Editorial

The way forward for big data analysis in ophthalmology and medicine

Stephen G. Odaibo[1,2], **David G. Odaibo**[1,3]

[1]Quantum Lucid Research Laboratories, Dubuque, Iowa, USA; [2]Medical Associates Clinic, Dubuque, Iowa, USA; [3]University of Alabama Birmingham, Birmingham, USA

Correspondence to: Stephen G. Odaibo, M.S.(Math), M.S.(Comp. Sci.), M.D., Chief Scientist and Founder. Quantum Lucid Research Laboratories, Dubuque, Iowa, USA. Email: stephen.odaibo@qlucid.com.

Submitted Feb 15, 2016. Accepted for publication Mar 01, 2016.
doi: 10.3978/j.issn.1000-4432.2016.03.02
View this article at: http://dx.doi.org/10.3978/j.issn.1000-4432.2016.03.02

It was the best of times; it was the worst of times. It was the season of population medicine; it was the season of personalized medicine—or was it? Charles Dickens' fictional depiction of late eighteenth century France walked a tense line that masterfully narrated the untenable socioeconomic disparity of its day. He described the forces that inevitably led to the French revolution. Today in the world of medicine we are on the verge of similar untenable tension. Unravelling of the human genome was indeed an epochal event that brought with it the promise of personalized medicine. Thirteen years later, this is a promise that is certainly yet to be fulfilled. We are unable to bear the massive weight of the unmet promise of personalized medicine. It is therefore little surprise that any glimmers of its potential realization get great attention and scrutiny. One such glimmer is the work of Dr. J. William Harbour and colleagues in which they have correlated gene expression profiles with prognosis in ocular melanoma (1). They classified (2) tumor gene expression profiles into two classes, those associated with a higher likelihood of metastasis and those associated with a lower likelihood of metastasis. Their clinically-relevant result is an example that rekindles hope that our massive investment in unravelling the human genome may yet yield a return. Our team at Quantum Lucid Research Laboratories consists of ophthalmologists, mathematicians, computer scientists, physicists, and engineers, and has a unique appreciation for how such work bridges ostensibly disparate worlds. Data from the human genome project is an example of big data. Other examples of big data include the troves of imaging and clinical laboratory data which we are accumulating in our health centers and other institutions. It is vital that such big data not lay dormant, but instead be translated into better means for diagnosing and treating disease.

Our prescription for the way forward for big data in ophthalmology and medicine includes the following three interrelated tenets: empiricism, data, and computing. We call it the Odaibo Big Data Framework.

Empiricism

Empiricism is the notion that knowledge is discovered through experiment and experience. In the seventeenth century, Galileo Galilei and his contemporaries popularized this millennia-old school of thought and formally introduced it to the Western world. Empiricism did transform and inform the scientific process, and today needs to be revisited in the area of big data. Empiricism is the executive and ideological arm of the Odaibo Big Data Framework. A wholehearted subscription to empiricism is necessary to make headway. We must acknowledge how little we understand and simultaneously realize how relatively little we actually need to understand to make progress. Big data is at odds with the way science has been done for the last couple of centuries. Science had and continues to be done from small labs studying one thing at a time in relative isolation. This type of "cottage science" can and has yielded great rewards. However it is by itself grossly inadequate to help us translate existing data into better treatments and cures for diseases. Cottage science often seeks to know *why*, while big data often seeks to know *what*. The two are not mutually exclusive since the *what* typically gives birth to the *why*. The 1965 Physics Nobel Laureate Julian Schwinger once said that the bicycle could never have been designed from first principles. In his view, it could never have been designed from knowledge of *whys*. Instead it required trial and error. It required experimentation—a sequence of *whats*. In the case of ocular melanoma, no one currently knows exactly *why* certain gene expression profiles are associated with more aggressive forms of the disease, while other profiles are associated with less aggressive forms. The exact mechanisms are certainly worth investigating. However, we must recognize that the mechanism questions did not exist till the big data analysis gave rise to them. In the Odaibo Big Data Framework, the investigator embraces the role of listener. Here, the investigator consciously allows nature tell its story through the data.

Data, big data

Empiricism and data are intrinsically-linked. Empiricism says knowledge arises from experience and experimentation. The outcome of an experiment is data. As a scientific community our capacity to gather large amounts of data has far outpaced our capacity to organize and make sense of this data. It has also far outpaced our ability to read this data and hear what story it tells. The way forward must be integrated in such a way that data collection and storage are done in a manner that facilitates accessibility, collaboration, and analysis. Systems must be designed with that end in mind. Data must be purposefully categorized in a hierarchical way that aims to allow modular and near-arbitrary a posteriori querying. The importance of big data in clinical practice quality improvement has long been known and is increasingly recognized (3-6). The Diabetic Retinopathy Clinical Research Network, United Kingdom National Ophthalmology Database, EyePACS, and the Intelligent Research in Sight (IRIS) registries are a few of the early repositories with potential to appropriately adapt. These existing clinical-practice focused infrastructures are encouraging, yet further organization and standardization is greatly needed. Progress has been slow in translating big data into clinically-relevant improvements in our understanding of disease. This can be remedied using the Odaibo Big Data Framework. Here, data collection and storage systems are purposefully built to facilitate accessibility, collaboration, and big data analysis.

Computing

Algorithm and software development is the effector arm of the Odaibo Big Data Framework. This arm of the framework transforms data into understandable and useful form. Here, useful information is extracted from otherwise overwhelming and unintelligible data. Sophisticated mathematical models serve as the translators between expert knowledge and computer code. The mathematical models link our understanding of medical science and disease with our understanding of how to build algorithms that run on a computer. This is where the rubber meets the road. It is where interdisciplinary collaboration is explicitly necessary—and can no longer be avoided. There is some encouraging progress here as well. For instance, our group is currently competing in the ongoing challenge to use Deep Learning Neural Network techniques to determine end-systolic and end-diastolic volume from cardiac MRI images. The competition uses 3D cardiac MRI images from the National Heart Lung and Blood Institute as input. Also, between February and July 2015, the California Healthcare Foundation sponsored a competition to attempt to use pattern recognition, image classification, and machine learning techniques to diagnose and grade diabetic retinopathy. The competition used color fundus photograph images from the EyePACS database as input.

Today there is tension surrounding big data analysis in ophthalmology and in medicine. Not the type of political tension that racked eighteenth century France, but tensions from a big promise in science that is yet to be fulfilled. We promised our patients that we will be able to use big data from the human genome project, the other omics sciences, and from imaging, to develop better and personalized treatments for them. They were enthused by the idea and it was widely funded. Today the data is available and is growing exponentially or faster, but we are not optimally managing and benefiting from it. We are on the verge of a data avalanche—one that can only be averted by a structured revolution in the way we gather, store, and analyze big data. This problem naturally calls for seamless cross disciplinary collaboration and a structured approach as we have outlined in this paper.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Onken MD, Worley LA, Ehlers JP, et al. Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death. Cancer Res 2004;64:7205-9.
2. Jolliffe I. Principal component analysis. John Wiley & Sons, Ltd., 2002.
3. Chiang MF, Boland MV, Margolis JW, et al. Adoption and perceptions of electronic health record systems by ophthalmologists: an American Academy of Ophthalmology survey. Ophthalmology 2008;115:1591-7; quiz 1597.e1-5.
4. Coleman AL. How Big Data Informs Us About Cataract Surgery: The LXXII Edward Jackson Memorial Lecture. Am J Ophthalmol 2015;160:1091-1103.e3.
5. Boland MV. Big Data, Big Challenges. Ophthalmology 2016;123:7-8.
6. Coleman AL, Morgenstern H. Use of insurance claims databases to evaluate the outcomes of ophthalmic surgery. Surv Ophthalmol 1997;42:271-8.

Cite this article as: Odaibo SG, Odaibo DG. The way forward for big data analysis in ophthalmology and medicine. Eye Sci 2016;31(1):4-6. doi: 10.3978/j.issn.1000-4432.2016.03.02