



Laptop Pro

BIG DATA IN MEDICINE

BY HEATHER ZEIGER, MS, MA
RESEARCH ANALYST

The Rise of Big Data in Research

At The Center for Bioethics & Human Dignity's 2015 summer conference, Dr. Jimmy Lin, founder and president of Rare Genomics Institute and director of Clinical Genomics at the National Institutes of Health National Cancer Institute, discussed how his group is helping to make genomics a clinical reality. Robert Stone, one of his first patients and now a teenager, has been confined to a wheelchair since he was one year old. His parents spent thousands of dollars trying to figure out why their healthy baby suddenly lost motor function. They eventu-

“Big Data rightfully placed within the context of a research program can serve as a valuable tool in the scientist’s toolbox.”

ally found Dr. Lin's group who connected the Stone family to Johns Hopkins and Baylor College of Medicine to get Stone's genome sequenced and analyzed by specialists. By comparing his genome to thousands of other genomes, they discovered that Stone has a mutation in the PRPKA gene, known as *Dystonia 16*. He is one of only nine patients in medical history to have this disease.¹ Now Dr. Lin's group is helping the Stone family to connect with specialists who work on therapies for genetic diseases.

A more traditional route, in which a single doctor or institution had investigated Stone's case, would have left the family in the dark because they would not have had the data pool available to compare Robert Stone's genome to others. Companies like Rare Genomics take a Big Data approach to solving medical puzzles. This approach emphasizes large data sets, algorithm-based analytics, and collaboration. While collecting data, even large amounts of data, is not new, as David Bollier of the Aspen Institute points out, the growing scale, sophistication, and ubiquity of data-crunching to identify novel patterns of information and inference is new.² Now companies

and research organizations can collect petabytes (=1,000,000 gigabytes or 1,000 terabytes) of data that can be sifted and analyzed with ever more sophisticated algorithms.

According to a recent *Nature* article, the Human Genome Project instigated this new approach as the first large-scale, government-funded Big Data endeavor, or what the article calls “consortium science.”³ At the time, their ambitious goal was to sequence all 3 billion base pairs of the human genome in hopes that this would provide clues to the genetic causes of certain diseases. In order to accomplish this, they had to work with a diverse

team of researchers hailing from various fields. Their work was as much about creating the technology as collecting the data.

Today there are several Big Data projects that include initiatives such as the 1,000 Genomes Project, the Cancer Genome Atlas, the Human Microbiome Project, the U.S. Precision Medicine Initiative, and the U.S. BRAIN Initiative. All of these involve acquiring massive amounts of data, and, at times, developing the technology to store and analyze the data concurrent with its collection. Many of these projects hope to use these databases to identify diseases and develop therapies, just as Dr. Lin's did with the Stone case. They also hope to use the data for many other future studies.

Aside from consortium-based projects, electronic medical records (EMRs) are another area where Big Data is changing medicine. Sifting through EMRs has allowed researchers to re-purpose drugs. Furthermore, personal device trackers, like FitBit, provide daily health data to help with preventative medicine and tracking diseases. These innovative projects offer many benefits but they also

raise bioethical concerns over privacy and informed consent.

How Is Big Data Being Used in Medicine?

Critics caution against an overly optimistic view of what Big Data can do. There is a cadre of people who believe large data sets coupled with sophisticated algorithms can replace clinical trials or the scientific method altogether.⁴ This appears to overestimate what Big Data is capable of doing; however, Big Data rightfully placed within the context of a research program can serve as a valuable tool in the scientist's toolbox.

Dr. Peter Yu, president of the American Society of Cancer Oncology, has expressed optimism about what Big Data can do to help cancer research.⁵ While clinical trials are still the “gold standard,” Big Data can answer some questions by revealing correlations, saving researchers time and money. One example of this is in breast cancer research, where tumor databases such as The Cancer Genome Atlas or the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) help identify genetic biomarkers that distinguish one tumor from another. There are many types of breast cancers, and by treating the cancer based on the type of tumor rather than its location on the body, doctors have seen better overall results in treatment.⁶

Another way that researchers are making use of a Big Data approach in medicine is in drug development. Often it is prohibitively expensive to research and develop specialty drugs that would be used for only a small subset of the population. However, using electronic health records (EHRs) and data analytics, researchers can find correlations between a certain therapeutic effect and a drug that has already been FDA-approved for another purpose.

One of the first cases to use EHRs to find a secondary use of a drug was a 2014 study at Vanderbilt University on metformin and cancer. Doctors noticed a correlation between a decreased incidence of cancer and people who were taking metformin,

a drug typically prescribed to regulate type 2 diabetes. An analytics study of EHRs found that diabetic patients on metformin had a 23% increased survival rate after being diagnosed with several types of site-specific cancers compared to the non-diabetic population.⁷ This correlation was verified by comparing the Vanderbilt data to EHR data from the Mayo Clinic and by independent review by thoracic nurses who examined the charts to determine drug exposure. Additional analysis showed that, indeed, patients on metformin saw a decrease in mortality compared to both non-diabetic patients and diabetic patients that were not on metformin.

What Are the Ethical Concerns Associated with Big Data?

The Human Genome Project set aside funds to investigate the bioethical issues surrounding the project, but many consortium-based research projects do not have such programs, even though Big Data projects pose important bioethical questions when it comes to privacy and informed consent. While anonymizing genetic data was at one time a possibility, current technologies make it impossible to have truly anonymous data.⁸ Furthermore, Big Data projects typically try to collect data that will have multiple uses, including future experiments that have not been thought of yet. Additionally, Big Data projects are collaborative, and, while sharing data helps with scientific discovery, it also means that data is accessible to more people, including hackers, and raises issues of data security.

In talking with Dr. Lin about how his group deals with the issue of privacy, he points out that there is always a balance between public interest and individual privacy. His group collaborates with research facilities around the world. One way that his group deals with privacy is to give patients complete control over their data, and to make them partners in the research process. The patient requests his or her data and provides it to the various care facilities and research institutions. This prevents institutional competition

and data hoarding, and it lets the patient decide who receives their data.

For many patients who are uncomfortable with sharing their genomic data, Dr. Lin says it is often an issue of risk versus benefits. Patients who are already sick have less to lose, and are more willing to share their data. Healthy patients are the

While anonymizing genetic data was at one time a possibility, current technologies make it impossible to have truly anonymous data.

ones who are often hesitant to share their genetic data for fear that they will face discrimination either by insurance companies or by their employers. Notably, they fear many of the things against which GINA (The Genetic Information Nondiscrimination Act of 2008) was meant to protect.

Medical data hacking is a growing problem. Any time data is shared over the Internet, it is at risk for hacking. Hospitals and insurance companies have been recent targets for hackers because medical data has a large black market value. Unlike credit card data, which is only good until the card is reported stolen, stolen medical data can be misused for much longer. It is often used for identity theft, obtaining expensive medical procedures without insurance, and for blackmail. This can go on for months until a patient realizes someone has stolen his or her data. One recent high-profile example of hacking involved Anthem, the company that owns Blue Cross Blue Shield. Last year they reported a hack that compromised 80 million customers' and employees' data. To put this in perspective, Anthem reports stopping about 200 hacking attempts per day.

The Common Rule and Informed Consent

Finally, Big Data research is running into problems with informed consent. The current policy in the United States for government-funded research involving human subjects, known as the Common Rule, does not require consent for second-

ary use of biological specimens, which is why it is currently being revised. Author Rebecca Skloot brought to light one of the most egregious cases of secondary biospecimen use without consent in her New York Times bestseller, *The Immortal Life of Henrietta Lacks*. In the case of Henrietta Lacks, doctors removed some cells from her cervical cancer biopsy

and, after growing them in the lab, found that they were able to grow indefinitely. At the time, doctors were not required to obtain informed consent from Mrs. Lacks. But, as it turned out, these cells, known as HeLa cells, became one of the first immortal human cell lines and have been used in countless studies and publications since the 1950s.

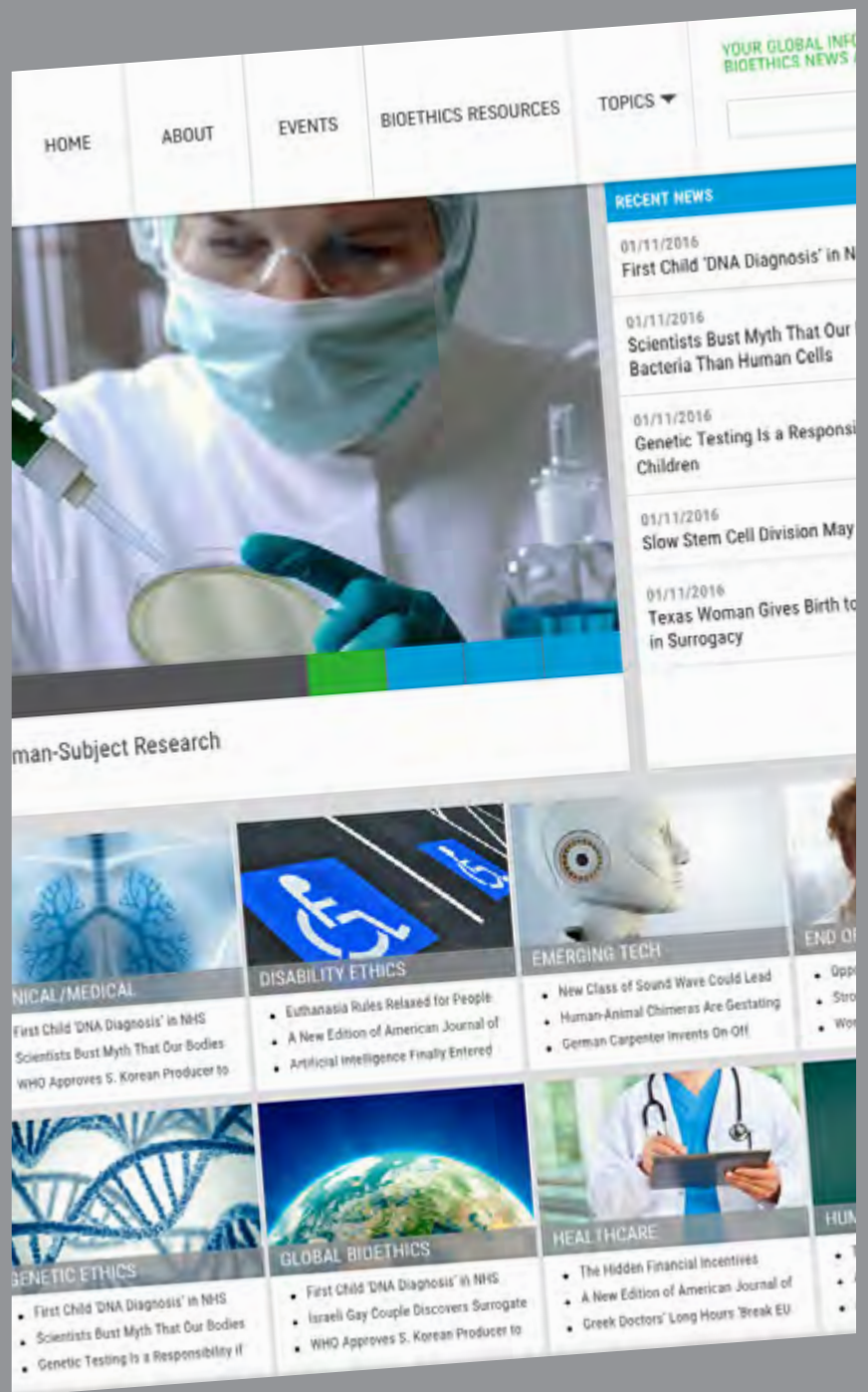
Flash forward to the 2000s, when people are still using and sequencing HeLa cell for research purposes so they can compare their results to prior studies. HeLa cells are the only cell lines currently in use that are still identified by the patient's name, and with current technologies, people can find out genetic information regarding Mrs. Lacks' children and grandchildren. Her story is an important lesson because today there are many biospecimens that have been collected over the years without the patient knowing that his or her biomaterial will be used for research purposes. Even when certain genetic markers have been removed, the biospecimens can still be re-identified.

Many Big Data projects involve bio-banking and collaboration across multiple institutions. The current wording in the Common Rule does not adequately address patients' concerns for consent and privacy. For this reason, the Common Rule is being updated to include consent for the collection and use of biospecimens for research purposes, and it includes changes to the institutional review process to accommodate research across multiple institutions.⁹

Big Data can serve as a powerful tool in the researcher's tool box to solve difficult puzzles like Robert Stone's mutation or determining why some breast cancers respond differently to treatment than others using large genomic databases. It can help drive down the cost of pharmaceutical research and development by investigating secondary uses of old medicine using millions of electronic medical records. But, just as scope and its accessibility are two of the major advantages to Big Data, they are also the two areas that are cause for concern for protecting patient privacy and informed consent. ●●●

- 1 Jimmy Lin, in discussion with the author, December 2015; Jimmy Lin, "Solving the Mystery of Rare Diseases with Technology and Crowdfunding," *TEDx-MidAtlantic*, Apr. 22, 2014, <https://www.youtube.com/watch?v=evC3HGI-Lnc>.
- 2 David Bollier, *The Promise and Peril of Big Data* (Queenstown, MD: Aspen Institute, 2010).
- 3 Eric D. Green, James D. Watson, and Francis S. Collins, "Human Genome Project: Twenty-Five Years of Big Biology," *Nature* 526, no. 7571 (2015), <http://www.nature.com/news/human-genome-project-twenty-five-years-of-big-biology-1.18436>.
- 4 Chris Anderson "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, June 23, 2008, <http://www.wired.com/2008/06/pb-theory/>.
- 5 Gabriel Miller "ASCO President Peter Yu, MD, on Big Data, Big Themes for Upcoming Annual Meeting," *Medscape*, May 27, 2015, <http://www.medscape.com/viewarticle/845246>.
- 6 Jill U. Adams, "Genetics: Big Hopes for Big Data," *Nature* 527, no. 7578 (2015), http://www.nature.com/nature/journal/v527/n7578_supp/full/527S108a.html#close.
- 7 Hua Xu et al., "Validating Drug Repurposing Signals Using Electronic Health Records: A Case Study of Metformin Associated with Reduced Cancer Mortality," *Journal of the American Medical Informatics Association* 22, no. 1 (2014): 1–10, doi:10.1136/ami-jnl-2014-002649.
- 8 Jennifer Cousin-Frankel, "Trust Me, I'm a Medical Researcher," *Science* 347, no. 6221 (2015), <https://www.sciencemag.org/content/347/6221/501.summary?related-urls=yes&legid=sci;347/6221/501>.
- 9 Christine Grady et al., "Broad Consent for Research with Biological Samples: Workshop Conclusions," *American Journal of Bioethics* 15, no. 9 (2015) 34–42, doi:10.1080/15265161.2015.1062162. The U.S. government has recently extended the comment period for the Common Rule: Notice of Proposed Rulemaking to January, 2016.

VISIT Bioethics.com



Your global source for
bioethic news and issues