identified data to the computing services giant, again "Dr. Google." (In fact, the data received by Google was not de-identified, but Dr. Feinberg later said that those reviewing Ascension's datasets had undergone healthcare privacy training.)

It's my argument that equity in health care can be addressed in different ways. And data like that collected by the Nightingale Project can be used to achieve aims related to equity. Even a skeptical view of what Google and Ascension-Seton are doing, it now seems, is hopelessly outdated, like the kind of medicine that Pat Nixon practiced back in the day. Just three years after revelations of the existence of Nightingale, a better question today, post-pandemic or at least as we experience what may be the final throes of COVID-19, is whether the two companies, Google and Ascension Health, have gone far enough in cooperation and in service to Big Data.

What will be transferred from Ascension to Google? In brief, the data (Brush, 2020) that will be transferred includes patient name, DOB, lab results, diagnoses and vital signs; address, family members, allergies, immunizations, pre-existing conditions and radiology scans. Imagine what Dr. Nixon could have done with all of that! While including family members and marital status as exported data may seem dubious, an invasion of privacy, some medical conditions *do* run in families and presumably an argument can be made for inclusion. Address is important because as Dr. Feinberg noted in his SXSW talk, the biggest single determinant of a patient's health status in America is, literally, zip code: where one lives, goes to school, works or plays.

Ascension-Seton and Google are both holding their cards close to their chests but among the uses of this data, radiology exams will serve as input for machine learning and artificial intelligence diagnostics, and patient DNA, which has not been explicitly mentioned by either company, will presumably inform cancer treatments.

In addition, per the companies—this is not a trade secret—Ascension-Seton will receive from Google in return (Copeland, 2019) wrangled data that will be used to review individual treatment plans. Individual test results will be reviewed in the context of aggregate tests results, in order to flag dangerous trends and/or potentially-ineffective treatment plans. For reasons that remain unclear, data will also be collected on physicians who are members of treatment teams and when those members change; apparently, data is therefore being used to grade providers in some way. Data on narcotics use and narcotics prescriptions will be reviewed by Nightingale, apparently in deference to the opiate crisis in the U.S. Finally, in deference to bottom lines, analytics will look at what Ascension is charging patients or insurance companies for care and if that needs to be massaged in light of the numbers, or completely changed.

Everything is speculation at this point, unless one has access to the Congressional investigation, but as has been noted by one data scientist (Stephens-Davidowitz, 2017) there is the data that an enterprise like Google knows it is looking for when the company enters into an agreement, to inform its popular health searches for example, Dr. Google, as doctors themselves say. Presumably, the company will also have its eyes peeled for *anything else of value*. Perhaps the biggest prize for Google will be the huge datasets now being constructed everywhere in the world related to COVID-19 symptoms, and COVID treatments, and which Google, as a tech company, would presumably not otherwise have access to. One might ask

why Ascension would choose Google among suitors (it's not been made public if money is changing hands) and the answer appears simple. Google has computing resources for wrangling huge datasets that rival the ability even of many governments.

What will be done with the data? HIPAA, the Health Insurance Portability and Accountability Act of 1996, is the legal standard by which the data transfer will be judged. In that respect, Google seems to have done its homework. The major exception to HIPAA is that data from the provider (Ascension-Seton in this case) can be transferred to another business entity or contractor *for further refinement* or data-cleaning. Which is apparently exactly what Dr. Google intends. It is difficult to be sure how everything will unfold in practice. Some of the company's designs are already known because they are similar to those of every other use of Big Data in health care. And the company has already been criticized in that regard: Google was sued a few years ago by patients over a past data deal with the University of Chicago healthcare system (Miliard, 2020) with the judge dismissing the case for failure to prove harm, and the British health authority instituted an inquiry into a previously-approved use of National Health Service data by the company. In a context that does not include ethics or in which ethical concerns are left to others to determine, the Nightingale data will profit, as may we all, by some attention from Google scientists. Again, it is my thesis at this time that whether Nightingale is right or wrong is less important than whether it works.

What's this going to be like for health care research in general? The most frequently used database for health care training and advancement today is probably MIMIC, created by MIT, comprised of 58,000 records of patients admitted to the critical care unit at Beth Israel Deaconess Medical Center in Boston, covering the years 2001 to 2012. The MIMIC dataset,

although very valuable to clinicians and educators alike, has certain limitations, not the least of which is its age. Health care was a rapidly-evolving environment even before COVID-19 and 10 years—the time since the MIMIC data was collected—is a long time in that regard. Also, MIMIC contains solely ICU data, whereas the Nightingale Project includes primary care visits and “ancillary” details of patients’ care, such as payments. Dr. Nixon would have been amazed. In his memoir he devoted considerable ink to his struggles to be *paid* by his patients.

The Nightingale/Ascension data will be continuously revised and includes, apparently, as mentioned above, the single most sought-after healthcare data in the world right now, related to COVID-19, including ventilator use and ventilator demand, successful vaccines and treatments that were rejected or were ineffective. The pandemic will likely radically change healthcare modalities, regarding both study and also outcomes. (A major question of the pandemic has been why outwardly similar patients can have such vastly different courses of illness, as well as the origins of the virus itself.) MIMIC (Johnson, A., Pollard, T., & Lu, S., 2016) does not include DNA markers, which Google has hinted, but perhaps not said explicitly, will be used to inform machine learning for cancer treatments.

The Nightingale dataset as described above is said to be almost 1,000 times larger than MIMIC (Copeland, 2019) with 50 *million* records, including multiple encounters with the same patients, which can be especially valuable by providing information on pre-existing conditions that may be useful for algorithm design. This reflects the importance of large datasets for informing diagnostic predictions. But the most important plus of Nightingale and Nightingale-like platforms is that the records are not static. As noted by Matheny et al. (2021), in *Artificial Intelligence in Health Care*, “Clinical AI performance can deteriorate within a site when

practices, patterns or demographics change over time.” In other words, the machine needs to keep learning. And to achieve that end, it needs new and ever more stats.

It's useful to speculate in general terms how Google will handle this statistical trove. Although there are vast differences between ordinary statistics and what Google is attempting, there are common threads and procedures. Speaking as a former Ascension-Seton nurse, who has used a hospital electronic health record or two—which is the computerized “brain” that runs the hospital and is repository of all its data—it's not a collection of pretty, comma-separated values in a file, or something that a researcher can just upload. What Google proposes *does* involve some heavy lifting on the company's part. Medication administration, pain scores, urination and bowel movements are all collected metrics in a modern hospital. Separating what is useful from what is not is a daunting task and one that most hospitals are not prepared to perform, because their statistical interest is the *present moment*, not retrospective analysis.

The Nightingale dataset will have to be divided for analysis among the individual records that constitute a patient's provider encounters, with entries sorted and holes filled. Machine learning especially does not like missing data—something that is included in one patient record but not in another—and wrangling any similar set of data will probably be difficult. A lot of computer wattage will be required, which Google presumably has access to. Indeed, considering the task at hand, it's not surprising that Google is the entity attempting it.

What is the direct clinical importance of Big Data in healthcare? “Data” in health care has many different meanings and has long existed. Perhaps the most ubiquitous example of the term “health data” are vital signs, heart rate, blood pressure and the like, which everyone is

familiar with from our primary care visits. Even in ancient times, physicians or healers paid attention to a patient's diet and exercise, as we do today, although the attention paid and reasoning may have been faulty, because in many cultures, illness and death were not seen as *natural* or measurable phenomena. That attention to what the patient had eaten and how he or she spent his or her day was data too, and might have included gods worshipped or gods offended, depending on the era. Jump to an age closer to our own. In Pat Nixon's era, data included—although he likely would not have used the term data—age, profession, medications taken, the patient's history of illness, symptoms, alcohol consumption, and many aspects of the healthcare dominion that are very similar to our own. Neighborhood perhaps but not zip code, race but not race *and* ethnicity. The physical exam by the physician him- or herself was paramount. As time has passed, however, what constitutes healthcare data has changed even more and expanded.

In our own era, the collection of healthcare data has been reconfigured even at the level of a simple primary care visit. Exposure to toxins and pollution, sexual practices—illicit drug use, smoking history—all of these factors and more constitute modern health data. Laboratory tests specifically intended to look for anomalies or deviations from normal values or to establish baselines are taken today, such as the ubiquitous complete blood cell count (CBC) and basic metabolic panel or (BMP.) Blood sugar is measured to judge probability of diabetes, and A1C to see how that blood sugar has trended over prior months. Apparently in history, up until the time of the beginning of Pat Nixon's practice, doctors and nurses were still *tasting* patient's urine for sweetness, because the technology had not caught up with the need for diabetes treatment (Roberts, 2015). Weight is tracked down to the pound, especially among children.

Certain conditions such as sexually transmitted diseases must also be tracked and reported to county public health officials. What kind of health insurance the patient has is considered data, as well, because availability of care—“good insurance” versus “bad insurance” or no insurance at all—is a determinant of health. Race is more important than in the past because we now recognize that in our culture, minorities have worse outcomes on a wide number of health measures. In recent decades the importance of data in care has been recognized time and time again.

What data *is* has also been categorized and quantified. Health data may be cardinal, like blood tests with “normal” values. There are actual numbers that can be ranked. Other, unstructured data like physician’s or nurses’ notes also exists, and which may have uses in artificial intelligence or machine learning. Traditionally, the form the data takes—or its locale of storage—has been *local*, in the doctor’s office or even in the possession of the physician. In recent years that record has taken shape as an EHR—electronic health record—that allows a patient’s past outcomes and interventions, as well as prior baseline data, to live on *electronically* or at least until the next visit. So that the patient history may be referred to by providers the next time the patient is seen more easily. It is there, at the level of the EHR, that the federal government has made itself heard as well, beginning in 2009 when President Obama signed the American Recovery and Reinvestment Act. Its HITECH provisions (Health Information Technology for Economic and Clinical Health) have offered lavish financial inducements to providers to modernize and digitize patient records and make the various hospital or clinical operating systems *interoperable*.

The first requirement of HITECH was that hospitals and clinics use EHRs—instead of the old, ubiquitous paper charts—a mandate that was easily enough achieved.

“But,” as Reisman (2017) pointed out, “the other key HITECH goal was to achieve interoperability across all settings of care—in other words, to connect EHR systems so that physicians can easily share their patients’ records with other providers regardless of the software being used—and this task has proved to be more challenging than anticipated.” In the not too distant past, different hospitals in the same company or the same hospital system might not have been able to communicate EHR data between themselves about a given patient they shared. The goal of HITECH was not only for the two facilities in the same system to be able to exchange data but also hospitals across *different* hospital systems to exchange the patient’s data, as that patient moves from location to location or provider to provider. The difficulty has been varying standards of different providers of electronic health records and the required signing on by a variety of interested parties, including physicians, patients, government (to establish standards or for reimbursement purposes) and software vendors. The natural question is why bother or why spend as much time, effort and money on data that (1) have largely been unneeded in the past and (2) may violate patient privacy by being sharing? Two trends have emerged in health care in recent years, one on a micro level and the other macro, and both involve expanded use of data in health care, our subject here.

What is Personalized Medicine? With the decoding of the human genome, and as medicine has become more sophisticated in terms of treatments and drug modalities, a new term began to be heard in health care two decades ago: personalized medicine or “precision medicine,” as a similar concept is also called. This does not mean, according to the National

Research Council (2011), creating meds and designing specific procedures for any given individual (medicine is far too busy an endeavor and the costs to the patient would be prohibitive to customize at *that* level.) Instead, the idea is to reduce patients to *smaller* subgroups or smaller *populations* that have certain genetic characteristics or predispositions in common, in order to better treat them. This is at what can be called the micro level. On the macro scale, health care has new techniques that still affect individual patients and are empowered by computer science, such as machine learning and artificial intelligence. It's here that we enter the realm of the ubiquitous "Big Data," involving very large datasets whose very power is their size.

Big Data in Health Care Writ Large: In 2019, just before the pandemic struck, the *European Journal of Public Health* (Pastoriono, R. et al.) published a study detailing the ways that Big Data for health was being used in the European Union. "The potential of Big Data in health care relies on the ability to detect patterns and to turn high volumes of data into actionable knowledge for precision medicine and decision makers. In several contexts, the use of Big Data in health care is already offering solutions for the improvement of patient care and the generation of value in healthcare organizations. The approach requires, however, that all the relevant stakeholders collaborate and adapt the design of their systems. They must *build the technological infrastructure* [italics added] to house and converge the massive value of healthcare data, and to invest in the human capital to guide citizens into this new frontier of human health and well-being."

The authors of the paper cite the '3V's' of Big Data: volume, high velocity and high variability, which are all factors that were unknown just a century ago in medicine, in the age of

the good Dr. Nixon, but are part of Nightingale-like projects today. Indeed, the European authors add a fourth 'V,' that of veracity, or how smart the data is and the depth of the knowledge that the details can provide. This is exemplified by genetic profiles and DNA, which are especially useful in oncology. In sum, per the *European Journal of Public Health* paper, "A specific definition of what Big Data means for health research was proposed by the Health Directorate of the Directorate-General for Research and Innovation of the European Commission: *Big Data in health encompasses high volume, high diversity biological, clinical, environmental, and lifestyle information collected from single individuals to large cohorts, in relation to their health and wellness status, at one or several time points.*" What's amazing here is that European standards are more restrictive and privacy safeguard-heavy than the corresponding American standards. Yet even in Europe what is being implemented sounds very much like Nightingale or any Nightingale-like projects.

The Business of Health Care In the University of Texas' Dell Medical School class of 2020, of forty-nine students who graduated, sixteen received Masters of Business Administration degrees as well as medical diplomas, the university reported in response to an open records request. That is almost exactly one-third of the new graduates.

Certainly, an argument can be made that money is a major concern of young doctors, as is true of everyone else. There are school loans to be paid off—in the case of MDs, the total owed can be hundreds of thousands of dollars. Medical practices need to be begun or to be bought into, and doctors' families need to be fed and educated, again true of everyone. An MBA can help with juggling the financial responsibilities of both a new practice and a new home. So, too, there is wealth acquisition—physicians place number one, year after year,

among the ranks of the highest-paid professionals in the United States and presumably the world. But it's not all about money and that appears to be true of some medical students who seek MBAs.

As is the case with engineering courses taught in medical school—or that are required as prerequisites—an MBA signifies *a way of knowing*. The University of Pennsylvania's Wharton School, for example, one of the premier business programs in the country (Wharton, 2022) lists five benefits of a business degree, none of which is directly related to money, although that is presumably an ultimate goal. All are applicable to health care: 1) Greater awareness of a global market. 2) Improved communication skills. 3) Expanded professional network. 4) Increased job opportunities. 5) Better time management.

Incumbent United States Surgeon General Vivek Murtha is a MD/MBA. “Many of the greatest challenges in health care today are business problems,” Evan Rachlin, an MD/MBA told *The Atlantic* (Viswanathan, 2014). One study cited by the magazine found that doctors with a business background have a “greater tolerance for ambiguity” than do other physicians. That’s certainly a plus in modern health care. “When you read a business case for the first time, there really is no right answer, and initially it can be really frustrating having gone through trying to solve every problem and follow all the rules like medical school taught us,” said [one MD/MBA student]. “You learn how to think about uncertainty.”

[The same student] “finds this sort of thinking relevant in his work as a primary care provider, since he is often the first line of contact for patients coming in with a problem. For example, he says, “You have to think about what one piece of data you want instead of ordering every test.” My argument here is not regarding the importance of MBAs to medical

science but, instead, the importance of *business models and business thinking*, especially regarding technology and data utilization, to health care. Business practices and advances are often *mediators* of medical advances. Money counts.

The stethoscope was invented in 1816 in France, supposedly because a physician was not comfortable placing his ear against a woman's bosom to hear her heart, as had been the practice previously. Almost forty years passed before a commercial market was established and production of something like the stethoscopes we know today. This is the way business practices have been integrated into health care throughout history. Steel had to be invented and production methods established before fine surgical instruments could be produced en masse. Alexander Fleming discovered penicillin in 1928 but it wasn't until the Second World War that mass production could begin to be achieved. As noted in the case of Dr. Nixon's era, radiology was invented far before the modality of the medical x-ray as we know it came about, with the evolution of hospitals which *had the capital necessary to invest* in these costly machines. Physicians and nurses today might still be tasting urine if someone had not invented a cheap way of measuring glucose. All these examples have one thing in common, demand and supply, in other words, a market. There had to be a need, there had to be someone to invent a tool or process to supply that need, there had to be investors to fund production and a way to reach the customer. Today, data is another tool of health care—one of the most important tools, one might argue. Isn't it reasonable that there is a market for health data?

What does Google really want? While much of what Google is capable of remains to be seen, there is nonetheless a potential guide to what the company wants. Once again, it involves Austin, not Ascension-Seton but Big Pharma not Big Tech. Big Pharma, for those who are not

familiar with the term, is the loosely defined confederation of companies that produce medicines and vaccines worldwide, companies like Johnson & Johnson and Pfizer in the United States, among others—the Swiss giant Roche—and GlaxoSmithKline of Great Britain. In a relatively short period of time, roughly the post-Vietnam War era, these companies have managed to become as hated as Big Oil or any defense industry, both for bad practices and bad karma. To say nothing of high medication costs.

Five years ago, one of Big Pharma's big players (Lomax, 2017) Merck made a move in Austin. Merck is maker, among other products, of Januvia diabetes medicine and Gardasil human papilloma vaccine, used to vaccinate young women in Texas, and the company announced new plans to open a "data center" in Austin. Big Pharma has long complained of a major deficit in its existing databases, which companies rely upon to design products. Almost all databases in the U.S. used by Big Pharma contain records on what are, by heritage, Northern European whites. Which is of limited use in designing pharmaceuticals because it's too restrictive. Merck was interested in Austin, per press reports, not just because of the city's high-tech sheen, and the willingness of Ascension-Seton to partner, but because Texas has *a diverse population* including significant Latino, black and Asian-American cohorts. In the end Merck received bad publicity and withdrew from the project without ever really getting started. But as Dean Johnston of the UT Dell Medical School said at SXSW, modern health care is dependent upon one factor, big datasets. Suppose there were datasets for sale, legally, in a healthcare information market?

Nor in the rush to condemn or question Google, should we ignore exactly what the company has done and must do. Writing healthcare algorithms is not for the faint-hearted. In

Artificial Intelligence in Health Care the authors describe an effect called “causality leakage,” which can take place when attempting to design AI/machine learning, or doing inferential statistics. The example given involves sepsis—what was the number one or two killer of patients in hospitals *before* the appearance of COVID-19—and involves, basically, bacteria from a focal infection, such as on a leg, entering the bloodstream and leading to a generalized and often fatal reaction.

Sepsis is diagnosed based in large part upon a blood test called a lactic acid level. A recently revealed problem with AI comes in the hospital context and is pertinent. Most hospitals have algorithms that track a patient’s vital signs and his or her lab values in order to issue sepsis alarms to nurses to check patients, and begin protocols, to prevent this deadly complication. Nurses and physicians can also employ the algorithm on their own, based upon suspicions, in order to check a patient’s condition. So-called *algorithm creep* has had an unfortunate side effect, however. Since the majority of the time that providers checked the algorithm, the patient did indeed have sepsis, the algorithm has learned not just what vital signs or what lab results may be associated with sepsis but has also “learned” that the mere asking of the algorithm (“Does this patient have sepsis?”) is associated many times with sepsis. And that too becomes reason for the algorithm to report back sepsis. Which was not the intention of the protocol as originally devised. And becomes a deceptive indicator for patient care.

In causality leakage, which is not limited to lactic acid levels, an algorithm viewing the data, instead of looking at the actual level of lactic acid in the patient’s body, instead learns by

pairing the mere ordering of the test with a high frequency of sepsis diagnoses and then diagnoses sepsis.

An algorithm doesn't know right from wrong, in other words, the algorithm just knows probability, which may not be what the doctor intended. It is for that reason that the work that Google is doing is impressive, separating the wheat from the considerable chaff, so to speak, or the noise from the signal. Google's entire business model involves algorithms in fact, while Ascension-Seton's experience in this regard is much less. Moreover, data analytics as regards AI and machine learning has changed recently. Prior use of patient data in the hospital included Clinical Decision Systems that sent alerts to nurses, for example, of trends in vital signs. Now the effort is much more complex and related not just to diagnosing who has a disease but who *will become sick* of a given condition later, by use of DNA profiles or medical chart-mining or whatever is being developed. Again, most hospitals lack that ability. Even a system as large as Ascension-Seton with 2600 facilities across the country is not a player in the data business, except as a supplier. And again, many of these concerns were pre-COVID-19. Post COVID-19, or what we hope is post-Covid, the concern is *prediction* which is much harder.

What does the pandemic mean for the Nightingale Project? Vice President Kamala Harris was the first high-ranking member of the Biden Administration to say aloud, shortly after her inauguration (You Tube, 2021) what many of us are thinking: not *if* the next pandemic, but when. And without saying it, the Vice President raised the concomitant issue of how to address issues of privacy and fears that Big Tech is already big enough. The world of health care has changed in only a year. A better question now is how quickly Big Pharma and Big Tech can pivot to meet the needs of a potentially-endangered species, the human race.

Pre-COVID there was already discussion in the medical literature that the provisions of HIPAA and patients having final say regarding their own data were obsolete. Privacy advocates were aghast. My own view as a healthcare worker and as a member of a potentially vulnerable minority group, before COVID-19, was that Big Pharma was terrible and that Big Tech was too big. Today—one hopes that we are post-pandemic or getting there—and after having received four Pfizer vaccines—Big Pharma is my best friend. Privacy is not outmoded but must be *revised*. But how?

Nor can we forget that the most important physician in the world today is Dr. Google, including at the hospital that employs me. For example, if a physician hasn't heard of an obscure condition that a patient is reporting or a little-known med that the patient is taking, and only needs a high-level view, the doc may well google it. That is the reality of modern health care. A nurse who has to perform a procedure that he or she hasn't done recently may refer to YouTube, which is owned by Google as well. How vital is it that Dr. Google remain up to date? That's a good question. One obvious solution is that de-identified health data, like The Nightingale Project, should be "public goods," free and available to everyone, not just Big Pharma or Big Tech. If that happens, which seems unlikely, equity remains an issue. Whose data will be made public? One idea is the novel concept of advancing health equity by using patient data to *improve* minority outcomes. Minority patients will still be ripped off, so to speak, but we'll get something out of it.

Recently a study by the University of Texas Southwestern School of Medicine was released that took that approach (Murphy, C, Lupo, P., et al., 2021), using minority patient data but using it to advance minority health care. Its highlights:

“

- STUDIED COLORECTAL, TESTICULAR, KIDNEY CANCERS & NON-HODGKINS LYMPHOMA
- BLACK MEN WITH NON-HODGKINS LYMPHOMA HAD A 57% SURVIVAL RATE COMPARED TO 75% FOR WHITE MEN
- SURVIVAL DECREASED AS POVERTY INCREASED FOR THESE HIGHLY TREATABLE CANCERS AMONG ALL RACE AND ETHNIC GROUPS
- THE SURVIVAL RATE FOR BLACK PATIENTS WITH TESTICULAR CANCER WAS 88.7% COMPARED WITH 96.6 FOR WHITES
- MEN WITH PRIVATE INSURANCE HAD SURVIVAL RATES 20 PERCENT HIGHER FOR TESTICULAR, COLORECTAL AND KIDNEY CANCERS, AND NON-HODGKINS LYMPHOMA THAN FOR MEN WITH NO INSURANCE OR WITH MEDICAID ”

These powerful results were achieved with a study of just 32,000 male patients and 55,000 females, with data gathered from the Texas Cancer Registry. Imagine what could be achieved with millions of records, as is possible with Nightingale or Nightingale-like endeavors. At this point in our collective history, can we really afford to ignore or leave unused patient data? Especially data that may improve healthcare analytics regarding disease? My premise would be no.

What are the downsides? There are, nonetheless, several aspects of the Nightingale Project that are problematic, even if one accepts its potential utility and its ethical framework—which many people, frankly, do not. Among the major issues regarding use of *any* dataset, not just Nightingale, are the danger inherent in inferential statistics (Stephens-Dadowitz, 2017) and the danger that is just now becoming manifest regarding AI/machine learning. Statistical inference, for example, has long been a fraught subject in data science because “cleaning up” a dataset, which is what Google will do in Nightingale, and deciding what conclusions can be drawn from it, or running new queries that may not have been intended to be run when the data was collected, can introduce error. Statistical inference also involves decisions made by scientists who are probably drawn from the majority culture; in fact that is one of the operant

descriptions of the majority culture, not just being a Caucasian in America, in this instance, but being a Caucasian scientist. And whose word may go unchallenged. Very likely one's own assumptions, biases and ignorance will also play a part in decisions made. Which may have a higher likelihood of disenfranchising minorities or special populations, even unintentionally.

The "conclusions" therefore that one draws about any subject, including a dataset, are necessarily part of who is drawing the conclusion. Which is why special populations need to have a seat at the table and why positionality statements may be asked of researchers. None of which are we likely to get here, with Google, in Project Nightingale, be that as it may.

The Ethical Conundrum We all know something about the multinational Google whose motto is "Do No Evil" or "Don't Be Evil" depending on the era, which is different—one might add—from "Do No Wrong" and may be context-specific. Obviously, considering the company's size and history, corporate ethics are fair game for doubters. So, too, the issue of what can be done to assure the accuracy of machine learning and artificial intelligence algorithms that are fueled by Big Data sets, like those that are part of Nightingale, has only begun to be explored and may cause *concern*, despite the best intentions of algorithm-writers. Such as the faulty sepsis protocol mentioned above and algorithm creep. An algorithm only knows what it's taught, or learns, and in some instances it learns the wrong thing. Or the algorithm may have had the wrong person for a teacher, one might say, very often a member of the majority culture, as when a Facebook algorithm was reported last year to associate mention of "primates" with black people, to Facebook's most recent chagrin.

Datasets can also be queried to produce fields that are not part of the original design. Which leads to the question—is that a useful add-on or unintended creep? So, too, conclusions

drawn from datasets are always prone to error and presumably Google intends to draw inferences from the data that the company receives. Artificial intelligence is prone to any number of errors based upon the predilections of the researchers and the designs themselves, in other words, and the mere fact that we as a society determine that healthcare data is too important not to be used, especially in an era of health crises, is not the same thing as saying that racially-challenged or overtly racist design is not an issue. Google or whoever the data-user may be must instead redouble efforts, at using the information responsibly. And probably under government supervision.

Stephens-Dadowitz (2017) makes note of what he calls the “curse of dimensionality” in using Big Data sets to draw causation. For example the ultimate Big Data set may be DNA, with tens of thousands of loci for any individual gene. And perhaps the worst example of bad use of data in medicine has been past attempts to draw conclusions about intelligence based upon race (eugenics) or bumps on the skull (phrenology.) Considering all the data points on a strand of DNA, one of them *will* likely correlate with intelligence, in the same way a groundhog’s appearance may, some years, correlate with the arrival of spring—if only by chance. But that is not causation or even prediction.

Agreeing, therefore, that Google can have the data legally does not mean that the company should have carte blanche in its use. This is less of a concern with Ascension Seton than with Google because obviously the hospital had a right to collect the information in the first place and there are already safeguards in place, regarding use of patient data by healthcare providers. And AI and machine learning have real and valuable uses in health care, on the Ascension side of the agreement. Should we toss out those uses—like AI’s ability to read

an x-ray for breast cancer better than a radiologist can—merely because someone has also created an app to determine if one’s social media postings are indicative of mental illness? My answer would be no, *especially* in the age of pandemic.

There are other issues that may arise that society needs to be prepared to discuss and regulate, regarding datasets in public health, without throwing out the proverbial baby with the bath water. Healthcare providers and regulators must be concerned with *what Google finds in the data* that may be unknown now but proves to be of value to Google or may help control business costs for Ascension-Seton. This may be the case because data wrangling, which is what Google will do with the Ascension-Seton material, can involve finding something new. Will an expensive zip code and/or a listing of a particular insurance type be used to push medication advertisements on websites of users whom Google or Seton-Ascension have identified? That would be inappropriate use, it seems. Genetic data is particularly problematic. There is the continuing concern that providers will leak to insurance companies the genetic markers that point to a patient’s long-term risks for expensive illnesses, so that the insurance company will not have to pay, if the company refuses to provide coverage in the first place. Anonymized data may turn out not to be anonymous after all, as the threat—if not the fact—of tracing back de-identified data exists. Any system created by humans runs the risk of being gamed, in this case using tracebacks, if there is value to be had in doing so.

Which special populations are most at risk? The *Financial Times* reported two years ago (Munschi, 2020) another aspect of expanded access to healthcare data that can easily devolve into exploitation of special populations. Apparently the most attractive genetic data for researchers involves Africans and African-ancestry individuals, because of greater variability at

DNA loci of black people. That means that African Americans for example are at greater risk for having their data appropriated than whites, which adds to the possible abuse of Negroes in this country that, historically, has already taken place. Henrietta Lacks, a black woman, had her genetic profile and genetic ancestry stolen by Johns Hopkins, the most prestigious medical institution in the country, to create the first immortal human cell line and that cell line is *still* a resource in world health care today. Johns Hopkins's misappropriation of the Lacks' genome, in Johns Hopkins' capacity as a preeminent research hospital, seems especially egregious. But at the time, during Dr. Nixon's era of medicine, no patient consent was required for use of research samples, which points to the need for ethical standards that keep up with medical technology. And Google, lest we forget, has no such pretensions to philanthropy, unlike Johns Hopkins. Google is a business and a very profitable one that wants to stay that way.

Often in cases in which a special population or a population of study subjects is disadvantaged by participation there may be a way to make amends. There are opt-outs or participatory remedies that make joining a study more tolerable or even beneficial. There is a specific argument, like in the UT Southwestern colorectal study cited above, that there are specific and tangible benefits to the subject. In other cases, none of those outcomes or possibilities seems possible. The Google-Ascension approach involves data extraction and data inference that will be far removed from the patient's awareness or influence on outcomes. In Nightingale, patients will not have the option to choose not to participate. Indeed, as originally designed they were not even in a position to know their patient information was being used, until the *Wall Street Journal* report. Opportunities to participate in design are obviously limited as well.

In the case of other “captive” study subject participants, prisoners for example, one remedy has been payment, of \$20 or \$50 into a prisoner’s commissary account. With Nightingale the idea of offering any kind of meaningful financial inducement to the individual owners of 50 million patient records seems far-fetched, even for a company of Google’s size and capitalization. The very size of the datasets make any individual compensation unlikely. What we may be looking at instead is the least concrete and perhaps most illusory promise possible, that giving up the data is altruistic and will help humanity. For some people, that may still not *be enough*.

Privacy and the Market for Healthcare Data The pandemic has already brought changes to both the practice of medicine and medical research, including the use of data. In the former case, COVID-19 has turned the peer-review system of research publication on its head. In some cases, there has been no peer-review at all. A medical journal in Croatia (Chirico, F., Teixeira da Silva, J. A., & Magnavita, N., 2020) reported being “hit by a storm of open-access pandemic-related manuscripts through preprint platforms, but also through accelerated review processes.” The same article reported, “An analysis of some of the most important servers (Google Scholar, bioRxiv, medRxiv, and ChemRxiv) indicated that 60% of studies were preprints, meaning that they reported non-peer-reviewed information.” In other cases, peer-review was curtailed, or rushed, or left until later, after the pandemic, all in the name of addressing a healthcare crisis. Some changes in the research pipeline are likely to remain post-pandemic, as journals review lessons learned, just as providers decide what changes to their practices should remain. Nor was that all. In the realm of data and privacy, increased use of telehealth, in order to avoid direct patient contact, if not necessary, presented its own set of issues, including the

possibility of insecure networks if, for example, a patient was being examined virtually at home and required use of his or her own Wifi, which might or might not involve unencrypted transmission. But the biggest change that pandemic may have brought to health care, and which is also represented by the Nightingale Project, regards privacy. And there it seems that Nightingale may not be so novel after all.

An interesting article in this regard was published last year in the New England Journal of Medicine, "HIPAA and the Leak of 'De-Identified' EHR Data," in which the authors (Mandl, K. & Perakslis, E.) note, as mentioned previously, that HIPAA does allow uses like those by Nightingale of de-identified data by a third party or a contractor who further refines data, like Google's stated role with Nightingale. (The difference being that in Nightingale identified data will also be used, when it is returned to the hospital after wrangling.) Furthermore, per these authors, "Organizations that don't qualify as business associates under HIPAA may also gain access to and use deidentified data sets. Such policies have enabled the rise of a multibillion-dollar industry comprising dozens of health-data aggregation companies and hundreds more companies producing tools and technologies that aggregate, link, and monetize EHR data." These are, however, unorganized, individual markets or individual marketers of health data. My own view is that organized markets with well-formulated privacy protections, mandated by government, is a logical next step. As with other healthcare changes brought on by COVID-19, some of these data practices are likely here for good. And a historical precedent already exists for the potential value of making them permanent. According to the U.S. Commodity Futures Trading Commission (CFTC, undated) during the late 19th and early 20th centuries, as the Chicago grain exchanges/markets were being established, the Illinois legislature stepped in and

set standards for various kinds and grades of grain, in order to protect the public and the Midwest's farmers.

The China Solution Last November federal intelligence officials issued a warning (Barnes, 2021) that the People's Republic of China has tried and likely will continue to try to steal large healthcare datasets from the United States, in order to fuel China's own healthcare enterprise. Although China has a population four times larger than the U.S., the difficulty for the PRC is that China is much more homogenous than the U.S. On the mainland, 94% of the Chinese are ethnic Han (U.S.-China Economic and Security Review Commission, 2021). Even though there are over fifty minority groups, those groups do not for example include black people. Among partner states of Beijing or countries that are friendly to the Chinese government there are doubtless significant minority demographic groups but they are in countries where, unlike the U.S., *the data has not already been collected*. Half the U.S. population, on the other hand, is minority, and genetic data has already been collected or can be easily collected from millions of individuals, because the infrastructure to do so already exists. There are for example over five million ethnic Chinese living in the U.S., which may mean that American health data are more complete than many Asian countries. There is therefore a lot to steal in American datasets that are controlled by government, hospitals, universities and the like.

Another frequent U.S. adversary (James & Scherer, 2020) that has also been accused of stealing American hospital data is Russia and the Russians face the same challenge as the Chinese. Over 80% of the country's population is "European Russian," and hence of less value for data mining. All of this is important, in the context of considering healthcare data as a "public good," because the U.S. government may be unlikely—even for population health

purposes and to fight pandemic—to make American data more available to governments that might otherwise just take it. The question becomes how to ensure that data is used freely but still kept under American control.

The key may be found in one country and it is not the United States. If one believes that using healthcare data—especially COVID data and genomic registries—in an effective manner in times of crisis is incumbent upon society, that is dependent in some sense in what other countries do, as well. Last year’s report, cited above, from the U.S.-China Economic and Security Review Commission, an independent governmental think tank that reports to Congress and the President, addressed this issue. In the white paper on China’s healthcare challenges (which in some respects mimic America’s, including chronic disease incidence, an aging population and a rural-urban health status divide) the Commission paid particular attention to the kind of data that the Nightingale Project is designed to gather. The Commission quoted the PRC government as describing healthcare data as a “fundamental, strategic national resource.”

The report continues, “The Chinese government’s efforts to collect healthcare data extend to the United States, whose healthcare data are particularly valuable due to the ethnic diversity of the U.S. population.” If that is correct, the PRC’s view of healthcare data is much like everyone else’s—including Google—something that must be obtained, and perhaps by means that have not been tried in the past, as in revising privacy rights. The Commission’s report notes that while Americans see the right of the individual to control his or her hospital data, that is not necessarily what is at play in the People’s Republic. There are other barriers: “While Chinese firms aggressively pursue the collection of healthcare data abroad, the Chinese government places strict limitations on the sharing of Chinese healthcare data with foreign

companies or researchers, in contrast to the open research environment of the United States. Some of the harshest provisions are found in China's human genetic resource regulations, which state that foreign parties cannot independently collect, store, use, transfer, or export human biospecimens obtained in China."

CONCLUSION

While having so many differences in views on healthcare data may seem to be an insurmountable barrier, differing beliefs may actually be grounds for compromise and cooperation in a world in which disease knows no borders. Three possible scenarios appear possible here, one altruistic and unlikely; one transactional and to be hoped for; and one antagonistic and on-going. What Google has done with the Nightingale Project, even if as designed it never comes to pass, this kind of more robust use of hospital data may be less an anomaly or a privacy-rights nightmare, and more a necessary wave of the future.

The first possibility, however unlikely, is that de-identified healthcare data will be made a “free” or “public good,” as mentioned before. This would emulate the data collection practices of the Nightingale Project but would eliminate many of the advantages enjoyed by Google. In this scenario, the U.S. government would maintain databanks for researchers and for entrepreneurs and this accessibility would effectively level the playing field against Big Tech and would have the added benefit of making data available even to foreigners, should we face another pandemic and the healthcare industry—including firms abroad—needs to react quickly. But as long as the world’s largest country does not release its data, this seems a kind of altruism that is unlikely to fly in Washington.

A second possibility is to do what Americans do well—that is, create a market. In this case for healthcare data. If this information really is as valuable as is believed, monetizing the datasets could serve a couple of purposes, including assuring access by foreign governments like China’s—if the PRC pays for it—and giving American exports a needed boost. This also

reduces the possibility of illegal transfers. The possibility of hacking is reduced, although not eliminated, if foreign enterprises can just buy what they want instead of declaring cyber war. In some sense this seems to be the most favorable outcome—data markets, like the Chicago grain exchanges of years past, or the futures/derivatives exchanges today, in which healthcare data can be bought, sold or traded. The third option is what we have now, a closed system for each country, inviting theft and hacking, in which healthcare data is weaponized or ignored entirely. Which provides limited added value to public health or private enterprise.

During the last two years, quarrels over vaccinations and masks have taught us how strongly-held are beliefs about the healthcare domain. Yet survival of the species is a potent argument for exactly the kind of change in healthcare data usage, as has happened since pandemic. Google was ahead of its time and may be onto something important, but should be surveilled closely and other options explored. There is in Big Data still a conflict whose prolongation does not bode well for the future and may require new thinking but which is, thankfully, beyond our scope here.

BIBLIOGRAPHY

Barnes, J. (2021, October 22). U.S. warns of efforts by China to collect genetic data. *The New York Times*. <https://www.nytimes.com/2021/10/22/us/politics/china-genetic-data-collection.html>

Brush, K. (Last updated, February 2020). Project Nightingale. *SearchHealthIT*. <https://www.techtarget.com/searchhealthit/definition/Project-Nightingale>

Burns, C. (1995). *Handbook of Texas*. Texas State Historical Association.

Chirico, F., Teixeira da Silva, J. A., & Magnavita, N. (2020). "Questionable" peer review in the publishing pandemic during the time of COVID-19: implications for policy makers and stakeholders. *Croatian medical journal*, 61(3), 300–301. <https://doi.org/10.3325/cmj.2020.61.300>

Cohen, I., Mello, M. (2019). Big data, big tech and protecting patient privacy. *Journal of the American Medical Association*, 322(12) doi:10.1001/jama.2019.11365, 1141-1142.

Copeland, R. (2019, November 11). Google's "Project Nightingale" gathers personal health data on millions of Americans. *The Wall Street Journal*. <https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790>

Copeland, R., & Needleman, S. (2019, November 12). Google's "Project Nightingale" triggers federal inquiry. *The Wall Street Journal*. <https://www.wsj.com/articles/behind-googles-project-nightingale-a-health-data-gold-mine-of-50-million-patients-11573571867>

Harvard Medical School. (2022). Prerequisite courses. <https://meded.hms.harvard.edu/admissions-prerequisite-courses>

Howell, Joel D. (2016). Early clinical use of the X-ray. *Transactions of the American Clinical and Climatological Association*, 127, 341-349. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5216491/>

James, W. & Scherer, S. (2020, July 16). Russia trying to steal COVID-19 vaccine data, say UK, U.S. and Canada. *Reuters*. <https://www.reuters.com/article/us-health-coronavirus-cyber/russia-trying-to-steal-covid-19-vaccine-data-say-uk-u-s-and-canada-idUSKCN24H236>

Johnson, A., Pollard, T., Shen, L. *et al.* (2016). MIMIC-III, a freely accessible critical care database. *Sci Data* **3**, 160035. <https://doi.org/10.1038/sdata.2016.35>

Lomax, L. (2017, May 17). The gentrification zone. *Texas Mostly*.
<https://texasmostly.blogspot.com/2017/05/the-gentrification-zone.html>

Los Angeles Times. (2021, April 26). *VP Harris: we must be prepared for the next pandemic* [Video]. You Tube. https://www.youtube.com/watch?v=o4lxx-NR_os

Mandl, K., Perakslis, E. (2021). HIPAA and the Leak of “De-identified” EHR Data. *New England Journal of Medicine*, *384*:2171-2173
doi: 10.1056/NEJMp2102616

Matheny, M., S. Thadaney Israni, M. Ahmed, and D. Whicher, Editors. 2019. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise*

Miliard, M. (2020, September 9). Judge dismisses data privacy suit against university of Chicago and Google. *HealthcareITNews*. <https://www.healthcareitnews.com/news/judge-dismisses-data-privacy-suit-against-university-chicago-google>

Munshi, N. (2020, March 4). How unlocking the secrets of African DNA could change the world. *Financial Times*. <https://www.ft.com/content/eed0555c-5e2b-11ea-b0ab-339c2307bcd4>

Murphy, C. C., Lupo, P. J., Roth, M. E., Winick, N. J., & Pruitt, S. L. (2021). Disparities in Cancer Survival Among Adolescents and Young Adults: A Population-Based Study of 88 000 Patients. *Journal of the National Cancer Institute*, *113*(8), 1074–1083.
<https://doi.org/10.1093/jnci/djab006>

National Research Council. (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*.
<https://pubmed.ncbi.nlm.nih.gov/22536618/>

Nixon, P.I. (1979). *Pat Nixon of Texas: Autobiography of a Doctor*. Texas A&M University Press.

Pastoriano, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Riccardi, W., Boccia, S. (2019). Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *European Journal of Public Health*, *29*(3), 23-27.
<https://pubmed.ncbi.nlm.nih.gov/31738444/>

Reisman, M. (2017). EHRs: The challenge of making electronic data usable and interoperable. *Pharmacy and Therapeutics*, *42*(9): 572–575.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5565131/>

Roberts, J. (2015, December 8). Sickening sweet. *Science History*.
<https://www.sciencehistory.org/distillations/sickening-sweet>

Skloot, R. (2010). *The Immortal Life of Henrietta Lacks*. Crown.

Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data and what the internet can tell us about who we really are*. Dey Street Press

SXSW. (2019). David Feinberg & Clay Johnston on the future of health care.
https://www.youtube.com/watch?v=e7Wkv_YITZM

U.S. Commodity Futures Trading Commission. (Undated). History of the CFTC.
https://www.cftc.gov/About/HistoryoftheCFTC/history_precftc.html

U.S.-China Economic and Security Review Commission. (2021). China's healthcare system: Addressing capacity shortfalls before and after COVID-19.
<https://www.uscc.gov/research/chinas-healthcare-system-addressing-capacity-shortfalls-and-after-covid-19>

Viswanathan, V. (2014, September 29). The rise of the M.D./M.B.A. degree. *The Atlantic*.
<https://www.theatlantic.com/education/archive/2014/09/the-rise-of-the-mdmba-degree/380683/>

Wharton School. (2022). Should you get an MBA? <https://mba.wharton.upenn.edu/why-get-an-mba/>

VITA

Lucius Lomax is a journalist by original training. He has worked as a nurse for the last twenty years, in addition to practicing journalism. He has a Bachelor of Science of Nursing, Highest Honors, from the University of Texas Medical Branch, and is currently enrolled as a master's student in nursing informatics at Texas Tech Health Sciences Center School of Nursing.