

Health Care Data Science for Quality Improvement and Patient Safety

October 1, 2016

Rajkomar A. Health Care Data Science for Quality Improvement and Patient Safety. PSNet [internet]. 2016. <https://psnet.ahrq.gov/perspective/health-care-data-science-quality-improvement-and-patient-safety>

Perspective

Background

Having spent countless sleepless nights on call during my internship hunched over coffee-stained tables writing notes by hand, I eagerly anticipated implementation of a state-of-the-art electronic health record (EHR) during my final year of residency. As a regular user of Amazon and Netflix, I assumed that merely adopting such technology would enhance my productivity in the hospital, the way it improved my life outside the hospital. Yet, I soon confronted the same clunky user interfaces and unintuitive EHR designs that have frustrated health care practitioners everywhere.⁽¹⁾ Optimistically, I assumed that even if the user interfaces were subpar, at least the data we so painstakingly entered into the system could be used to improve patient care.

Although the log of all EHR activity is dumped into large databases daily, the process of piecing together analyzable datasets is done by report writers. This new class of health information technology (IT) professionals has direct access to stores of data packed in large spreadsheets. To find a piece of data, they look up the aisle and shelf where it is stored and retrieve it for you, analogous to a librarian finding a single book in an enormous library. Yet clinical data is complex and contextual—a heart rate may be listed under the formal vital sign table or under nursing documentation, where it is listed as a pulse. A report writer without clinical background may not appreciate that a request for heart rate should actually include data from both tables, and a clinician requesting data would not even know to give that level of specification.

After failing a few times to get valid datasets pertaining to a specific patient population from report writers, I hypothesized that I could do it faster and more accurately if I extracted the data myself. Therefore, I went through a lengthy training process (24 hours of in-person classes, a 4-hour project, and 4 hours of supervised testing) to get direct access to EHR database. I was only the second of more than 2400 UCSF faculty members to do so.

Complex Health Care Data Requires Multidisciplinary Teams to Understand

Once I had direct access to billions of rows of data spread across thousands of tables, I realized that I could theoretically answer nearly any clinical question I had. Did clinicians at my institution transfuse blood in a way consistent with modern guidelines? Could we predict which patients would have avoidable emergency department visits? I had expected that extracting data to answer these questions would be similar to the process of examining laboratory results for my hospitalized patients—as long as you know where to look, the process would be transparent.

However, harnessing EHR data along with sensor, app, and patient-reported outcomes to improve patient safety and quality is challenging. The data hardcodes nearly every detail of the work done by other health professionals. Therefore, understanding the data requires close collaboration among multiple health professionals. A data analyst who is not comfortable seeking out nurses, pharmacists, or laboratory technicians would have an even harder time piecing together the raw data in a meaningful way.

As an example, a week after I had finished caring for patients in the hospital, I wanted to review my use of antibiotics. I had assumed that knowing when and how to order a medication were sufficient to analyze the data on how I prescribed medications. Instead, I found an incredibly complex trail of data showing every step in which multiple health care professionals took care of countless details of formulating medications, transferring them to the right unit, and administering them. As a clinician, I ordered medications, but I soon realized that I only initiated a Rube Goldberg–like machine that handled all the details for this seemingly simple order to be carried out. Moreover, I noticed an unanticipated consequence of poor user interfaces. With multiple ways to order the same laboratory test or medication, clinicians frequently must cancel and reorder services. Rather than reflecting the clinician's intention, the data faithfully encodes the ostensibly capricious computer instructions of clinicians trying to get the system to do what they want.

Emerging Role for a Clinician Data Scientist

After observing this phenomenon time and again, I recognized a glaring need for a clinician–data translator. The core skills would be domain expertise in clinical systems, ability to extract data from large electronic stores, and thorough understanding of how to rigorously analyze large datasets. In real life, these skills are intertwined because if a researcher wants to add a variable to analysis, the translator would need to assess the benefit of that addition, whether the data exists within the database, and if it can be extracted successfully. In other fields, the combination of a domain expert, a computer scientist, and a statistician is referred to as a data scientist.

The benefits of a clinician data scientist extend far beyond simply accessing large, accurate datasets, although that is the first step. Consider the real-world problem of credit card fraud detection, where suspicious purchases are flagged and occasionally credit card orders are declined to prevent fraudulent purchases. For example, PayPal found that purchases from multiple parts of the world may indicate fraud, but adopting this rule universally would unnecessarily flag the purchases from a pilot. To prevent this, humans review the results from computer purchasing algorithms. Now imagine that, instead of looking for fraud, we are looking for adverse events, such as a strange antibiotic pattern or constellation of vital signs, which could prompt closer review by a clinician to catch events early or even before they occur. With a foot

in both the clinical and data science realms, clinician data scientists are best positioned to determine which questions warrant significant investment to answer and, just as importantly, how to integrate them into workflows. This role is distinct from that of a clinician–statistician; biostatisticians have traditionally focused on the design of clinical studies that have defined protocols of data collection and analysis.

In other domains where prediction is applied—like which movie or product you may enjoy—the consequences of acting on an inaccurate recommendation are limited. In medicine, accepting a suggested clinical intervention generated by a flawed algorithm carries significant risk to patients and even clinicians, who might be operating in a busy environment where second guessing every decision is impractical. Clinician data scientists can help bridge the gap of where to apply novel algorithms and how to design safeguards to prevent mistaken applications of algorithms.

There is no current pathway to train such clinician data scientists. Many data science training pathways and degrees are domain agnostic: designed to create professionals who can work in a variety of fields. The training focuses on methodology and technical skills that are broadly applicable but, frankly, technical. However, learning how to leverage distributed file systems for batch analysis does not appeal to most clinicians, who may be interested in knowing enough to work productively with data scientists but not in creating a complex pipeline of data infrastructure. Just as clinician researchers must know enough about logistic regression and survival analysis to work productively with biostatisticians, clinician data scientists should know enough of the technical details of the data flows and programming to be conversant with the data scientists they will collaborate with, although they do not need to become expert programmers themselves. New training programs must be created that blend the technical training of data scientists with particular emphasis on applications to the health care domain, which requires collaboration with multiple health care professionals.

As the work of clinician data scientists becomes more prominent, rank-and-file clinicians will also need additional training to work with data products. Recently, a representative of an EHR vendor recounted the story of one physician who was shown an algorithm that predicted a high readmission risk for a patient driven by the high number of medications prescribed. He asked a data scientist if he should prescribe fewer medications to reduce this risk. The doctor failed to appreciate that the machine learning algorithm found a significant correlation between the number of medications and readmission risk, but a high number of medications did not cause the patient to be readmitted.

Although the call for clinician data scientists has largely been in the context of precision medicine (focused on "-omic" data), it should be supplemented with a call for clinician data scientists who can harness clinical datasets to improve quality and safety. EHR datasets contain valuable information that can provide insight on delivering better care, whether through retrospective analysis or enabling prospective trials. This work will depend on involving clinician data scientists intimately in the process as a bridge between raw data and the clinical activity to be understood and optimized.

Alvin Rajkomar, MD Assistant Professor Division of Hospital Medicine University of California, San Francisco

References

1. Rosenbaum L. Transitional chaos or enduring Harm? The EHR and the disruption of medicine. N Engl J Med. 2015;373:1585-1588. [\[go to PubMed\]](#)