The Outing of the Medical Profession: Data marathons to open clinical research gates to frontline service providers.

LSE blogs.lse.ac.uk/impactofsocialsciences/2014/07/15/the-outing-of-the-medical-profession/

Could greater data transparency across the medical field solve the problem of unreliable evidence? **Dr. Leo Anthony Celi** *charts the efforts to improve the publicly available MIMIC database, a creation of the public-private partnership between MIT, Beth Israel Deaconess Medical Center and Philips Health-Care, through a series of data marathons. Data scientists, nurses, clinicians and doctors are coming together to collaborate and answer clinically relevant questions to establish a better system of cross-validation and replication.*

The biggest phenomenon that is impacting healthcare is not the data revolution or the open data movement. It is the outing of the medical profession and the biomedical research enterprise which commenced with the Institute of Medicine's report "To Err is Human" [pdf] in 1999. In 2013, we were dealt back-to-back blows with the National Academy of Sciences' report "Shorter Lives, Poorer Health" [pdf] and The Economist piece "Unreliable research: Trouble at the lab". The medical profession (and the entire healthcare industry for that matter) has not been functioning as well as we thought it did. "Half of what we know might be wrong, and the other half useless," is perhaps the most damning appraisal of the state of medical knowledge that came from Professor John Ioannidis in his editorial "How Many Contemporary Medical Practices are Worse than Doing Nothing or Doing Less?".

> *Many people think that doctors make their recommendations from a basis of scientific certainty, that the facts are very clear and there's only one way to diagnose or treat an illness. In reality, that's not always the case. Many things are a matter of conjecture, tradition, convenience, habit.*

These are the words of Dr. Arnold Relman, who led the New England Journal of Medicine for 23 years. One of the biggest critic of the health system, Dr. Relman died last week on his 91st birthday. Much has been written about open data and data sharing as a potential solution to the problem of unreliable research. But we believe that the solution lies not only in complete data transparency, but more importantly, a culture of deeper collaboration among investigators is necessary. Some may argue that competition, as opposed to cooperation, is the engine that drives scientific discovery.

For example, one might maintain that fierce competition accelerated the process of Watson and Crick's solution of the structure of deoxyribonucleic acid (DNA). However, the competitive process can become counterproductive when secrecy transcends honest collegiality and an 'end justifies any means' approach is adopted. Open collaboration with Pauling, Franklin and Wilkins may well have *shortened* the discovery process. In addition, the glory would have been distributed differently or at least, more widely shared. Watson and Crick feared the possibility of Pauling's latching onto the solution first far more than they welcomed the potential contribution of Pauling's genius. Indeed, it wasn't until Watson and Crick obtained unauthorized and questionably ethical access to the crystallographic work of Rosalind Franklin that they were able to correctly deduce the double helical structure.

Image courtesy of author

The basic premise of a data marathon is to bring together frontline providers – nurses, pharmacists, doctors – with data scientists to answer clinically relevant questions over the course of a weekend. Participants will work with a large open-access database called MIMIC, a creation of public-private partnership between MIT, Beth Israel Deaconess Medical Center and Philips. MIMIC has attracted data scientists and clinicians who have collaborated on studies that have included examination of treatment effect heterogeneity, comparativeness effectiveness research, cost analysis and predictive modeling, among others. The next MIMIC database Critical Data Marathon will be held 5-7 September 2014 but the registration deadline is 31st July.

> *It ought to be remembered that there is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things. Because the innovator has for enemies all those who have done well under the old conditions, and lukewarm defenders in those who may do well under the new.*
> *-Niccolo Machiavelli*

One of the aims of the data marathon is to open the research gates to front line clinicians, data scientists and students who have not typically engaged in clinical investigations and empower them to contribute and become part of a data-driven learning system. There is increasing concern that this approach will only further augment the noise resulting from biases and problems that currently plague the scientific literature. We share these concerns and we propose to establish a better system of cross-validation and replication between groups working on similar problems. Competing laboratories become partners. But this will only work if there is a more lateral distribution of investments, grant funding, and credit for scientific discoveries. The added accuracy of the scientific findings is only one of the benefits of the systematization of data interrogation. Another will be the enhanced ability of individuals of every educational level and area of expertise to thrust themselves into the fray and contribute to science.

For information on the Critical Data London-based event in September, please contact @tompollard.

Featured image courtesy of Leo Celi.

Note: This article gives the views of the authors, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our Comments Policy *if you have any concerns on posting a comment below.*

About the Author

Dr. Leo Anthony Celi *is an internist, an intensivist, and an infectious disease specialist, who has practiced medicine in 3 continents, giving him broad perspectives in healthcare delivery. In addition, he pursued a master's degree in biomedical informatics at Massachusetts Institute of Technology (MIT) and a master's degree in public health at Harvard University. He founded and directs Sana (*http://sana.mit.edu*) at the Computer Science and Artificial Intelligence Laboratory at MIT. His research interests are in the field of clinical data mining, health information systems and quality improvement. He holds a faculty position at Harvard Medical School as an intensivist at the Beth Israel Deaconess Medical Center and is the clinical research director for the Laboratory of Computational Physiology at MIT (*http://lcp.mit.edu/*). Finally, he is one of the course directors for HST.936 at MIT – health information systems to improve quality of care in resource-poor settings (*http://sana.mit.edu/education/2014hst936/*).*