

Using Electronic Record Data to Encourage Better Care: Where We Are and Where We Could Be

Shira Fischer, MD, PhD

RAND Corporation, Boston, MA, USA.

J Gen Intern Med 32(7):728–9

DOI: 10.1007/s11606-017-4035-9

© Society of General Internal Medicine 2017

My children live in a world that a few decades ago was in the realm of science fiction. They are unfazed by—or more accurately, expect—video to accompany all phone calls, whether conducted on the sidewalk or in a moving car. They don't think twice about being able, with a simple keyword search, to summon up a picture of any animal or cartoon character that comes up in conversation. They set timers for baking cookies by asking Siri, the voice recognition on my smartphone, to remind us in 10 min. They pick YouTube videos from recommendations customized based on their viewing history.

Yet little of this kind of technology is being used in medicine. That's not to say medicine has not made great headway in the past decade: targeted chemotherapy drugs save lives, better imaging has allowed for earlier detection of disease, and developments in genomics permit, for example, running a panel of tests for more than 100 inherited diseases for \$149.¹ But many technologies that we accept as routine in everyday life—the “big data” that make YouTube and Amazon suggestions useful and enable accurate search, smart recommendations, voice recognition, and more—are not being used to their best advantage in medicine.

Obviously, diagnosing cancer is more complex than coming up with restaurant recommendations. However, components of the approach could be applied directly. Other folks who liked that Italian place in the North End also enjoyed a new bistro downtown. Likewise, patients who had symptoms like yours were more likely to test positive on this screening test, so perhaps you should get that test done too. Farfetched? Not at all. Risk scores are used in medicine all the time, but they're usually developed using historical data, refined with extensive work, and then locked. The Framingham Risk Score, developed in 1998,² has been updated twice—perhaps sufficient for a score that is calculating 10-year risk, but failing to account for new treatments or specific sub-populations.

The mathematics and the technical capability to enlist such tools exist. More than 7 years ago, Reis and colleagues used longitudinal patient history information from claims data to

accurately predict a patient's future risk of a diagnosis of domestic abuse.³ The predictions were eerily accurate, identifying victims 10 to 30 months before the diagnosis showed up in their charts. And yet no emergency room is using such an algorithm at the point of care today; our electronic tools have only the simplest calculators built in. We are many steps away from using the vast data we have to their maximal capability.

In this issue of JGIM, Meyer et al. describe an electronic health record (EHR)-based algorithm designed to address the problem of delayed follow-up of abnormal thyroid stimulating hormone (TSH) test results.⁴ The researchers developed an algorithm based on retrospective analysis of 1 year of EHR data from the Department of Veterans Affairs and validated the results using chart review. This approach is logical: using electronic record data and documented outcomes, we can develop algorithms to improve those outcomes for future patients.

The findings are encouraging for a number of reasons. The intervention, once programmed to a high level of accuracy and implemented, has little marginal cost. It offers an easy way to identify patients to prevent a potential adverse event. Slightly modified, it could be expanded to other hospitals. Refinement of the algorithm should be possible as new data come in. And it can serve as a model for other interventions.

However, the potential impact of this one specific algorithm is quite limited. Of the almost 300,000 patients who had a TSH test result in the study's data set, only 1250 were identified as having an abnormally elevated result that potentially needed follow-up. Of these, only 163 (0.06%) had a correctly identified delayed follow-up (and another 108 who were incorrectly identified as being delayed). Thus, while the intervention could be useful for these patients, the yield is small, and there are almost as many false positives as true positives, a concern in the era of alert overload.

Thus this solution is only a first step. The high false-positive rate will cause alert fatigue and invites refinement of the algorithm. And unless applied to additional clinical situations beyond hypothyroidism, this innovation will not affect large numbers of patients. Still, the method represents a clear beginning and can be improved, for example, by adding recent thyrogen injection as an exclusion criterion. This kind of automatic electronic surveillance should be happening more often. In an ideal world, we would have a tool that could check on follow-up for all lab results, not just for the specific case of hypothyroidism, and that would be subject to ongoing

refinement as new data are collected. Like the study of claims data predicting domestic violence, we should be analyzing historical data constantly to find and refine the best predictors of outcomes, to allow earlier diagnosis or, even better, prevention.

We are not there yet, for a number of reasons. Healthcare data are hard to collect, though it has gotten easier over time as institutions develop bigger repositories. Outcomes are hard to track, especially for information stored in free text. Databases like the MIMIC Critical Care database⁵ and contests like the Kaggle prediction challenge⁶ have made data available to the public for analysis and testing. Such data sets, however, even when available, are incomplete and limited. Like any data set collected for a non-research purpose, claims data have systematic problems for use in research. Start and stop dates for medications are notoriously inaccurate, and there is much complexity with regard to medications not prescribed by the institution collecting the data. In addition to issues of data accuracy, there is missingness in medical data that is not random, since there are a large number of fields that are only populated in some circumstances. There are privacy issues in making data easily accessible for algorithm development, and data are cordoned into different data “banks,” as patients receive care from multiple institutions or locations or submit claims to different payers. Combining the same person’s records from multiple sources is actually quite difficult because of the lack of a unique patient identifier, and much of the critical information in each record is still stored in free text, or even in PDFs, which can make the data difficult to retrieve.

We are slowly tackling those issues, one at a time. As we do so, algorithms like Meyer’s will become increasingly useful, and we do see many researchers taking up the effort. A research team led by Schiff recently took a similar approach, looking at more than 700,000 outpatient records. Rather than limiting their analysis to a few specific medications, they examined all outliers, looking for medication errors. The system was implemented prospectively, and Schiff et al. found relatively few false positives (75% of the chart-reviewed alerts were accurate). And rather than programming a specific threshold, the system was actively learning and identifying outliers by taking into account results for the same lab test obtained over time.⁷

One could envision algorithms like the ones by Meyer or Schiff being implemented for every laboratory test, taking into account patient diagnoses, time since last test, medications, allergies, and more. In addition, a true learning system would have a feedback element where a user could, in real time, rate an alert as appropriate or inappropriate, which would dynamically improve accuracy and relevance automatically without

the need for direct human modification of the algorithm. This could lead to the inclusion of factors in the algorithm that were not even considered at first—perhaps distance from the testing center plays a role in delayed follow-up, or outcomes are split on demographics that weren’t even considered.

Some individuals worry that the EHR is being asked to replace the physician. That, of course, is not the goal. The goal is safe, effective, patient-centered, timely, efficient, and equitable care.⁸ Smart algorithms, well-designed reminders, and data-driven information can help us get closer to these goals.⁹

With much room for expansion and improvement, the solution implemented by Meyer and colleagues still demonstrates the fundamental theorem of informatics, to wit: “A person working in partnership with an information resource is ‘better’ than that same person unassisted.”¹⁰ Or, more simply, *brain + computer > brain*. Which is something my kids already know. They ask me questions, and if I tell them I don’t know the answer, they immediately suggest, “Let’s ask Siri!”

Corresponding Author: Shira Fischer, MD, PhD; RAND Corporation, 20 Park Plaza, Suite 920, Boston, MA 02116, USA (e-mail: sfischer@rand.org).

Compliance with Ethical Standards:

Conflict of Interest: The author declares that he/she does not have a conflict of interest.

REFERENCES

1. JScreen. JScreen - FAQs. 2014. Available at: <https://jscreen.org/faq/>. Accessed Feb 2017.
2. Wilson PW, D’Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837–1847.
3. Reis BY, Kohane IS, Mandl KD. Longitudinal histories as predictors of future diagnoses of domestic abuse: modelling study. *BMJ*. 2009;339:b3677.
4. Meyer AN, Murphy DR, Al-Mutairi A, et al. Electronic detection of delayed test result follow-up in patients with hypothyroidism. *J Gen Intern Med*. 2017. doi:10.1007/s11606-017-3988-z.
5. MIT Lab for Computational Physiology. MIMIC-III. 2015. Available at: <https://mimic.physionet.org/>. Accessed Feb 2017.
6. Kaggle. Practice Fusion Analyze This! 2012 - Prediction Challenge. 2017. Available at: <https://www.kaggle.com/c/pf2012>. Accessed Feb 2017.
7. Schiff GD, Volk LA, Volodarskaya M, et al. Screening for medication errors using an outlier detection system. *J Am Med Inform Assoc*. 2017.
8. Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. National Academies Press; 2001.
9. HealthIT.gov. Benefits of EHRs. 2014. Available at: <https://www.healthit.gov/providers-professionals/improved-diagnostics-patient-outcomes>. Accessed Feb 2017.
10. Friedman CP. A “fundamental theorem” of biomedical informatics. *J Am Med Inform Assoc*. 2009;16(2):169–170.